Forecasting the wind speed process using higher order statistics and fuzzy systems

J. Antari, R. Iqdour et A. Zeroual

Department of Physics, Cadi Ayyad University, Faculty of Sciences
Semlalia, B.P. 2390, Marrakesh 40001, Morocco

(reçu le 01 Octobre 2006 - accepté le 21 Décembre 2006)

Abstract - *This investigation has two main objectives. The first one is to propose a statistical method, based on the fourth order cumulants, to identify single-input single-output 'SISO', finite impulse response 'FIR' system using non gaussian input, zero mean and independent identically distributed signals. The second objective is to search, as an application, a model for forecasting the wind speed time series and to compare the obtained results with those obtained using the Takagi-Sugeno 'TS' fuzzy techniques. The prediction results obtained by the proposed method show that the sequences of generated values have the same statistical characteristics as those really observed and better than those obtained using 'TS' fuzzy systems. Additionally, the model developed on the basis of the statistical method fits well wind speed time series and can be used for forecasting purpose with an accuracy of 94 % and above.*

Résumé - *Cette recherche a deux principaux objectifs. Le premier objectif est de proposer une méthode statistique, basée sur les cumulants d'ordre quatre, pour identifier le premier entrant et le premier sortant 'SISO' du système de réponse d'impulsion fini 'FIR', utilisant l'entrée non gaussienne, le zéro moyen et en signaux indépendants identiquement distribués. Le deuxième objectif est de rechercher, comme application, un modèle pour prédire la série chronologique de la vitesse de vent et de comparer les résultats obtenus à ceux obtenus en utilisant les techniques floues de Takagi-Sugeno 'TS'. Les résultats de prévision obtenus par la méthode proposée prouvent que les séquences valeurs générées ont les mêmes caractéristiques statistiques que ceux réellement observés et sont meilleurs que ceux obtenus en utilisant les systèmes flous de 'TS'. De plus, le modèle développé sur la base de la méthode statistique s'adapte correctement à la série chronologique de la vitesse de vent et peut être employé dans le but de prévision avec une précision de 94 % et plus.*

Keywords: Cumulants - Forecasting - FIR - Higher order statistics - Modelling - Wind speed time series.

1. INTRODUCTION

In the rapid pace of development of the civilization, the demand for energy has been steadily growing. A strong need has been felt to look at renewable sources to augment the power production with the known conventional sources that would otherwise perish in a few more decades. Wind has always attracted attention as a potential source of energy with its free availability and negligible impact on the environment [1-4].

The increased use of energy and the depletion of the fossil fuel reserves combined with the increase of the environmental pollution have encouraged the search for clean and pollution-free sources of energy. One of these is wind energy. This is a clean, inexhaustible and a 'free' source of energy that has served the mankind for many centuries by propelling ships, driving wind turbines to grind grains and for pumping water. Despite the high cost of wind power this may become a major source of energy in the years to come. This is so because the severe pollution of the planet originating from the burning of the fossil fuels.

The predicted variations of meteorological parameters such as wind speed, relative humidity, air temperature, etc. are needed in the renewable industry for design, performance analysis, and running cost estimation of these systems [5, 6]. Therefore the climatology is defined as a set of probabilistic statements on long-term weather conditions [7], and wind climatology as that branch of climatology that specializes in the study of winds, from which information on extreme winds is provided to structural designers. Such information is also needed for wind energy producers and engineers who design coastal civil structures, for example breakwaters. From a structural engineering point of view, forecasting the maximum wind speed that is expected to affect a structure during its lifetime is important to the designer. On the other hand, in coastal engineering practices, not only the magnitude but also the directionality of wind becomes important. The duration of wind, in addition to its magnitude and direction, is also required in wind energy production systems, and the amount of energy that can be produced depends upon it.

Several methods used in literature for solving the problem of the generating of the wind speed time series such as: alternative approaches used in the generation of simulated wind speed time series were used by Kaminsky *et al.* [8]. J. Horstmann *et al.* [9], Thiria *et al.* [10] and Richaume *et al.* [11] applied the Neural Networks to wind retrieval from spaceborne scatterometer (SCAT) data and European Remote Sensing ERS-1 SCAT data, Sfetsos [12] examined adaptive neuro-fuzzy inference systems and neural logic networks and compared them to the traditional autoregressive moving average (ARMA) models. Dukes and Palutikof [13] employed the Markov chain in order to estimate hourly mean wind speed with very long return periods. Another Markov chain based study was conducted by Sahin and Sen [14]. Castino *et al.* [15] coupled autoregressive processes to the Markov chain and simulated both wind speed and direction. I.A. Pérez *et al.* [1] was applying the autocorrelation function in atmospheric research and P. Ramirez, J.A. Carta [2] put the comparison between the maximum entropy principle and the Weibull distribution.

The wind speed, which is characterized by the presence of a stochastic component, is a non-stationary process, the analysis of wind speed and temperature is significant to describe momentum and heat exchanges and for wind energy applications. Accordingly, using Second Order Statistics (SOS) without performing a transformation on original data to have Gaussian and stationary processes may engender errors in prediction. For this reason we use the techniques of Higher Order Statistics (HOS) to develop a model able to predict the wind speed time series mainly from a random process. The HOS constitute a powerful tool in modelling non-stationary processes when the output signal of a system is known and corrupted with an additive non Gaussian noise. In order to evaluate the obtained model we compare it with the (TS) fuzzy model using the root mean square error (RMSE) and the index of agreement (d) between observed and predicted values.

2. MODEL AND ASSUMPTIONS

We consider the single-input single-output (SISO) model of the finite impulse response (FIR) system described by the following relationships:

Noise free case
$$x(n) = \sum_{i=0}^{q} h(i)e(n-1) \tag{1}$$

With noise
$$S(n) = x(n) + v(n) \tag{2}$$

where $\{e(n)\}$ is the input sequence, $\{h(i)\}$ is the impulse response coefficients, q is the order of FIR system, $\{x(n)\}$ is the output of system and $\{v(n)\}$ is the noise sequence.

The principal assumptions made on the model can be presented as follows:

A1: The input sequence $\{e(n)\}$ is independent and identically distributed (i.i.d) zero mean, the variance is $\sigma_e^2 \cong 1$, and non Gaussian.

A2: The system is causal, i.e. $h(i) = 0$ for $i < 0$ and $i > q$, and where $h(0) = 1$.

A3: The measurement noise sequence $\{v(n)\}$ is assumed to be zero mean, (i.i.d), Gaussian and independent of $\{e(n)\}$ with unknown variance.

3. BASIC RELATIONSHIPS

In this section, we present the general fundamental relations which permit to identify the FIR linear systems (MA model) using Higher Order Cumulants (HOC).

The m[th] order cumulants of the $\{x(n)\}$ can be expressed as a function of impulse response coefficients $\{h(i)\}$ as follows [16]:

$$C_{mx}(t_1,...,t_{m-1}) = \gamma_{me} \sum_{i=0}^{q} h(i)h(i+t_1)....h(i+t_{m-1}) \tag{3}$$

with γ_{me} represents the m[th] order cumulants of the excitation signal at origin.

If we take $m = 2$ into Eq. (3) we obtain the second order cumulant (AutoCorrelation Function (ACF)):

$$C_{2x}(\tau) = \gamma_{2e} \sum_{i=0}^{q} h(i)h(i+\tau) \tag{4}$$

For $m = 4$, Eq. (3) becomes:

$$C_{4x}(\tau_1,\tau_2,\tau_3) = \gamma_{4e} \sum_{i=0}^{q} h(i)h(i+\tau_1)h(i+\tau_2)h(i+\tau_3) \tag{5}$$

The Fourier transforms of the 2[nd] and 4[rd] order cumulants are given respectively by the following equations:

$$S_{2x}(\omega) = TF\{C_{2x}(\tau)\} = \gamma_{2e} \sum_{i=0}^{q} \sum_{\tau=-\infty}^{+\infty} h(i)h(i+\tau)\exp(-j\omega\tau) = \gamma_{2e} H(-\omega)H(\omega) \tag{6}$$

with $H(\omega) = \sum_{i=0}^{+\infty} h(i)\exp(-j\omega\tau)$.

$$S_{4x}(\omega_1,\omega_2,\omega_3) = TF\{C_{4x}(\tau_1,\tau_2,\tau_3)\} = \gamma_{4e} H(-\omega_1-\omega_2-\omega_3)H(\omega_1)H(\omega_2)H(\omega_3) \tag{7}$$

So, if we take $\omega = \omega_1 + \omega_2 + \omega_3$, the equation (6) becomes:

$$S_{2x}(\omega_1+\omega_2+\omega_3) = \gamma_{2e} H(-\omega_1-\omega_2-\omega_3)H(\omega_1+\omega_2+\omega_3) \tag{8}$$

then, from the Eqs, (7) and (8) we construct a relationship between the spectrum, the bispectrum and the parameters of the output system:

$$S_{4x}(\omega_1,\omega_2,\omega_3)H(\omega_1+\omega_2+\omega_3) = \mu_{(4,2)} H(\omega_1)H(\omega_2)H(\omega_3)S_{2x}(\omega_1+\omega_2+\omega_3) \tag{9}$$

with $\mu_{(4,2)} = (\gamma_{4e} / \gamma_{2e})$.

The inverse Fourier transform of the Eq. (9) is:

$$\sum_{i=0}^{q} C_{4x}(t_1-i,t_2-i,t_3-i)h(i) = \mu_{(4,2)} \sum_{i=0}^{q} h(i)h(t_2-t_1+i)h(t_3-t_1+i)C_{2x}(t_1-i) \tag{10}$$

Based on the relationship (10) we can develop the following algorithm (& 4) based on the Higher Order Statistics (HOC).

4. IDENTIFICATION METHODS

4.1 Proposed method based on HOC [17]

If we take $t_1 = t_3$ into Eq. (10) we obtain:

$$\sum_{i=0}^{q} C_{4x}(t_1-i, t_2-i, t_3-i)h(i) = \mu_{(4,2)} \sum_{i=0}^{q} h(i)h(t_2-t_1+i)h(i)C_{2x}(t_1-i) \quad (11)$$

If we use the ACF property of the stationary process (such as $C_{2x}(t) \neq 0$ only for $-q \leq t \leq q$ and vanishes elsewhere) and if we suppose that $t_1 = 2q$ the Eq. (11) becomes:

$$\sum_{i=0}^{q} C_{4x}(2q-i, t_2-i, 2q-i) = \mu_{(4,2)} h^2(q)h(t-q)C_{2x}(q) \quad (12)$$

If Eq (12) is causal (i.e. $h(i) = 0$ for i < 0 and i > q), the choice of t_2 imposes that $(t_2 \geq q)$. So, this implies $0 \leq t_2 - q \leq q$. For this reason, we have $t_2 = q, q+1, ..., 2q$.

If we take $t_1 = t_2 = -q$ into the Eq (11), we obtain the following equation:

$$\sum_{h=1}^{q} C_{4x}(-q-i, -q-i, -q-1)h(i) = \mu_{(42)} \sum_{i}^{q} h(i)h(i)h(i)C_{2x}(-q-i)) \quad (13)$$

According to the ACF property the relation (13) becomes:

$$C_{4x}(-q,-q,-q)h(0) = \mu_{(4,2)} h^3(0)C_{2x}(-q) \quad (14)$$

with $h(0) = 1$ we obtain:

$$C_{4x}(-q,-q,-q) = \mu_{(4,2)} C_{2x}(-q) \quad (15)$$

Using the property of the cumulants: $C_{4x}(t_1, t_2, t_3) = C_{4x}(-t_1, t_2-t_1, t_3-t_1)$. So the equation (15) becomes:

$$C_{4x}(q,0,0) = \mu_{(4,2)} C_{2x}(q) \quad (16)$$

So, we based on Eq. (16) for eliminating $C_{2x}(q)$ in Eq. (12), we obtain the equation constituted of only the fourth order cumulants:

$$\sum_{i=0}^{q} C_{4x}(2q-i, t_2-i, 2q-i)h(i) = h^2(q)h(t_2-q)C_{4x}(q,0,0) \quad (17)$$

To simplify the Eq. (17), we consider equation (5) with $\tau_1 = \tau_2 = q$ and $\tau_1 = \tau_2 = q, \tau_3 = 0$, we obtain respectively the relationships:

$$C_{4x}(q,q,\tau_3) = \mu_{(4,2)} h^2(q)h(\tau_3) \quad (18)$$

$$C_{4x}(q,q,0) = \mu_{(4,2)} h^2(q)h(0) \quad (19)$$

From Eqs. (18), (19) and $\tau_3 = q$, we obtain:

$$h(q) = \frac{C_{4x}(q,q,q)}{C_{4x}(q,q,0)} \quad (20)$$

From Eq. (20) we obtain the following form:

$$h^2(q)C_{4x}(q,0,0) = \left(\frac{C_{4x}(q,q,q)}{C_{4x}(q,q,0)}\right)^2 C_{4x}(q,0,0) = \alpha \tag{21}$$

Using the Eqs (21) and (17) we obtain the proposed algorithm based only on fourth order cumulants:

$$\sum_{i=0}^{q} C_{4x}(2q-i, t_2-i, 2q-i)h(i) = \alpha h(t_2-q) \tag{22}$$

The system of Eq. (22) can be written under the matrix form as follows:

$$\begin{bmatrix} C_{4x}(2q-1,q-1,2q-1) & . & . & . & C_{4x}(q,0q) \\ C_{4x}(2q-1,q-1,2q-1)-\alpha & . & . & . & . \\ 0 & . & . & . & . \\ . & 0 & . & . & . \\ 0 & . & 0 & . & C_{4x}(q,q,q) \end{bmatrix} \begin{bmatrix} h(1) \\ . \\ . \\ . \\ h(q) \end{bmatrix} = \begin{bmatrix} \alpha - C_{4x}(2q,q,2q) \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \tag{23}$$

Or in more compact form, the Eq. (23) can be written as follows:

$$A\,\theta = b \tag{24}$$

With A the matrix of size $(q+1,q)$ elements, θ a column vector of size $(q,1)$ and b is a column vector of size $(q+1,1)$. The Least Square (LS) solution of the system of equation (24) is given by:

$$\theta = \left(A^T A\right)^{-1} A^T b \tag{25}$$

with $(.)^T$ represents the transpose of $(.)$.

4.1.1 Algorithm test

In this subsection we test th performance of the proposed method (Eq. (25)) before to be applied to the real data (wind speed time series), for this reason we use for example the simulation of the MA(2) model given by the following equations: $x(n) = e(n) + 1.75e(n-1) - e(n-2)$, in noise free case, $S(n) = x(n) + v(n)$ with noise.

The simulation results are illustrated in Table 1 using different sample sizes (N = 300, 600, 900, 1200) with signal to noise ratio (SNR = 40 dB) and for 40 Monte-Carlo runs.

The SNR is defined by: $SNR = 10\log_{10}\left(\sigma_x^2 / \sigma_v^2\right)$. σ_x^2 and σ_v^2 represent respectively the variance of the output system and noise signal.

Table 1: Estimated parameters using proposed method (Eq. 25) for SNR = 40 dB, different sample sizes and for 40 Monte Carlo runs

N	Estimated parameters ± std	Proposed method
300	$\hat{h}(1) \pm$ std	1.7205 ± 0.3148
	$\hat{h}(2) \pm$ std	–0.9478 ± 0.4127
600	$\hat{h}(1) \pm$ std	1.7746 ± 0.2547
	$\hat{h}(2) \pm$ std	–1.0825 ± 0.1487
900	$\hat{h}(1) \pm$ std	1.7687 ± 0.0460

	$\hat{h}(2) \pm$ std	-0.9745 ± 0.0214
1200	$\hat{h}(1) \pm$ std	1.7401 ± 0.0238
	$\hat{h}(2) \pm$ std	-0.9857 ± 0.0214

True values are: $h(1) = 1.75$ and $h(2) = -1$

From **Table 1** we can note that:

The estimate parameters using the proposed method (Eq. (25)) are near to the true parameters and the values of the standard deviation (std) demonstrate the small fluctuation around the mean parameters. So, we can apply the proposed method (Eq. (25)) for generating the real data, in this study we try to predict the wind speed time series (&.5).

4.2 Takagi-Sugeno fuzzy systems

Lets $Z = \{z_t\}$ to be the database representing the set of the available observations $z_t = (x_t, y_t)$ $(t = 1,2,...,N)$. The Takagi-Sugeno fuxzzy model (TS) consists of aggregating of a fuzzy rules R_k $(k = 1,2,...,c)$ with the following structure [18]:

$$R_k : \textbf{If } x_t \text{ is } A_k \textbf{ Then } \hat{y}_{t,k} = \beta_0 + x_t' \beta_k \quad k = 1,2,...,c \text{ and } \quad t = 1,2,...,N \tag{26}$$

R_k $(k = 1,2,...,c)$ indicates k^{th} fuzzy role, x_t is the input variable ($x_t \in R^n$), $\hat{y}_{t,k}$ is the output of the rule k relative to the input x_t and A_k is a fuzzy set and $\beta_k = (\beta_1, \beta_2, ..., \beta_n)$.

The output $\hat{y}_t$ relative to the input x_t obtained after aggregating of c TS fuzzy rules, can be written as a weighted sum of the individual conclusions:

$$\hat{y}_t = \sum_{k=1}^{c} \pi_k(x_t)\hat{y}_{t,k} \qquad \text{where} \qquad \pi_k = \frac{\mu_{A_k}(x_t)}{\sum_{j=1}^{c} \mu_{A_j}(x_t)} \tag{27}$$

where μ_{A_k} is the membership function related to the fuzzy set A_k.

The membership functions are selected Gaussian types [19]:

$$\mu_{A_k}(x_t) = \exp\left(-\frac{1}{2}\| x_t - m_k \|_{S_k}^2\right) \tag{28}$$

$$\| x_t - m_k \|_{S_k}^2 = (x_t - m_k)' S_k' S_k (x_t - m_k) \tag{29}$$

The centers m_k and matrix S_k are initialized by projection of the partition obtained from GK algorithm:

$$S_k = \left(F_k^{(x)}\right)^{-1/2} \qquad \text{and} \qquad m_k = V_k^{(x)} \tag{30}$$

$F_k^{(x)}$ and $V_k^{(x)}$ are the projections of the variance covariance matrix, and cluster centers, k respectively on the input space.

The identification of the TS fuzzy systems requires two types of tuning [19, 18]:

Structural tuning: concerns the determination of the number of rules c and the fuzzy sets A_k to be used in the fuzzy system. For that we have used the Gustafson-Kessel (GK) fuzzy clustering algorithm with the following fuzzy validity criterion:

$$S(c) = \sum_{t=1}^{N} \sum_{k=1}^{c} (\mu_{kt})^m \left(\| z_t - v_k \|^2 - \| v_k - \bar{z} \|^2 \right) \quad (31)$$

Where z_t is the t^{th} data point, v_k is the centre of the k^{th} cluster, $\bar{z}$ is the average of data and m is the fuzzification exponent. So the optimum number of clusters is determined by increasing the c parameter and looking for the minimum of the fuzzy validity criterion $S(c)$.

Parametric tuning: the model parameters (linear and non linear) are estimated. The goal of the parameters optimization is to find the 'best' approximation $\hat{y}_t$, to the measured output y_t. The linear parameters β_k are identified using the Global Least Squares (GLS) algorithm, while the Levenberg-Marquardt (LM) algorithm is using to estimate the non linear parameters (S_k and m_k).

5. APPLICATION TO MODELING THE WIND SPEED TIME SERIES

5.1 Data analysis

In this section, we use the proposed algorithm (i.e., Eq. 25) for modelling the wind speed times series $\{y_j\}$, $j = 1,2,....,400$.

From the (Fig. 1) we note a linearity progression; this demonstrates that the process is non-stationary.

Before analyzing the time series $y(j)$ and in order to use the proposed algorithm based on fourth-order cumulants, it is necessary to verify if the time series is non Gaussian. For this reason, we plot the Normal Probability (Fig. 2) of the time series $y(j)$.

From the curve (Fig. 2), we can conclude that the wind speed times series is non-Gaussian and then the proposed algorithm can be used to identify the process $y(j)$.

In order to make this phenomenon in evidence, we calculate the AutoCorrelation Function (ACF) (Fig. 3) of the wind speed time series. The low decrease of the ACF confirms that the process is non-stationary [20]. So, it is necessary to transform the time series $\{y(j)\}$ to a stationary process using differentiation operator ∇: $z(j) = \nabla y(j) = (1 - B)^d y(j)$, where B is the backward shift operator and d is the order of differentiation. The transformed time series $z(j)$ is plotted in (Fig.4) (in our case $d = 1$) has the behaviour of stationary process.

5.2 Model selection for forecasting wind speed times series

The ACF (Fig. 5) and the fourth order diagonal cumulant (Fig. 6) of the time series $\{z(j)\}$ obtained after application of the operator ∇ (paragraph 5.1) demonstrate that the time series $\{z(j)\}$ seems to be stationary.

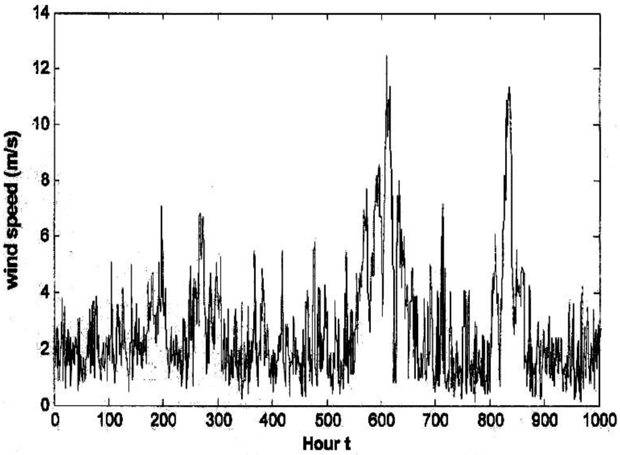

Fig. 1: Wind speed times series $y(j)$

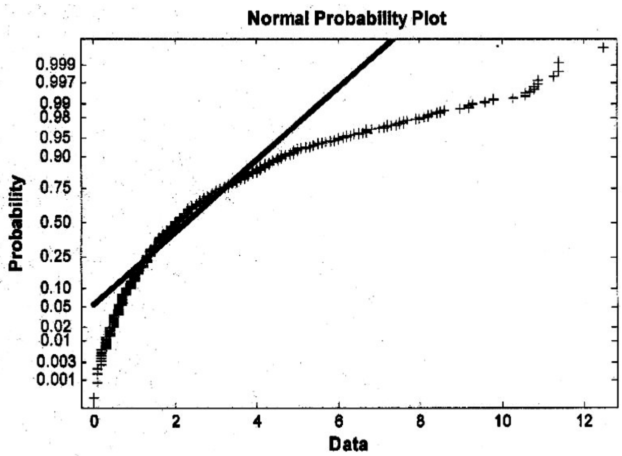

Fig. 2: Normal probability of the wind speed times series $y(j)$

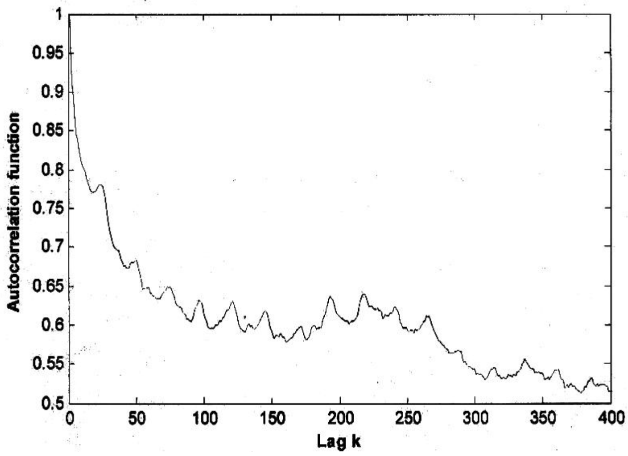

Fig. 3: ACF of the wind speed times series $y(j)$

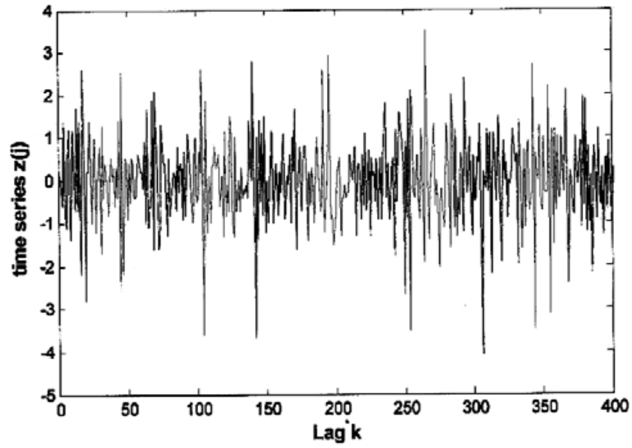

Fig. 4: Transformed wind speed times $z(j)$

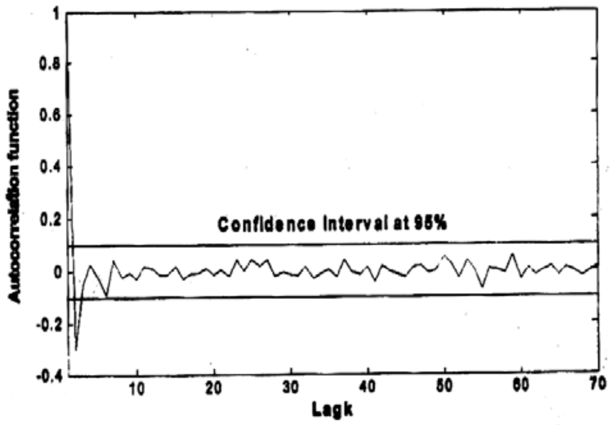

Fig. 5: ACF of the transformed wind speed times series $z(j)$

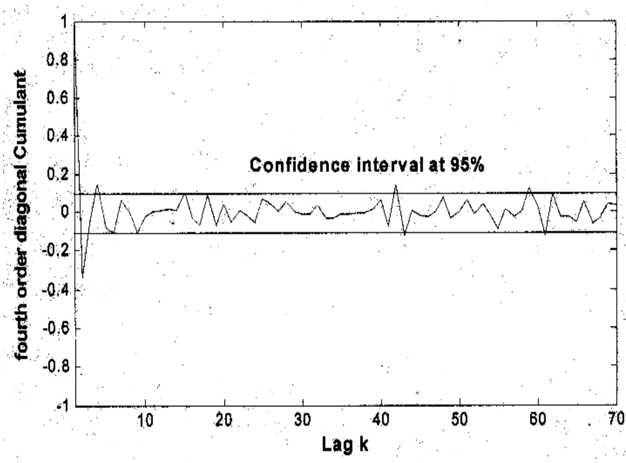

Fig. 6: Fourth order diagonal cumulant of the transformed wind speed times series $z(j)$

The selection of the appropriate model able to represent the time series $\{z(j)\}$ can be made by the following procedure:

1. The two Figures (5 and 6) of the ACF and the fourth order diagonal cumulant obtained using the proposed algorithm (Eq. (25)) show that the model which can represent the transformed wind speed times series $z(j)$ is a MA model with an order 2 or 3.

2. The selection of the signal input variance is done by the relation combining the excitation variance and the autocorrelation of the output (Eq. (3)). We have $\sigma_{2e} = \frac{C_{2z}(0)}{\sum_{j=0}^{q} h^2(j)}$, then $\sigma_{2e} < C_{2z}(0)$. In our case $C_{2z}(0) = 1.2463$. The procedure begins from the input excitation values equal to $\sigma_{2e} = 1.2380$, then we search the optimal variance as follows:

We estimate $\hat{h}(j)$ parameters of the model (**Table 2**), and from $\hat{h}(j)$ we:

- Generate the transformed wind speed time series $\hat{z}(j)$.
- Estimate the ACF of the generated time series.
- Estimate the fourth order diagonal cumulant of the generated time series.

Table 2: Estimated parameters of the model MA (3)

$\hat{h}(1)$	$\hat{h}(2)$	$\hat{h}(3)$	$\hat{\sigma}_e^2$
0.4690	1.5978	-0.0297	1.5231

The Figs 7 and 8 represent respectively the ACF and the fourth order diagonal cumulant using the proposed algorithm (Eq. (25)). The results show that the transformed wind speed times series $z(j)$ and generated $\hat{z}(j)$ time series are not markedly different.

The Fig. 9 represents the ACF of the residuals between the transformed $z(j)$ and generated $\hat{z}(j)$ time series by the model MA (3) (**Table 2**). So, the model is statistically acceptable because the residual time series can be described by a while noise.

In Fig. 10 we represent the histograms of the transformed $z(j)$ and generated $\hat{z}(j)$ time series obtained by the model MA (3) (**Table 2**). It is important to note that there is a slight difference between them. So, the model appears to be correct and have an acceptable accuracy for the description of the wind speed lime series.

The scatter diagram plotted in Fig. 11 presents a comparison between measured and predicted time series obtained by MA (3) which constitutes another mean to test the performance of the model. So, the distribution of the values around the bisectrix informs us on the validity of the model MA (3). Furthermore, the Fig. 12 shows the evolution of the measured and the predicted (based on the predictor method given in [21]) wind speed time series using the proposed model MA (3). We observe that there is almost a complete agreement between the two time series, this implies that the MA (3) can be used for forecasting the wind speed time series.

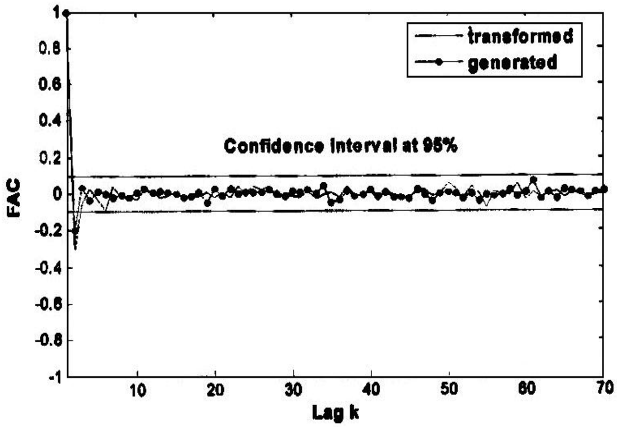

Fig. 7: ACF of the wind speed transformed and generated time series

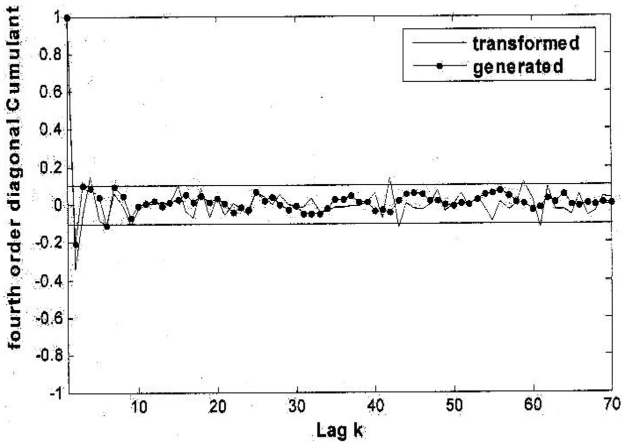

Fig. 8: Fourth order diagonal cumulant transformed and generated wind speed time series by the proposed algorithm

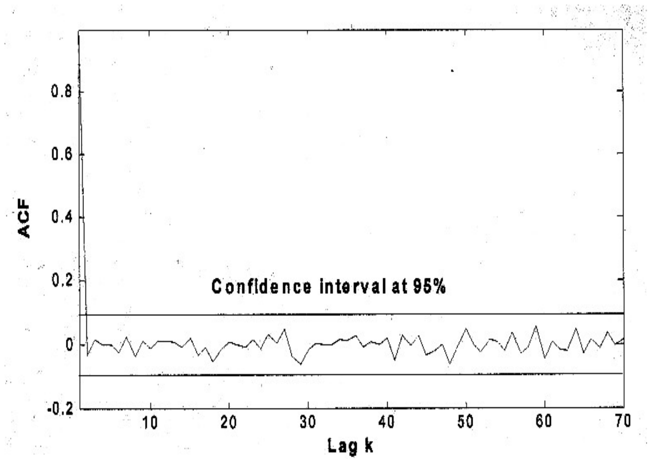

Fig. 9: Autocorrelation function of the residual time series

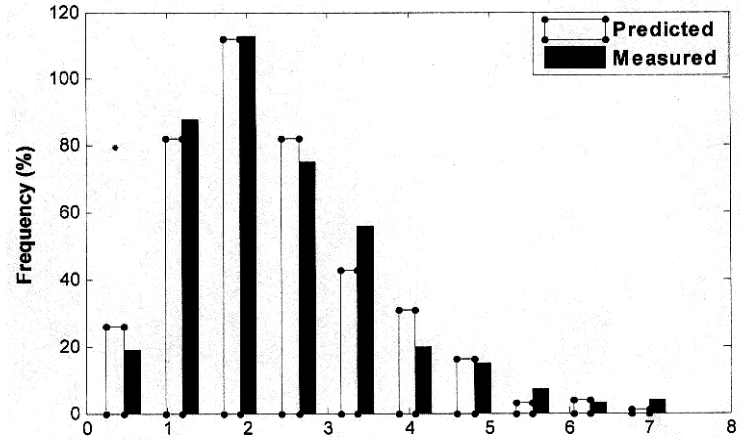

Fig. 10: Histogram of measured and prediced wind speed time series

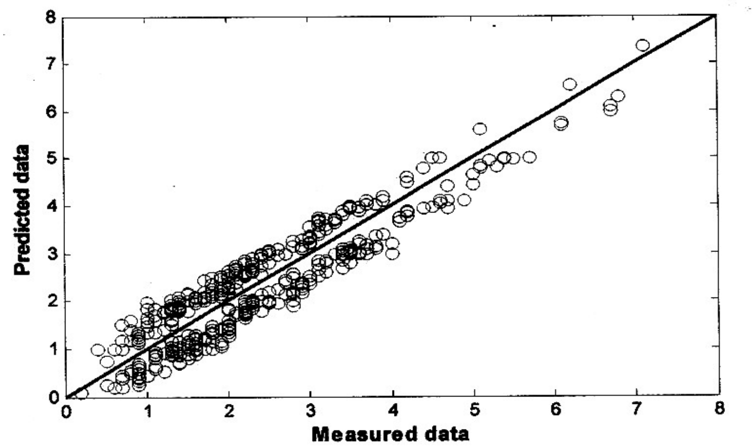

Fig. 11: Scatter diagram of measured and predicted wind speed time series

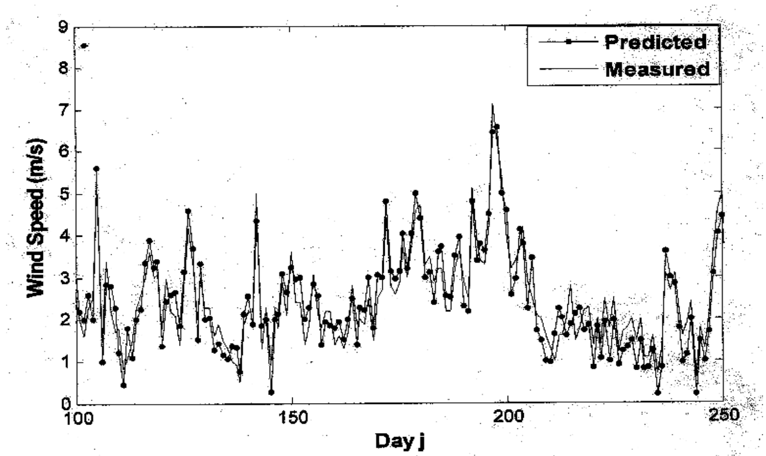

Fig. 12: Measured and predicted the wind speed time series

5.3 Comparison of the proposed model and the TS fuzzy model

In this section we measure the statistical indicator: Root Mean Square Error (RMSE) which is the difference between observed and estimated values. It is used to evaluate the performance of the proposed model and the TS fuzzy model during the training phase.

RMSE is computed by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_t - \hat{y}_t\right)^2} \tag{32}$$

where y_t is the original time series, $\hat{y}_t$ is the predicted time series and N is the number of pattems.

In the test phase, the performances of the identified models were evaluated by calculating the RMSE. In addition, an index of agreement between observed and predicted values (d) is calculated by:

$$d = 1 - \frac{\sum_{i=1}^{N}\left(\hat{y}_t^{'} - y_t^{'}\right)^2}{\sum_{i=1}^{N}\left(\left|\hat{y}_t^{'}\right| - \left|y_t^{'}\right|\right)^2} \tag{33}$$

where $\hat{y}_t^{'} = \hat{y}_t - \overline{y}_t$ and $y_t^{'} = y_t - \overline{y}_t$.

The index d can be any value between 0 and 1, the nearer d is to 1 the better is the agreement between observed and predicted values.

Table 3: Values of the RMSE and d criterion in the test

	RMSE	d (%)
Proposed algorithm	0.8425	94.32
Fuzzy systems	0.9426	92

From **Table 3**, we can note that:

The RMSE values are 0.8425 m/s and 0.9426 m/s respectively for the proposed model and the TS fuzzy model.

The adequacy of fit is also assessed using the agreement index, which returns the percentage of similarity between the measured and the predicted data of the wind speed. Variations in agreement index (d) from 94 % to 92 % are obtained. These results are considered as an indicator of the adequate correctness of the prediction, which strengthens the robustness of the proposed and TS fuzzy models. But it is important to note that the identified proposed model is more accurate than the developed TS fuzzy model.

6. CONCLUSION

There were two main objectives of this investigation. The first was to develop a new algorithm based on the fourth order cumulants for estimating the parameters of the model MA, using q+1 equations to estimate q parameters and we test the performance of the proposed algorithm using different sample sizes over 40 Monte-Carlo runs. The second was to test the accuracy of the proposed algorithm for forecasting the wind speed time series and comparing it with the TS fuzzy systems. The obtained results show that the MA (3) is a satisfactory model for a good prediction of wind speed time series using non Gaussian input signal with fixed variance. Moreover, the MA (3) is efficient and gives RMSE values and d better than the TS fuzzy model.

Finally, the model developed can be used to generate a wind speed time series having the same statistical features as the measured data, but requires:

1. A purely random variable $e(j)$ (with non Gaussian noise), with no correlated samples and with zero mean;

2. The coefficient of the moving average model describing the stochastic component e (j);

3. An integration of the predicted time series $\hat{z}(j)$.

REFERENCES

[1] I.A. Pérez, M.A. Garcia, M.L. Sanchez and B. de Torre, '*Fit of Wind Speed and Temperature Profiles in the Low Atmosphere from Rass Sodar Data*', Journal of Atmospheric and Solar-Terrestrial Physics, Vol. 68, pp. 1125 – 1135, 2006.

[2] P. Ramirez and J.A. Carta, '*The Use of Wind Probability Distributions Derived From the Maximum Entropy Principle in the Analysis of Wind Energy, A Case Study*', Energy Conversion and Management, Vol. 47, pp. 2564 – 2577, 2006.

[3] Y.Q. Xiao, Q.S. Lia, Z.N. Li, Y.W. Chow and G.Q. Li, '*Probability Distributions of Extreme Wind Speed and its Occurrence Interval*', Engineering Structures, Vol. 28, pp. 1173 – 1181, 2006.

[4] M.M. Shokrieh and R. Rafiee, '*Simulation of Fatigue Failure in a Full Composite Wind Turbine Blade*', Composite Structures, Vol. 74, pp. 332 – 342, 2006.

[5] S. Katogirou, C. Neocleous, S. Michaelides and C. Schizas, '*Artificial Neural Networks for the Generation of Isohyets by Considering Land Configuration*', Proceedings of the Engineering Applications of Neural Networks (EANN'98) Conference, Gibraltar, pp. 383 – 389, 1998a.

[6] S.A. Kalogirou, C.C. Neocleous and C.N. Schizas, '*Artificial Neural Networks in Modelling the Starting-up of a Solar Steam Generation Plant*', Applied Energy, Vol. 60, N°2, pp. 89 – 100, 1998.

[7] E. Simiu and R.H. Scanlan, '*Wind Effects on Structures*', New York, John Wiley & Sons, 1986.

[8] F.C. Kaminsky, R.H. Kirchhoff, C.Y. Syu and J.F. Manwell, '*A Comparison of Alternative Approaches for the Synthetic Generation of a Wind Speed Time Series*', Transactions of the ASME, Vol. 113, pp. 280 -289, 1991.

[9] J. Horstmann, H. Schiller, J. Schulz-Stellenfleth and S. Lehner, '*Global Wind Speed Retrieval from SAR*', IEEE Transactions on Geosciences and Remote Sensing, Vol. 41, N°10, October 2003.

[10] S. Thiria, F. Badran, C. Mejia, and M. Crépon, '*A Neural Network Approach for Modelling non Linear Transfer Functions: Application for Wind Retrieval from Spaceborne Scatterometer Data*', J. Geophys. Res., Vol. 98, pp. 22827 - 22841, 1993.

[11] P. Richaume, F. Badran, M. Crepon, C. Mejia, H. Roquet and S. Thiria, '*Neural Network Wind Retrieval from ERS-1 Scatterometer Data*', J. Geophys. Res., Vol. 105, pp. 8737 -8751, 2000.

[12] A. Sfetsos, '*A Comparison of Various Forecasting Techniques Applied to Mean Hourly Wind Speed Time Series*', Renewable Energy, Vol. 21, pp. 23 – 35, 2000.

[13] M.D.G. Dukes and J.P. Palutikof, '*Estimation of Extreme Wind Speeds with Very Long Return Periods*', Journal of Applied Meteorology, Vol. 34, pp. 1950 – 1961, 1995.

[14] A.D. Sahin and Z. Sen, '*First-Order Markov Chain Approach to Wind Speed Modelling*', Journal of Wind Engineering and Industrial Aerodynamics, Vol. 89, pp. 263 – 269, 2001.

[15] F. Castino, R.Festa and C.F. Ratto, '*Stochastic Modelling of Wind Velocities Time Series*', Journal of Wind Engineering and Industrial Aerodynamics, Vol. 74-76, pp. 141 – 151, 1998.

[16] D. Boss, B. Jelonnek and K.D. Kammeyer, '*Decision-Feed Back Eigenvector Approach to Blind ARMA Equalization and Identification*', Proceedings IEEE-SP/ATHOS Workshop on Higher Order Statistic, Gerona, Spain, 12 - 14 June, 1995.

[17] J. Antari, A. Zeroual and S. Safi, '*Blind Estimation of non-Gaussian Finite Impulse Response (FIR) Systems Using Higher Order Statistics*', Second International Symposium on Communications, Control and Signal Processing ISCCSP, pp. 13–15, Marrakech, Morocco, March 2006.

[18] T. Takagi and M. Sugeno, '*Fuzzy Identification of Systems and its Application to Modelling and Control*', IEEE Trans. Systems Man Cybernet, Vol. 15, pp. 116 - 132, 1985.

[19] A. Fiordaliso, '*Systèmes Flous et Prévision de Séries Temporelles*', Hermès Science, 1999.

[20] P.J. Brokwell and R.A. Davis, '*Time Series Theory and Methods*', 2nd Ed. Springer-Verlag, New York, Berlin, Heidelberg, London, 1990.

[21] G. Box and G. Jenkins, '*Time Series Analysis Forecasting and Control*', Holden-Day, Oakland, California, 1976.