Forecasting Wind Turbines Failure Data Using Advanced Fuzzy Time Series

João L. R. Silva
P&D MSc. Researcher
PPGEE - UFMG

Gabriel A. C. Lima
P&D PhD. Researcher
Aremas

Fernanda C. Marinho
Maintenance Engineer
Brookfield Renewable

Augusto Prestes
Maintenance Eng. Manager
Brookfield Renewable

***Abstract*—This paper proposes the study of advanced fuzzy time series to forecast wind turbines times-to-failure data. Failure data predictions makes significantly influence to the availability and reliability of an electrical system. The standards reliability models usually used to predict stochastically these events are mathematical and statistical complex and the use of fuzzy time series brings new fields to reduce this complexity and time processing.**

***Index Terms*—Wind Turbines, Failure Data, Fuzzy Time Series, Forecasting Models, Reliability Theory.**

I. Introduction

The continuous growing of power system in the world brings lots of challenges in the fields of power systems stability, power quality, system protections and other fields. To deal with this challenge, some reliability evaluation methods are developed to study and predict failures in an electric system. The reliability evaluation methods are today's state of the art to predict failure events [1] [2] [3].

The use of fuzzy time series for solving problems which data can be interpreted as vague, linguistic, fuzzy values or numbers was introduced first by Song and Chissom in 1993 [4] [5] [6] [7]. The advantages of using fuzzy time series lies in the possibility to be implemented by algorithms, are easy to improve algorithms to increase forecast accuracy and no need to have strong background in mathematics and statistics.

The set of times-to-failure data from onshore wind turbines are forecasting using the fuzzy time series and then compared with state of the art reliability models usually used to predict them. The results obtained in this work shows that the use of advanced fuzzy time series reduces the mathematical and statistics complexity of the models, brings less machine time consumption and better forecast accuracy than standard reliability models.

II. Wind Turbines Reliability

The technology of wind turbines for generating electricity dates back to the end of nineteenth century. Modern large wind turbines developments date back to work in Europe and the United States, later stimulated by oil price rises after the 1973 Yom Kippur War [8].

Wind turbines failures are normally resulting of an unacceptable operational condition, such as an over-temperature, over-speed or pitch problem. Exceptions may occur due to gearbox, generator or blade failures [8]. The basic structure of a modern three-blade, upwind horizontal axis wind turbine is exemplified by figure 1.

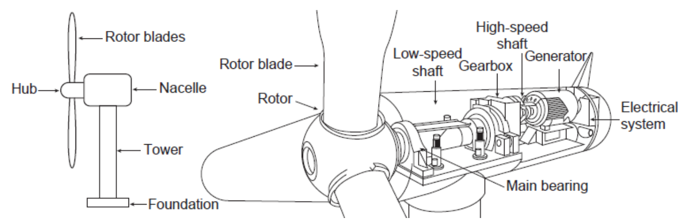

Fig. 1. Wind turbine layout and terminology.

The stochastic nature of mechanical or electrical equipment failures, significantly influence the availability of a wind turbine. The importance of knowing this availability has motivated several studies of reliability, since consumers, mainly commercial and industrial sectors, are increasingly sensitive to profit losses associated with unavailability of electrical system. An accurate method to predict failure events is the use of probability distributions for the times-to-failure data vector (1). Time series are also an alternative method and in this paper it will be studied the use of fuzzy time series with reliability evaluation to better predict these times-to-failure events.

$$T_E = [T_{E_1}, T_{E_2}, ..., T_{E_n}] \tag{1}$$

III. Reliability Theory

Life data analysis is a field of reliability engineering concerned to determine the probability of failures for a given set of failure data. Life data can be lifetimes of equipments or products in the marketplace and can be expressed as the time equipment/product operated successfully or the time that equipment/product operated before it failed. These lifetimes can be measured in days, hours, miles, cycles-to-failure or any other metric with which the life or exposure of a product can be measured [9].

The basis of life data analysis is supported in the study and knowledge of the statistical distribution and its parameters. There are several distributions functions used in life data analysis and reliability engineering, the most used are exponential, normal, lognormal and Weibull. The probability density function of the Weibull distribution (2) is one of the most used in reliability engineering. The reason for that is due to the versatility of this distribution in relation to the others distributions such as exponential, normal and lognormal [9].

The probability density function, cumulative distribution and reliability functions for Weibull distribution are respectively given by (2) and (3):

$$f(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{2}$$

$$F(t) = 1 - R(t) = 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{3}$$

where:

$f(t)$:	probability density function;
$F(t)$:	cumulative distribution function;
$R(t)$:	reliability function;
t:	time;
β:	shape parameter;
η:	scale parameter;
γ:	minimum life.

The Weibull shape parameter are correlated with the various stages of equipment life, i.e, early life ($\beta < 1$), useful life ($\beta = 1$) and Wear-out life ($\beta > 1$). The scale parameter provides information on the occurrence of a failure when it happen with probability of 63.2%.

To obtain the parameters for a given distribution and a set of data, the maximum likelihood method:

$$\Lambda = \ln(L) = \sum_{i=1}^{R} \ln f(x_i; \theta_1, \theta_2, ..., \theta_k) \tag{4}$$

from the statistical point of view, is considered the most robust method to estimate these parameters. The main idea about this method stands in determining the most probable parameters values of functions that fits a given set of data [9] [10] [11]. The parameters are obtained when the maximum value for Λ is reached using some class optimizations algorithms like Nelder-Mead, Ellipsoidal and others [12] [13] [14].

IV. Fuzzy Time Series

Time series is a set of data indexed in time order or a collection of observations made sequentially in time [15]. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are the mean daily temperature of a city, crude oil price per barrel and so on. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time series forecasting is the use of a model to predict future values based on previously observed values [16]. Time series could be divide in two types, i.e., i-Deterministic Time Series, where the pattern of the component of time series almost fixed as time passed, and ii-Stochastic Time Series where the pattern of the component of time series changing during time [15]. Many methods to predict time series were developed during decades, the main methods are the Naive model, Mean method, Moving Average Smoothing, Weighted Mean method, Exponential Smoothing, Additive Holt-Winters method, ARMA and ARIMA models.

Fuzzy time series (FTS) combines the fuzzy set theory developed by Zadeh in 1965 with time series analysis. According [17], the fuzzy set theory is being applied into wider and wider areas, such as decision making, planning, logic, systems theory, artificial intelligence, economics, control theory and so on. The main reason for using fuzzy sets instead of crisp sets is the advantage to deal with formal, powerful and quantitative framework to cope with the vagueness of human knowledge as it is expressed by means of natural languages. The methodology around FTS are discussed briefly during the explanation of these FTS methods described in items from IV-A to IV-D. The results obtained will be discussed in details on section V.

A. Yu FTS Algorithm

The Yu FTS algorithm or Weighted FTS is very similar to the Chen FTS algorithm [15] [18]. A complete description of this algorithm could be viewed in [19]. The steps to implement the algorithm are presented below.

Step 1. Define the universe of discourse:

$$U = \left[D_{min} - D_1, D_{max} + D_2\right] \tag{5}$$

based on historical data, min and max values of the data set and proper numbers D_1 and D_2 to adjust the range of data. Another way to define the universe of discourse could be implement using k-means clustering as presented in [20].

Step 2. Partition the universe of discourse (5) into n-intervals with equal length:

$$U = \left[u_1, u_2, ..., u_n\right] \tag{6}$$

The number of intervals could be defined by experience of the analyst or by some analytical, e.g., Huarng method or evolutionary algorithms that find the optimum number of intervals [15].

Step 3. Define fuzzy sets on the universe U. The fuzzy sets:

$$A_i = f_{Ai}(u_1)/u_1 + f_{Ai}(u_2)/u_2 + ... + f_{Ai}(u_b)/u_b \tag{7}$$

could be defined using some linguistic values such as A_1 = (not many), A_2 = (not too many), A_3 = (many), A_4 = (many many), A_5 = (very many), A_6 = (too many), A_7 = (too many many). In this paper the fuzzy sets are defined according to:

$$\begin{aligned} &A_1 = \{u_1/1, u_2/0.5, ..., u_{n-1}/0, u_n/0\} \\ &A_2 = \{u_1/0.5, u_2/1, ..., u_{n-1}/0, u_n/0\} \\ &\vdots \\ &A_n = \{u_1/0, u_2/0, ..., u_{n-1}/0.5, u_n/1\} \end{aligned} \tag{8}$$

Step 4. Fuzzify historical data. In some FTS methods the assign values of the fuzzy sets is subjective. In this paper a modular function was adopted as presented in:

$$f_A(m,n) = \begin{cases} 1.0, & if\ D(m) \geq U(1,n) \wedge \\ & D(m) \leq U(2,n); \\ 0.9, & if\ (\Delta > 0.50\Delta u) \wedge \\ & (\Delta \leq 0.75\Delta u); \\ 0.8, & if\ (\Delta > 0.75\Delta u) \wedge \\ & (\Delta \leq 1.00\Delta u); \\ 0.7, & if\ (\Delta > 1.00\Delta u) \wedge \\ & (\Delta \leq 1.25\Delta u); \\ 0.6, & if\ (\Delta > 1.25\Delta u) \wedge \\ & (\Delta \leq 1.50\Delta u); \\ 0.5, & if\ (\Delta > 1.50\Delta u) \wedge \\ & (\Delta \leq 1.75\Delta u); \\ 0.4, & if\ (\Delta > 1.75\Delta u) \wedge \\ & (\Delta \leq 2.00\Delta u); \\ 0.3, & if\ (\Delta > 2.00\Delta u) \wedge \\ & (\Delta \leq 2.25\Delta u); \\ 0.2, & if\ (\Delta > 2.25\Delta u) \wedge \\ & (\Delta \leq 2.50\Delta u); \\ 0.1, & if\ (\Delta > 2.50\Delta u) \wedge \\ & (\Delta \leq 2.75\Delta u); \\ 0.0, & if\ (\Delta > 2.75\Delta u). \end{cases} \tag{9}$$

to automate this process. This function uses the n-interval length of U:

$$\Delta u = U(2,1) - U(1,1) \tag{10}$$

and the distance between the m-actual data, $D(m)$, with respect to the mean of the n-interval length.

$$\Delta = \left| D(m) - \frac{\Delta u}{2} \right| \tag{11}$$

Step 5. Establishing fuzzy logical relationships (FLR). This process find the sequence of fuzzy sets that is equal to 1, i.e., A_i = 1:

$$\begin{aligned} &A_1 \rightarrow A_1 \\ &A_1 \rightarrow A_1 \\ &\vdots \\ &A_6 \rightarrow A_6 \\ &A_6 \rightarrow A_7 \end{aligned} \tag{12}$$

Step 6. Forecast all the right hand side of the fuzzy data in FLR according to:

$$F(i) = [M_1, M_2, ..., M_i] \times [w_1, w_2, ..., w_i]^T \tag{13}$$

with the weighted matrix defined by (14), (15) and (16).

$$W(t) = [w_1^{'}, w_2^{'}, ..., w_k^{'}] \tag{14}$$

$$W(t) = \left[\frac{w_1}{\sum\limits_{h=1}^{k} w_h}, \frac{w_2}{\sum\limits_{h=1}^{k} w_h}, ..., \frac{w_k}{\sum\limits_{h=1}^{k} w_h} \right] \tag{15}$$

$$\sum_{h=1}^{k} w_h^{'} = 1 \tag{16}$$

where:
$i = k$: total number of fuzzy sets in the right hand side of FLR;
M: midpoint of the n-intervals in the universe of discourse U;
W: calculated weights for each midpoint Mi's.

B. Exponential FTS Algorithm

The Exponential FTS algorithm makes some corrections in the weights calculated in Yu FTS algorithm (15) and (16). The steps to implement the exponential FTS are the same that Yu FTS with some differences when calculate the weights (17). The adjustment of parameter c makes great difference in fuzzy forecast, usually the value c is set as 1.2.

Step 1. Implement the steps from 1 to 5 according Yu FTS algorithm.

Step 2. Forecast all the right hand side of the fuzzy data in FLR according to (13) with the weighted matrix defined by (14) and:

$$W(t) = \left[\frac{1}{\sum\limits_{h=1}^{k} w_h}, \frac{c}{\sum\limits_{h=1}^{k} w_h}, ..., \frac{c^2}{\sum\limits_{h=1}^{k} w_h}, \frac{c^{k-1}}{\sum\limits_{h=1}^{k} w_h} \right] \tag{17}$$

where:
$i = k$: total number of fuzzy sets in the right hand side of FLR;
c: weight constant with $c \geq 1.2 \leq i \leq k$;
M: midpoint of the n-intervals in the universe of discourse U;
W: calculated weights for each midpoint Mi's.

C. Transformation FTS Algorithm

Transformation FTS algorithm is commonly used to remove noisy effects of data and therefore increases the forecastability of time series [15]. The Box Cox transformation is applied to the actual data series. The steps to implement the Transformation FTS algorithm are presented as follow:

Step 1. Apply Box Cox transformation to the actual data series and then obtain the λ parameter and Z vector:

$$Z_t^{\lambda} = \begin{cases} \frac{Z_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \ln(Z_t) & \lambda = 0 \end{cases} \tag{18}$$

Step 2. Calculate d_S vector according to:

$$d_s = \frac{Z_S^{\lambda} - Z_{S-1}^{\lambda}}{Z_S^{\lambda}} \tag{19}$$

Step 3. Calculate Δ_q as a difference of the third and first quartile, according to:

$$\Delta_q = Q_3 - Q_1 \tag{20}$$

Step 4. Calculate parameter l according to:

$$l = \frac{2 \times \Delta_q}{n^{1/3}} \tag{21}$$

Step 5. Obtain the estimated vector d_S using Yu FTS algorithm.

$$\hat{d}_S = [M_1, M_2, ..., M_i] \times [w_1, w_2, ..., w_i]^T \tag{22}$$

Step 6. Calculate the estimated y_S parameter according to:

$$\hat{y}_S = \frac{1}{1 - \hat{d}_S} \tag{23}$$

Step 7. Apply:

$$\hat{Z}_S^{\lambda} = \hat{y}_S Z_t^{\lambda} \tag{24}$$

to the vector Z obtained in (18).

Step 8. Obtain the forecast data series according to:

$$F = \hat{Z}_S = \left(\lambda \hat{Z}_S^{\lambda} + 1\right)^{1/\lambda} \tag{25}$$

D. High-Order FTS Algorithm

In some cases the data to be forecast are not related to the last state, thus they are related to some sequence of former states. According [15] the sequence of former states can be done by some of these options below:

- by expert knowledge;
- by data discovery;
- by applying try and error procedure;
- by ACF (Auto-correlation function) analyzing;
- by some evolutionary algorithm.

High-Order FTS are very usefull for seasonal and multiseasonal series. The steps to implement High-Order FTS algorithm are related below:

Step 1. Implement the steps from 1 to 4 according Yu FTS algorithm.

Step 2. Choose the sequence of former states according strategies presented before. For example:

$$F(t-2), F(t-1) \rightarrow F(t) \tag{26}$$

presents the states according example in [15].

Step 3. Establishing fuzzy logical relationships (FLR) according the former states. In:

$$\begin{aligned} &A_1, A_1 \rightarrow A_1 \\ &A_1, A_1 \rightarrow A_2 \\ &\vdots \\ &A_6, A_7 \rightarrow A_7 \\ &A_7, A_7 \rightarrow A_6 \end{aligned} \tag{27}$$

the FLR are composed of two former states [15].

Step 4. Eliminate the recurrence of fuzzy logical relationship (27) and defines the fuzzy logical relationship group (FLRG) for the former states:

$$\begin{aligned} &A_1, A_1 \rightarrow A_1, A_2 \\ &A_1, A_2 \rightarrow A_3 \\ &\vdots \\ &A_6, A_6 \rightarrow A_7 \\ &A_7, A_7 \rightarrow A_6 \end{aligned} \tag{28}$$

Step 5. Forecast all the right hand side of the fuzzy data in FLR according to:

$$F(i) = \frac{\sum_{j=1}^{i} M_j}{i} \tag{29}$$

V. Results

The wind turbine times-to-failure data, analyzed by both reliability and FTS algorithms methods, was obtained from measurements realized in a large onshore wind farm located at Brazil. The failure model for electric motor fan was obtained using maximum likelihood method and Weibull distribution function, the estimated parameters were $\beta = 1.70$ and $\eta = 702,3$ days. These distribution parameters values were also used to generate 30 points for the time series forecasting. Figure 2 presents the fuzzy time series which best forecast the times-to-failure data, where blue dots represents the generated sample and red dots the forecast values. Figure 3 compares best FTS forecast with reliability evaluation method used to predict times-to-failure data.

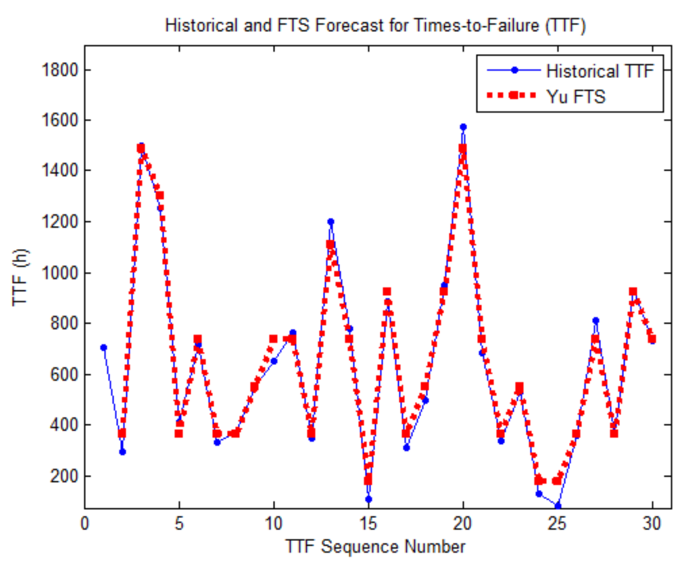

Fig. 2. Yu FTS forecast for times-to-failure.

The evaluation methods for fuzzy time series applied where obtained using forecast errors measures, i.e, MAPE - Mean

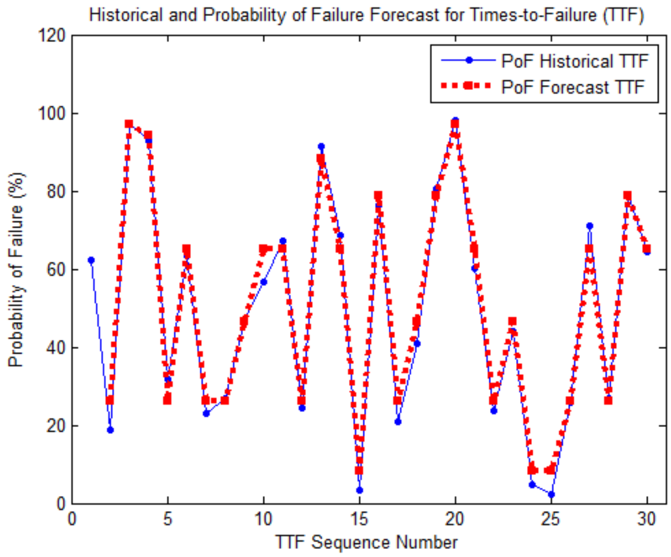

Fig. 3. Probability of failure (PoF) comparisons between FTS and reliability predictions.

Absolute Percentage Error (referred as mean error), MAD - Mean Absolute Deviation, MSE - Mean Squared Error and RMSE - Root Mean Squared Error. For all these methods the smaller value are the better method that fits the model.

FTS Algorithms Comparisons

The mean error for times-to-failure forecasting using Yu and Exponential were 13%. The Transformation FTS algorithm implemented obtain a mean error of 16% and for High-Order FTS algorithm, mean error was 17%. The sequence of former states used in High-Order FTS was chosen according to:

$$F(t-3), F(t-2), F(t-1) \rightarrow F(t) \tag{30}$$

Table I presents the summary of statistics estimated errors comparisons between fuzzy forecast methods.

TABLE I
FTS Forecast Error Comparison

Forecast Method	MAPE	MAD	MSE	RMSE
Yu FTS	13	40	2.40×10^3	49
Exponential FTS	13	40	2.40×10^3	49
Transformation FTS	16	113	23.3×10^3	153
High-Order FTS	17	60	8.20×10^3	91

VI. Conclusion

From the results obtained in section V it is possible to verify that both Yu and Exponential FTS were the best fuzzy time series method to fit times-to-failure of a wind turbine electric motor fan. It also can be viewed in Figure 3 that reliability evaluation fails to predict times-to-failure when the probability (PoF) are less than 70%. The use of FTS algorithms instead of standard reliability evaluation to predict times-to-failure has the advantages that mathematical and statistics background necessary in FTS is much more simple and easy to understand and implement than the standard reliability evaluation prediction methods.

Acknowledgments

The authors of this research paper were financial supported by Brookfield Renewable and ANEEL (Agência Nacional de Energia Elétrica).

References

[1] M. H. J. BOLLEN. *Understanding Power Quality Problems - Voltage Sags and Interruptions*. IEEE Press Series on Power Engineering. New Jersey, 2000.
[2] R. BILLINTON. *Power System Reliability Evaluation*. Gordon and Breanch Science Publishers, Inc. New York, 1970.
[3] R. E. BROWN. *Electric Power Distribution Reliability*. Marcel Dekker, Inc. New York, 2002.
[4] Q. SONG, B. S. CHISSOM. *Forecasting Enrollments With Fuzzy Time Series*. 16 p. Annual Meeting of the Mid-South Educational Research Association, Nov. 91, Lexington, KY.
[5] Q. SONG, B. S. CHISSOM. *Fuzzy Time Series and Its Models*. Fuzzy Sets and Systems 54 (1993) 269-277. North-Holland. Elsevier.
[6] Q. SONG, B. S. CHISSOM. *Forecasting Enrollments With Fuzzy Time Series - Part I*. Fuzzy Sets and Systems 54 (1993) 1-9. North-Holland. Elsevier.
[7] Q. SONG, B. S. CHISSOM. *Forecasting Enrollments With Fuzzy Time Series - Part II*. Fuzzy Sets and Systems 62 (1994) 1-8. North-Holland. Elsevier.
[8] P. TAVNER. *Offshore Wind Turbines - Reliability, availability and maintenance*. The Institution of Engineering and Technology. London, United Kingdom, 2012.
[9] RELIASOFT CORP. *Life Data Analysis Reference*. ReliaSoft Publishing. Tucson, AZ, USA, 2003.
[10] RELIASOFT CORP. *System Analysis Reference - Reliability, Availability & Optimization*. ReliaSoft Publishing. Tucson, AZ, USA, 2003.
[11] R. L. KASHYAP, *Maximum Likelihood Identification of Stochastic Linear Systems*. IEEE Transactions on Automatic Control, vol. Ac-15, No.1, Feb. 1970.
[12] F. G. GUIMARÃES. *Multiobjective Optimization Class Notes*. Universidade Federal de Minas Gerais. Course of Multiobjective Optimization - EEE910.
[13] F. G. GUIMARÃES, L. S. BATISTA, E. G. CARRANO. *Introduction to Multiobjective Optimization - Class Notes*. Universidade Federal de Minas Gerais.
[14] R. R. SALDANHA. *Optimization in Electric Engineering - Class Notes*. Universidade Federal de Minas Gerais. Course of Optimization in Electric Engineering - EEE948, 2014.
[15] H. J. SADAEI, F. G. GUIMARÃES. *Advanced Fuzzy Time Series Class Presentations and Notes*. Universidade Federal de Minas Gerais. Course of Advanced Fuzzy Time Series - EEE933, 2016.
[16] WIKIPEDIA: Time series. Available in: <https://en.wikipedia.org/wiki/Time_series>. Last Access in June 19, 2017.
[17] H. J. SADAEI, R. ENAYATIFAR, M. H. LEE, M. MAHMUD. *A Hybrid Model Based on Differential Fuzzy Logic Relationship and Imperialist Competitive Algorithm for Stock Market Forecasting*. Applied Soft Computing 40 (2016) 132-149. Elsevier.
[18] S. M. CHEN. *Forecasting Enrollments Based on Fuzzy Time Series*. Fuzzy Sets and Systems 81 (1996) 311-319. Elsevier.
[19] H. K. YU. *Weighted Fuzzy Time Series Models for TAIEX Forecasting*. Physica A 349 (2005) 609-624. Elsevier.
[20] Z. ZHANG, Q. ZHU. *Fuzzy Time Series Forecasting Based on K-Means Clustering*. Open Journal of Applied Sciences. 2012.