Short-Term Wind Power Forecasting Based On Quantile Regression

José F. M. Pessanha, Valk L. O. Castellani, Victor A. Andrade
Departamento de Otimização Energética e Meio Ambiente
Centro de Pesquisas de Energia Elétria Cepel
Av. Horácio de Macedo 354, Ilha do Fundão, Rio de Janeiro, Brazil
francisc@cepel.br, valk@cepel.br, andrade@cepel.br

ABSTRACT

Short-term wind power forecasts are fundamental information for the safe and economic integration of wind farms into an electric power system. In this work we present a Generalized Additive Model to predict the wind power quantiles (Quantile Regression) from which we obtain a prediction of the wind power production probability density function in a wind farm. The methodology was implemented in the VENTOS Program. In order to illustrate the application of the methodology as well as the VENTOS Program this work presents the results achieved by a computational experiment based on real data from a wind farm located in Galicia, Spain.

Keywords: *WInd power, Quantile Regression, Probabilistic Forecasting.*

INTRODUCTION

In order to enable the integration of wind power into the power systems, operation centers require new tools for electric power management, among which the wind power forecast [1]. In general, forecasting methods provide the prediction of the expected value of the wind power production without, however, accurately quantifying and reporting forecast uncertainty. Because of the intermittent wind speed characteristics, the expected value of wind power production without adequate uncertainty quantification does not fully meet the needs of the power system operation. For example, in power systems with a large share of wind power sources, the operating reserve should be dimensioned to cover uncertainties in wind power and load forecasting [2].

The probability density contains all the information of a random variable, so to deal with the uncertainty, we must consider the predictions of probability density of wind power production provided by probabilistic forecasting methods [3], for example, the quantile regression [4], the non-parametric density estimation [5] and the Gaussian process [6].

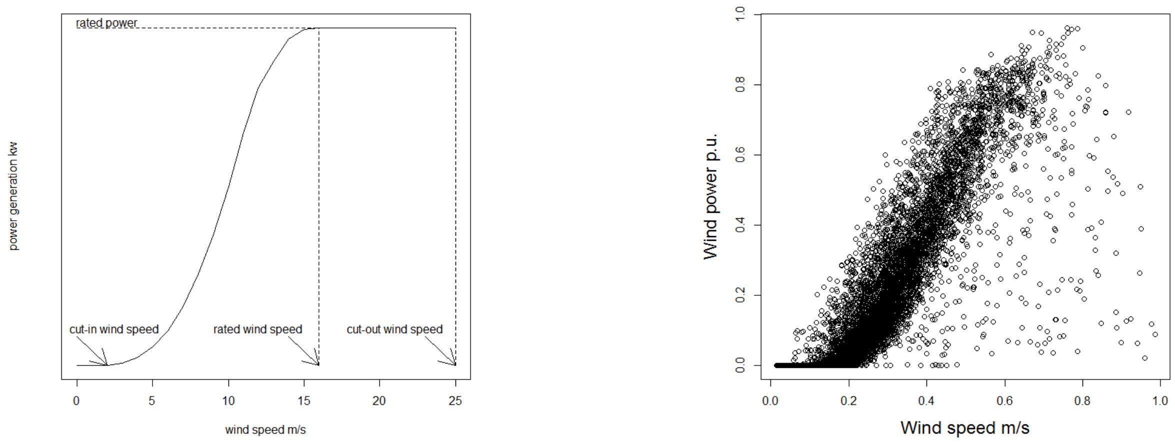
This article presents the general guidelines of a probabilistic forecasting methodology of the wind power production, useful in unit commitment scheduling and real-time operation. The methodology is based on the quantile regression [4] specified as a generalized additive model [7]. The option for this approach lies in its relative ease of computational implementation. The forecast methodology was implemented in C ++ and is available in the VENTOS Program, the wind power forecasting module that integrates the ENCAD Energy Modeling System developed by CEPEL.

In its first version, the VENTOS program provides punctual forecasts (medians) of the wind power production, an input data for the unit commitment optimization carried out by DESSEM model. The methodology is illustrated through the results of a computational experiment with real data coming from a wind farm located in Galicia, Spain.

WIND POWER

Under ideal conditions, the relationship between wind speed v and energy production P is defined by the power curve $P(v)$ shown in Figure 1a. The energy production starts at wind speeds between 2 and 3 m/s (cut-in wind speed). Then, energy production increases rapidly with increasing wind speed until rated power is reached, when the speed reaches a value between 12 and 17 m/s (rated wind speed). For speeds above 25 m/s (cut-out wind speed), the wind turbine must be turned off to avoid risk of damage to the rotor [8]. Thus, a simple way to compute wind power production forecasting is to predict wind speed and find the corresponding power production in the power curve. However, as illustrated by the scatter plot in Figure 1b for a given wind speed, the power production may present different values. Therefore, the relationship between wind speed and power production is stochastic and a punctual forecasting may not be sufficient.

The uncertainty can be quantified by predicting the probability distribution (probabilistic forecasting) of the wind power production in the forecasting horizon. The probability density function contains all the information of a random variable and the quantiles are a simple way of describing it. The quantiles can be achieved by quantile regression models [4,9].



(a) Theoretical power curve (b) Empirical power curve (scatterplot)

Figure 1 – Power Curve.

QUANTILE REGRESSION MODEL

In a linear regression model, the fitted regression equation estimates the expected value of the response variable *y* as a function of the explanatory variables *X*. In a quantile regression model, the regression equation estimates the quantile τ (τ = 0.5 for the median) of the response variable as a function of the explanatory variables. In the quantile regression the estimation procedure seeks to minimize the objective function in the Linear Programming Problem (LPP) in Table 1, whose solution provides the estimates of the *K*+1 regression coefficients β. For τ=0,5 the LPP (1) correspond to the Least Absolute Deviations (LAD).

Table 1 Quantile regression estimation

Linear Programming Problem (LPP)	Iteratively Reweighted Least Squares (IRLS)
$\underset{u,v,\beta}{Min} \quad \tau\sum_{i=1}^{n} u_i + (1-\tau)\sum_{i=1}^{n} v_i$ s.t. $y_i - X_i^T\beta = u_i - v_i \forall i = 1, n$ $u_i \geq 0 \,\forall i = 1, n$ observations $v_i \geq 0 \,\forall i = 1, n$ observations $\beta \in R^{K+1}$	Step 1) Let h=0 (iteration count), $\beta_h = 1_{K+1}$ and $\Omega = I_{(K+1)(K+1)}$ Step 2) Let h=h+1 and compute $\beta_h = \left(X^T\Omega X\right)^{-1} X^T\Omega Y$ Step 3) Stop if the convergence criterion is reached Step 4) Compute the residuals $\hat{\varepsilon} = Y - X\beta_h$ and weight vector w If $\hat{\varepsilon}_i < 0$ then $w_i = \tau \cdot \hat{\varepsilon}_i$ else $w_i = (1-\tau)\cdot\hat{\varepsilon}_i \;\forall$ *i=1,n* observations Step 5) Let $\Omega_{ii} = 1/w_i$ and return to Step 2.

Alternatively, the quantile regression equations can be fitted by the Iteratively Reweighted Least Squares (IRLS) shown in Table 1. The set of regression equations fitted for different values of τ allows us to characterize the probability distribution of the response variable conditioned to the values of the explanatory variables.

In general, wind speed forecasts are reported in the form of polar coordinates: speed module (V) and direction (θ). The direction is an angle, a circular random variable [10], and for this reason it is more convenient to work with cartesian coordinates of wind speed: $u=V\cos(\theta)$ and $w=V\sin(\theta)$.

In the VENTOS program, the quantile regression model has the same specification proposed in [4] where the wind power production depends on the wind speed cartesian coordinates (u, w). The quantiles estimated by the set of regression equations describe the conditional density probability function $f(P|u,w)$. In order to capture the nonlinearities in the relations between the power production and the wind speed, the regression equation include as explanatory variables the coefficients of the cartesian components u and w in ten B-Splines basis totaling 20 explanatory variables [4]. The resulting model corresponds to the following Generalized Additive Model (GAM) [7] for a given quantile τ:

$$\hat{Q}_t(\tau)=\beta_0(\tau)+\sum_{j=1}^{10} b_j\left(u_t^*\right)\beta_j^u(\tau)+\sum_{j=1}^{10} b_j\left(w_t^*\right)\beta_j^w(\tau)+\varepsilon_t \tag{2}$$

where ε_t denotes a random error, $\beta^u(\tau)$ and $\beta^w(\tau)$ denote the regression coefficients for the quantile τ and $b(u^*)$ and $b(w^*)$ represent the explanatory variables, in this case, the coefficients of the wind speed cartesian components in the B-Splines basis.

The model in (2) should be fitted separately for each quantile τ, for example, for each τ in the set {5%, 10%, 25%, 50%, 75%, 90%, 95%}.

The power production is confined to the interval between zero and the wind farm capacity. However, the procedures to fit the regression equations admits that the response variable is unconstrained. Thus, the application of the estimation procedure directly to the power production data can result in negative quantile or quantile above the wind farm capacity. In order to avoid these situations, the power production data y is previously transformed by the hyperbolic arc tangent function:

$$y_t^{**} = 0.5\log[(1+y_t^*)/(1-y_t^*)] \tag{3}$$

where $y_t^* = (y_t - \min(y))/(\max(y)-\min(y))$ is the normalized value of the power production.

The prediction of a power production quantile in MW (Q) is obtained by applying the hyperbolic tangent function to the estimated quantile (*q*) from regression equations and scaling the resulting value *q** to the wind power production range:

$$q^* = [\exp(q)-\exp(-q)]/[\exp(q)+\exp(-q)] \tag{4}$$

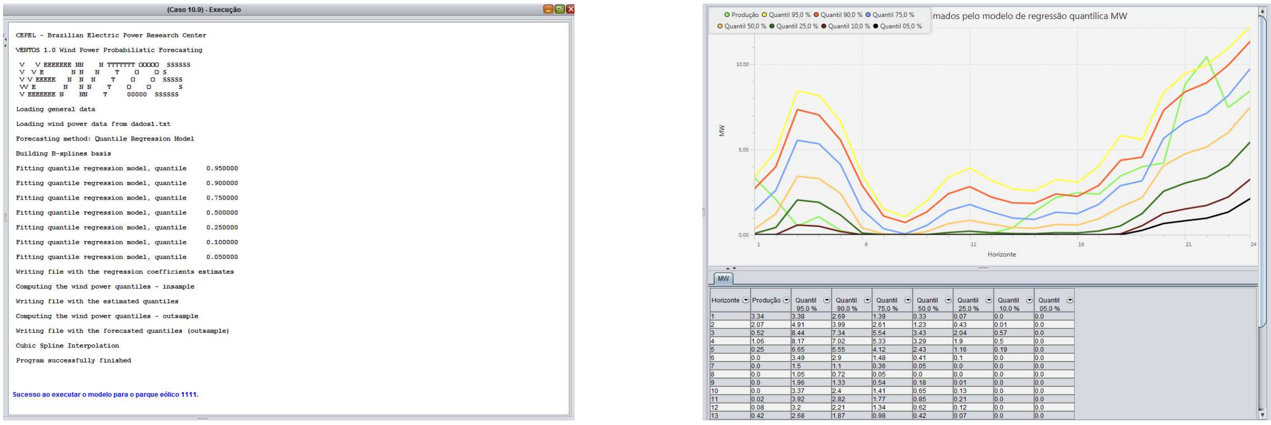
$$Q = (\max(y)-\min(y))\cdot q^* + \min(y) \tag{5}$$

VENTOS PROGRAM

The VENTOS program is the module responsible for providing wind power production forecasts to the ENCAD system developed by Cepel. In the current version the program provides forecasts with hourly and half-hourly temporal resolution. Basically, the program has two execution modes: fitting models and prediction. In both cases, the wind power data (power production and wind speed/direction) must be provided on hourly basis.

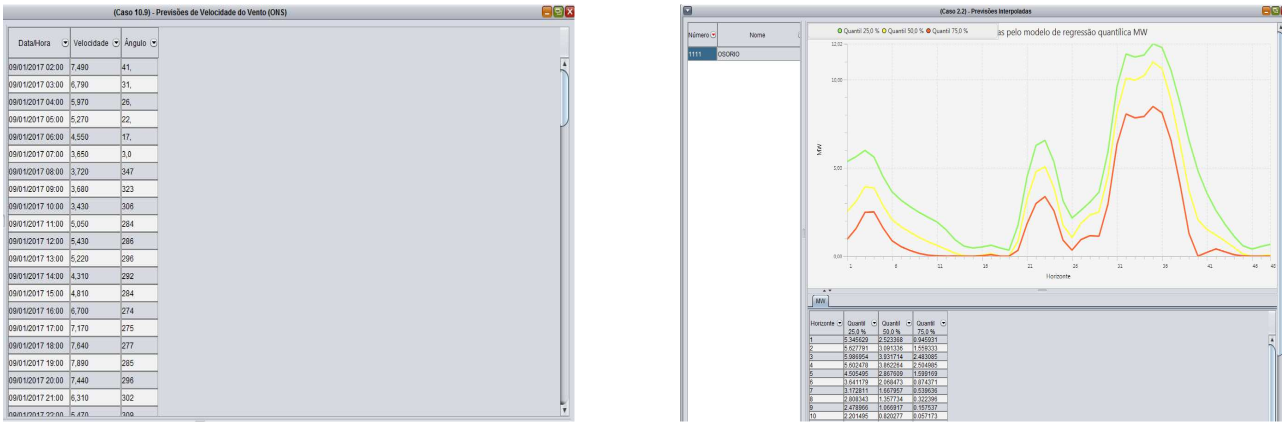
In the fitting mode (Figure 2), the quantile regression equations are fitted on the same wind power data for each quantile $\tau \in$ {5%, 10%, 25%, 50%, 75%, 90%, 95%}. The equations are fitted separately and the set of fitted equations can provide predictions for the quantiles of the power production as a function of wind speed/direction. The dataset required to fit the quantile regression equations includes the past observed values of the power production and the wind speed/direction, both in hourly basis. In addition, we can consider the wind speed/direction from past weather forecasts instead of their observed values.

In the prediction mode (Figure 3), the previously fitted regression equations can be applied to the wind speed/direction forecasts from a NWP model (Numerical Weather Prediction) in order to compute the quantiles forecasts of the power production. For example, the VENTOS Program can predict the quantiles of the wind power production on hourly basis up to five days ahead, based on the hourly weather forecasts available by the National System Operator (ONS) at http://www.ons.org.br/operacao/previsao_horaria_vento.aspx. In addition, the VENTOS Program apply cubic splines interpolation [11] to the hourly wind power generation forecasts in order to achieve half-hourly forecasts.



(a) Fitting quantile regression equations (b) Observed values and quantiles predictions

Figure 2 – VENTOS Program (fitting mode).



(a) wind speed/direction forecasts from http://www.ons.org.br/operacao/previsao_horaria_vento.aspx

(b) half-hourly power production forecasting

Figure 3 – VENTOS Program (prediction mode).

COMPUTATIONAL EXPERIMENT

In order to illustrate the prediction methodology adopted in the VENTOS program a computational experiment was carried out with data from a wind farm located in Galicia, Spain. The dataset available cover the period from January 2014 to January 2015 and include five hourly time series: predicted and observed wind speed/direction and observed power production. The wind farm has a installed capacity of 17.56 MW distributed in 24 wind turbines of 5 technologies and 9 different manufacturers. The diversity of technologies and the complex terrain

[12,13] explain the large dispersion of the observations shown in Figure 1b. Table 2 shows some statistics from the hourly averages of the wind speed and power production. The statistics show a low capacity factor of the wind farm (21%) and a great variability of power production (coefficient of variation equal to 1,14).

Table 2 Wind power statistics

Statistics	Wind Speed m/s	Power Production MWh/h
Maximum	24.6	16.37
Median	6.3	1.80
Minimum	0	0
Average	7.1	3.43
Standard Deviation	3.8	3.92

The probability density function of the power production at each hour in the forecasting horizon was characterized by the following quantiles: 5%, 10%, 25%, 50%, 75%, 90% and 95%. The regression equations were fitted with hourly data over the period between 1 January 2014 and 9 November 2014 (insample period).

The observations for the period from 10 November 2014 until 14 November 2014 constitute the outsample set. The power production forecasts were obtained by applying the fitted regression equations to the forecasts from a NWP model.

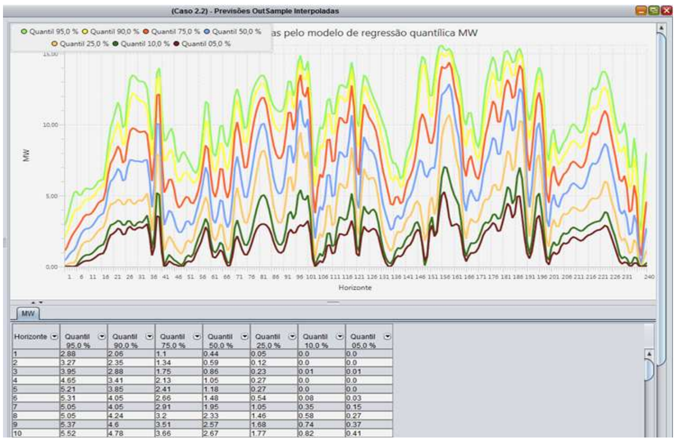
The forecasting horizon is 24 hours ahead with hourly resolution and the wind speed/direction forecasts from NWP were updated only once time a day.

Table 2 shows the root mean square error (RMSE) and the mean absolute deviation (MAD) based on the deviations between the observed hourly aggregated wind power output and the predicted medians ($\tau = 0.5$). In addition, Table 2 shows the actual frequencies by which observed power production is below a given quantile (Reliability scores).

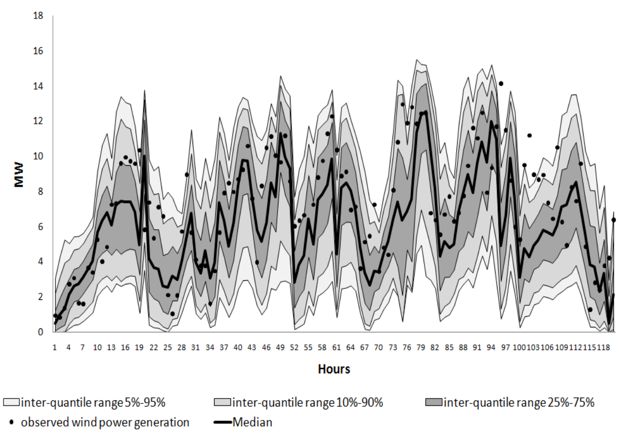
The main result produced by the VENTOS program are spreadsheets and plots containing the hourly and half-hourly quantile forecast of the wind power production (Figure 4a). The spreadsheets can be easily exported to the MS Excel for the purpose of building other reports and charts as shown in Figure 4b.

Table 2 Goodness of fitting statistics

Goodness of fit statistics	Insample	Outsample
MAD	1.49 MW	1,98 MW
RMSE	2.25 MW	2,60 MW
Reliability 95%	94.89%	96.87%
Reliability 90%	89.88%	91.67%
Reliability 75%	74.92%	63.54%
Reliability 50%	50.00%	37.50%
Reliability 25%	25.07%	27.08%
Reliability 10%	10.13%	11.46%
Reliability 5%	5.15%	8.33%



(a) plot produced by the VENTOS Program



(b) plot Excel based on results from VENTOS Program

Figure 4 – Quantiles forecasts for the half-hourly power production.

CONCLUSIONS

This work presents a methodology for probabilistic forecasting wind power production based on the quantile regression. The regression equations are specified as a generalized additive model (GAM) and provide quantiles forecasts of the wind power production. Thus, this approach provides forecasts of the wind power production probability density function beyond of the forecasts of expected value (median). This approach is ideal to handle with the wind intermittency and uncertainty in wind speed forecasts. The methodology was implemented in the VENTOS Program and the good results from computational experiments show its potential. In spite of the satisfactory results, further investigations must be carried out, particularly with data from the Brazilian wind farms.

REFERENCES

[1] Estanqueiro, A., Castro, R., Flores, P., Ricardo, J., Pinto, M., Rodrigues, R., Peças Lopes, J. (2008). How to prepare a power system for 15% wind energy penetration: the Portuguese Case Study. Wind Energy. v. 11, 75-84.

[2] Matos, M.A., Bessa, R.J. (2011). Setting the Operating Reserve Using Probabilistic Wind Power Forecasts. IEEE Transactions on Power Systems. 2, May, vol. 26, 594-602.

[3] Zhang, Y., Wang, J., Wang, X. (2014). Review on probabilistic forecasting of wind power generation. Renewable and Sustainable Energy Reviews. vol. 32, 255-370.

[4] Nielsen, H.A., Madsen H., Nielsen, T.S. (2006). Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy. April, vol. 9, 95-108.

[5] Bessa, R.J., Mendes, J., Miranda, V., Botterud, A., Wang, J., Zhou, Z. (2011) Quantile-copula density forecast for wind power uncertainty modeling, IEEE Trondheim PowerTech, Norway, Jun

[6] Chen, N,. Qian, Z., Nabney, I.T., Meng, X. (2014) Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction, IEEE Trasactions on Power Systems, v. 29, n. 2, March, pp. 656-665.

[7] Wood, S.N. (2005). Generalized Additive Models: an introduction with R. Chapman & Hall.

[8] Custódio, R.S. (2009). Energia eólica para produção de energia elétrica. Energia eólica para produção de energia elétrica. Rio de Janeiro: Eletrobras.

[9] Anastasiades, G., McSharry, P. (2013). Quantile Forecasting of Wind Power Using Variability Indices. Energies, vol. 6, 662-695.

[10] Pewsey, A., Neuhauser, M., Ruxton, G.D. (2013) Circular Statistics in R, Oxford.

[11] Dubinsky, A., Elperin, T. (1997). A method for calculating a load curve using average values of load over time intervals. Electrical Power & Energy Systems. vol. 19, 6, 393-395.

[12] Marti, I., Kariniotakis, G., Pinson, P., Sanchez, I., Nielsen, T.S., Madsen, H., Giebel, G., Usaola, J., Palomares, A. M., Brownsword, R., Tambke, J., Focken, U., Lange, M., Sideratos, G., Descombes, G. (2006) Evaluation of Advanced Wind Power Forecasting Models – Results of the Anemos Project, European Wind Energy Conference, Athènes, Greece, February.

[13] Pinson, P. (2006) Estimation of the uncertainty in wind power forecasting, These, Ecole des Mines de Paris.

BIOGRAPHIES

José Francisco Moreira Pessanha – Rio de Janeiro, 1971. B.Sc. in Statistics from Ence (1992), B.Sc. in Electrical Engineering from Uerj (1994), M.Sc. (1999) and D.Sc. (2006) in Electrical Engineering from Coppe and PUC-Rio respectively. He is professor at Rio de Janeiro State University and Researcher at Brazilian Electric Power Research Center (Eletrobras Cepel).

Valk Luiz de Oliveira Castellani – Volta Redonda (RJ), 1975. Technologist in Data Processing from FAA (1998), Post Graduated in Database Technology (2001) and Computer System Engineering (2002) from Estácio, M.Sc. in Computer Science from UFRJ. He is Researcher at Brazilian Electric Power Research Center (Eletrobras Cepel).

Victor Almeida Andrade – Rio de Janeiro, 1987. B.Sc. Electrical Engineering from UFRJ (2013) and M.Sc. (2017) in Electrical Engineering – Enphasis on Artificial Intelligence – from COPPE/UFRJ. He is Researcher at Brazilian Electric Power Research Center (Eletrobras Cepel).

.