Short-term wind power prediction based on AR models

Mario J. Durán, Daniel Cros and Jesus Riquelme
Electrical Engineering Department
Universidad de Sevilla

***Abstract*— The wind power penetration increase and the trend for the wind farms to enter the market, makes necessary the development of new prediction tools. Prediction tools have been fully verified and used for demand forecast and, more recently, to predict the market prices. Among the different prediction methods proposed, the initial and more deeply verified were the statistical ones. This is the reason for considering, at a first stage, time series statistical methods. In the present work, the use of AR models which include information regarding historical wind power data is explored and the improvement over persistent model is assessed. It is shown, for different wind farms, comparative results for different models, chosen data and for the 6, 12 and 24 hours forecast which can be useful for the daily and hourly markets.**

***Index Terms*— Wind power, short-term prediction, time series analysis, AR model.**

NOMENCLATURE

B backshift operator

Φ Order of AR model

ϕ Polynomial function for the predicted variable

Θ Error term order in ARMA model

θ Polynomial function for the error term

P_t Wind farm generated power at time t

E_{tot} Mean error for all predictions

E Mean error for a prediction

ϵ Absolute hourly error for a certain time

t_{hor} Time horizon in hours

t_d Delay in hours from calculation to prediction

t First hour after prediction calculus

t_{ei} Estimating period initial time

t_{ef} Estimating period final time

t_{vi} Verification period initial time

t_{vf} Verification period final time

$WF1...WF4$ Wind farms considered in the study

SUBSCRIPTS

$real$ value taken from real wind farm data

max Maximum value in a time series

min Minimum value in a time series

SUPERSCRIPTS

$\wedge$ Estimated value

Per result of the persistent model

AR result from the AR model

I. INTRODUCTION

WIND energy has become a mature technology and has spectacularly increased in the last years [1]. This has been both because of technology improvements and due to ambitious European Union objectives motivated by the Kyoto protocol requirements. Only in Spain, the installed wind power for 2011 is aimed to be 13.000 MW, so that it will reach 11 % of the total electric energy production. This number has already exceeded in other countries such as Denmark or some parts of Germany [2]. Wind farms are also attractive, apart from environmental issues, due to low initial cost and very short time between site selection and operation.

Nevertheless, the wind power production increase, the competitive deregulated markets and the mature technology, has made the requirements to increase. The idea about the wind energy as a small part of the power system not contributing to its operation and out of the market, is gradually changing. In the Spanish market, the 2004 updated law allows the wind farm owners to freely sell their energy in the market, but assuming the same conditions as the conventional energy producers. Consequently, they have to provide complementary services and are allowed to participate in the secondary reserve. Furthermore, even if the wind farms are out of the market, their owners have to pay for deviations over 30 %.

So, the uncertainty of the wind power production affects not only the system operator, but also the wind farms operators (market, maintenance). The need for an accurate estimate of the next day wind power production is so becoming more and more urgent.

Among the different approaches for the wind power prediction, two main groups can be considered. On the one hand, there are methods that carry out a wind speed prediction in a first stage and, afterwards, estimate the generated energy using a wind farm power curve. These group of techniques are mainly based on physical considerations. Prediktor, developed in Risø [1], or Previento, developed in the University of Oldenburg, are representative of this methods. They are based on numerical weather predictions (NWP), that provide a grid data for a certain region, and then the local wind is estimated using several software methods (WAsP for Prediktor, Forewind for Ewind, etc.). Once the wind is calculated for the wind farm location, the power curve is used to estimate the generated power. Although some statistical treatment is included in this methods, they are mainly based on physical considerations and so they need much more detailed information than other approaches.

On the other hand, there are methods that carry out the estimation by using a statistical methods and real time wind farm data. The WPPT, developed in the Denmark technical University (DTU), is representative of this approach and uses some selected wind farms and weather reports from the HIRLAM model (in the same way as Prediktor).

Although there are differences in the quantities, it is generally accepted that statistical methods are more effective in the short term while the methods including more physical information are more accurate for a longer period. Physical models main disadvantages are the need for detailed information and the inaccuracy of the real data fitting the power curve for different conditions [2].

The objective of the paper is to use statistical models, already successful for the load, demand and prices forecast [3], [4], [5] to predict the power generated for the case of four Spanish wind farms located in a nearby region. The time horizon used for all the different cases is 6, 12 and 24 hours so that the estimation can be useful for the market pool. The models are based on wind power time series [6].

In section 2 the initial models that consider only the wind power time series (AR) are compared showing the different absolute, maximum and minimum errors depending on the model order, selected data and type of model. Section 3 considers a fixed order and amount of data to build the model extracted from section 2 and shows comparative results. All results are compared to the persistent (Naive-1) model which is a benchmark for short-term prediction. Finally, section 4 discuss the more relevant conclusions.

II. Models based on wind power time series

In this section both AR models are built and the results are extracted and compared with the persistent model.

The general expression for the Auto Regressive Moving Average (ARMA) models is [6]:

$$\phi(B)P_t = \theta(B)\epsilon_t \tag{1}$$

being P_t the wind farm power at time t, $\phi(B)$ and $\theta(B)$ functions of the backshift operator B, $B^l P_t = P_{t-l}$, and ϵ_t the error term. Functions $\phi(B)$ and $\theta(B)$ contain factors of polynomial functions of the form $\phi(B) = 1 - \sum_{l=1}^{\Phi} \phi_l B^l$ and $\theta(B) = 1 - \sum_{l=1}^{\Theta} \theta_l B^l$ where ϕ_l and θ_l are the ARMA model coefficients and Φ and Θ are the model orders that are determined while building the model.

If only the power terms are considered, then it becomes just an AR model and so it predicts the power at time t based on the past power data up to $t-\Phi$ and the model coefficients ϕ_l. Some of the coefficients can be set to zero if there is clue about periodicity in the time series data [2], but this is not the case for the present application, so all the coefficients up to Φ are considered.

There are some variables to select before obtaining the results from a model. Regarding the kind of prediction, it is necessary to set:

- Wind farm (WF1, WF2,WF3 or WF4)
- Time horizon t_{hor}
- Delay in the estimation t_d

With respect to the chosen model, it is also necessary to fix the following parameters:

- Type of model (AR, ARMA, etc)
- Model order Φ,Θ (1-∞)

And if the type of selected data to carry out the forecast is concerned, the following parameters should be established:

- Training data period $[t_{ei}, t_{ef}]$
- Verification data period $[t_{vi}, t_{vf}]$
- Estimation data

The training data period is the amount of data used to determine the model coefficients ϕ_l (e.g. September 2002 to September 2003), while the verification data period is used to compare the estimated with the real power to evaluate the accuracy of the method.

The verification period is set to a month (September 2003 to October 2004) for all the next cases, while the model order, the training data period and the type of data for estimation of an AR models are discussed in the following sections.

The time horizon t_{hor} is the amount of hours to be predicted, and it is set to 24 for the daily market and 6 and 12 for the intradaily market. The delay in the estimation t_d is included to take into account that the bid is not done for the next hour, but after a certain time t_d. A delay of 12 hours is considered for the daily market (t_{hor}=24 h), and 2 hours for the intradaily market (t_{hor}=6 or 12 h).

The estimation data choice refers to the fact that you can consider only past real data (up to the time you perform the prediction) or consider real and estimated data for future time.

Previous to get the results, some of the above mentioned variables have to be established.

For comparative purposes, three errors are considered:

- Hourly error inside a prediction $\Rightarrow \varepsilon(t) = P(t) - \hat{P}(t)$
- Mean error for a prediction $\Rightarrow$

$$E(t) = \frac{\frac{\sum_{i=t+t_d}^{t+t_d+t_{hor}} \left(P(i) - \hat{P}(i)\right)}{t_{hor}}}{\frac{\sum_{i=t_{vi}}^{t_{vf}} P(i)}{t_{vf}-t_{vi}}} \times 100 \tag{2}$$

- Mean error for all predictions in the verification period $\Rightarrow$

$$E_{tot} = \frac{\sum_{i=t_{vi}}^{t_{vf}-t_{hor}-t_d} E(i)}{t_{vf} - t_{hor} - t_d - t_{vi}}$$

Relative errors have been avoided due to periods when the wind farm is not producing any power, and so the error for a prediction $E(t)$ is the absolute error rated with the mean power for the whole verification period. The mean error for all predictions E_{tot} provides an idea of the accuracy as a percentage during the whole verification period, and so it is chosen to compare different possibilities.

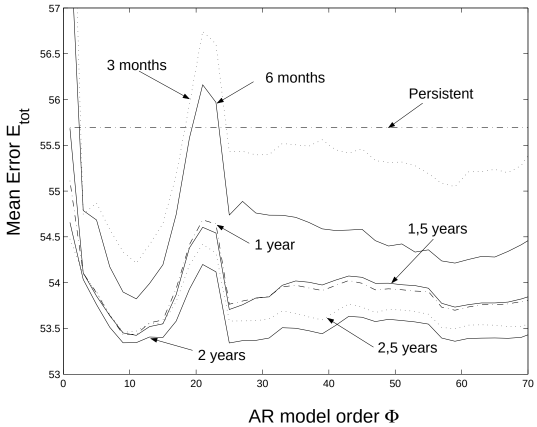

Fig. 1. Mean error for different AR model orders and training periods

II-A. Order of the model

The wind power time series data are taken with an interval of 1 hour. Before building the model, it is necessary to choose how many hours backwards are going to be considered.

In order to select the proper order of the AR model, some tests have been carried out, showing that the optimum order is fairly independent of the training period, as it is shown in Fig.1 for the wind farm WF1, using real and estimated data and a time horizon of 6 hours. Furthermore, the error is not very sensitive to small changes in the order model around the optimum order that provides minimum error.

Nevertheless, the optimum order depends on the wind warm and the type of estimation (daily or intradaily), so different orders have been considered for the different cases. The selected order also changes with the instant when the prediction is being carried out t, so this parameter has to be updated to obtain optimum prediction.

Low model order leads to inaccurate results and there is a minimum error which in this case is achieved for an order of 11. Although this case is just one of the possibilities, different cases have been considered regarding the time horizon, the wind farm, the training data period, etc. showing a similar behavior with some variations in the model order that provides the minimum error. So, for the next results the model order will be set to the optimum value for all the cases. It should be pointed out that this order depends on the wind farm but is independent of the time horizon or time delay.

II-B. Training period

The data set for the AR model coefficients also influences the accuracy of the estimation as it is shown in Fig.1

If the chosen period is too long, the data can not be representative if there has been any kind of change in the wind farm, and so too old data is not so valuable. On the other hand, if the period is too short, the method is not able to extract the proper information while building the model. The results for the different wind farms and time horizons are summarized in table I considering the following training periods:

a) A year (September 2002 to September 2003)
b) 6 months (March 2003 to September 2003)
c) 3 months (July 2003 to September 2003)

Results with time periods longer than a year provided similar accuracy to the case of a year and so saturation is reached. For training period longer than 2 years, though quite similar, the error was slightly higher, due to some too old data.

TABLE I

INFLUENCE OF THE TRAINING PERIOD ON THE ESTIMATION ACCURACY

	$t_{hor}=6h$			$t_{hor}=12h$			$t_{hor}=24h$		
	a	b	c	a	b	c	a	b	c
WF1	53.4	53.8	54.2	59.3	59.8	60.4	78.2	78.2	78.6
WF2	37.1	37.1	37.2	42.0	42.0	41.9	55.6	56.0	56.0
WF3	48.3	48.3	48.4	58.8	58.8	58.8	86.3	86.2	86.2
WF4	38.1	38.2	38.3	47.3	47.3	47.6	74.2	74.4	74.4

Among the three possibilities, it is noticeable that the highest error is obtained if the period is too short (3 and 6 months), while the best results are obtained for the case of a year in most of the cases, and so this is the chosen period for the next results.

II-C. Estimation data

Once both the model order and the training data have been chosen, it is time to explore the adequacy of the different possibilities for the estimation data as the input for the AR model. If the prediction starts at time $t+t_d$, with a time horizon t_{hor}, three approaches can be considered:

a) Begin with real data always starting at $t-1$
b) Begin with the last estimated value

The two approaches are shown in Fig.2 showing also the training and verification periods.

In the third possibility real data and estimated data are mixed in the estimation process. Table II summarizes the resulting errors for different time horizons.

In all the cases the third approach yields more accurate results, and so this type of data is chosen for the next results.

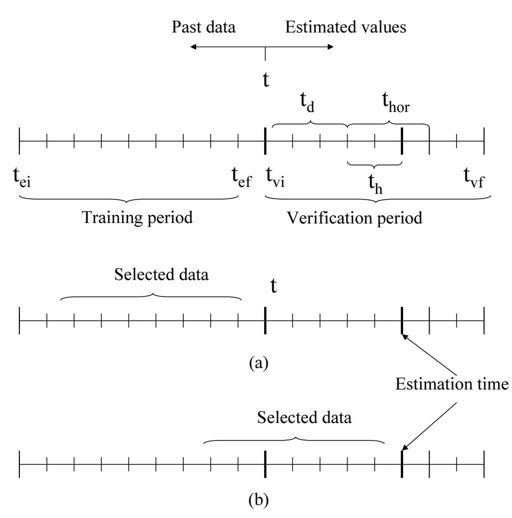

Fig. 2. Estimation data possibilities

TABLE II

INFLUENCE OF THE TRAINING PERIOD ON THE ESTIMATION ACCURACY

	$t_{hor} = 6h$		$t_{hor} = 12h$		$t_{hor} = 24h$	
	a	b	a	b	a	b
WF1	54.1	53.4	60.4	59.3	85.6	78.2
WF2	38.7	37.1	45.4	42.0	63.9	55.6
WF3	48.8	48.3	60.7	58.8	95.4	86.3
WF4	39.0	38.1	50.0	47.3	87.1	74.2

III. AR MODEL RESULTS

In order to show the model accuracy, the results using the previously mentioned characteristics, are compared with the persistent model which is the benchmark for short-term prediction. The comparison is carried out calculating, apart from the mean error E_{tot}, the standard deviation σ and the maximum E_{max} and minimum E_{min} errors both for the persistent and AR model. The results are summarized in table III for the three time horizons and the four wind farms considered in this study.

TABLE III

ESTIMATION ACCURACY FOR PERSISTENT AND AR MODEL

	AR model				Persistent model			
	E_{tot}	σ	E_{max}	E_{min}	E_{tot}	σ	E_{max}	E_{min}
	$t_{hor} = 6h$							
WF1	53.4	43.0	264	0.01	55.7	50.1	330	0
WF2	37.2	34	205	0	39.1	35.5	200	0
WF3	48.3	45.5	307	0	49.2	47.5	307	0
WF4	38.1	34.8	206	0	39.8	35.6	190.8	0
	$t_{hor} = 12h$							
WF1	59.3	43.0	261.8	0.5	62.3	49.8	296.8	0
WF2	42	37.6	233	0	45.8	39.3	223	0
WF3	58.8	50.4	309	0	61.2	53.42	314	0
WF4	47.3	40.9	216	0	50.8	41.4	215	0
	$t_{hor} = 24h$							
WF1	78.2	56.2	304	7.9	88	65.9	334	4.5
WF2	55.6	51.4	260	1.3	64.5	54.8	267	0
WF3	86.3	59.2	328	0	96.4	69.4	335	0
WF4	74.2	48.1	262	11	88	61.4	277	4.8

As expected, the error increases as the time horizon is higher and shows an improvement over persistent model in all the cases for the mean error, ranging from 1.8 % to 15.6 % of improvement. The standard deviation is also lower in the AR model in all the cases indicating smoother results compared to persistent case.

Apart from the mean error for the whole verification period E_{tot}, it is interesting to know the mean error of each prediction E_t during the verification period. This error is shown, for a time horizon of 24 hours and for the wind farms WF1 and WF4, in Fig.3 and 4, identifying the time for the maximum E_{max} and minimum E_{min} error. From the figures it can be observed that the AR model provides a smoother and lower error than persistent model.

In order to obtain representative results for the hourly error $\epsilon(t)$, two different periods 1 and 2 have been chosen, each one with an error similar to the mean error. For the wind farm WF4, the period called 1 with an error next to 74.2 % has been chosen to see the real power P^1_{real} and the AR model prediction P^1_{AR} and the period called 2 with an error next to 88 % has been chosen to shown the real power P^2_{real} and the persistent model prediction P^2_{per}. These results are shown in Fig.5 and are representative of the errors for the 24 hours horizon with a delay of 12 hours for WF4. Persistent model maintains constant the predicted power showing a worse fit while the AR model decreases the predicted power during the period providing a lower error.

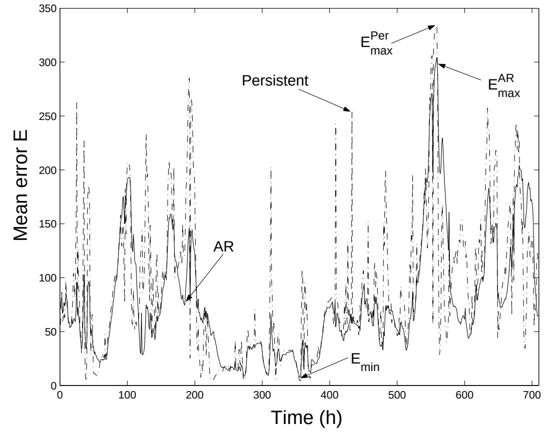

Fig. 3. Mean error for persistent and AR model for WF1

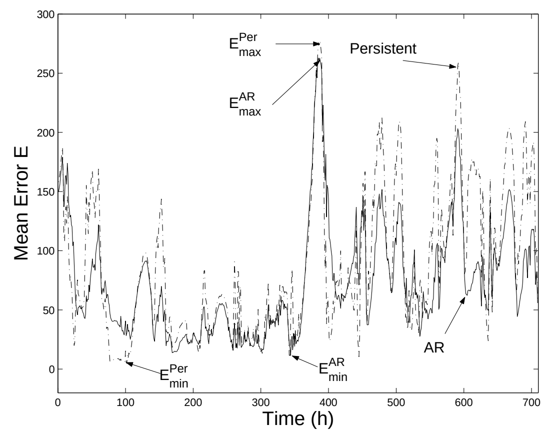

Fig. 4. Mean error for persistent and AR model for WF4

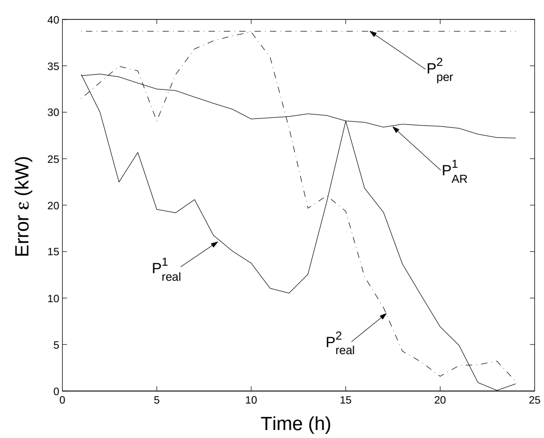

Fig. 5. Mean error for persistent and AR model for WF4

IV. Conclusions

The present paper has developed AR models to forecast the wind power prediction of different wind farms for different time horizons. The absolute error has been obtained as a function of the model order selecting the optimum value. Concerning the training period, results have shown that a year is time enough to extract the information from the time series. The selection of older data is not useful since saturation is reached and older data can even spoil the final result. As long as the estimation data is concerned, it has been shown that the better results are obtained if the selected data is not only real data, but a mixture of real and estimated values.

Finally, the comparison between persistent and AR model has shown that the AR model can provide a best prediction although the improvement is limited, obtaining a maximum improvement of 15.6 %. The development of statistical model including the wind speed as an exogenous variable is the next step to increase the improvement over persistent model.

References

[1] L. Landberg "Shor-term prediction of local wind conditions," *Dissertation*, Risø-R-702(EN), Risø National Laboratory, Roskilde, Denmark 1994, ISBN 87-550-1916-1.

[2] S. Li, D. C. Wunsch, E. O'Hair y M. G. Gieselman "Using neural networks to estimate wind turbine power generation," *IEEE Trans. on Energy Conversion*, Vol. 16, No. 3, pp. 276-282, September 2001.

[3] J. Contreras, R. Espinola, F. J.Ñogales y A.J. Conejo "ARIMA models to predict next-day electricity prices," *IEEE Trans. on Power Systems*, Vol.18, No.3, pp. 1014-1019, August 2003.

[4] Gross, G. and Galiana, F.D. "Short-term Load Forecasting", Proceedings of IEEE, Vol. 75, Nº12, Decembre 1987, pp. 1558-1573

[5] Papalexopoulos, A. and Hesterberg, T., "A Regression-Based Approach to Short-Term System Load Forecasting", IEEE Transaction on Power System, Vol. 5, Nº 4, November 1990, pp. 1535-1550.

[6] G. E. Box y G. Jenkins "Time series analysis. Forescasting and control," *Holden Day*, 1976.