

# **A new Approach for Uncertainty Estimation in Wind Power Predictions**

Ulf Gräwe, Jens Tambke, Lüder von Bremen, Nadja Saleck  
ForWind – Centre for Wind Energy Research  
Institute of Physics, University of Oldenburg, 26111 Oldenburg, Germany  
[ulf.graewe@forwind.de](mailto:ulf.graewe@forwind.de)

## **Summary**

For the grid management and trading of wind power generation, it is necessary to have on the one hand the accurate deterministic forecast and on the other hand the knowledge about the underlying probability distribution. In recent years much progress was achieved in the development of advanced deterministic forecast methods [1]. To give uncertainty estimates, state of the art models are ensemble predictions or power curve mapping of the error distribution of the wind forecast.

In this work we develop a new method to estimate the probability distribution of the forecast error. Here, only the characteristics of the predicted power time series determine the uncertainty prediction. We have defined five different measures for these characteristics to relate the forecast error to these quantities. Two of them are physically motivated (time series energy and gradients), two are probabilistic measures (entropy and jump-entropy), and also the Lyapunov exponent of the time series is used. By combination of these five measures, we can estimate the time evolution of the standard deviation of the forecast error. The Fokker Planck equation with the appropriate boundary conditions is solved for a look-ahead time of up to 72 hours to describe the evolution of the error distribution.

The developed algorithms were tested in three case-studies: First for an offshore point forecast at the FINO1-mast, second for the onshore EWE-Region with an installed wind power capacity of approx. 2 GW and at last for the whole German wind power prediction, with an installed capacity of approx. 17GW. The validation of this new method has shown that on all three test cases the confidence intervals are estimated with an error of  $\pm 2-5\%$  of the empirical coverage.

It was shown that the developed tools allow for a good estimate of the trust levels of wind power predictions with a small amount of input information, which is an important end-user requirement.

**Key words:** short-term wind power prediction, forecasts, uncertainty prediction, grid integration

## **1. Introduction**

For the management or trading of wind power generation, it is necessary to have on the one hand accurate point forecasts and on the other hand the knowledge about the underlying probability distribution. A common method to estimate the uncertainty is to use ensemble predictions (the European Centre for Medium-Range Weather Forecasts (ECMWF) offers a 52 member ensemble prediction). To avoid these costs, we assumed that a power forecast does not only carry the deterministic part but also the uncertainty information. Therefore by looking on the shape of the power prediction the underlying probability distribution can be estimated.

## **2. Data**

To simulate the wind power prediction for the three test cases we used wind forecast data of the ECMWF. Wind speeds from vertical high resolved model were integrated. 00UTC and 12 UTC model runs are used up to forecast step 72h, for FINO1 only the 48h forecast were used. Simulated and measured wind power time series for Jan.-Dec. 2004 were utilized. Because for FINO1 only wind data are available, we simulated a wind power output by multiplying the wind data with a power curve of a (3MW) wind turbine.

## **3. The error measure**

The ideas to define the following error measures are heuristically based. The challenge is that we only have 24 forecast steps. With this quite small amount of information we want to say something about the time evolution of a chaotic system. Therefore the aim is to extract as much information out of this data as possible (in a physical and also in a probabilistic sense). Due to the limitation in the amount of data conventional analysis tools like correlation functions, spectra or the autocorrelation are not applicable. Thus we think with the chosen error measures to cover the space of possible error ratings.

### **3.1 The Energy**

The energy is a measure to quantify the energy content of the time series. By this we calculate the area under the time series.

$$E_N = \sum_{i=2}^N (P_{i-1} + P_i) \frac{\Delta t}{2} \quad (1.1)$$

Here  $N$  is the number of forecast steps,  $\Delta t$  is the forecast time step size and  $P_i$  is the power prediction at time step  $i$ .

The forecast is integrated via the trapezoidal rule, and we get  $E_N$  for every prediction step. For the first forecast step we use as integration rule the right side approximation.

### **3.2 Gradients (Roughness)**

The roughness is a measure to describe the gradients in the wind power time series.

$$R_N = \sum_{i=1}^N \left| \frac{\Delta P_i}{\Delta t} \right| \quad (1.2)$$

We use the discrete version of the derivative and sum up only the absolute values, because we are only interested in the magnitude of the gradients and not in the sign of the slope. As differentiation rule we use the central difference and at the boundaries the left (right) side approximation.

One can expect that high values of  $R$  corresponds to high prediction errors, because there are strongly changing weather conditions which the numerical weather prediction system do not capture correctly in many cases.

### **3.3 The Entropy**

The definition of the entropy from an information theoretical point of view is defined as

$$H_S = -\sum_{i=1}^N p(P_i) \ln p(P_i) \quad (1.3)$$

there  $p(P_i)$  is the probability of getting the power value  $P_i$  and

$N$  the number of events. The entropy gives a number which characterizes the amount of uncertainty that one has concerning which outcome will result.

The motivation to use an entropy measure is to get information about the probability (and uncertainty) of the occurrence of a specific realisation of the power prediction. Accordingly a power prediction that is for 72 hours at the upper power limit can be regarded in a different way than a prediction with an average power generation.

### 3.4 The Jump Entropy

The definition of the jump entropy is similar to the definition of the entropy. But now we ask for the probability to get a power prediction  $P_i$  at time  $t$  under the condition  $P_{i-1}$  at  $t - \Delta t$ . So we are looking for the conditional probability

$$J_i = p(P_i, t | P_{i-1}, t - \Delta t) \quad (1.4)$$

This conditional probability can be extracted out of measured power time series. With this trained statistic we are going into the forecast time series and calculate the jump probability for every time step and compute out of this the jump entropy.

$$H_J = -\sum_{i=1}^N J_i \ln J_i \quad (1.5)$$

With the jump entropy measure we try to characterise the conditional probability of two following power predictions. One can for instances imagine that jumps which largely separate to prediction steps (in the occurrence of atmospheric fronts) lead to less predictability. Otherwise small jumps, in the case of stable situations, can characterise good prediction quality.

### 3.5 The Lyapunov Exponent

In the theory of dynamical systems a convenient indicator of the sensitivity to small orbit perturbations is the Lyapunov exponent  $\lambda$ . The Lyapunov exponent  $\lambda$  of a system gives the average rate of divergence (or convergence) of infinitesimal nearby initial conditions. This reads in the finite version

$$\lambda = \frac{1}{T} \sum_{n=1}^T \ln \left| \frac{\Delta x_n}{\Delta x_0} \right| \quad (1.6)$$

Here  $T$  is the averaging time,  $\Delta x_n$  the separation of the trajectories at step  $n$  and  $\Delta x_0$  is the initial distance. If  $\lambda$  is positive, this means an exponential divergence of the trajectories, i.e. chaos. Convergence and stable fixed points occurs for negative  $\lambda$ .

To get an idea how we calculate the Lyapunov exponent, consider the following scenario (for simplicity we use only the 00 UTC forecast). If we get the forecast for the 10.10.2004, then we also have the forecast from the 9.10.2004 (24-hours before) and also the forecast from the 8.10.2004 (48-hour before) available. This means that we have for prediction steps between 3 and 24 hours an overlap of these three predictions. For prediction steps between 27 and 48 hours we only have an overlap of the actual and the day before forecast. So the actual forecast is likely the most accurate, because it is the closest one to the 'future'. The

forecasts from the 'past' can only say something about the convergence or predictability.

We use this additional information to calculate the Lyapunov exponent. In the case of a fully deterministic, non sensitive system these three forecasts would give the same value and the overlap would not yield new information and  $\lambda = -\infty$ . But in weather forecasting one has to deal with a chaotic system. Additionally every forecast run starts with new initial conditions. The divergence/convergence of these runs gives a measure of the predictability.

To calculate the Lyapunov exponent the following definitions are made:

- $T$  = number of overlapping forecast
- $\Delta x_0 = 1\text{MW}$  (only for normalization purposes, to get rid of the dimension)
- $\Delta x_n$  = difference between the actual forecast and the additional forecast (the information from the past)

The Lyapunov exponent gives information how the prediction from the past converge/diverge to the actual power prediction. Because each run of the NWP starts with new initial start conditions, that are adapted to the actual weather system, we can extract information how stable the actual weather conditions are.

## 4. Estimation of the standard deviation

Now we have defined five measures to characterise every step in a power prediction. Next we combine these five measures to estimate the standard deviation of the forecast error

### 4.1 Combination of error measure

A combination of the error measures seems reasonable to account for different performance of the error measures at different time steps. The simplest way is multiple linear regression:

$$err = a_1 E + a_2 R + a_3 H_S + a_4 \lambda + a_5 H_J \quad (1.7)$$

The challenge is now to estimate those five coefficients. As a first guess we used least square estimation by solving an over determined system of equations. These five coefficients were used to initialise an optimisation routine for fine tuning of the parameter. This procedure gave the following weighting.

	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
FINO1	32	14	11	39	4
EWE	29	4	39	26	2
Germany	36	1	30	5	28

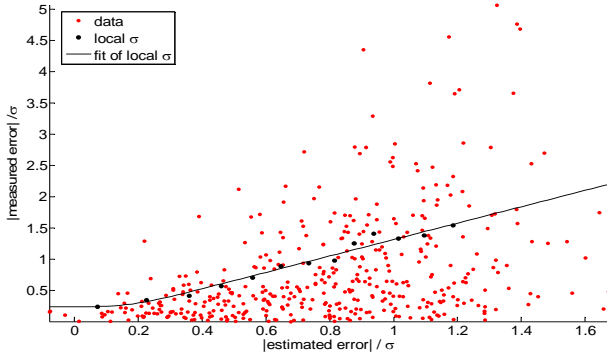
**Table 1: Impact in percent of the single error measure on the estimated forecast error after optimisation was done**

Clearly visible is that the energy has a high weight for all three test cases whereas the gradient measure is of less importance. Interesting is that for every test case the single measure have different influence. Another feature is that the importance of the chaos measure decreases, whereas the size of the power prediction area increases. The next step is now to estimate the standard deviations out of this combined error measure.

#### 4.2 The estimation procedure

The measured and the estimated forecast error for a prediction step of 30 hours is plotted in fig. 1. Visible is the triangular shape of the scatter plot, means that for small errors the estimated error remains small, whereas the spread of the estimated error increases for increasing measured error. To this scatter a moving standard estimation was applied (resulting in the black dots). Next the function of (1.8) was fitted to this local standard deviation to get a continuous description of the standard deviation.

$$\sigma(x) = \frac{a}{b} \left( bx + c - \ln(1 + \exp(-bx + c)) \right) + d \quad (1.8)$$



**Figure 1: fit of the standard deviation of the normalised estimated forecast error to the normalised measured one at a forecast horizon of 30 hours (EWE area)**

Now we have a computational method to estimate the standard deviation of the forecast error of a time series for every time step.

1. take a training time series to estimate the coefficients for (1.7) and (1.8)
2. compute the estimated forecast error
3. calculate the standard deviation out of this estimation

This procedure enables us to estimate the standard deviation to a given deterministic power forecast.

Now we have the std. dev., but we don't know the underlying probability distribution (which is far away from being Gaussian, see also [2]). By considering the physical bounds (the maximal possible power output as an upper bound and zero power as a lower bound) one can also say the distribution gets more skewed if it gets close to these bounds. Thus we need a description of the time evolution of the forecast error distribution.

#### 5. Time evolution of the forecast error distribution

The proper tool to estimate this time evolution is the Fokker Planck equation. The equation describes the time evolution of a probability density function  $p(x, t)$  under the influence of the drift term  $A(t)$  and the diffusion term  $B(t)$ .

$$\partial_t p(x, t) = \partial_x \left( -A(t)p(x, t) + \frac{B(t)}{2} \partial_x p(x, t) \right) \quad (1.9)$$

To solve this equation we also need two boundary conditions. The appropriate boundary conditions for this problem are no flux or reflecting boundaries at  $x = 0$  and  $x = L$ . This means that the probability to find a power prediction outside the physical bounds (the maximum possible produced wind power as upper limit at  $x = L$  and zero produced power as lower limit at  $x = 0$ ) is equal zero.

$$\left[ -A(t)p(x, t) + \frac{B(t)}{2} \partial_x p(x, t) \right]_{x=0, L} = 0 \quad (1.10)$$

At least the initial condition to start the time evolution is needed. Here a probability profile is given.

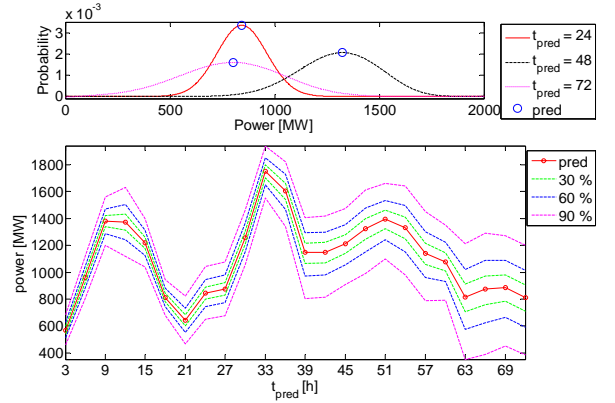
$$p(x, 0) = f(x) \quad (1.11)$$

In numerical experiments we estimated that this initial distribution follows a power law and can be described best by a Cauchy distribution:

$$p(x) = \frac{1}{\pi} \frac{a}{a^2 + (x - \mu)^2} \quad (1.12)$$

where  $a$  characterises the width of the distribution and  $\mu$  the mean.

The drift term  $A(t)$  can be computed from the deterministic power forecast and the diffusion term  $B(t)$  can be estimated from the calculated standard deviation. With all these ingredients the Fokker Planck equation can be integrated. From this final solution we extract the distribution profiles for the corresponding forecast time steps (see fig. 2). From these profiles we can easily compute the appropriate confidence levels (the quantiles) and get by this the uncertainty information. With this information one can choose an advisable visualisation to access at a glance the quality and reliability of the power prediction time series.



**Figure 2: upper picture: probability distribution for different time steps; lower picture: translation of these profiles into a contour plot of the confidence levels**

#### 6. Validation

The evaluation of probabilistic forecasts is more complicated than evaluating deterministic ones. While it is easy to say that a point forecast is false because the deviations between predicted and real values is of practical magnitude, an individual probabilistic forecast cannot be deemed as wrong. To judge probabilistic forecast the relevant measure will be defined.

##### 6.1 Reliability

Necessary to define the reliability is the indicator variable

$I_{t,k}^{(\alpha)}$ , following [3], which is defined as follow

$$I_{t,k}^{(\alpha)} = \begin{cases} 1, & \text{if } p_{t,k} \in [B_{t,k}^{(\alpha)}, B_{t,k}^{(1-\alpha)}] \\ 0, & \text{if } p_{t,k} \notin [B_{t,k}^{(\alpha)}, B_{t,k}^{(1-\alpha)}] \end{cases} \quad (1.13)$$

where  $p_{t+k}$  is the deterministic power prediction at base time  $t$  and forecast step  $k$ , and  $B_{t,k}^{(\alpha)}$  and  $B_{t,k}^{(1-\alpha)}$  are the lower and upper bound of the stated confidence level  $\alpha$ . Therefore the indicator variable tells if the outcome of the actual measurement lies in the prediction interval ('hit') or not ('miss'). To estimate now the reliability all valid hits are counted

$$r_k^{(\alpha)} = -\alpha + \frac{1}{N} \sum_{t=1}^N I_{t,k}^{(\alpha)} \quad (1.14)$$

Consequently we sum up valid hits and normalise with the total number of tries  $N$ . This procedure gives the deviation from the 'perfect prediction'. The comparison allows to summarise the calibration of several quantiles or intervals and thus to see at a glance if a given method tends to systematically underestimate (or overestimate) the uncertainty.

## 6.2 Sharpness

When dealing with sharpness, attention is given to the width of the prediction intervals

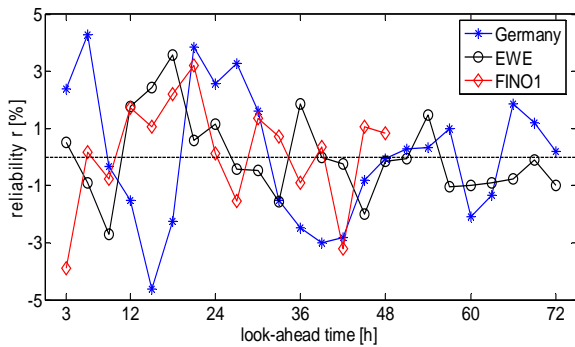
$$\delta_k^{(\alpha)} = \frac{1}{N} \sum_{t=1}^N (B_{t,k}^{(1-\alpha)} - B_{t,k}^{(\alpha)}) \quad (1.15)$$

thus we estimate the average size of the prediction interval for a given horizon  $k$ .

The sharpness is a secondary validation criterion. At first, the forecast must be reliable and second the width of the underlying distribution should be as small as possible. Hence one would prefer an algorithm with a higher sharpness in the case of equal reliability.

## 6.3 Validation of the three test cases

In the following, the three test cases are compared in one figure. To make them comparable the sharpness is normalised by the total installed power. Reliability and Sharpness are computed for five confidence levels (10%, 30%, 50%, 70%, and 90%) but only the mean is compared.

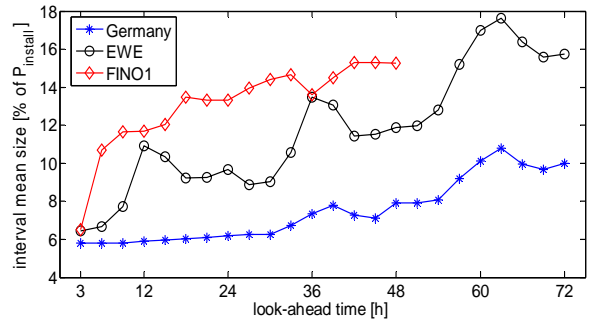


**Figure 3: reliability diagram for the test cases Germany, EWE<sup>1</sup> and FINO1**

<sup>1</sup> EWE is a multi-service energy company, headquartered in the federal state of Lower Saxony. Most customers of EWE are based in Lower Saxony in the Northwest.

In fig. 3 the validation of the reliability is shown. It is visible that for all test cases the deviations from a 'perfect prediction' do not exceed  $\pm 5\%$  (means in only 5% of all samples we could not estimate the distribution right) and that they are for a horizon above 48h not bigger than  $\pm 3\%$ . Hence the proposed method can estimate the underlying error distribution. If one increases the sample size (2-3 years of power time series) we expect a convergence of the statistics. It is also shown that the estimation of the forecast error distribution is independent of the size of the prediction area.

If one compares the sharpness, the average interval size depends on the prediction area. In fig. 4 is shown that for Germany the estimation method offers the narrowest intervals whereas for FINO1 the average width of the distribution is the broadest. Also visible is that the method can adapt to the daily cycle especially to the EWE-region. For the whole area of Germany due to averaging effects this effect is not strongly pronounced. For the offshore test case this adoption is slightly reversed.



**Figure 4: plot of the average interval size normalised by the total installed capacity**

## 7. Conclusion

In this work we showed that a wind power forecast does not only carry the deterministic part but also some probabilistic information. With the help of the proposed method it is possible to extract enough uncertainty information to estimate the underlying probability distribution.

## 8. References

- [1] J. Tambke, L. v. Bremen, G.Kariniotakis et al.: Short-Term Forecasting of Offshore Wind Farm Production – Developments of the Anemos Project. Proceedings of the European Wind Energy Conference EWEC 2006, Athens, Greece
- [2] M.Lange, On the Uncertainty of Wind Power Predictions-Analysis of the Forecast Accuracy and Statistical Distribution of Errors, Trans. ASME, J. Solar Energy Eng., 2005, Vol.127, 177-184
- [3] P.Pinson, G.Kariniotakis, H.Nielsen, T.Nielsen, H.Madsen: Properties of Quantiles and Interval Forecast of Wind Generation and their Evaluation (Proceedings of the European Wind Energy Conference EWEC 2006, Athens, Greece)