Simulation scenarios and prediction intervals in wind power forecasting with the Beta distribution

Álvaro Jaramillo Duque[1], Ismael Sánchez[2], Edgardo Castronuovo[1] and Julio Usaola[1]
[1]Department of Electrical Engineering
[2]Department of Statistic
Carlos III of Madrid University
Campus of Leganés – 28911 Madrid (Spain)
Phone:+34 91 624 9999, Fax:+34 91 624 9430
Email: alvaro.jaramillo@uc3m.es, ismael@est-econ.uc3m.es, castronuovo@ieee.org and jusaola@ing.uc3m.es

Abstract. A methodology for the simulation of the wind power scenario for a short term horizon (one or two days in advance) is proposed. The covariance of the historical errors and the wind power forecast are use to generate a conditional random variable that represent the power wind production as a scenario. With the information provided by the scenario simulation, the energy deviation during a period and the prediction interval for each hour are obtained. The Beta distribution is used to represent the behaviour of the wind power production due to its better performance.

With the results, it is possible to quantify the uncertainty of wind energy production. Finally, comparing the covariance and correlation of the simulated errors with historical errors, the procedure of the methodology is validated.

Key words

Wind power, error analysis, forecast wind production, simulation scenarios, prediction intervals.

1 Introduction

When working with wind energy, one of the most difficult tasks is dealing with the uncertainty associated to the future production. From the technical point of view, all the forecast wind power errors must be compensated to supply the power demanded by the load. In the energy market context, these uncertainties increase the risks that the utilities are exposed to, since the deviations in the production can be penalized economically by the System Operator (SO). The energy cost is then increased, and this is the main reasons for seeking to quantify the uncertainty of the forecast wind power.

The agents use the wind forecasting tools, either meteorological, statistic or a combination of both, to estimate the wind power production for a short time horizon. This information is generally required for SO and must be provided by the utility. Moreover, it is an indispensable tool for the design of market strategies.

The information provided by the prediction tools is the power wind estimation and some levels of reliability associated with the prediction. However, this information is not enough to know the order of possible deviations of the energy. The analysis of the historical data provides valuable information that can be used to estimate the wind production uncertainty.

In [1] the wind power and the load demand predictions is analysed, to achieve the uncertainty and quantify it. The study case is a small island, where the proportion of wind energy is relevant compared with the total load demand. In this work, the error is not related to the prediction horizon and the prediction deviations are calculated in absolute terms. The methodology is validated on typical days, profiling and quantifying the economical deviation when dealing with different scenarios and levels of uncertainty.

An alternative approach for calculating the prediction error is made in [2] using a dependent function of the prediction horizon and the amount of installed power. In [3], an empirical approach is used to calculate the prediction error as a proportion of the total production, based only on the prediction time horizon.

In [4], a methodology that relates the prediction errors at different prediction horizons through a covariance matrix is used. In this work, the wind power production is represented as a Non-Parametric function to simulate the behaviour of the variable in scenarios.

The representation of the wind power prediction as a Beta distribution is proposed in [5], used to calculate the probability function of errors in the prediction. The prediction horizon is used to determine the parameters of the probability distribution.

The motivation of the present work is accurately calculate the possibles deviations of the energy, considering the interdependence among the errors and the wind power at different time horizons. A joint probability distribution will be the ideal solution to achieve the possible energy errors during a future period, but the non-linear and the non-Normality of the wind power production made this task complex. The possibles deviations of the energy and the prediction intervals are necessary to achieve the proper operation of the electrical system, through dimensioning of the storage elements, development of reserves programs and market strategies, amount other operational problems.

The historical forecast errors at different horizons are used to calculate the covariance matrix, and this matrix is use to generate a multidimensional random variable that simulates the prediction errors, maintaining the interdependence structure. A statistics transformations is applied before obtaining the conditioned variable that represent the possible production scenarios for future horizon. Finally, energy deviations and the prediction intervals are calculated.

2 Proposed Methodology

A data set of a real wind farm is use as study case; the information for two consecutive years is available for analysis (this data set is explained with more details in next sections).

With these data, the forecast errors are calculated as the difference between power productions and predictions. The information required for the analysis consists of wind power production, hour by hour, and the forecasts made in the past with the predictive tool for every hour, performed h hours earlier:

$$\hat{e}_{t+h|t} = p_{t+h} - \hat{p}_{t+h|t} \qquad \forall\ t, h$$

$$\varepsilon_t = \left[\hat{e}_{t|t-1}, \hat{e}_{t|t-2}, \cdots, \hat{e}_{t|t-H}\right]^T \qquad \forall\ t, h \tag{1}$$

Where the notation $t+h|t$ refers to the predictions made for hour $t+h$ at time t; and $h = [1,\ 2,\ \ldots,\ H]$.

In (1), errors can be represented as a normal distribution [1, 6] but the conditional errors in the wind power prediction can not be represented in that way [7]. The information of the prediction errors can be concentrated in a covariance matrix Σ_t, as shown in equation 2.

$$\Sigma_t = (\varepsilon_t \varepsilon_t^T) \tag{2}$$

This covariance matrix concentrates all the available information present on the historical errors and the interdependence of them. The covariance matrix for this single case is stationary, but can be update using newer values like is done in [4]. Figure 1(a) represents the covariance matrix calculated with all the historical data. Figure 1(b) represents the correlation matrix of the forecasting errors, enabling to see the linear relationship between each pair of prediction time horizons.

The strategy proposed here uses the simulation of the wind energy production in the short term to obtain the behaviour through the conditional distribution. The simulated scenarios are analysed to find the possible energy deviations during a period and the prediction intervals for each hour.

The simulation scenarios must reproduce the behaviour of the wind power production. Therefore, different Beta distribution are used in each hour, instead of Gaussian distributions, because the variable behaviour using Beta distribution is better represented, [8, 9, 7, 5].

Using a Multivariate Gaussian random number generator a variable of dimension H is created, which simulates the forecasting errors behaviour, $X_h^{(s)}$, (3); with mean zero and a covariances matrix, Σ_t. A set of possibles predictions errors are generated s times, achieving the structure and uncertainty in each random variable.

$$X_h \sim \mathcal{N}_h(\mu_0, \Sigma_t) \tag{3}$$

For simplicity, the prediction errors are assumed stationary of second order and with mean value null. According to the previous assumption, a statistical transformation is applied to $X_h^{(s)}$, from a joint Gaussian distribution to a conditional Beta distribution. This process requires two steps, from the joint Gaussian distribution to the Uniform distribution, and then, from the Uniform distribution to the conditional Beta Distribution 4.

$$Y_h^{(s)} = F(X_h^{(s)} \mid \mu_h, \sigma_h) \qquad \forall\ s, h \tag{4}$$

Where μ_h is a vector of zeros and σ_h is the vector that contains the standard deviation of the prediction errors, $\texttt{std}(\hat{e}_{t|t-h})$. The Uniform variable can be considers as a random seed coming from the simulation errors (joint Gaussian distribution) to generate the conditional Beta Distribution.

The simulated wind power production scenarios for the future horizon, $p_{t+h|t}^{(s)}$, are represented using Beta distributions, obtained through:

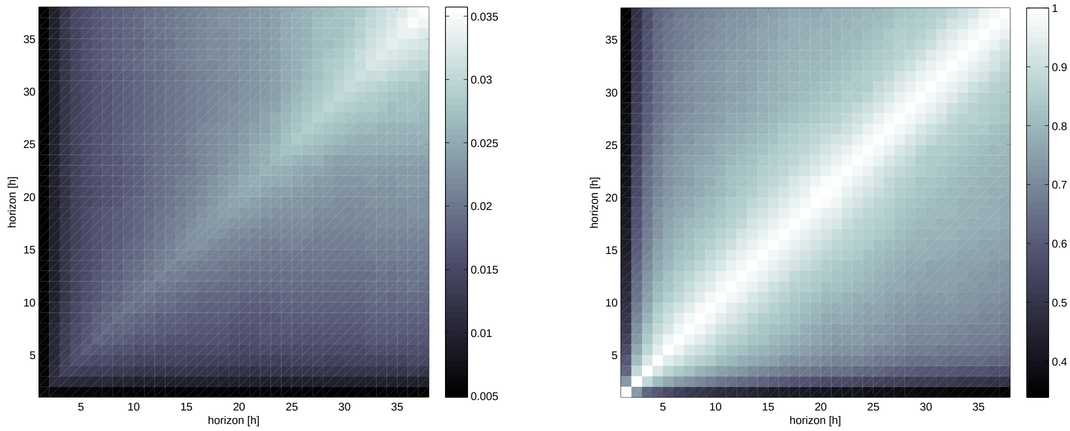
$$p_{t+h|t}^{(s)} = F^{-1}(Y_h^{(s)} \mid \hat{\alpha}_{t,h}, \hat{\beta}_{t,h}) \qquad \forall\ s, h \tag{5}$$

In (5), the parameters of the conditional Beta distribution $\hat{\alpha}_{t,h}$ and $\hat{\beta}_{t,h}$ are calculated in function of $\hat{\mu}_{t,h}$ and $\hat{\sigma}_{t,h}^2$, as it is shown (6) and (7).

$$\hat{\alpha}_{t,h} = \hat{\mu}_{t,h} \left(\frac{\hat{\mu}_{t,h}\,(1-\hat{\mu}_{t,h})}{\hat{\sigma}_{t,h}^2} - 1 \right) \qquad \forall\ h \tag{6}$$

$$\hat{\beta}_{t,h} = (1-\hat{\mu}_{t,h}) \left(\frac{\hat{\mu}_{t,h}\,(1-\hat{\mu}_{t,h})}{\hat{\sigma}_{t,h}^2} - 1 \right) \qquad \forall\ h \tag{7}$$

Where the conditional mean value is considered as the expected wind power production for the near future, $\hat{\mu}_{t,h} = p_{t+h|t}$, and the conditional variance is approximately the variance of historical errors, $\hat{\sigma}_{t,h}^2 = \texttt{Var}(\hat{e}_{t|t-h})$. A conditional variance, more accurate, could also consider the expected production $\texttt{Var}(\hat{e}_{t|t-h}|\hat{p}_{t|t-h})$, which will be included in future releases.



(a) Covariance matrix.

(b) Correlation matrix.

Fig 1: Historical Analysis of the errors.

3 Results

A Study Case

The data set for the study case consists of the production of wind power, hour by hour, and power forecasts for each hour, made h hours earlier, from March 2001 to April 2003. The prediction of wind power was obtained by using the forecasting tool SIPREOLICO [10, 11], developed by the University Carlos III of Madrid for Red Eléctrica de España, the operator of the electrical system. With this data set, the prediction errors and correlation matrix are calculated.

All the data are normalized with the nominal capacity of the wind farm. There are 18,937 power measures and predictions for each hour. However, not all the data set is available because technical interruptions (either of the wind farm or the forecasting tool), so the data was filtered in order to exploit as much information as possible.

For the time t the covariance matrix of the errors, Σ_t, and the wind power prediction, $\hat{p}_{t+h|t}$ are available. The horizon has been chosen as $H = 38$ hours, taken as reference the Spanish regulation, where utility must inform to the SO the estimated production of one day (24 hours) 14 hours before [12]. For the proposed analysis, 10,000 simulated scenarios are simulated to quantify the uncertainty of the prediction. The proposed method is not CPU-time demanding, so for other types of analysis the number of scenarios could be increased.

Based on the simulated scenarios, the possible energy deviations and the prediction intervals are calculated.

B Scenarios Simulation

Figure 2 shows only thirty of the simulated scenarios, $p^{(s)}_{t+h|t}$, and the wind power prediction, $\hat{p}_{t+h|t}$. Each scenario follows a path that inherits the sequence of the errors, which means that each point of the scenario depends on its previous value, the historical errors variance and the wind power prediction.

Figure 3 plots the prediction intervals obtained from all the scenarios simulated, represented in bands; in addition, the wind power forecasting, $\hat{p}_{t+h|t}$ and power production, p_{t+h} are drawn. Each band represents the probability that the wind power production has to be within the range covered by this probability range at future time h. The band with the bigger size corresponds to 90% and within this, is the 80% band, being a little narrower and darker. The scale continues in this way until the 10% band, which is the narrowest and darkest. For a smaller band range, less is the probability for the wind power production to be within it.

The main advantage of the proposed methodology is the simplicity to achieve the simulated scenarios and the ability to quantify the uncertainty in the wind power predictions, taking into account the relationship between the errors at different time horizons and the conditioned behaviour of the wind power prediction.

C Energy Deviations

Due to the uncertainty in the wind power forecast, it is necessary to reserve an amount of energy to cover the energy deviation between programmed and real productions. In some cases, the amount of energy reserved is oversized or calculated on a heuristic way, with the intention of avoiding operation problems. However, these conservative strategies means that resources are not exploited in a sustainable way and in some cases are underutilized.

The advantage of the simulated scenarios is the possibility to calculate the energy deviation, as the integral of the difference between the wind power forecast and the wind simulated scenarios, (8). In the present work, the magnitudes of the deviations and the frequency in which they occur are calculated. The last 24 hours of simulated time horizon are consider, $h_1 = 15$ and $h_2 = 38$ (as mentioned above, using as reference

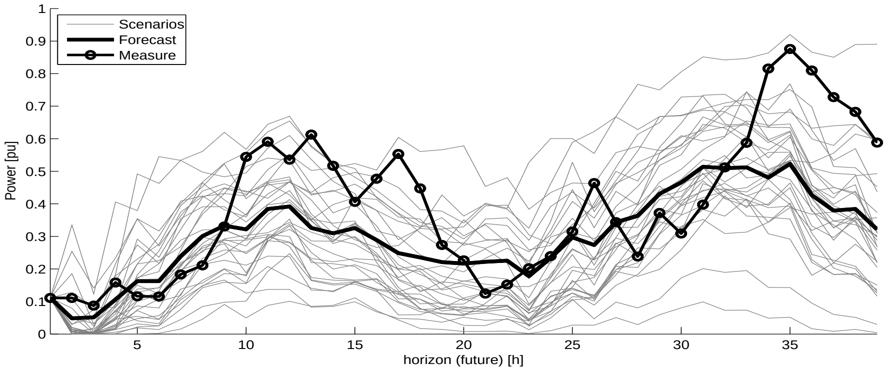


Fig 2: Possibles scenarios and forecast

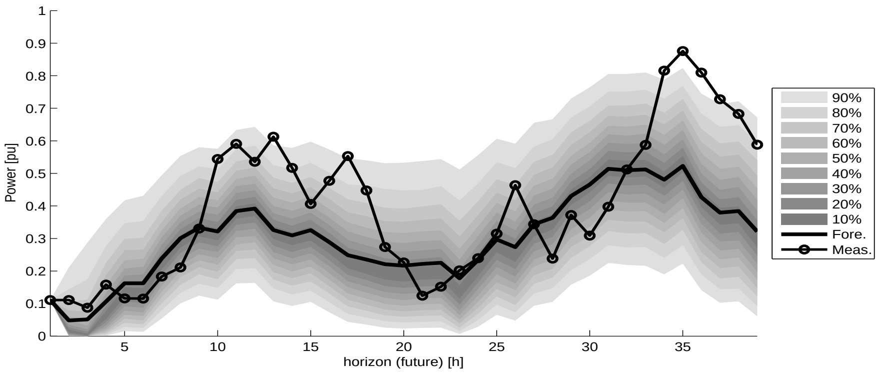


Fig 3: Scenarios prediction intervals

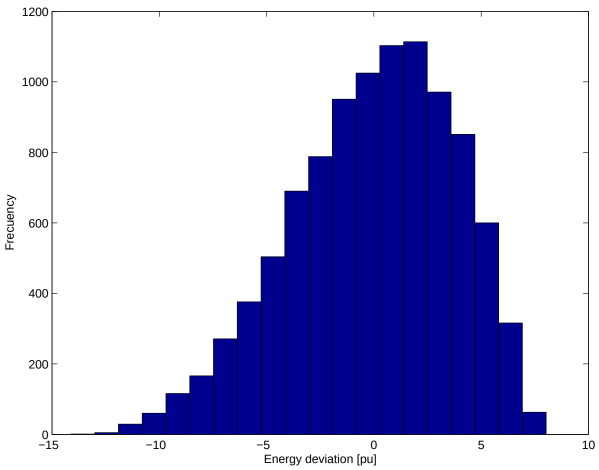


Fig 4: Energy deviation for all scenario, $E^{(s)}$

the Spain regulation). Figure 4 shows the histogram of the energy deviations. With this curve it is possible to chose the amount of energy related with a probability level of errors.

$$E^{(s)} = \int_{h_1}^{h_2} \left(\hat{p}_{t+h|t} - p^{(s)}_{t+h|t} \right) dh \qquad \forall\ s \qquad (8)$$

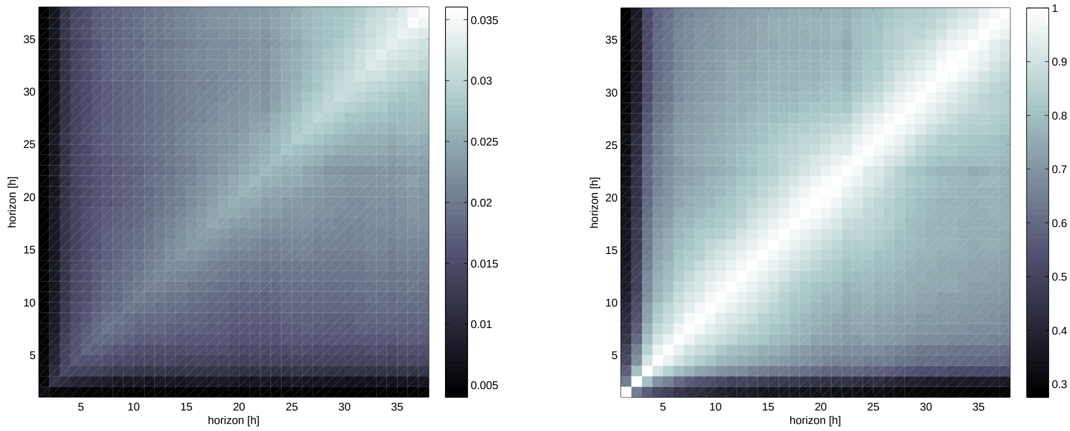
D Monitoring the Errors Structure

As a validation of the method procedure, the possibles errors are calculated as the difference of the simulated scenarios, $p^{(s)}_{t+h|t}$, and the real power production, $\hat{p}_{t+h|t}$. The covariance and correlation matrices of the simulated data are shown in Figures 5(a) and 5(b) respectively. These matrices are similar with the matrices obtained from historical data, Figures 1(a) and 1(b). It is possible to see that both covariance and correlation matrices maintains the same structure and values during the steps in the transformations.

4 Conclusions

The method presented here allows the calculation of the errors in the prediction for the wind power production through the simulated scenarios, using the interdependence of the errors presented in the historical data.

The uncertainty behaviour of the wind power production was replicated with simulation scenarios. The main advantage of the methodology is that each point in the prediction path of each the scenarios depends of the values of the previous points and is conditional for the wind power prediction value, recreating the nature of the original variable. This makes possible to model the non-linear and asymmetrical behaviour of



(a) Covariance matrix.

(b) Correlation matrix.

Fig 5: Analysis of the scenarios simulation errors

the wind power prediction variable.

The calculation of simulated trajectories is very promising because is it possible to obtain the possible energy deviation and the prediction errors intervals, making easy to understand the behaviour of the wind energy production and the relationship with its prediction. The information obtained is useful to calculate the optimal operation of the control centre for renewable energy, to design investment strategies, to calculate the optimal operation of systems with storage capacity and others operations problems.

References

[1] A. Tsikalakis, Y. Katsigiannis, P. Georgilakis, and N. Hatziargyriou, "Determining and exploiting the distribution function of wind power forecasting error for the economic operation of autonomous power systems," in *Power Engineering Society General Meeting, 2006. IEEE*, p. 8 pp., 2006.

[2] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *Power Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 587–595, 2005.

[3] G. Koeppel and M. Korpås, "Increasing the network in-feed accuracy of wind turbines with energy storage devices," in *The Sixht World Energy System Conference*, (Torino, Italy), jul 2006.

[4] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.

[5] H. Bludszuweit, J. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 983–991, 2008.

[6] R. D. Cook and S. Weisberg, *Residuals and Influence in Regression.* 1982.

[7] J. Bjørnar, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 3–11, 2006.

[8] S. Bofinger, A. Luig, and H. G. Beyer, "Qualification of wind power forecasts," in *Proceedings of the Global Wind Power Conference*, 2002.

[9] A. Fabbri, T. GomezSanRoman, J. RivierAbbad, and V. MendezQuezada, "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market," *Power Systems, IEEE Transactions on*, vol. 20, pp. 1440–1446, Aug. 2005.

[10] I. Sánchez, J. Usaola, O. Ravelo, C. Velasco, J. Domínguez, M. G. Lobo, G. González, F. Soto, B. Diaz-Guerra, and M. Alonso, "Sipreólico-a wind power prediction system based on flexible combination of dynamic models. application to the spanish power system," in *World Wind Energy Conference in Berlin, Germany*, 2002.

[11] I. Sánchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.

[12] J. M. Angarita and J. G. Usaola, "Combining hydrogeneration and wind energy biddings and operation on electricity spot markets," *Electric Power Systems Research*, vol. 77, no. 5-6, pp. 393–400, 2007.