diction_using_fuzzy_random_auto-regression_time_series_model.pdf

by

Maximum-minimum temperature prediction using fuzzy random auto-regression time series model

Conference Paper · August 2017

DOI: 10.1109/ISCBI.2017.8053544

4 authors:

Riswan Efendi
Universiti Pendidikan Sultan Idris (UPSI)

Noor Azah Samsudin
Universiti Tun Hussein Onn Malaysia

Nureize Arbaiy
Universiti Tun Hussein Onn Malaysia

Mustafa Mat Deris
Telkom University

Maximum-Minimum Temperature Prediction Using Fuzzy Random Auto-Regression Time Series Model

Riswan Efendi[1,2], Noor Azah Samsudin[1], Nureize Arbaiy[1], Mustafa Mat Deris[1]
[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia,
Batu Pahat, Malaysia
[2]Mathematics Department
UIN Sultan Syarif Kasim
Pekanbaru, Indonesia
e-mail: riswan@uthm.edu.my; azah@uthm.edu.my, nureize@uthm.edu.my; mmustafa@uthm.edu.my

***Abstract*—Many models have been suggested to predict the weather and temperature data. Most of them used the single point data in building prediction equations. Besides that, the randomness, the vagueness and possibility of the temperature data are also not much concerned. In this paper, we introduce the minimum-maximum procedure for daily temperature modeling based on fuzzy random auto-regression time series. The proposed procedure was able to cover the variability of the temperature in nature. The result showed that mean square error of proposed model is smaller than the existing models.**

Keywords-component; fuzzy random variable, temperature, min-max procedure, auto-regression model

I. INTRODUCTION

Temperature prediction is very essential activity in human life, such as agriculture, tourism, transportation, and others. Many models have been presented to predict the temperature data. However, the forecasting accuracy is still issue and not easy task to predict the temperature because its trend data is very complex and fully uncertainty. In fuzzy time series, some models have been introduced to handle the temperature trend using two-factor time variant [1], [2], [3] and the hybrid techniques [4]. However, the randomness, the vagueness and the possibility of input data are not considered in the models.

In this paper, we introduce a minimum-maximum (min-max) procedure to model the variability and trend of temperature using fuzzy random auto-regression time series [6], [7]. This procedure is very appropriate to handle the double inputs of the time series data, such as, temperature, stock market, exchange rate, and others. Because these data have the small gap between minimum and maximum values. The detail of theories, proposed procedure, and implementation are discussed in Sections II, III, and IV, respectively.

II. BASIC THERORIES OF FUZZY RANDOM MODEL

A. Fuzzy Random Variable (FRV)

From [5], [6], the fuzzy random variable Y, with possibility distribution μ_Y, the possibility, necessity, and credibility of event $\{Y \le r\}$ can be written as:

$$Pos\,\{Y \le r\} = sup\,\mu_Y(t),\, t \le r, \tag{1}$$

$$Nec\,\{Y \le r\} = 1 - sup\,(t),\, t \ge r, \tag{2}$$

$$Cr\,\{Y \le r\} = ½\,(1 + sup_{t\le r}\mu_Y(t) - sup_{t\ge r}\mu_Y(t)). \tag{3}$$

From Eq. (3), we note that the credibility measure is an average of the possibility and the necessity measures, i.e., $Cr\{.\} = \frac{Pos\{.\}+Nec\{.\}}{2}$. Credibility measure is to develop a certain measure, which is a sound aggregate of the two extreme cases. That is the possibility (which expresses a level of overlap and highly optimistic in this sense) and necessity (that articulates a degree of inclusion and is pessimistic in its nature). Based on credibility measure, the expected value of fuzzy variable is presented as follows [5].

Definition 1. Expected value of fuzzy variable [6]
Let Y be a fuzzy variable. The expected value of Y is defined as:

$$E(Y) = \int Cr\,\{Y \ge r\}\,dr - \int Cr\,\{Y \le r\}\,dr, \tag{4}$$

under the condition that the two integral are finite. Assume that $Y = [a^l, c, a^r]_T$ is triangular fuzzy variable or number (TFV = TFN). Making use of Eq. (4), we determine the expected value of Y to be

$$E(Y) = \frac{(a^l + 2c + a^r)}{4}. \tag{5}$$

Definition 2. Fuzzy random variable [6]
Suppose that (Ω, Σ, Pr) is a probability space and F_v is a collection of fuzzy variables defined on possibility space $(\Gamma, \mathcal{P}(\Gamma), Pos)$. A fuzzy random variable is a mapping $X : \Omega \to F_v$ such that for any Borel subset B of R, $Pos\{X(\omega) \in B\}$ is a measurable function of ω. Let X be a fuzzy random variable on Ω. From the previous definition, we know, for each ω ϵ Ω, that $X(\omega)$ is a fuzzy variable. Furthermore, a fuzzy random variable X is said to be positive if, for almost every ω, fuzzy variable $X(\omega)$ is positive almost surely. The expected value of the fuzzy variable $X(\omega)$ is denoted by $E(X(\omega))$, for any fuzzy random variable X on Ω, for each ω ϵ Ω. This has been proved to be a measurable function of ω [5], i.e., it is random variable. Hence, the expected value of the fuzzy random

variable X is defined as the mathematical expectation of the random variable $E(X(\omega))$.

Definition 3. Expected value of fuzzy random variable [6]

Let X be a fuzzy random variable defined on probability space (Ω, Σ, Pr). Then, the expected value of X and variance of X are defined as

$$E(X) = \int \Omega \left[\int Cr\{\xi(\omega) \geq r\}\, dr - \int Cr\{\xi(\omega) \leq r\}\, dr \right] Pr(\omega), \quad (6)$$

$$Var(X) = E(X - e)^2, \quad (7)$$

where $e = E(X)$ is given by Eq. (7).

B. Fuzzy Random Auto-Regression Model (FR-AR)

For time series data points, autoregressive or AR(p) model can be written as:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t, \quad (8)$$

where $\emptyset_1, \ldots, \emptyset_p$ are coefficients of $Y_{t-1}, \ldots, Y_{t-p}$, respectively, e_t is an error models at time-t. While, based on [6], the input and output data $\tilde{Y}_{t-p}$ for all $p = 0, 1, 2, \ldots, n$ are fuzzy random variables, which are written as:

$$\tilde{Y}_t = \cup_{i=1}^{n} \left[\left(Y_{it}^{l}, Y_{it}^{c}, Y_{it}^{r} \right)_T, P_{it} \right], \quad (9)$$

where $\tilde{Y}_t$ is a time series data at time-t and its formatted as a triangular fuzzy number [left, l; center, c; ; right, r]. Based on Eq. (10), all values given as fuzzy numbers with probabilities, P_{it}. These data, $\tilde{Y}_t$ also can be presented in Table I.

TABLE I. FUZZY RANDOM INPUT-OUTPUT DATA

Sampl e	Output	Input		
1	$\tilde{Y}_t$	$\tilde{Y}_{t-1}$	$\tilde{Y}_{t-2}$	$\cdots$ $\tilde{Y}_{t-k}$
2	$\tilde{Y}_{t-1}$	$\tilde{Y}_{t-2}$	$\tilde{Y}_{t-3}$	$\cdots$ $\tilde{Y}_{t-(k+1)}$
…	…	…	…	…
n	$\tilde{Y}_{t-n}$	$\tilde{Y}_{t-(n+1)}$	$\tilde{Y}_{t-(n+2)}$	$\cdots$ $\tilde{Y}_{t-(k+n)}$

III. DATA PREPARATION

In real situations, non-stationary time series data are frequently occurred in our daily life, such as, electricity load consumption, stock market prices, temperature, airline passenger data, others. It can be measured in many different ways, such as, low-high, minimum-maximum, and closed-open values. In this section, we highlights the classification of the temperature input: min-max inputs. It is important to emphasize on classification of the data before deciding appropriate procedure for preparing the data towards building an FR-AR model. A minimum-maximum (min-max) procedure are proposed for double input. This preparation procedure is presented in Section III*A*.

A. Minimum-Maximum Procedure

Generally, the existing models are limited to application with single input for building univariate forecasting models. Apparently, the most of data are obtained from secondary sources. Since the data are obtained from secondary sources, we may have to confront with validity, biasness, and representation issues. Additionally, these data may also be presented in single input form that is not considering variability issue. Consequently, these limitations contribute to building less accurate forecasting models.

On the other hand, in our proposed procedure, we are using the largest and smallest values in the data that is called min-max procedure. The benefit of using this procedure is useful to handle variability in the data. Consequently, a more accurate forecasting model can be achieved. There are existing works using the smallest and largest values in the data to build a forecasting model in stock market and electricity applications [6],[7]. However, the data preparation procedures to obtain the min-max values from the data were not well discussed. The range of min-max values obtained is consistent throughout the data. Consequently, the min-max data presented are lack of originality and not natural. In our procedure, we will consider to capture variability in the data. The temperature behavior pattern is a result of combination between trend and seasonal time series. The fluctuation in the temperature pattern can be observed using min-max procedure. The min-max procedure may allow us to capture the variability in the consumption behavior which leads to better recognition in the pattern. Moreover, the gap between minimum and maximum values in the consumption can represent a range of data. By using the values in the range of data, the variability can be determined. Since the min-max values fluctuate every day as a result the range values will also fluctuate. Besides the min-max values, the fluctuation may be caused by other unpredicted factors. Interestingly, the min-max values can be transformed into triangular fuzzy number (TFN) directly. The result of the transformation is shown in Figure 1.

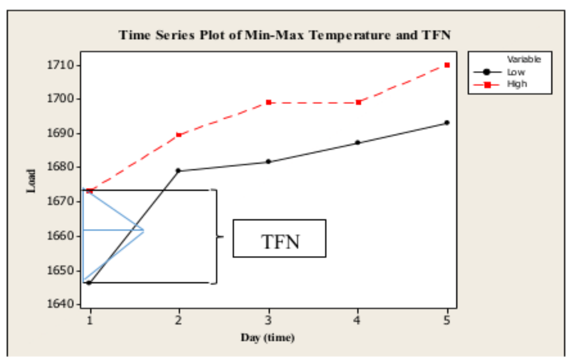

Figure 1. An example of transformation data into TFN

Figure 1 shows the min-max data being presented in fuzzy format. Thus, its TFN can be obtained directly without left-right spread into consideration. Based on Figure 1, the equivalence of min-max data with TFN can be represented as:

Min data (Y_t^{min}) $\equiv$ Left spread of TFN (Y_t^l)
Midpoint (Y_t^c) $\equiv$ Center of TFN (Y_t^c)
Max data (Y_t^{max}) $\equiv$ Right spread of TFN (Y_t^r)

TABLE II. MIN-MAX DATA AND FUZZY DATA

Time	Min	Max	Fuzzy data (FD)
1	Y_1^{min}	Y_1^{max}	$[Y_1^l, Y_1^r]$
…	…	…	…
n	Y_n^{min}	Y_n^{max}	$[Y_n^l, Y_n^r]$

Therefore, all min-max values in the temperature data can be transformed into fuzzy values as presented in Table II. The FD data in this table are sufficient to build the desire FR-AR model. In other words, the temperature data has successfully been prepared using the min-max procedure.

B. *Building FR-AR Model Using Algorithm*

In this section, we will present the algorithm details in building FR-AR model [6], [7]. Indeed, the transformation step for the type of data presented in Sections *A* will be included. Additionally, the parameter estimate using linear programming (LP) approach will be discussed too. The details of the algorithm are presented in Steps 1-7 as follows:

Step 1: Transform the actual time series into TFN based on the procedures in Section *A*.

Step 2: Determine fuzzy data (FD) using the TFN results in Step 1.

Step 3: Divide FD into two groups of fuzzy random data, FRD_1 and FRD_2.
The length (l) for each FD can be defined as the difference between upper (b_i) and lower data (a_i), such that $i = 1, 2, 3, \ldots, n$. By using value of l, the fuzzy random time series data can be divided into FRD_1 and FRD_2.

Step 4: Calculate the expected value $E(Y)$ and $Var\ (Y)$ of FRD.
Before building FR-AR with parameter estimates, a confidence-interval that is induced by $E(Y)$ and $Std.Dev(Y)$ of a fuzzy random variable (FRD), then we consider the one-sigma confidence $(1 \times \sigma)$ interval (CI) of each FRD [5]. The CI of FRD can be expressed as:

$$CI \cong [E(Y) - Std.Dev\ (Y), E(Y) + Std.Dev\ (Y)].$$

Note that, the CI being adopted in this step is different from the one that we used to estimate sample mean. The corresponding CI (output) results for each FRD_1 (input) and FRD_2 (input).

Step 5: The general FR-AR model can be expressed as:

$$Y_t = [\widehat{\emptyset}_1^l, \widehat{\emptyset}_1^r]\tilde{Y}_{t-1} + [\widehat{\emptyset}_2^l, \widehat{\emptyset}_2^r]\tilde{Y}_{t-2}, \qquad (11)$$

The linear equation system is to be developed using the input (FRD_1, FRD_2) and output (FRD) to estimate parameters in Eq. (11). There are two linear equation system to be developed for left-right inputs. Then, both linear equation systems will be solved using linear programing (LP), namely, simplex approach to estimate the parameters of FR-AR model.

Step 6: Determine the predicted FR-AR model from Step 5.

$$\hat{Y}_t = [\widehat{\emptyset}_1^l, \widehat{\emptyset}_1^r]\tilde{Y}_{t-1} + [\widehat{\emptyset}_2^l, \widehat{\emptyset}_2^r]\tilde{Y}_{t-2}, \qquad (12)$$

which $\hat{Y}_t$ is a predicted fuzzy random time series at time-t, $\widehat{\emptyset}_1^l, \widehat{\emptyset}_1^r$ and $\widehat{\emptyset}_2^l, \widehat{\emptyset}_2^r$ are pair predicted parameters of fuzzy random time series ($\tilde{Y}_{t-1}, \tilde{Y}_{t-2}$), respectively. The uniqueness of the estimated model parameters is that left-right values are always the same. Therefore, interestingly Eq. (12) is capable to forecast three different values, low, medium and high as follows:

$$\hat{Y}_t = \begin{cases} [\widehat{\emptyset}_1^l]\tilde{Y}_{t-1} + [\widehat{\emptyset}_2^l]\tilde{Y}_{t-2}, & \text{forecast of low data} \\ [\widehat{\emptyset}_1^r]\tilde{Y}_{t-1} + [\widehat{\emptyset}_2^r]\tilde{Y}_{t-2}, & \text{forecast of high data} \\ \frac{\hat{Y}_t^l + \hat{Y}_t^r}{2}, & \text{forecast of medium data} \end{cases}, \qquad (13)$$

Unlike existing models which are limited to forecast single value only.

Step 7: Evaluate and interpret the min-max-average width of the possibility of model by based on the following criteria:

$$W_i = \min, \max(\hat{Y}_i^l - \hat{Y}_i^r), \qquad (14)$$

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i^h - \hat{Y}_i^r). \qquad (15)$$

Eq. (14) can be used to determine the minimum and maximum width of model's possibility. Meanwhile, the average can be determined using Eq. (15). Essentially, the smaller the average is, the smaller the ambiguity in the proposed model.

IV. IMPLEMENTATION

In this section, we present the implementation of proposed algorithm to predict (training and testing) the daily Changi temperature (min-max and average) from 01/01/2015 – 31/12/2015. By following steps given in Section III*B*, FR-AR model can be written as:

$$\hat{Y}_t = 0.6671873(Y_{t-1})_T + 0.3446873(Y_{t-2})_T. \qquad (16)$$

Based on Eq. (16), we eventually obtained same parameter estimates for left and right values ($\widehat{\emptyset}_1^l = \widehat{\emptyset}_1^r$ and $\widehat{\emptyset}_2^r = \widehat{\emptyset}_2^l$) for each model. Essentially, these parameters can be used to measure the contribution of FRD_1 and FRD_2 to $\hat{Y}_t$ (Output). For example, 66.718% of FRD_1 ($\tilde{Y}_{t-1}$) and 34.468% of FRD_2 ($\tilde{Y}_{t-2}$) contribute to $\hat{Y}_t$ of temperature. Then, the mean square error (MSE) of the models are compared with the previous models (ARIMA and fuzzy time series) in Table III.

TABLE III. MSE COMPARISON

Model	Temperature	
	Training	Testing
ARIMA	115.232	97.889
FTS	227.346	150.654
Proposed FR-AR	70.968*	73.393*

Due to the limitation in ARIMA and FTS models, we use medium values only for comparison. Recall that, the data are divided into training and testing data. The MSE for the FR-AR models are smaller than ARIMA and FTS models for both, training and testing. The MSE of FR-AR model is the smallest for all the data sets. The MSE values of proposed model imply that the forecasting error can be reduced significantly. In other words, the FR-AR model benefits from the use of min-max values in proposed procedure becuase the estimated parameters can reduce the range of data. Thus, the variation will be minimise too and can be controled. As a result, the predicted values tend to be near to the medium data. Thus the accuracy is improved.

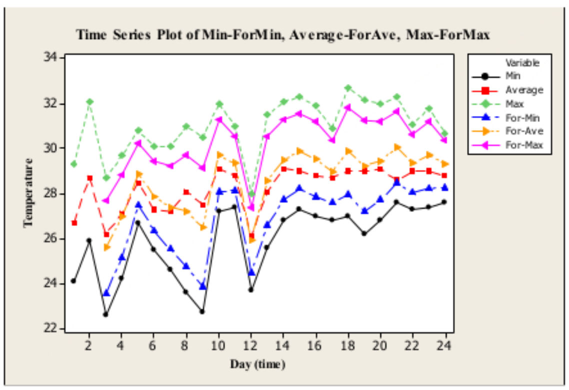

Figure 2. Actual and forecasted values using FR-AR model

Due to the performance of the FR-AR model, we apply this model to forecast three different values simultaneously: min, average and maximum of Changi (Singapore) temperature. Figure 2 shows the fitting of min-max of actual data with min-max of forecast values for several days. Note that, the graphs for the predicted min-max values are very close to the graphs of the actual data. Specifically, the width of the possibility of the proposed model is 4.169 on average, 1.55 at the minimum and 6.948 at the maximum. Meanwhile, the width of the possibility of actual min-max data is considerably large that is 6.18 on average, 2.3 at the minimum and 10.3 at the maximum. Significantly, the width of FR-AR model is much smaller than the width of the possibility of actual min-max data. As a result, the ambiguity of FR-AR model can be minimized. In the case of our data, the gap between the minimum and the maximum values is considerably small on average. Due to the small gap between the values, thus, applying min-max procedure is more appropriate. It is found that the average width of the possibility of FR-AR model for all data sets is smaller than the average width of the possibility of the actual data. Due to the small width of the possibility, the ambiguity in the FR-AR models are also lessen.

V. CONCLUSION

In this paper, we build an improved FR-AR model by employing min-max procedure in data preparation. While this procedure is more appropriate to prepare data with small gap between minimum and maximum. Additionally, most existing statistical models require differencing technique to ignore seasonal and trend components. In contradiction, such differencing technique is not required in the proposed procedure. In the comparison of MSE, we considered non-fuzzy and fuzzy models, namely, ARIMA and FTS. Based on application, our proposed FR-AR model with new input type outperforms other both existing models. Based on our perspective, this study can be implemented in handling the non-stationary time series data from various domain problems.

ACKNOWLEDGMENT

This study is supported by Ministry of Higher Education, Malaysia. Fundamental Research Grant Scheme, Vot 1609 and in part by a great from Research Gates IT Solution Sdn. Bhd.

REFERENCES

[1] S. M. Chen, and J. R. Hwang, "Temperature prediction using fuzzy time series," IEEE Transaction on Systems, Man and Cybernatics, Part B, vol. 30, pp. 263-275, 2000.

[2] L. W. Lee, L. H. Wang and S. M. Chen, "Temperature prediction and TAIFEX forecasting based on fuzzy logical relationship and genetic algorithms," Expert Systems with Applications, vol. 33, pp. 539-550, 2007.

[3] L. W. Lee, L. H. Wang and S. M. Chen, "Temperature prediction and TAIFEX forecasting hig-order fuzzy logical relationships and genetic simulated annealing techniques," Expert Systems with Applications, vol. 34, pp. 328-336, 2008.

[4] Y. Sharma and S. Sisodia, "Temperature prediction based on fuzzy time series and MTPSO with automatic clustering algorithm," Computational and Business Intelligence (ISCBI), pp. 101-105, 2014.

[5] J. Watada, S. Wang and W. Pedrycz, "Building confidence-interval based fuzzy random regression model," IEEE Trans. Fuzzy Syst, vol. 17, pp. 1273-1283, 2009.

[6] L. Shao, Y-H. Tsai, J. Watada and S. Wang,"Building fuzzy random auto-regression model and its application, Intel. Decision Tech. (SIST), vol. 14, pp. 24-30, 2012.

[7] R. Efendi, N. Arbaiy and M. M. Deris,"Estimation of confidence-interval for yearly electricity load consumption based on fuzzy random autor-regression," Advances Intel. Syst. Compt. (CIIS), vol. 1, pp. 61-68, 2017.

diction_using_fuzzy_random_auto-regression_time_series_model.pdf

ORIGINALITY REPORT

18%	%	18%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Computational Intelligence in Information Systems", Springer Nature, 2017 Publication	9%
2	Lu Shao, You-Hsi Tsai, Junzo Watada, Shuming Wang. "Chapter 16 Building Fuzzy Random Autoregression Model and Its Application", Springer Science and Business Media LLC, 2012 Publication	4%
3	Refik Arıkan. "An unknown Roman bridge on sangarius and ancient road system around it", New Trends and Issues Proceedings on Humanities and Social Sciences, 2016 Publication	2%
4	"Intelligent and Interactive Computing", Springer Science and Business Media LLC, 2019 Publication	2%
5	"Recent Advances on Soft Computing and Data Mining", Springer Science and Business	2%

Media LLC, 2018
Publication

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		