

REDUCTION OF WINDPOWER PREDICTION ERROR BY SPATIAL SMOOTHING EFFECTS

Ulrich Focken, Matthias Lange, Kai Mönnich, Hans-Peter Waldl

Department of Energy and Semiconductor Research, Faculty of Physics,
Carl von Ossietzky University of Oldenburg, D-26111 Oldenburg,
Fax ++49-441-798-3326, email: office@ehf.uni-oldenburg.de, www.physik.uni-oldenburg.de/ehf

Hans Georg Beyer, Armin Luig

Department of Electrical Engineering, University of Applied Sciences (FH) Magdeburg, D-39114 Magdeburg

We discuss the statistical effects of predicting the power output of spatially distributed wind farms. Our forecasting procedure provides the expected power output for a time horizon up to 48 hours. It is based on the large scale wind field prediction which is generated operationally by the German weather service.

In this paper we focus on the reduction of the forecast error for the aggregated power output of wind farms in a spatially extended region. Due to spatial smoothing effects the error decreases considerably compared to a single site. This reduction strongly depends on the size of the region rather than on the number of wind farms it contains. We investigate the spatial smoothing effect using measured data from 30 sites in Germany. To generalize the result we consider several model ensembles of wind farms and the current distribution of wind turbines in Germany based on a statistical approach.

Keywords: Forecasting Models - 1, Utility-Integration - 2, Dispersed Turbine Systems - 3

1 Introduction

The development of wind energy use has led to a noticeable contribution to the energy supply in Germany. At the moment, for some regional utilities the installed capacity of wind turbines is of the order of magnitude of the minimal load (approx. 30 % of max. load). The feed in of electricity by wind energy acts as a negative load leading to an increase in fluctuations of net load patterns. The insecurity of the temporal development of wind speed may have consequences for the operation of conventional power plants or load management, respectively. For a time scale from some hours to two days additional conventional reserves have to be kept ready to replace the wind energy share in case of decreasing wind speeds.

In this paper we concentrate on the reduction of the error of a wind power prediction by spatial smoothing effects continuing our work in [1]. We focus on two major variables determining the magnitude of this statistical effect, namely the spatial extension of the region and the number of wind farms it contains. For a large region the mean distance between the sites is larger than for a small region such that on average the correlation of the prediction error is weaker. Thus, the regional error is expected to decrease with increasing size of the region. Moreover, we look at the influence of the number of sites on the regional prediction error.

We use data from 30 wind farms in Germany to form typical regions with different extensions corresponding to a medium and large utility supply area and sum up the according measured power output. Fictitious model ensembles together with the correlation function based on the measured data allows us to shed some light on the general statistical behaviour of distributed wind farms regarding the prediction error. Our investigation is concluded by calculating the error reduction for the distribution of all wind farms in Germany.

2 Forecasting method

The wind power prediction method we use and its performance are described in detail in [2,3]. The principle scheme of the prediction system can be seen in figure 1. As input the result of an operational numerical weather prediction model is used. The German weather service (DWD) currently operates the "Lokalmodell" which replaced the "Deutschlandmodell" in November 1999. Our calculations are based on the windspeed and direction forecast up to 48 hours. The resolution of the data is $14 \times 14 \text{ km}^2$, i.e. rather sparse, so a spatial refinement is necessary to predict the wind power at a specific site. We calculate the windspeed at hubheight under consideration of roughness, orography and farm effects.

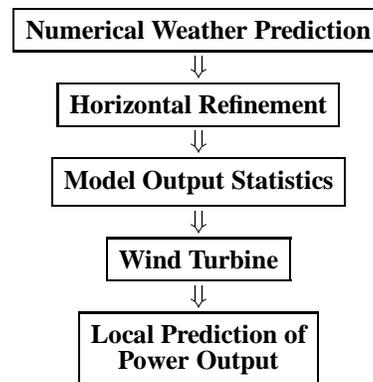


Figure 1: Principle of the spatial refinement of the numerical weather prediction leading to a local prediction of wind conditions.

3 Prediction error of single sites

The quality of the power prediction for a single site is determined by comparing the results of the locally refined prediction and measured data [3].

For this purpose archived prediction data for the years 1996 to 1999 was provided by the German weather service. In particular, we use the 6, 12, 18, 24, 36, and 48 hours predictions from the 00 UTC run. The measured data was collected from the same period of time in the framework of the German Scientific Measuring and Evaluation Programme (WMEP) carried out by ISET, Kassel [4].

Figure 2 shows a comparison between prediction and measurement for the power output of a wind turbine in the North German coastal region. In general the predicted and the measured time series correspond rather well. Significant differences can be seen mainly for the 36 and 48 hours prediction. In particular, the beginning of a storm on day 326 is not correctly predicted and on day 330 the prediction shows a time shift of several hours.

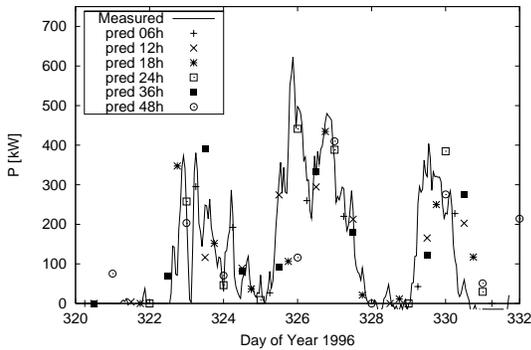


Figure 2: Typical timeseries of measured and predicted power output for one site.

In order to quantify the difference between power prediction and measurement we use the root mean square error (rmse) normalized to the installed power P_{inst} of the wind turbines in the period of time to be considered (equation (1)).

$$rmse_{norm} = \sqrt{\frac{1}{M} \sum_{i=1}^M (P_{pred} - P_{meas})^2 / P_{inst}} \quad (1)$$

P_{pred} is the predicted power output, P_{meas} the measurement and M the number of data points.

Table 1 summarizes the results of the comparison between measured data and predictions for single sites. The rmse rises from 13% for the 6 hours prediction to 19% for 48 hours. The increase of the prediction error with increasing time horizon might be due to the growing systematic error in the numerical weather forecast for longer prediction times.

prediction time [h]	6	12	18	24	36	48
rmse [% $P_{installed}$]	13	15	15	15	20	19

Table 1: Rmse between prediction and measurement of power output normalized to installed power (1996). The values are averaged over 30 sites in Germany.

4 Spatial smoothing

Under operational conditions a prediction for the combined power output of many wind farms distributed over a large region is needed, e.g. the supply area of a utility. By integrating over a region the errors underlying the measurement

and the forecast at single sites cancel out partly. These statistical smoothing effects lead to a reduced prediction error for a region compared to a local forecast. The size of the region and the number of sites it contains are the main parameters that influence the magnitude of the error reduction. The analysis of measured data shows this effect but is constrained to a fixed ensemble of sites. To generalize our findings we use model ensembles which require a statistical description of the regional prediction error in terms of spatial correlations.

4.1 Ensemble of Measurement Sites

Our first approach is to investigate the spatial smoothing effect using data from an ensemble of 30 wind farms in the Northern part of Germany. The sites are divided into regions of two different types according to typical areas covered by a medium and a large utility. The smaller regions with a diameter of approximately 140 km (see figure 3) contain three to five measurement sites each. The bigger regions are about 350 km in diameter with five to seven sites each. For comparison we form a very large region containing all sites which has a size of about 730 km.

The predicted and measured power output of a region is calculated by adding up the time series for every wind farm located in the region and dividing them by the number of wind farms. The rmse between these two ensemble time series gives the regional prediction error. Table 2 shows the results for the different region sizes and various prediction times. The rmse of the ensemble, i.e. the regional prediction error, is normalized to the mean rmse of the single sites and averaged over regions of the same size. For the given ensemble this ratio decreases with increasing region size, e.g. the six hours prediction gives an average ratio of 0.77 for the 140 km region, 0.65 for the 350 km region, and 0.49 for the 730 km region. In all cases the reduction of the regional prediction error is less pronounced for larger prediction times.

prediction time [h]	6	12	36	48
140 km region	0.77	0.78	0.83	0.82
350 km region	0.65	0.64	0.72	0.74
730 km region	0.49	0.46	0.58	0.62

Table 2: Ratio between regional error and mean error of single site ($rmse_{ensemble}/rmse_{single}$) for different regions and forecast horizons.

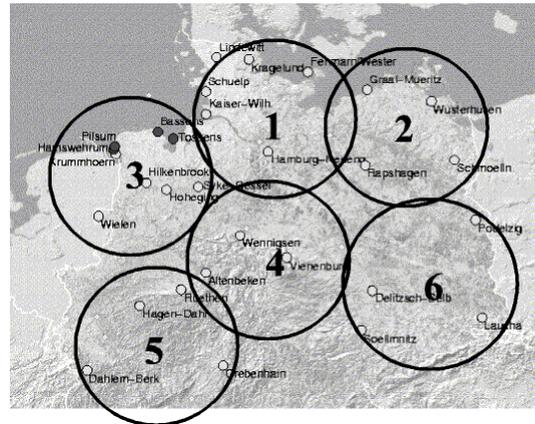


Figure 3: Regions with 140 km in diameter. The points denote the measurement sites.

4.2 Model ensembles

The analysis for the specific set of measurement sites shows a significant decrease of the prediction error compared to a

single site. In order to draw general conclusions about other configurations of wind farms we use random ensembles of sites. This allows us to vary the size of the regions and the number of wind farms over large ranges to see how the reduction of the error depends on these parameters. For this purpose we need a statistical description of the regional prediction error.

The key element connecting the spatial distribution of sites with the regional prediction error is the crosscorrelation function r_{xy} of the difference between prediction and measurement, i.e. $P_{pred}(t) - P_{meas}(t)$, for the single sites. If r_{xy} is known, the standard deviation $\sigma_{ensemble}$ of the differences between measurement and prediction, i.e. the rmse centred with the mean bias, can easily be calculated using the σ_x of the individual sites by

$$\sigma_{ensemble}^2 = \frac{1}{N^2} \sum_x \sum_y \sigma_x \sigma_y r_{xy} \quad (2)$$

where N is the number of sites in the region. $\sigma_{ensemble}$ will now play the role of the regional prediction error.

At first the crosscorrelation of the measured data is determined. For each pair of the 30 wind farms r_{xy} is calculated and ordered according to the distance between the two sites x and y . Figure 4 shows crosscorrelation versus distance for the 36 hours forecast where the pairwise data points have been averaged over 25 km bins. The curve decreases rather rapidly for distances below 100 km.

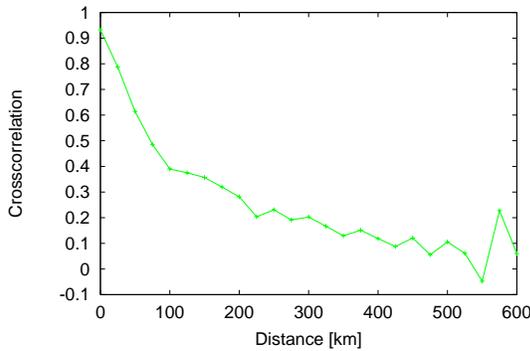


Figure 4: Spatial correlation of prediction deviations for 36 hours forecast.

We obtain a proper correlation function allowing the application of equation (2) by fitting analytic functions of the form $r_{xy} = a \cdot e^{-d/b}$ (a and b are fitparameters and d is the distance between the two sites) to the crosscorrelation derived from the measured data. It turns out that piecewise exponentials lead to a suitable fit to the data points.

The geographical coordinates of the model ensembles are chosen randomly. Each result given in the following represents an average value over ten realizations of ensembles with fixed size and number of sites.

With the correlation function r_{xy} based on the fitted data we can now use equation (2) to calculate the prediction error $\sigma_{ensemble}$ of the model regions. We set the σ_x of the wind farms to one which means that they all have the same weight. Figure 5 shows the ratio between the regional error and the mean of single sites $\sigma_{ensemble}/\sigma_{single}$ for two regions with different sizes versus the number of sites in the region. Obviously, $\sigma_{ensemble}/\sigma_{single}$ approaches a saturation level for increasing number of wind farms. This limit is already reached for a rather small number of wind farms. After that the error reduction does practically not depend on the number of sites, e.g. for the size of a typical large utility (approx. 370 km) less than 50 sites are sufficient to tell the constant level of 0.63.

The saturation level decreases with increasing size of the region. This is illustrated in figure 6 where the limit values for regions with different extensions containing 4000 sites are shown. There is a rapid decay for extensions below 500 km.

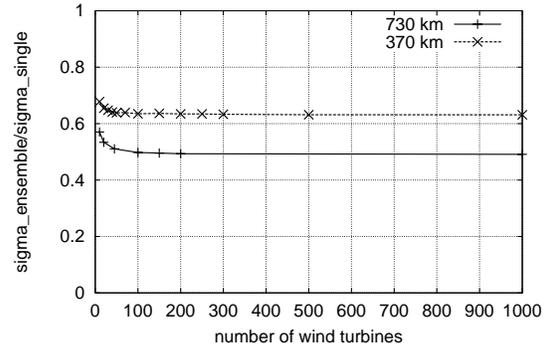


Figure 5: Ratio $\sigma_{ensemble}/\sigma_{single}$ versus number of sites for 36 hours forecast.

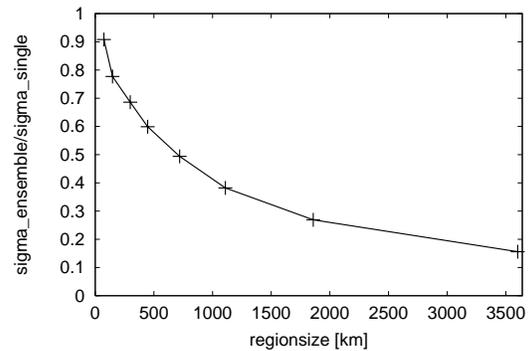


Figure 6: Saturation values of $\sigma_{ensemble}/\sigma_{single}$ (4000 sites) for the 36 hours forecast.

4.3 Distribution of German wind farms

Finally, we consider the real distribution of the wind farms in Germany (in 1999) as a special model ensemble and calculate the ratio between the regional error to a single site as above. For the 36 hours prediction this gives $\sigma_{ensemble}/\sigma_{single} = 0.43$. Note that this ratio for an equivalent region of the size of Germany with randomly distributed wind farms would be lower because the real distribution shows a strong imbalance of sites in the North and South (figure 7).

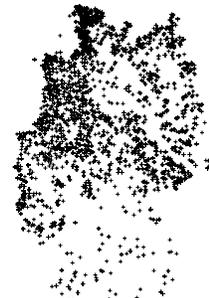


Figure 7: Distribution of wind turbines in Germany in 1999.

5 Analysis of the temporal structure of the forecast errors

For a further improvement of the forecast quality methods to correct the actual forecasts using knowledge of previous errors (model output statistics) are discussed (see e.g. [3]). A basic approach in this context is the analysis of the autocorrelation structure of the errors. A typical scatter diagram for the single site forecast errors of consecutive days is shown in figure 8 (left) The respective inspection shows that the autocorrelation coefficient is about 0.2 only. Thus the application of a simple linear correction procedure based on the previous error will not lead to any remarkable improvements of the forecast.

Looking again for the change of the structure of the forecast errors when going from single sites to ensemble data gives the result that the temporal correlation of the errors is slightly but remarkably increased. In figure 8 (right) the scatterplot for the errors of ensemble forecast for all sites under investigation is represented. For this set a correlation coefficient is increased to about 0.4. It has to be remarked, that due to missing values in the data sets these do not always refer to identical ensembles.

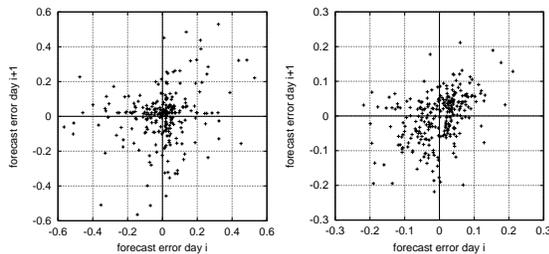


Figure 8: Left: Example for a scatter diagram of normalized single site forecast errors for pairs of 2 consecutive days. Right: Corresponding scatter diagram for normalized ensemble data. Each point refers to the ensemble output of 15-30 sites, depending on data availability.

To approve the range of the autocorrelation value for the ensemble the same type of statistical modelling as applied above for the inspection of the standard deviation of the forecast errors as presented in the previous section is used. As additional parameter the crosscorrelation value for a time lag of one day for pairs of stations has to be used here. From this information the autocorrelation of the forecast errors for the ensemble of 30 sites is recalculated. For the 6 hours forecast this value is in a good accordance with the respective parameter gained from the scatter plot as mentioned above. The autocorrelation values for the ensemble resulting for the various forecast horizons is given in figure 9 together with a typical result for a single site.

Summing up these findings, it may be stated that, again due to the levelling out of purely stochastic contributions to the single site forecast errors, at least for the 6 hours forecasts a somewhat stronger linear link between the ensemble forecast errors of subsequent days exists. For the basic ensemble of 30 sites, the respective autocorrelation coefficients are however small (below 0.5). As this is still a small value the procedures for an exploitation of this effect have to be refined ones.

6 Resume

We investigate the statistical smoothing effects that arise if a wind power prediction is made for a region with spatially distributed sites. As expected we find a reduction of the prediction error of the aggregated power prediction compared to a single site. For an ensemble of wind farms where the

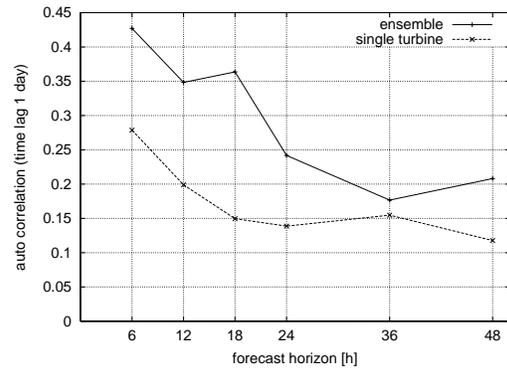


Figure 9: Autocorrelation coefficients of single site (lower line) and ensemble errors for forecast horizons of 6-48h. The data for the ensemble are based on pairwise correlation characteristics.

analysis is based on measured data the improvement of the prediction is noticeable even for rather small regions and only few sites. Using model ensembles with randomly chosen locations allows us to generalize the results to identify the impact of the two main parameters, namely the spatial extension of the region and the number of sites it contains. We find that the magnitude of the reduction does strongly depend on the size of region, i.e. the larger the region the larger the reduction. Concerning the number of sites contained in the area we observe a saturation level which is already reached for a small number of wind farms. This means that only few sites are sufficient to determine the magnitude of the improvement of the power prediction.

With the results of our analysis it is now possible to estimate the regional smoothing effect of the wind power prediction error very easily by just considering the size of the region in question.

As an additional effect of regarding regionally averaged power forecasts a somewhat stronger link between the forecast errors of consecutive days can be identified. This may be beneficially used in the refinement of procedures to correct the actual ensemble power forecast using knowledge on previous errors.

7 Acknowledgements

We like to thank ISET and DWD for providing the data. Parts of this work have been funded by the EU with project JOR3CT980272, the federal state Sachsen-Anhalt, the Bundesstiftung Umwelt, and the Hans-Böckler-Stiftung.

8 References

- [1] U. Focken, M. Lange, K. Mönnich, H.P. Waldl: *Räumliche Ausgleichseffekte bei der Vorhersage der Leistungsabgabe von Windkraftanlagen*. DEWEK 2000, Wilhelmshaven.
- [2] H.G. Beyer, D. Heinemann, H. Mellinghoff, K. Mönnich, H.-P. Waldl: *Forecast of regional power output of wind turbines*. EWEC 1999, Nice.
- [3] K. Mönnich: *Vorhersage der Leistungsabgabe netz-einspeisender Windkraftanlagen zur Unterstützung der Kraftwerkseinsatzplanung*. Dissertation, Oldenburg, 2000
- [4] ISET: *Wind Energy Report Germany 1999/2000 - Annual Evaluation of WMEP*, Kassel, 2000