Wind Time Series Modeling for Power Turbine Forecasting

Silvia Nunnari, Luigi Fortuna, and Antonio Gallo
Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica, Catania, Italy
Email: {silvia.nunnari, luigi.fortuna, antonio.gallo}@dieei.unict.it

***Abstract*—This paper addresses the problem of predicting the average wind speed at different prediction horizons ranging from 6-hours to 1-day based on wind velocity recorded at a point. The problem is relevant in several application fields and recently appears of particular interest for operators of electrical wind turbine plants and/or for optimisation of conventional power plants. Exogenous inputs are not taken into account in this preliminary work, so that the problem is set as pure time series identification and carried out by considering NAR (Non-linear Auto Regressive) models. Therefore model performance are solely related to the degree of autocorrelation of the considered time series and to some extent on the kind of non-linear approximation basis function taken into account. Different data set were considered for training and validation purposes in order to assess the model generalization capabilities. The role of model order was evaluated on the space of representative performance indices. Results show that while the forecasting performances are remarkable for the 6 and 12 hours prediction horizons, they look no so good for the 1-day prediction horizon. Work is still in progress in order to overcome these shortcomings.**

***Index Terms*—wind speed, time series, NARX models, sigmoid networks, smart grids**

I. Introduction

Wind energy has always provided the driving force for several human activities. The use of this kind of energy has subsequently fallen into disuse with the deployment of electrical energy from fossil fuels. However, the recent attention paid to climate changes, promoted a renewed interest for the production of electrical energy from renewable sources and therefore also from wind. This type of energy, in comparison to other renewable energies, requires lower investment and generally is largely available almost everywhere. Unfortunately wind power is affected by strong uncertainty. Indeed the strength and direction of wind changes on a scale of days, hours, or even minutes, depending on weather conditions. Furthermore nowadays wind turbine for generation of electricity can be very large (even hundred of MW) and are connected to the so-called smart grids, i.e. complex power interconnected systems were power can be more efficiently generated, transmitted and consumed. This implies for the wind turbine operators the need to forecast the performances of their plants as accurately as possible in order to avoid destabilization of the overall grid. Furthermore in case the operators are not able to deliver their traded amount of energy are subjected to pay fines.

Bearing in mind that the power output in a wind turbine depends on the cube of wind velocity, it is easy to understand that the challenges of accurately predicting the power output in a wind turbine are strongly dependent on the ability to identify reliable models for the wind speed stochastic process.

The topic of predicting wind power has been widely addressed in literature. A review on 30 years of history of the wind power short-term prediction was given in [1] while a more recent and comprehensive review was given in [2] as a deliverable of the ANEMOS. plus European Project. In this latter contribution wind models are classified into two broad classes depending on the fact that a Numerical Weather Prediction model (NWP) is involved or not. Usually models that include NWP are referred to as physical models, while models of the latter class are referred to as statistical or time series approach.

The choice of involving a NWP model or not depends on the forecast horizon. Models that make use of NWP are considered for long time horizon while time series approach are considered for short time horizon (less the 24 hours). This latter kind of models can be very useful for optimisation of conventional power plants, where reasonable prediction horizons can vary between 3 to 10 hours depending on the size of the system and the type of conventional units included. In order to clarify this aspect it could be useful to consider, as an example, that for systems including only fast conventional units, such as diesel gensets or gas turbines, the horizon can be below 3 hours [2]. Another aspect that characterizes the two classes of models is the space scale. Usually physical models are considered for large space scale prediction while time series approaches are more appropriate for prediction at a point. Indeed physical models, when considered for point wise prediction involve a quite complex process of statistical downscaling [3].

Since a discussion about the features of the two described model classes is beyond the scope of this paper, here a new NAR time series modelling approach, based on the use of sigmoid functions will be presented.

Manuscript received August 20, 2013; revised November 12, 2013.

doi: 10.12720/ijoee.2.2.112-118

II. Wind Speed Data Analysys

Wind speed time-series is a random process with a marked seasonal component as shown for instance in Fig. 1.

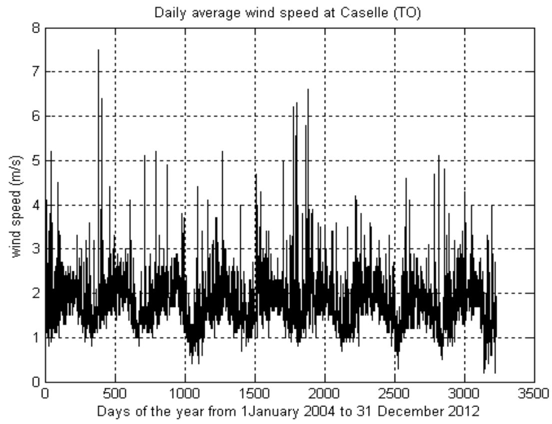

Figure 1. Daily average wind speed time series recorded at Caselle (Turin, Italy) from January 2004 to December 2012.

The spectrum of a the daily mean data set shown in Fig. 1 is reported in Fig. 2, where it is possible to recognize a peak with a period of 1 year.

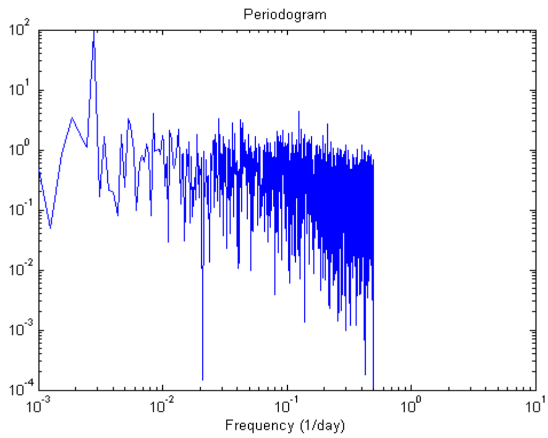

Figure 2. Spectrum of the daily wind speed time series. Frequency in abscissa is expressed in $(\text{day})^{-1}$.

At a lower time scale, for instance in the order of 1 hour, wind speed exhibits a typical scattered behavior, as shown in the sample data given in Fig. 3.

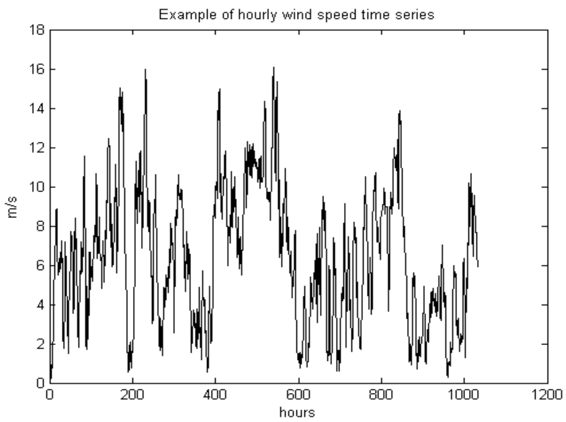

Figure 3. A record of hourly average wind speed.

The mean and variance of the considered data set during the 4 seasons are reported in Table I.

TABLE I. Mean and Variance of Wind Velocity of the Considered Data Set

Season	Mean (m/s)	Variance (m/s)
Winter	2.4938	1.0517
Spring	2.9776	0.6113
Summer	2.7918	0.4628
Autumn	2.1575	1.0517

As it possible to see the mean of the daily mean values are higher during spring and summer season while, on the contrary, the variance is lower. In any case, the variability of the mean and variance of the time series with seasons indicates a non-stationary feature.

In order to get some a-priori insights about the possible success of linear models to predict wind velocity, it is possible to consider the autocorrelation function. As an example, for the daily mean wind time series represented in Fig. 1, the autocorrelation function is shown in Fig. 4.

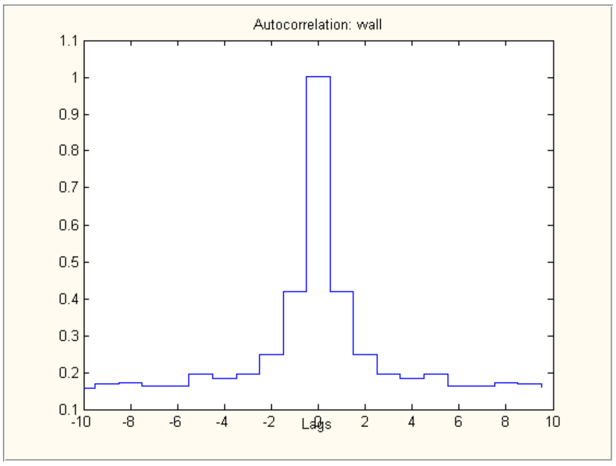

Figure 4. Autocorrelation of daily mean speed time series.

Form Fig. 4 it appears that the autocorrelation decays rapidly as early as the first lag values, thus confirming that the daily mean wind time series is very scattered. Moreover, since the higher autocorrelation coefficient (at lag 1) is about 0.42 even the possibility to correctly predict the mean value for the day after (i. e. 1 day ahead) appears to have a limited chance of success by using simple linear autoregressive models.

For hourly mean wind speed, such that shown in Fig. 3, the autocorrelation function is shown in Fig. 5.

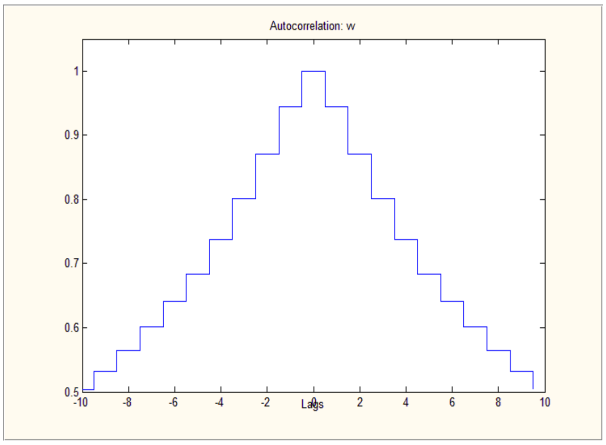

Figure 5. Autocorrelation of hourly mean speed time series.

From Fig. 5 it appears that autocorrelation takes about 10 lags to decay to values lower than 0.5 and thus the prediction problem without using exogenous inputs seems realistic in the horizon of hours even using linear models.

III. Wind Speed Time Series Modeling Approaches

Several methods have been proposed in literature for modelling wind speed time series. The most studied approaches are autoregressive moving average (ARMA) [4], autoregressive integrated moving average (ARIMA) [5], fractional ARIMA (fARIMA) [6], hybrid ARIMA [7-8]. Since, as mentioned in the previous section, linear models may have limited chance of success for modelling wind time series, we explore in this paper the possibility to use NAR (Non-linear Autoregressive) models. The rationale for this is that random inputs are not the only source of irregular behaviours of the system output. Indeed nonlinearities or chaos can produce very irregular data even with purely deterministic equation. So it is better try to explain irregularities in a given time series to both the presence of random inputs and nonlinearities [9]. The literature about non-linear modelling techniques of wind speed time series is very rich. For instance, various Data-Mining approaches, such as the Cluster Center Fuzzy Logic (CCFL), Multi Layer Perceptron (MLP) neural networks, k-nearest neighbour model (k-NN) and Adaptive Neuro-Fuzzy Inference (ANFIS) have been considered in [10]. Others non linear approaches have been presented in [11]. A quite general way to express a non linear model is the so-called NARX form, as shown in expression (2):

$$y(t) = F\big(y(t-1), \cdots, y(t-n_a), u(t-n_k), \cdots, u(t-n_k-n_b)\big) \tag{1}$$

Here $F(\boldsymbol{x})$ is an unknown non-linear function of the vector argument x, $y(t)$ is the scalar system output, $u(t)$ the related input variable (e.g. meteorological). Of course the input variable can be more than one and in this case expression (1) must be according modified. When exogenous inputs u is not available, expression (2) becomes a NAR model. Several of the most powerful time series modern techniques such as the MLP artificial neural networks, the Fuzzy and the Neuro-Fuzzy techniques can be considered for approximating the unknown function F given an appropriate set of measured data. In this paper we have referred to the nlarx procedure, available in the MATLAB© Identification Toolbox [12], where several non linear estimation options are available such as wavelet based functions, sigmoid function etc. However it is not the aim of this paper to propose a comparison among these kinds of estimators. How we will discuss in the next section, by a trial and error approach we have found that for the considered application, good results can be obtained by using the sigmoid function. This means that the F function is expanded as a series of terms of the basic function expressed in (2)

$$f(z) = \frac{1}{(e^{-z}+1)} \tag{2}$$

IV. Performance Indices

In order to objectively evaluate the goodness of a prediction model several performance indices can be taken into account. Such indices can be roughly grouped into two separate sets: a) global fit indices, i.e. those indices that give measures of the fit of the overall time series (i.e. for instance the RMSE error), and b) those that give a measure of the capability of a given model to predict critical episodes (i.e. for instance the SP index), referred to here as exceedance indices.

A. Global Fit Indices

The definition for the most common global indices is given below. Let us indicate as O and P the observed and predicted time series than we have the following definition.

The mean bias error, expressed by (3) is the degree of correspondence between the mean forecast and the mean observation. Lower absolute numbers are best. Values < 0 indicate under forecasting.

$$Mean\ Error = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i) \tag{3}$$

The Mean Absolute Error (MAE) expressed by (4) is the mean of the absolute value of the residuals from a fitted model. Lower numbers are best.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \tag{4}$$

The Root Mean Square Error (RMSE) expressed by (5) is considered to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. It is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \tag{5}$$

B. Exceedance Indices

The target of predicting the exact value of the mean wind velocity model, even if in principle useful, is not realistic, due to wind stochastic nature. Thus it seems more realistic to assess the goodness of a prediction model evaluating its capability to forecast if the mean wind velocity will exceed a given threshold. On the other hand, the practical problem of wind turbine plant operator is not to predict the exact value of energy that the plant will be able to produce, but to know if the plant will produce at least a certain amount of energy. Bearing this in mind, we have borrowed a set of performance indices originally adopted by the European Environment Agency [13] to test the capabilities of a short term forecast model to predict exceedance of photochemical smog. These indices are defined according with the following standard *contingency table* (see Table II):

TABLE II. CONTINGENCY TABLE

Alarms	Observed		
Forecasted	Yes	No	Total
Yes	*a*	*f-a*	*F*
No	*m-a*	*N+a-m-f*	*N-f*
Total	*m*	*N-m*	*N*

where:

N is the total number of data points;

f is the total number of forecast exceedance

m is the total number of observed exceedance;

a is the number of correctly forecast exceedance.

Using these definitions, the following indices can be defined:

SP (the probability of detection) is the fraction of correct forecast critical events. Its values ranges from 0 to 100, 100 being the best value.

$$SP = \frac{a}{m}100 \tag{6}$$

SR (the percentage of predicted exceedances actually occurred) is the fraction of realized forecast critical events (range from 0 to 100 with a best value of 100).

$$SR = \frac{a}{f}100 \tag{7}$$

FA (the false alarm rate) express the percentage of instances when predicted exceedances did not actually occur. With respect to a good model FA should approach zero.

$$FA = 100 - SR \tag{8}$$

V. RESULTS

The data set considered in this paper is represented by mean values of wind velocity measured at Caselle (Turin, Italy) from January 2004 to December 2012. Data were measured with 1 hour sample time at 10 m height from ground level and have been reported at the turbine hub height, by using a model of the type represented by expression (9), as suggested by [14].

$$U_z = U_{10}\left(\frac{z}{10}\right)^p \tag{9}$$

In expression (9) U_z represents the wind speed at height *z*, z is the height from ground level, U_{10} the wind speed measured at 10 m from ground and *p* a coefficient depending from terrain conditions. For rural terrain p is assumed to be about 0.16 [14].

Three different classes of NAR models have been identified depending on the prediction horizon considered. Models were trained on a set consisting of about 80% of available data and validated on the remaining 20%, in order to assess the generalization capabilities. The role of model order is evaluated on the space of representative performance indices. Appropriate software was coded in MATLAB to this purpose.

A. Results for 6-Hours Ahead Prediction Models

Mean errors, MAE and RMSE computed for 6-hours prediction models with order ranging from 1 to 20 are shown in Fig. 6.

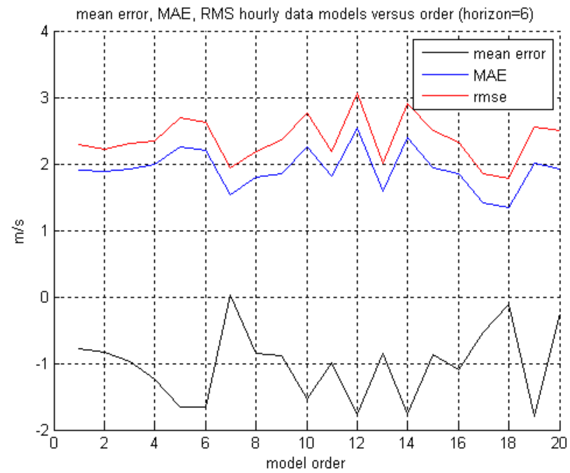

Figure 6. Mean error, MAE and RMSE (horizon=6 hours).

As it is possible to see, global errors are quite independent of the model order. In particular the mean error is < 0 indicating that the considered approach tends to under forecast. The SP and FA exceedance indices computed for order models ranging from 1 to 20 is shown in Fig. 7.

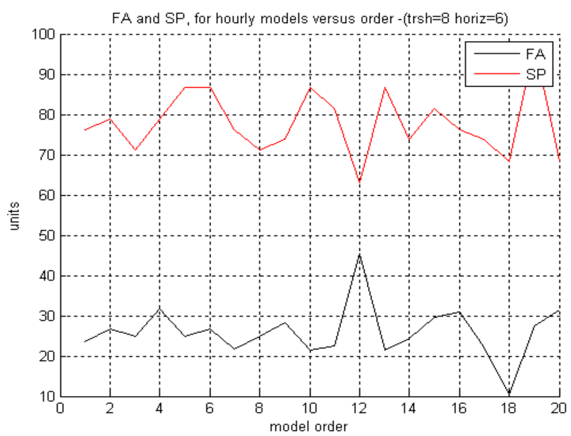

Figure 7. SP and FA for orders ranging from 1 to 20.

Fig. 7 shows that the percentage of events exceeding the threshold of 8 m/s correctly forecast, with a time horizon of 6 hours, is of about 80% while the number of false alarms is of about 30%. Furthermore it seems that low order models (say 6 - 7) perform better. A detail in terms of M, F and A is given in the following Fig. 8.

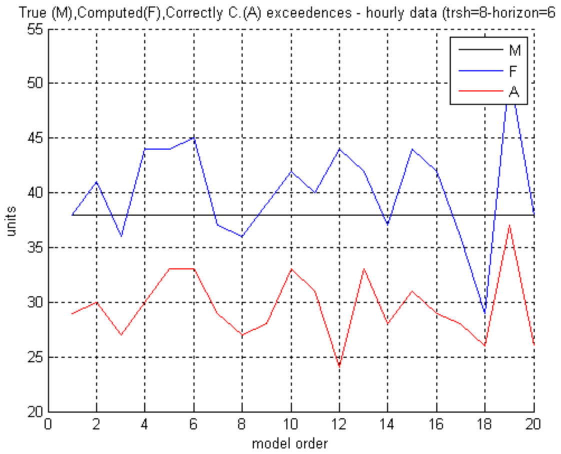

Figure 8. M, F and A for orders ranging from 1 to 20.

The horizontal line represents the level M of exceedances observed in the testing time series. Fig. 8 shows that models tend to forecast a number F of exceedances higher than those actually observed (M), while, of course, the number of exceedances correctly forecast A is lower than M. Furthermore it seems that high order models produce a scattered behavior in terms of F and A thus meaning that they are less stable in terms of prediction accuracy.

The time behavior of the 6-hours prediction model with order equal to 20 for the validation set is reported in Fig. 9.

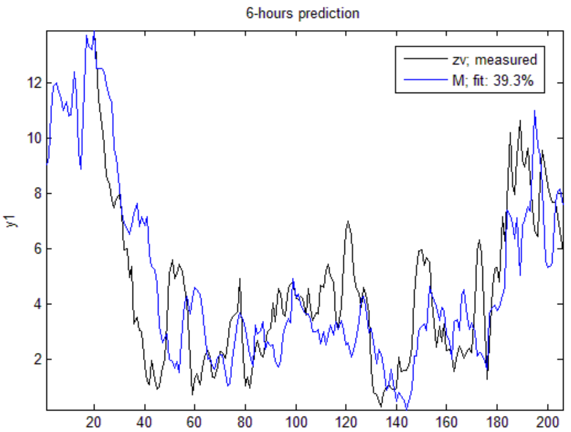

Figure 9. Time behavior of the 6 hour prediction model (validation set).

The black and blue lines correspond to model output and measured data respectively.

B. Results for 12-Hours Ahead Prediction Models

Mean errors, MAE and RMSE computed for 12-hours prediction models for a range of NAR model order ranging from 1 to 20, are shown in Fig. 10.

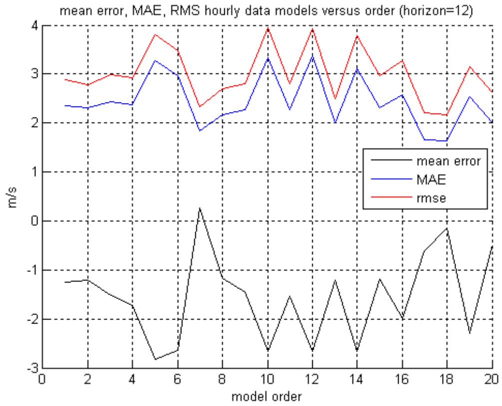

Figure 10. Mean error, MAE and rmse of NAR models of orders ranging from 1 to 20, for prediction horizon equals to 12 hours.

While performances in terms of exceedance indices are reported in Fig. 11.

It can be seen that the performance of 12-hours prediction model are lower that the corresponding 6-hours models, both in terms of global and exceedance indices, but still considered acceptable. Indeed even the number of FA may reach high values (even 60%), for appropriate model orders they can be holding lower than 30%.

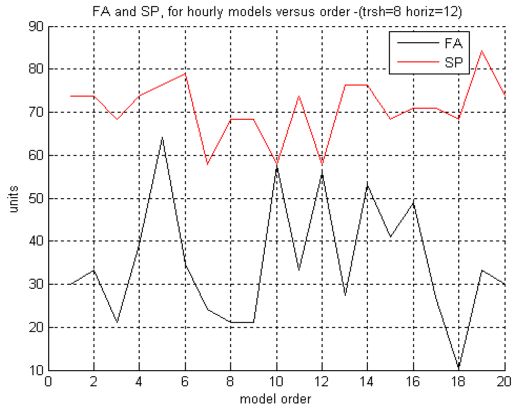

Figure 11. SP and FA for orders ranging from 1 to 20.

A detail of exceedances performances is given in the following Fig. 12, where the number of episodes in terms of M, F, and A exceeding the threshold of 8 m/s contained in the evaluation set are shown.

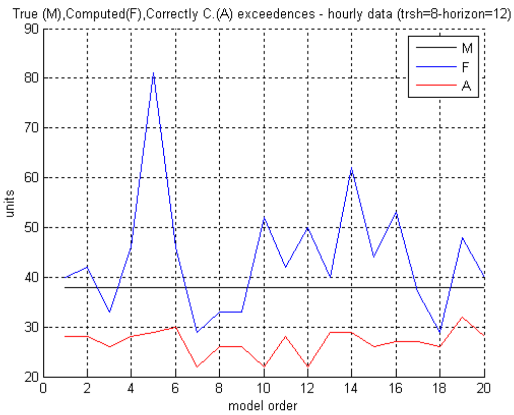

Figure 12. M, F and A for orders ranging from 1 to 20.

The time behavior of the 12-hours prediction model with order equal to 20 versus the validation set is reported in Fig. 13.

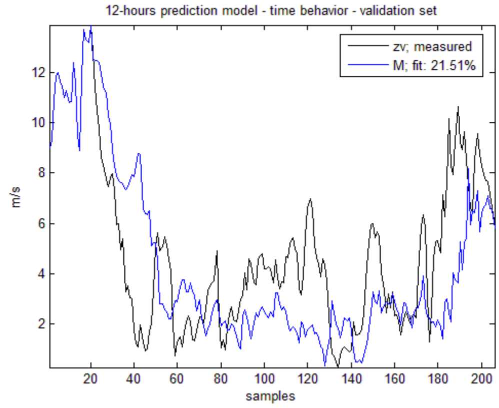

Figure 13. Time behavior of the 6 hour prediction model (validation set).

C. *Results for 1-Day Ahead Prediction Models*

For 1-day ahead prediction models the time series of daily mean wind velocity were considered. Thus models were identified as being 1-step prediction. As usually, the data set was divided in two parts, reserving the first 80% for the model training and the remaining 20% for the validation. Global errors indices obtained are shown in the following Fig. 14.

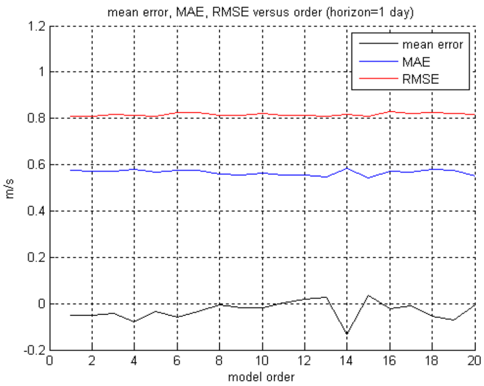

Figure 14. Mean error, MAE and rmse of NAR models of orders ranging from 1 to 20, horizon = 1-day.

Exceedance indices for a threshold of 3 m/s are shown in Fig. 15.

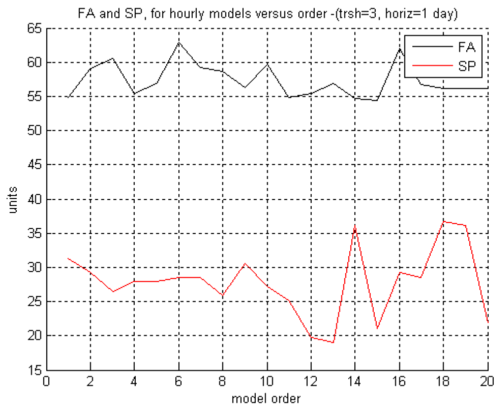

Figure 15. SP and FA for orders ranging from 1 to 20.

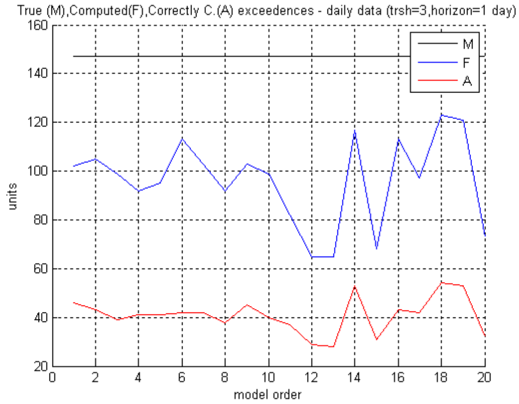

Figure 16. M, F and A for model orders ranging from 1 to 20.

Finally a detail in terms of T, F and A concerning the exceedance episodes contained in the validation data set is reported in Fig. 16. As can be seen for all identifies models the SP value is in average of about 30% while the FA, i.e. the number of false alarms is higher than 55%. Of course such performances are non useful for application purposes.

The time behavior of the 1-day prediction model with order equal to 20 versus the validation set is reported in Fig. 17. It is possible to appreciate that while the model is capable to fit the general trend of measured data it is not capable to predict the peak values of the daily mean wind velocity.

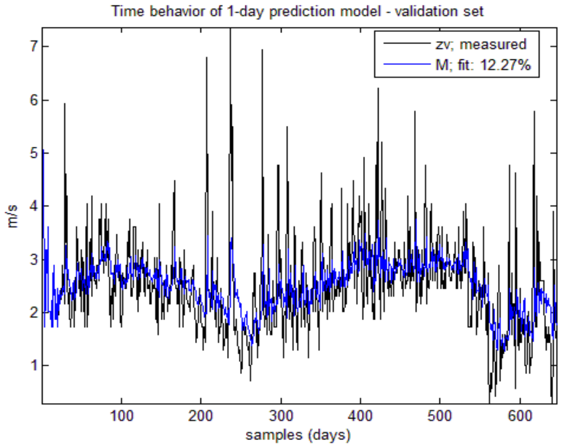

Figure 17. Time behavior of the 1-day prediction model (validation set).

VI. CONCLUSIONS

In this paper a new NAR time series modelling approach, based on the use of sigmoid functions has been presented. Results show that the identified models perform quite well to forecast wind speed velocity at a point for the 6 and 12 hours prediction horizons. Therefore, they can be reliably used for applications such as optimization of conventional power plants, where reasonable prediction horizons usually vary between 3 to 10 hours, depending on the size of the system and the type of conventional units included. Unfortunately results are not so good for the 1-day horizon. However, it is to be stressed here that in this preliminary work exogenous inputs were not deliberately taken into account. Work is still in progress in order to evaluate new opportunities of success by using appropriate exogenous inputs such as other meteorological variables and redundancies of spatial distribution of recording stations.

ACKNOWLEDGMENT

We thank the ARPA Toscana who provided wind data considered in this analysis.

REFERENCES

[1] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable & Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725-1744, 2008.

[2] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. (2011). The state of the art in short-term prediction of wind power–A literature Overview. [Online]. pp. 1-110. Available: http://www.anemos-plus.eu/

[3] I. Martí Perez., "Wind forecasting activities," presented at the First IEA Joint Action Symposium on Wind Forecasting Techniques, Norrköping, December 2002.

[4] R. Billinton, H. Chen, and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy," *Microelectron Reliability*, vol. 36, no. 9, pp. 1253–1261, 1996.

[5] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "ARIMA-based time series model of stochastic wind power generation," *IEEE Trans. on Power Systems*, vol. 25, no. 2, pp. 667-676, 2010.

[6] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renewable Energy*, vol. 34, no. 5, pp. 1388-1393, 2009.

[7] J. Hu, J. Wang, and G. Zeng, "A hybrid forecasting approach applied to wind speed time series," *Renewable Energy*, vol. 60, pp. 185-194, December 2013.

[8] J. Shi, J. Guo, and S. Zheng, "Evaluation of hybrid forecasting approaches for wind speed and power generation time series," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 3471-3480, 2012.

[9] H. Kantz and T. Schreibe., *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series 7, 1st ed. 1997, Cambridge University Press (reprinted 2000, 2002).

[10] M. Schlechtingen, I. Ferreira Santos, and S. Achiche, "Data-mining approaches for wind turbine power curve monotoring: A comparative study," *IEEE Trans on Sustainable Energy*, vol. 29, pp. 68-85, 2013.

[11] S. C. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y. Wu, and F. Wu, "Load/price forecasting and managing demand response for smart grids," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 68-85, 2012.

[12] Mathworks Documentation Center. [Online]. Available: http://www.mathworks.it/it/help/ident/ref/nlarx.html.

[13] R. M. Van Aalst and F. A. A. M. De Leeuw, "National ozone forecasting system and international data exchange in northwest Europe," *European Topic Centre on Air Quality*, pp. 1-50, 1997.

[14] G. Kiely, *Environmental Engineering*, McGrawHill, 1997, ch. 8, pp. 369-370.

Silvia Nunnari born in Catania on 1989. She received the M. Sci degree in Computer Engineering from the University of Catania (Italy) in 2012. At the present she is a PhD student at the University of Catania. Her scientific research interests include applications of ICT technologies to Smart Grids.

Luigi Fortuna, born in Siracusa on 27/05/1953, engineer, is Full Professor of System Theory since November 1994 at the University of Catania. He is the coordinator of the Ph.D. Course in Electronic and Automatic Engineering. He is author of more than 400 scientific publications, 7 books and 10 industrial patents. He is also IEEE Fellow, Chairman of the IEEE Committee on CNNs.

Antonio Gallo, born in Caltagirone (CT) on 1948, is Full Professor of Process Control since November 1990 at the University of Catania. He is co-author of hundreds of scientific papers and has coordinated several research activities.

