Stochastic Analysis for Wind Speed Forecasting

Ciprian Nemes
Technical University of Iasi
Faculty of Electrical Engineering
Iasi, Romania
cnemes@ee.tuiasi.ro

Mbuyamba Tshibanda
University of Johannesburg
Faculty of Engineering and the Built Environment
Johannesburg, South Africa
mtshibanda@uj.ac.za

Abstract—**The stability and availability required on the electrical power systems with wind sources are directly related to the accuracy of a short-term forecasting wind speed model. This paper presents a wind speed forecasting model based on one of the widely used time-series regression models, namely the Auto-Regressive Integrated Moving-Average (ARIMA). The method requires historical wind speed data for a given area, collected over a long time interval, in order to evaluate the required parameters of the wind speed ARIMA model.**

Keywords: wind power, time-series regresion techniques, wind speed forecasting.

I. INTRODUCTION

In the last decades, a growing interest in renewable energy resources has been observed. Unlike other renewable energy sources, wind energy has become competitive with conventional power generation sources and, therefore, the application of wind turbine generators has the highest growth among other sources. Wind is one of fastest growing energy source and it is considered as an important alternative to conventional power generating sources.

Wind generation brings a great amount of benefit to power system, such as a cheaper energy comparing with the conventional generation, a lower gas emission, a faster implementation of the wind systems projects, a higher availability of wind energy resources for large areas, and so forth. Meantime, wind generation brings a series of difficulties to the traditional power systems, such as the uncontrollability of power generation, intermittence of generated power, the wind speed presents irregularly fluctuating and intermittent behaviour and also a poor predictability.

The availability of short-term forecasting wind speed model for a particular area is essential for operation planning of wind energy systems. This paper presents an application of time-series regression models in order to establish a model for the mean daily wind speed for an area located in north-east of Romania.

II. WIND FORECASTING NECESSITY

The operation of power systems with renewable energy sources has to consider the stability of the electrical power system, which is based on a reliable power generation that is permanently balanced by the load. In power system operations, it is imperative necessary to ensure the adequate electric power capacity, having in view the continuously changing of operating conditions. In a traditional power system, the System Operator supervises all available controllable resources, in order to ensure a safety reserve in operation of power system. The wind generation is an uncontrollable generating resource, depending on wind availability, thus the integration of wind generation into an electric power system will have a major impact on the system operation process.

One of the most important requirements for wind generation planning and operation in the power systems is an accurate wind speed forecasting. The operation of wind generation should be carefully performed in order to mach the system's needs with the different forecasting time frames [1]. The applications of wind power forecasts over a different period of time can be classified into long-term and short-term intervals:

- Long-term wind speed forecasting usually covers few years ahead, considering the monthly and yearly values. Wind generation depends on geographical location and climatic condition, varying from season to season and from year to year. The long-term wind forecasting is used for development planning of wind power generation as well as the power system infrastructure, necessary for integration of the wind energy systems. The decisions of developing of power system infrastructure (including transmission line and substations) are made with looking ahead to meet demand growth and satisfy reliability requirement.
- Short-term forecasting requires information about the wind speed from one hour up to a few days, being necessary to schedule the available generating units based on the predicted load demand and wind generation, as well as other system conditions, such as transmission constraints and generating units maintenances. The hourly schedules of generating units, also known as real-time operation, are crucial information for an economical and safety operation of power systems with wind resources.

A literature survey [1], [2] indicates that other authors also define the concept of medium forecasting, which are usually used to estimate the wind speed from a week to a year time interval.

III. Time-Series Models

Time-series can be defined as a sequential set of data measured over a given time interval, commonly ordered in time, such as the hourly, daily or monthly measurements. The data term has the meaning of discreet values measured at equal intervals of time. The methodology which concerns with the analysis of this class of date is known as the analysis of time-series regression technique or the analysis of dynamical series [3]. The basic idea of time-series forecasting is to build an accurate matching pattern for available data, after that to use this pattern to obtain the forecasted value with respect of the time frame. A time-series model predicts the forecasted value as a response of time series to a linear combination of its own past values and current or past values of other time-series.

Different numbers of models are used to characterize the time- series regression technique. In this section are presented the main types of patterns used in time series analysis, which are widely used in a large number of practical applications [2],[3].

A. *Auto Regressive process (AR)*

In an auto regressive process by order *p*, noted as AR(p), the current value of the time series, *y(t),* is expressed as a linear relationship between its previous values, being described by the following relationship:

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + ... + \phi_p y(t-p) + e(t) \quad (1)$$

where ϕ_i represents autoregressive parameters, which considers the influence of values that range between *y(t-i)* and *y(t)*, while $e(t)$ represents a white noise series with mean zero and variance σ_e^2, which is independent by *y(t-i)* terms. Usually, AR(p) patterns are expressed using a $\Phi(B)$ operator, which describes the difference between first to p^{th} behind *y(t)* process.

$$\Phi(B) \cdot y(t) = e(t) \quad (2)$$

where: $\Phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 ..... - \phi_p B^p$.

Therefore, the process *y(t)* can be considered to be an output signal from a transfer filter, $\Phi(B)^{-1}$, having as the input signal a white noise sequence, *e(t)*, as is shown in Fig 1.

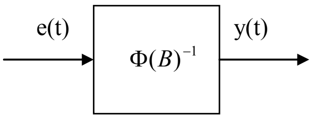


Fig. 1 - Transfer function for an AR model

B. *Moving Average process (MA)*

In a moving-average process by order *q*, noted as MA(q), the current value of the time series, *y(t)*, is expressed as a linear relationship between the current and previous values of a white noise series, $e(t), e(t-1), \cdots$, as in presented in follows:

$$y(t) = e(t) - \theta_1 e(t-1) - \theta_2 e(t-2) - .... - \theta_q e(t-q) \quad (3)$$

where θ_i represents the moving average parameters. A similar application of the operator Θ(B) on the white noise series, allows that previous equation to be written as:

$$y(t) = (1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q) e(t) = \Theta(B) e(t) \quad (4)$$

Thus, based on previous expression, the time-series process can be considered as an output signal from a linear filter, with the transfer function Θ(B) and with a white noise sequence, *e(t),* as an input signal, as is depicted in Fig 2.

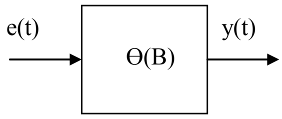


Fig. 2 - Transfer function for a MA model

C. *Auto Regressive Moving Average process (ARMA)*

If the previous processes (MA and AR) are combined together, a new class of process can be defined, namely the Auto-Regressive Moving Average (ARMA), this being one of the widely used time-series process. ARMA procedure analyzes and forecasts the time series equally spaced data, the mathematical representation of ARMA (p, q) being described by the following relationship:

$$\begin{aligned} y(t) &= \phi_1 y(t-1) + \phi_2 y(t-2) + ... + \phi_p y(t-p) + \\ &+ e(t) - \theta_1 e(t-1) - \theta_2 e(t-2) - .... - \theta_q e(t-q) \end{aligned} \quad (5)$$

where ϕ_i and θ_j are called the autoregressive and moving average parameters.

In this case, the *y(t)* process can be considered as a output signal of a linear filter whose transfer function is given by the rapport of two polynomials Θ(B) and $\Phi(B)$, having as a input signal a white noise sequence, *e(t)*, as is depicted in Fig 3.

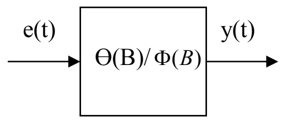


Fig. 3 - Transfer function for an ARMA model

D. *Auto Regressive Integrated process with Moving Average (ARIMA)*

The above time-series processes, like AR, MA or ARMA processes, are called stationary processes. This means that the mean of the series and the covariance between the observations for any of these processes don't change in time. Unfortunately, many of the practical time series are characterized by a given non-stationary behaviour, in which the trends and other qvasistationary characteristics vary in time. If the process is non-stationary, it is necessary a changing of the process into a stationary process [4]. This changing can be conducted in the case of non-stationary time series using a differencing process, symbolized by ∇ operator, where:

$$\nabla y(t) = y(t) - y(t-1) = (1-B)y(t) \tag{6}$$

Consequently, an order d differentiated time series is written as:

$$\nabla^d y(t) = (1-B)^d y(t) \tag{7}$$

A differentiated stationary series can be modelled using one of the AR, MA or ARMA processes, obtaining an integrated time series processes. For a series that needs to be d order differentiated and that can be modelled as a p and q AR and MA processes, the ARIMA(p,d,q) model can be written as:

$$\Phi(B)\nabla^d y(t) = \Theta(\mathrm{B})\mathrm{e(t)} \tag{8}$$

As a result of daily, weekly, yearly or other periodicities, many time series present periodic behaviours in response to one or more of these periodicities [5]. Therefore, a seasonal ARIMA model should be rewrite in accordance with following relationship:

$$\phi(B)\Phi(B^S)\nabla^d \nabla_S^D y(t) = \theta(B)\Theta(B^S)e(t) \tag{9}$$

The *y(t)* series represents an output signal from an unstable linear filter whose inside signal is a white noise sequence *e(t)* and whose transfer function is given by the combination of processes depicted in Fig 4.

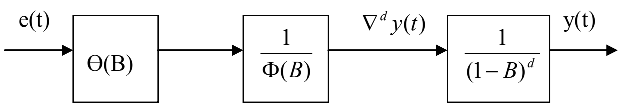


Fig. 4 - Transfer function for an ARIMA model

IV. Numerical Example

For a numerical analysis, a real wind speed measurement is used in the paper to forecast the wind speed values based on the time series technique. The wind data used for analysis was collected from the north-east area of Romania, over one year. The data collection was made at one hour interval, the hourly average values being recorded. The data was first examined for any missing values or outliers. The missing values have been replaced by the average of values from same day, the daily time-series values being shown in the Fig 5.

During the identification process, the first analyze is an evaluation of autocorrelations and partial autocorrelations functions. The plots of these autocorrelation functions show the degree of correlation between past values of the series, as a function of the number of days from the past, for that the correlation is computed.

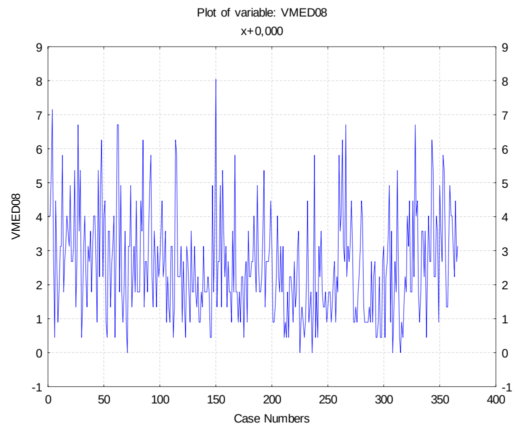


Fig. 5 Daily wind speed average data base for one year

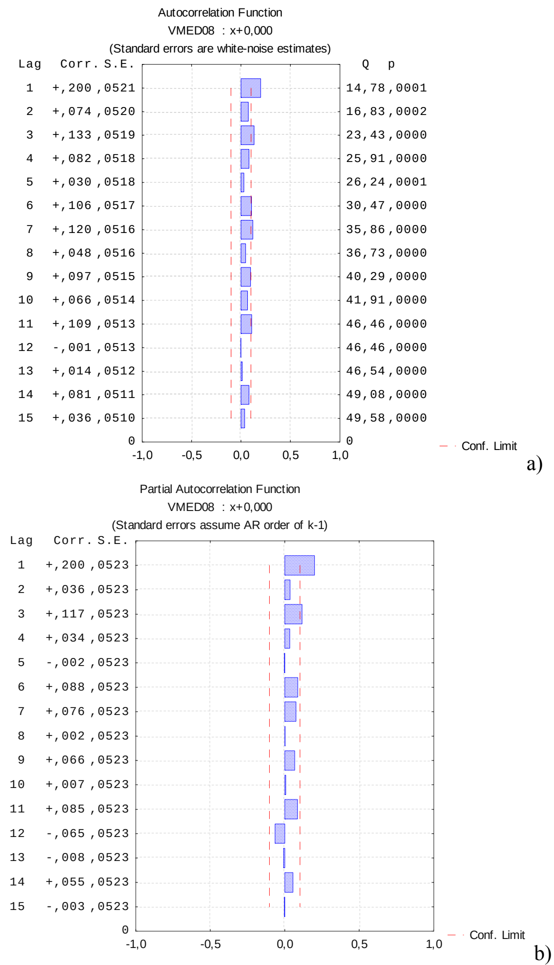


Fig. 6. Autocorrelation function (a) and partial autocorrelation function (b) of the wind speed original database

As can be seen from the Figs 6.a and 6.b, the autocorrelation and partial autocorrelation functions indicate that the time-series is slightly autocorrelated, with an order between 1 and 3, while rest of autocorrelation and partial autocorrelation values are between the confidence limits. These aspects involve the possibility of an Auto Regressive Moving Average model analysis. According to the autocorrelation function, it results that the autoregressive transformation parameter will be chosen to p=1. Analyzing the partial autocorrelation function of the wind speed series, it results a value of moving average parameter between q=1÷3. Based on the ARMA(1, 1÷3) analyses, it appears the necessity of a seasonal differentiation of original series. From the process analysis, it results the following optimal model for the original wind speed database, namely an ARIMA(1,1,1)(1,0,0), with a seasonal lag equal with 12. The estimation of autoregressive and moving average parameters have been conducted to p=0,76314 and q=0,66582, respectively. The residual values of original database after that ARIMA(1,1,1)(1,0,0) process has been applied, are plotted in Fig 7.

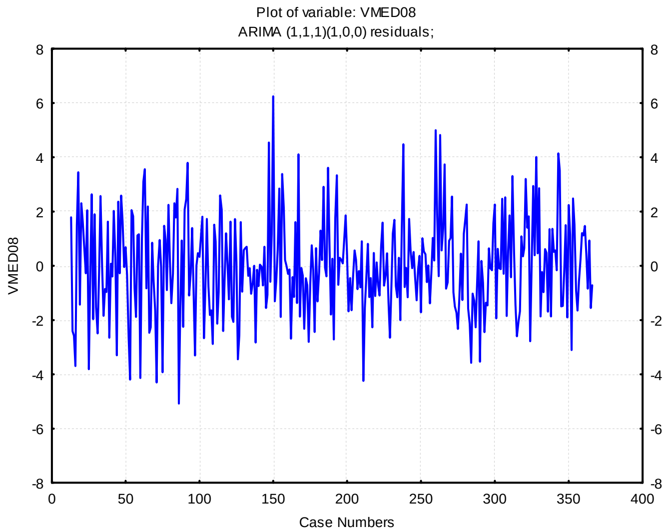
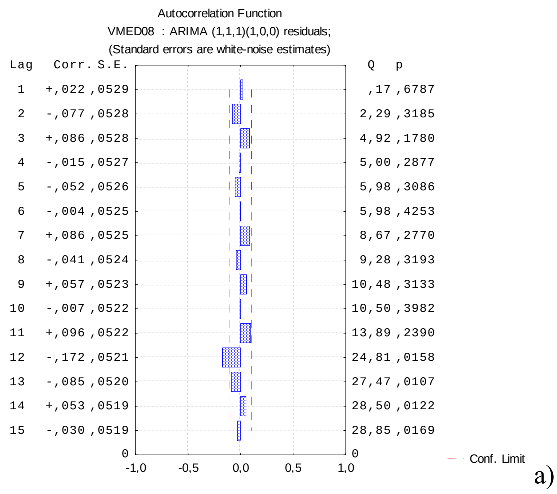
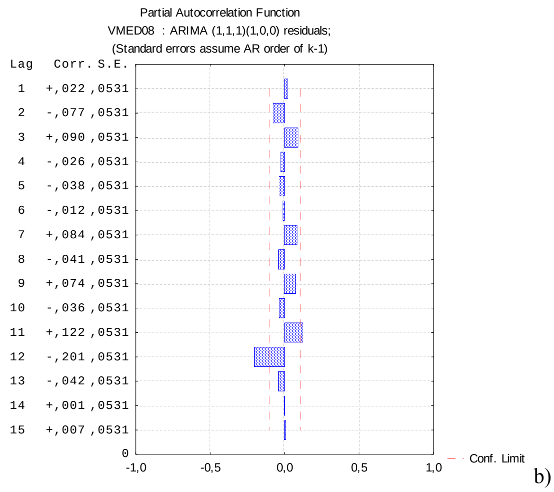


Fig. 7 Residuals ARIMA (1,1,1)(1,0,0) series

The autocorrelation and partial autocorrelation functions of residual series, indicate that the residuals are uncorrelated, as can be seen in Figs. 8 a,b. Moreover, a chi-square statistical test applied to the error autocorrelations indicates that the predicted error is very close to a white noise with a zero mean.



a)



b)

Fig. 8. Autocorrelation function (a) and partial autocorrelation function (b) of the residual series

From the statistical results can be concluded that the ARIMA (1,1,1)(1,0,0) is adequate to model this particular wind speed series. For estimation of speed wind values for the next 30 days, the data has been divided into two groups, namely: one data group from January 1 to December 15, for the prediction model and the second group from 16th to 31th of December for testing and validating the resulting regression model. In Fig 9, the dashed central line was created based on the estimated ARIMA (1,1,1)(1,0,0) model, while the continuous line are the original historical data. The extreme dashed line from Fig. 9 shows the limits of estimated values of wind speed, in a 0.9 confidence level interval.

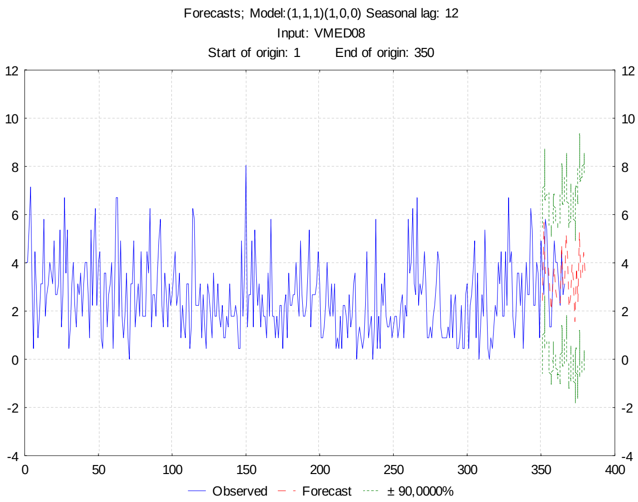


Fig. 9. Comparison between original and forecasted time-series

A numerical comparison between original and forecasted daily wind database are presented in Table 1. The forecasted values of speed wind are generate for the 30 days, from which 15 days are over the last days from 2008, and others 15 days are from beginning of 2009. For these values have been computed the residual values, they being tabulated in last column of table.

TABLE I. FORECASTED AND ORIGINAL VALUES OF WIND SPEED

Day No.	Forecast	Lower 90%	Upper 90%	Std.Err.	Observed	Residual
351	2,439109	-0,58915	5,467364	1,836189	2,222827	-0,216282
352	5,675669	2,63311	8,718230	1,844863	4,842877	-0,832792
353	3,549342	0,49848	6,600203	1,849896	3,612008	0,062666
354	3,770664	0,71498	6,826348	1,852821	3,914502	0,143838
355	3,833222	0,77473	6,891712	1,854522	3,259986	-0,573236
356	2,922760	-0,13736	5,982883	1,855512	3,518218	0,595458
357	2,026027	-1,03505	5,087100	1,856089	2,620609	0,594582
358	3,014286	-0,04734	6,075913	1,856424	2,995469	-0,018817
359	3,797367	0,73542	6,859316	1,856620	3,961013	0,163646
360	2,845871	-0,21627	5,908008	1,856734	2,933191	0,087320
361	2,404439	-0,65781	5,466685	1,856800	2,311085	-0,093354
362	2,433367	-0,62894	5,495677	1,856838	2,796262	0,362895
363	2,333283	-1,01482	5,681391	2,030133	2,039125	-0,294158
364	4,759747	1,40864	8,110854	2,031951	5,851340	1,091593
365	3,070129	-0,28272	6,422982	2,033010	3,001931	-0,068198
366	3,166855	-0,18701	6,520724	2,033626		
367	5,189266	1,83481	8,543727	2,033985		
368	4,287307	0,93250	7,642112	2,034194		
369	2,144318	-1,21069	5,499323	2,034315		
370	2,580753	-0,77437	5,935875	2,034386		
371	3,924387	0,56920	7,279577	2,034427		
372	3,254162	-0,10107	6,609393	2,034451		
373	1,562187	-1,79307	4,917440	2,034465		
374	3,819895	0,46463	7,175161	2,034474		
375	2,392645	-1,66884	6,454126	2,462688		
376	5,271071	1,20322	9,338921	2,466550		
377	3,337712	-0,73384	7,409267	2,468796		
378	3,503920	-0,56979	7,577631	2,470104		
379	4,432694	0,35773	8,507660	2,470865		
380	3,525966	-0,54973	7,601662	2,471308		

V. CONCLUSIONS

The effects of wind speed forecast errors, strictly evaluated as the difference between forecasted and real wind speed, will lead to abnormally planning and operation conditions. There are two cases that could be expected: first, the day-ahead forecasted values are less than the real-time actual values and, the second one, the day-ahead forecasted values are greater than the real-time actual values. In first case, in order to maintain optimal operation, some wind generators should be curtailed, therefore, the unnecessary operations will increase the additional operating cost. In the second case, it is necessary to generate power from other sources, power sources that may not be available and in this case the generating system may not be able to entirely meet the real-time load. So, an accurately forecast wind speed is imperative necessary.

In the paper, an example of ARIMA process was presented in order to forecast the day-ahead wind speed values. It should be noted, that the method can't be applicable to all situations. So, the used method should be chosen considering more factors, such as the time frame, pattern of the data, cost of the forecasting, desired accuracy, availability of the data, and also the experience of operator.

ACKNOWLEDGEMENT

This paper was supported by the project PERFORM-ERA "Postdoctoral Performance for Integration in the European Research Area" (ID-57649), financed by the European Social Fund and the Romanian Government.

REFERENCES

[1] Matthias Lange, Ulrich Focken, Physical Approach to Short-Term Wind PowerPrediction, Berlin Heidelberg, Springer 2006.

[2] J.H.Chow, F.F.Wu, J.A.Momoh, Applied Mathematics for restructured electric power systems, New York: Springer, 2005.

[3] Theodor Popescu, Serii de timp, Editura Tehnică Bucureşti 2000.

[4] G.Janacek, L.Swift, Time series: forecasting, simulation, applications, West Sussex: Ellis Horwood Limited, 1993.

[5] I. Moghram and S.Rahman, "Analysis and evaluation of five short-term load forecasting techniques," IEEE Trans. on Power Systems, vol.4, no.4, pp. 1484-1491, Oct.1989.