This paper describes different methods to estimate the uncertainty of wind power forecasts in terms of prediction intervals. The single methods and an ensemble average model have been applied to shortest-term wind power forecasts (forecast horizon = 1, 2, 4 & 8 h) of 62 spatially distributed wind farms in Germany to obtain intervals with a nominal reliability of 90, 95 and 98 %. Furthermore the “ISET online model” was used to calculate the prediction interval for the total wind power generation of Germany with a reliability exceeding 99 %. The skill of the resultant intervals is investigated with regard to reliability and sharpness. It was found that the skill depends on the quality of the underlying wind power forecast.

1. Introduction

Wind power is a fluctuating source integrated as a must-run unit in the power system. Therefore the integration of large amounts of wind power into the electricity supply system requires accurate and reliable forecasts of the electricity generated by wind turbines for the next hours to days ahead. In addition to power station outage and stochastic load variability the fluctuating wind power generation is the main reason for controlling and balancing the electrical grid. Hence, improved accuracy of wind power forecasts does not only lead to higher grid reliability but also to attractive cost savings.

Nowadays the most operating wind power forecast systems are based on (point) numerical weather predictions (NWPs) and predict one single value for each forecast horizon. Typically the uncertainty of such forecasts is represented by a set of global error values, e.g. the root mean square error (rmse). The disadvantage is that these values represent rather the climatological than the meteorological aspect is that these models can be adapted to dynamic prediction intervals but also for quantile estimations due to their ability to forecast a whole probability density function (PDF) of the expected power. Out of this distribution the “best” point power forecast and the corresponding uncertainty estimations can be extracted [4], [5]. This method mainly reflects the predictability of the weather situation and hence the uncertainty of the NWP.

The approach, presented in this paper, is based on the observed forecast errors of an existing wind power prediction system and the already processed (point) NWP. The advantage is that all input parameters for the uncertainty estimation models are already available and that the models can be seen as a simple add-on to the existing forecast system. There is no dependence to the more expensive ensemble NWPs. Another important aspect is that these models can be adapted to generate uncertainty estimations in terms of dynamic prediction intervals but also for quantile forecasts.

Six model approaches have been applied to estimate the forecast uncertainty of an existing wind power prediction system in terms of dynamic prediction intervals. They are introduced in section 3. A simple static approach serves as a benchmark (see section 2).

Two case studies are presented in section 4. In the first one, the models are applied to compute prediction intervals of the shortest-term forecasts for 62 German wind farms. The second one shows prediction intervals for the total German wind power generation.

2. Static prediction intervals

Two criteria determine the quality of a prediction interval. The first one is the reliability. The reliability ranges from 0 to 100 % and provide the percentage of measurements inside the interval. In other words the reliability represents the probability that a single value is inside the interval. The other quality parameter is the sharpness that is expressed by the
average interval size. A sharpness of 100% would present an interval covering the complete power range.

The first problem concerning the generation of prediction intervals is to find the best trade-off between reliability and sharpness. It is evident that the average interval size increases rapidly for an increasing reliability higher than ca. 90%. The second problem is that the observed reliability can differ from the nominal reliability which has specified during model development. Hence the model must be calibrated and it must be guaranteed that the calibration does not change.

The easiest solution to estimate a forecast uncertainty is the worst case expectation, i.e. the largest observed forecast error of the running system is applied to every forecast step as a constant uncertainty. This is a forecast uncertainty with 100% nominal reliability and correspondingly large sharpness. The example time series of forecast errors in Figure 1 shows that the size of the interval covers nearly the complete power range (see Fig. 1, green).

The simple approach to generate intervals with smaller sharpness is to use the statistic of the observed errors. As outlined in Fig. 1 the quantiles (cyan lines) have been derived in a way that 95% (nominal reliability) of all measurements are located inside the interval, 2.5% above and 2.5% beneath. As shown in the lower plot of Fig. 1 the resultant interval has a significantly smaller sharpness than the worst-case approach. Due to the simplicity in applying this approach it has been selected as benchmark for this work.

3. Models of dynamic prediction intervals

The aim of this work is the evaluation and comparison of models for estimating the forecast uncertainty in terms of dynamic prediction intervals. All models presented in the next subsections are based on already observed forecast errors of an existing wind power forecast system. The models estimate the upper and lower deviation from the underlying power prediction in a way that the measurements are within the interval with a specified probability i.e. nominal reliability.

![Fig. 1: Time series of forecast errors (top) and prediction intervals (bottom)](image)

### 3.1 Adaptive model

The adaptive model is based on the regular updated error distribution (pdf) with a higher weighting of the last g hourly forecasts errors (here g=48). The upper and lower uncertainty estimations (quantiles) $q_1$ and $q_2$ depending on the forecast time $t_i$ were calculated as follows:

$$q_i(t_i) = \frac{a_i*(pdf(t_{i-g},...,t_{i-1})+b_i*(pdf(t_{i-2},...,t_{i}))}{a_i+b_i}$$

$$q_i(t_i) = \frac{a_i*(pdf(t_{i-g},...,t_{i-1})+b_i*(pdf(t_{i-2},...,t_{i}))}{a_i+b_i}$$

The parameters $a_i$ and $b_i$ ($i=1,2$ and $a_i < b_i$) are the weighting factors and s is the forecast horizon.

### 3.2 Simple classification

Using the simple classification method the model to generate prediction intervals was developed by investigating historical weather data and the simultaneous observed forecast errors. The determination of the respective forecast error is based on a simplified classification of the weather situation based on wind speed and wind direction given by the NWP. Regarding this modelling method the error distribution and the consequent quantiles were computed depending on wind speed and wind direction. It is worth to note that this modelling approach leads to discontinuities of the interval that have to be avoided. Smoothing techniques or fuzzy set logics [3] can be applied to handle this problem.

### 3.3 Artificial neural networks (ANN)

Two artificial neural networks (ANN) are used to calculate the upper and lower border of the prediction interval. The ANN are trained with the historical time series of power forecasts. In the first step the respective time series of forecast errors is divided in one time series covering all positive errors and in a second with all negative errors. These error time series are scaled in way that the mean positive and negative errors are equal to the quantiles of the total error distribution corresponding to the specified reliability. Each of the both time series is used separately for the training of the two ANNs. As input for the ANN the same NWP parameters that have been used for the power prediction system, the predicted power itself and/or the recently observed forecast errors are potential predictor variables. The selection of the best input parameters concerning uncertainty estimation can vary from wind farm to wind farm. In operation the resultant ANN output in form of the upper and lower uncertainty has to be combined with the power prediction to get the prediction interval.

### 3.4 Linear quantile regression

The regression models are developed to reflect a relation between the observed forecast errors and the selected input parameters, like NWP data and/or the predicted power and recent forecast errors. Compared to linear regression models that approximate the conditional mean of the response variables, the quantile regression formalism relies on estimations of either the median or other
quantiles of the response variables, which is an advantage for the development of prediction intervals based on various quantiles of the error distribution [6]. That is why the specification and implementation of the nominal reliability is comparably easy for all quantile regression models.

3.5 Multi-linear regression
The setup of the multi-linear regression models requires a splitting of the observed errors in positive and negative errors and hence two separate model trainings. The positive and negative errors have to be scaled like done for the ANN training. The resulting set of linear equations can be solved easily using the (Moore-Penrose-) Pseudo inverse of the matrix including the selected input parameters [7]. In operation the regression models compute the upper and lower uncertainty.

3.6 Ensemble average
The ensemble average has been calculated by averaging arithmetically the resultant upper and as well the lower uncertainties of all models.

4. Evaluation of the models
Two case studies have been performed to validate the performance of the several models concerning reliability and sharpness depending on the quality of the underlying wind power prediction system.

4.1 Case study I: single wind farms
The models described in section 3 were applied to shortest-term forecasts of 62 German wind farms with forecast horizons of 1, 2, 4 and 8 hours. The quality of these 248 power forecasts in terms of the root-mean-square-errors normalized on capacity (nrmse) range between 4 and 19 %. As input parameters for the several models the weather forecasts that has already been used for the power prediction, the predicted power itself and the recent forecast errors have been selected. The main objective was to calculate dynamic prediction intervals with a nominal reliability of 90, 95 and 98%.

The applied data were adapted from the testing-phase of the corresponding wind power prediction systems including different periods from January 2004 until September 2006 and the complete time-series spanning the period from October 2006 to May 2007. For the assessment of the models the data of 2007 (01/01/2007 – 30/04/2007) have been reserved.

Fig. 2 and 3 present the model performance illustrated by the averaged quality parameters of the 248 prediction intervals concerning reliability and sharpness. It is obvious that the observed reliability of the quantile regression model is the only one which is nearly equal to the nominal reliability and hence presents the best performance related to the model reliability. The multi-linear, ANN and ensemble-average models are over-confident in average, i.e. their observed reliability is higher than the nominal reliability. On the other hand the simple-classification and adaptive models and as well the static approach are under-confident and consequently not reliable for decision-making problems. Generally, the mean interval sizes of the over-confident intervals are larger compared with under-confident intervals. Having a look on Fig. 3 it is obvious that all single models result in prediction intervals with a nearly equal averaged sharpness depending on the observed reliability. The ensemble average model presents the best performance and an improvement of about 11 % compared to the single models. A further analysis has revealed that this improvement is further increasing for increasing reliability.

The investigation of the model performance depending on the quality of the wind power prediction system itself has shown that for all models a linear relation between the sharpness and the nrmse of the power forecast exists (Fig. 4, left). After calibrating the static approach and the ensemble average model in a way that the observed reliability is equal to the nominal reliability (see Fig. 4, right) the improvement concerning the sharpness by using the ensemble average model becomes very clear. It is apparent that the improvement is further increasing for decreasing quality of the
corresponding wind power prediction system. With respect to 95%-prediction intervals the sharpness of the ensemble average model can be estimated depending on the rmse of the power forecast of a single wind farm by the following equation (as a result of a linear fit in Fig. 4, right):

\[
\text{Sharpness} \% = 3 \times \text{rmse} \% + 2.7
\]  

(2)

4.2 Case study II: German wind power generation

The objective of this case study was to calculate the prediction intervals of shortest-term forecasts (1, 2, 4 and 8 h) of the total German wind power generation with a reliability exceeding 99 %. The calculation method is based on the prediction intervals (adapted from the ensemble average models) of 62 wind farms distributed over Germany. Applying the transformation algorithm of the “ISET online model” [8] the 62 single intervals were used to calculate the prediction interval of the total German wind power generation. The study has been performed with the same data as used in case study one.

The analysis of the four final prediction intervals reveals that the reliability exceeds 99.7 %. The corresponding mean interval size depending on the forecast horizon is shown in Fig. 6. An improvement of about 25 % is obtained when using the up-scaled ensemble average prediction intervals compared to the static approach. A reliability of 99.7 % means that the forecast error is outside the forecast interval for less than 24 hours of one year. As shown in Fig. 5 significant forecast errors are covered by the prediction interval but situations during extreme events like hurricane “Kyrill” require further research.

5. Conclusion

Five different models to estimate dynamic prediction intervals of existing wind power forecast systems have been evaluated and compared. Their performance concerning reliability differs significantly but their sharpness is nearly equal. The results of all models have been combined in an ensemble model which results in a higher quality of forecast uncertainty estimation. Regarding the sharpness it leads to an improvement of about 11 % compared to the single models. Concerning the total German wind power generation an improvement of about 25 % is obtained when using the up-scaled ensemble average prediction intervals compared to the static approach. It was also shown, that the advantage to use the ensemble model instead of each of the single models increases for higher reliabilities and decreasing quality of the underlying power prediction system.

Acknowledgement

The work presented in this paper has partially been funded by the German Federal Ministry of Environment, Nature Conservation and Nuclear Safety in the project “Deutscher Windmonitor” (BMU FKZ 0327584).

References