

Sahin, Atakan and Kumbasar, Tufan and Yesil, Engin and Dodurka, M. Furkan and Karasakal, Onur (2015) An approach to represent time series forecasting via fuzzy numbers. In: Proceedings: Second International Conference on Artificial Intelligence, Modelling, and Simulation, AIMS 2014. IEEE, Piscataway, NJ., pp. 51-56. ISBN 9781479975990 , http://dx.doi.org/10.1109/AIMS.2014.36

This version is available at http://strathprints.strath.ac.uk/60315/

**Strathprints** is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<u>http://strathprints.strath.ac.uk/</u>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: <a href="mailto:strathprints@strath.ac.uk">strathprints@strath.ac.uk</a>

# An Approach to Represent Time Series Forecasting via Fuzzy Numbers

Atakan Sahin<sup>1,2</sup>, Tufan Kumbasar<sup>1</sup>, Engin Yesil<sup>1</sup>, M. Furkan Dodurka<sup>1,2</sup>, Onur Karasakal<sup>2</sup>

<sup>(1)</sup> Istanbul Technical University, Faculty of Electrical and Electronics Engineering,

*Control and Automation Engineering Department*, Maslak, TR-34469, Istanbul, Turkey <sup>(2)</sup>Getron Bilişim Hizmetleri A. Ş., Yıldız Teknik Üniversitesi Davutpaşa Kampüsü,

Teknopark Binası B1 Blok, Esenler, 34220, Istanbul, Turkey

{sahinata, kumbasart, yesileng}@itu.edu.tr; {furkan.dodurka, onur.karasakal}@getron.com

Abstract—This paper introduces a new approach for estimating the uncertainty in the forecast through the construction of Triangular Fuzzy Numbers (TFNs). The interval of the proposed TFN presentation is generated from a Fuzzy logic based Lower and Upper Bound Estimator (FLUBE). Here, instead of the representing the forecast with a crisp value with a Prediction Interval (PI), the level of uncertainty associated with the point forecasts will be quantified by defining TFNs (linguistic terms) within the uncertainty interval provided by the FLUBE. This will give the opportunity to handle the forecast as linguistic terms which will increase the interpretability. Moreover, the proposed approach will provide valuable information about the accuracy of the forecast by providing a relative membership degree. The demonstrated results indicate that the proposed FLUBE based TFN representation is an efficient and useful approach to represent the uncertainty and the quality of the forecast.

# Keywords-forecasting, fuzzy time series, fuzzy numbers, fuzzy estimator

# I. INTRODUCTION

In recent years, the research works on point forecast and prediction approaches using computational intelligence methods have been widely increased. It has been reported in [1] that there are two main problems with forecasting methods presented in literature: (i) the models become unreliable in the presence of uncertainty and (ii) no indication of accuracy of point forecasts is provided. The accuracy of the forecast is measured with performance indexes such as Mean Absolute Percentage Error (MAPE), Percentage of Error (POA), etc. [2]. Though, since there is always an error margin in predictions, there is a need to define error bounds in forecast like Confidence Interval (CI) and Prediction Interval (PI). The CIs handle with the accuracy of the prediction of the regression, i.e., of the mean of the target probability distribution. However, the PIs consider the accuracy with which can predict the targets themselves, i.e., they are based on estimates of the distribution [3]. A PI is constructed from lower and upper bounds that will cover the future unknown value with a prescribed probability called a confidence level  $((1-\alpha)\%, \alpha \in [0,1])$  [4]. The main motivation for the construction of PIs is to quantify the likely uncertainty in the point forecasts. Availability of PIs allows the decision makers to quantify the level of uncertainty associated with the point forecasts. PIs that are relatively wide indicate the presence of high level of uncertainties in the underlying system operation. This useful information can guide the decision makers to avoid the selection of risky actions under uncertain conditions. On the other hand, narrow PIs give the decision makers the opportunity to decide more confidently with less chance of confronting an unexpected condition in the future.

The construction of PIs has been a subject of much attention and has been implemented in temperature prediction, travel time prediction in baggage handling system, watershed simulation, solder paste deposition process, and time series forecasting [4]. In literature, different methods exist for the construction of PIs as delta technique [5], Bayesian technique [6], bootstrap method [7], mean-variance estimation [8], Lower and Upper Bound Estimation (LUBE) method [9].

In this study, a Fuzzy logic based Lower and Upper Bound Estimator (FLUBE) based Triangular Fuzzy Number (TFN) representation and generation approach is presented to estimate the uncertainty in the forecast. This approach will give the opportunity to the decision maker to quantify the uncertainty of the point forecasts with linguistic terms which might increase the interpretability. Moreover, the proposed approach will provide valuable information about the accuracy of the forecast by providing a relative membership degree with respect to the target data. The proposed approach consists of two main phases, the offline FLUBE design and the online TFN generation. Thus, we will start by presenting the design steps and internal structure of the FLUBE. Then, the online TFN generation and representation is presented in detail. To illustrate the proposed approach, the Australian monthly electrical consumption data set is handled. Based on the presented results, it will be concluded that proposed FLUBE based TFN representation is an efficient and useful approach to represent the uncertainty of the forecast.

This paper is organized as follows. In Section 2, we introduce the experiment data set In Section 3 presents the new method for construction of TFN. Section 4 includes the experimental results and discussion. Section 5 presents the conclusion and future works.

# II. THE AUSTRALIAN MONTHLY ELECTRICITY CONSUMPTION DATA SET

The Australian monthly electricity consumption data set [10] will be used to illustrate the proposed approach. This data set involves the electricity consumption values from January 1956 to August 1995, thus the data set has a total of 476 samples. As it can be clearly seen in Fig. 1, the consumption of the electricity has always increased with respect the time which shows the trend property of the data. Moreover, the data has a seasonality characteristic of 12 months. The seasonality property can be clearly seen from the zoomed out subplots presented in Fig. 1. Thus, the data has trend and seasonality

characteristics which are commonly encountered in time series analysis [11, 12].



Figure 1. The Australian monthly electricity consumption data set

#### III. FORECAST REPRESENTATION WITH TFN

In this section, we will firstly present a FLUBE to define the uncertainty in the forecast. Then, the uncertainty in the forecast will be represented with TFNs. Here, instead of the representing the forecast with a crisp value which is defined with PI, the level of uncertainty associated with the point forecasts will be quantified by defining TFNs within the uncertainty interval provided by the FLUBE. This will give the opportunity to handle the forecast as linguistic terms which will increase the interpretability.

#### A. Fuzzy Logic based Lower and Upper Bound Estimator

The proposed FLUBE consists of two Fuzzy Logic Systems (FLSs) which will define the uncertainty bounds of the point forecast error. The FLUBE is designed in an offline manner (as sketched in Fig. 2) in the following 3 steps.

## Step 1: Design of the Forecast Model for FLUBE

In the first step, the time series forecast model is constructed, by using simply the target data (S) to obtain the Forecast value (F). Here, the forecast model can be constructed from different structures such as time series regression, conditional means/variance (ARIMA, SARIMA, GARCH, etc.) and multivariate models. In this study, we will prefer a SARIMA forecast model for illustrative purposes. However, it is worth to underline that the proposed FLUBE can be also easily employed to other forecast structures.

Step 2: Training data generation for FLUBE.

In this step, a training data set is constructed where the input is the target data (S) and the output is error value (E) which is defined as follows:

$$E_k = F_k - S_k \tag{1}$$

where  $F_k$  is the forecast value at the  $k^{th}$  sample. In Fig. 3, this input/output data set is sketched for the Australian monthly electricity consumption data set. Once the training data set is constructed, a selection algorithm is used to find the lower and upper bounds of the input/output data, i.e.,  $E_k < 0$  ( $E_{min}$ ) and  $E_k > 0$  ( $E_{max}$ ). However, this will result with a conservative representation and this might not catch the variation of the error values. Therefore, we will select the boundary values of

the error terms  $(E_{min}, E_{max})$  from predefined number of subintervals (P). Therefore, the  $E_{max}$  and  $E_{min}$  and their corresponding S values  $(S_{min} \text{ and } S_{max})$  will be collected from each subinterval of S such that  $E_k < 0$   $(E_{min})$  and  $E_k >$ 0  $(E_{max})$ . When there does not exist any  $E_k$  satisfying these constraints, then the selection method will not assign any values of  $E_k$  to the  $E_{max}$  and  $E_{min}$  for this subinterval of S. Obviously, lower and upper bounds constructed with more selected points (more subintervals) will provide a general approximation while neglecting the variation of the error values. Thus, P is a design parameter and needs to be tuned for each handled data set. An illustration of the collected boundary values of error terms has been shown in Fig. 4 for P=80. Note that, different selection algorithms can be also used to obtain the  $E_{max}$  and  $E_{min}$  data set.



Step 3: FLUBE Design

In this step, the design of the FLUBE is performed. The FLUBE consists of two Fuzzy Logic Systems (FLSs) which will define the lower and upper bounds of the error terms, respectively. The fuzzy rule base structure of the Lower and Upper FLSs are as follows:

$$R_{\omega}^{L}$$
: If S is  $A_{\omega}^{L}$ , Then  $E^{L}$  is  $B_{\omega}^{L}$  (2)

$$R^{U}_{\omega}$$
: If S is  $A^{U}_{\omega}$ , Then  $E^{U}$  is  $B^{U}_{\omega}$  (3)

where  $A_{\omega}^{L}$  and  $A_{\omega}^{U}$ , are the antecedent membership functions,  $B_{\omega}^{L}$  and  $B_{\omega}^{U}$  are the consequent crisp sets and W ( $\omega = 1, ..., W$ ) is the total number of rules.

In the design of the FLUBE, we will use the data sets  $[S_{min}, E_{min}]$  and  $[S_{max}, E_{max}]$ , as explained in Step 2, to the train Lower FLS (LFLS) and Upper FLS (UFLS). In this

study, we will use the ANFIS toolbox /MATLAB to generate the fuzzy inference systems.



Figure 3. Consumption-error distribution of the electricity data



Figure 4. The collected data sets via the selection algorithm

As it has been mentioned the previous step, the number of subintervals (P) is a design parameter and its choice will directly affect the mapping of the FLUBE. The effect of the P value is illustrated in Fig. 5 for P = 80 and P = 320, respectively for the electrical consumption data set. It can be observed that relatively small values of the P parameter provided extreme bounds which cover almost all error terms via FLUBE. In spite of that, relatively big value of the P parameter presents a smoother bound characteristic that can be caused less coverage performance via FLUBE.

**Remark:** If the mean value of the error distribution is not around zero, then the mean value should be subtracted from the error terms ( $E_{min}$ ,  $E_{max}$ ) in Step 2 (training data generation phase). Thereby, the mean value should be added to the outputs of FLUBE.

#### B. TFN Generation

In this subsection, the novel online TFN representation and generation of the uncertainty in the forecast will be explained. This presentation tries to explain how the target values are similar to the forecast values by using linguistic terms, instead of the conventional PI representations in literature. Here, at each sample, the forecast uncertainty of the next sample will be approximately defined by TFN. Consequently, the TFN generation will be accomplished in an online manner as shown in Fig. 6.



Figure 5. The effect of the P value: (a) P=80, (b)P=320

As it has been mentioned in the preceding subsection, FLUBE has been designed with respect to the *S* target data. However, the  $S_k$  value will not be available for evaluating FLUBE, to generate the  $E_k^L$  and  $E_k^U$  (at the  $k^{th}$  sample). Thus, we will use the single point forecast value  $F_k$  as the input of FLUBE since the expected value of the error terms  $(E_k)$  is zero. In this context, the lower and upper uncertainty bounds  $(L_k, U_k)$  of the forecast are defined as follows:

$$L_k = F_k + E_k^L$$

$$U_k = F_k + E_k^U$$
(4)

Here, we will employ a TFN which is defined by a triplet  $(L_k F_k, U_k)$  to represent the forecast at the  $k^{th}$  sample. The membership degree of the TFN at the  $k^{th}$  sample  $(\mu_k^{TFN})$  is defined as follows [13]:

$$\mu_{k}^{TFN}(X') = \begin{cases} 0, & \text{if } X \notin [L_{k} \ U_{k}] \\ \frac{X' - L_{k}}{|E_{k}^{L}|}, & \text{if } X \in [L_{k} \ F_{k}] \\ \frac{U_{k} - X'}{|E_{k}^{U}|}, & \text{if } X \in [F_{k} \ U_{k}] \end{cases}$$
(5)

where X' is a crisp input. A generic TFN is illustrated in Fig. 7. It is worth to underline that, in comparison to other PI methods, TFN representation provides knowledge to the

decision maker by giving information about forecast accuracy degree  $(\mu_k^{TFN})$ .



Figure 7. Triangular Fuzzy Number representation

# C. Performance index.

Here, the performance of proposed method is examined with PI Coverage Probability (PICP) and PI Normalized Averaged Width (PINAW) which are commonly used in literature to evaluate the PI performance. [14] The PICP is measured by counting the number of target values covered by the constructed PIs. The PICP shows in which probability target values will be covered by the lower and upper bounds and thus is defined as:

$$PICP = \frac{100}{n} \sum_{k=1}^{n} c_k \tag{6}$$

where *n* is the number of samples and  $c_k$  is [11]:

$$c_k = \begin{cases} 1, & \text{if } y_k \in [L_i \ U_i] \\ 0, & \text{if } y_i \notin [L_i \ U_i] \end{cases}$$
(7)

The second measure is the PINAW which provides a measure about the width of the PIs and is defined as:

$$PINAW = \frac{100}{nR} \sum_{k=1}^{n} (U(X_k) - L(X_k))$$
(8)

where R is the range of underlying targets. Same as in the PICP measure, using the extreme target values as lower and upper bounds of PIs will result with 100% PINAW. From a practically point of view, it is important to have narrow PIs (relatively small PINAW value) with a high coverage probability (relatively high PICP value).

#### IV. RESULTS AND DISCUSSIONS

This section presents the experiment results where the proposed method is illustrated via the Australian monthly electrical consumption data set.

As it has been asserted in Section 3, in the first step of the design of the FLUBE, there is a need of a forecast model. Thus, since the handled data set has seasonality and trend characteristics, we have used a Seasonal AutoRegressive Integrated Moving Average (SARIMA) forecast model which is with constant mean c defined as follows [15,16]:

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D(S_k-c) = \theta(L)\Theta(L^s)a_k$$
(9)

where parameters corresponds as follows

$$\phi(L) = \left(1 - \phi_1 L - \dots - \phi_p L^p\right)$$
  

$$\theta(L) = \left(1 - \theta_1 L - \dots - \theta_q L^q\right)$$
  

$$\Phi(L) = \left(1 - \Phi_1 L^s - \dots - \Phi_p L^{sp}\right)$$
  

$$\Theta(L) = \left(1 - \Theta_1 L^s - \dots - \Theta_0 L^{sQ}\right)$$
  
(10)

where *L* described as lag operator with polynomial equation  $L^iS_t = S_{t-i}$ , *p* and *P* degree of AutoRegressive (AR) and Seasonal AutoRegressive (SAR) operator parameterized as  $\phi$  and  $\Phi$ , *q* and *Q* degree of Moving Average (MA) and Seasonal Moving Average (SMA) operator parameterized as  $\theta$  and  $\Theta$ .

For the experimental data set, the SARIMA(p,d,q) (P,D,Q)<sub>s</sub> model has been designed as follows SARIMA (2,1,2)(15,1,2)<sub>12</sub>. The performance values of the forecast model with respect to MAPE, POA, Symmetric Mean Absolute Percentage Error (SMAPE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) [2] are tabulated in Table 1. It can be clearly seen that the forecast model performed with a satisfactory performance [17] where it has provided relatively low MAPE and RMSE values.



In second step of the FLUBE design, the error terms  $E_k$  (given in Equation (1)) are calculated via SARIMA forecast model. The distribution of the error terms  $(E_k)$  with respect to the consumption values  $(S_k)$  is shown in Fig. 2. Once these data sets are obtained, the presented selection algorithm has been employed for P = 80 which provides a subinterval range of 200 m.kWh. As it has been mentioned in the "Remark", there is need to shift the error terms by its mean which is calculated as -3.12. Then, the required training data sets  $[S_{min}, E_{min}]$  and  $[S_{max}, E_{max}]$  are collected.

In the final step of the FLUBE design, the two FLSs (LFLS and UFLS) are generated by using the collected training data sets  $[S_{min}, E_{min}]$  and  $[S_{max}, E_{max}]$ . Here, the LFLS and UFLS are constructed with 10 Gaussian antecedent MFs and 10 linear consequent MFs, thus in total 10 rules (W = 10). It is worth to underline that the design of the FLUBE has been accomplished in an offline manner.

As it has been mentioned in Section 2, the FLUBE will provide an approximate bound ( $E^L$  and  $E^U$ ) on the uncertainty of the underlying system. We will now use this information to define the forecast with linguistic terms, i.e. TFNs. In the TFN generation, we will use the outputs of the FLUBE at the  $k^{th}$ sample  $(E_k^L \text{ and } E_k^U)$  to define the uncertainty bounds of the forecast value via Equation (4). For instance, the uncertainty bounds of the forecast  $(L_k \text{ and } U_k)$  are shown in Fig.8 for of the handled data set. Now, the TFN can be simply generated by using Equation (5). Since the TFN will naturally provide an extra dimension  $(\mu)$ , the forecast should be illustrated in a 3-D plot as shown in Fig. 9 where the online generated TFNs for each sample can be clearly seen. It can be observed that at each sample (k) the uncertainty interval of the forecast has different width with respect to the point forecast value. Thus nonsymmetrical TFNs will be generated. This phenomena can be clearly observed in Fig. 9 for sample k=390.

The performance of the novel FLUBE based TFN generation and representation has been evaluated with respect to defined measures PICP and PINAW. The calculated values of PICP and PINAW are found as 87.88% and 4.5099%, respectively. Since the proposed forecast representation resulted with a relatively small PINAW value and a high PICP value, thus the forecast uncertainty interval generated by the FLUBE provided a satisfactory prediction performance.



To analyze the forecast quality, the distribution of the membership grades ( $\mu$ ) (via Equation (5)) of the studied benchmark data is illustrated in Fig. 10. Since the membership grades are mostly distributed around the maximum membership grade value 1, it can be concluded that the quality of the forecast is relatively good. Thus, the proposed FLUBE based TFN representation is an efficient and useful approach to represent the uncertainty of the forecast.



Figure 9. Results in 3D for 10 samples of data

### V. CONCLUSION AND FUTURE WORKS

In this study, a novel representation of the forecast uncertainty through the construction TFNs (linguistic terms) has been presented. Since the uncertainty of the point forecast has been quantified with linguistic terms, the interpretability of the forecast has been relatively increased and a relative degree of accuracy could have been defined. The proposed approach consists of an offline design and an online generation phase. In the offline design, the design of the FLUBE, which provides the uncertainty interval for the training data, is accomplished. Then, in the online phase, TFNs are generated for each sample within the generated uncertainty interval provided by the FLUBE. The presented approach has been illustrated on a benchmark data set. Based on the presented results, it has been concluded that the proposed FLUBE based online TFN representation and generation is an efficient approach to represent the uncertainty and accuracy of the forecast.

Future work will focus on extending the proposed approach to datasets which inherit a higher level of uncertainty and complexity.

#### ACKNOWLEDGMENT

This research is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) (project number: 113E206) and also by TUBITAK Industrial Research Funding Program (Grant Number 7131341, awarded to GETRON Bilisim Hizmetleri A.S.). All of these supports are appreciated.

#### References

- A. Khosravi and S. Nahavandi, "An interval type-2 fuzzy logic systembased method for prediction interval construction," Applied Soft Computing., vol. 24, no. 9, pp. 222–231, 2014.
- [2] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, pp. 679-688, 2006.
- [3] T. Heskes, "Practical confidence and prediction intervals," in Advances in Neural Information Processing Systems, vol. 9, T. P. M. Mozer and M. Jordan, Eds. Cambridge, MA: MIT Press, 1997, pp. 176–182.
- [4] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," IEEE Trans. Neural Netw., vol. 22, no. 9, pp. 1341–1356, Sep. 2011.
- [5] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," IEEE Trans. Power Syst., vol. 25, no. 3, pp. 1496–1503, Aug. 2010.
- [6] D. Srinivasan and M. Lee, "Survey of hybrid fuzzy neural approaches to electric load forecasting," in Proc. IEEE Int. Conf. Syst., Man Cybern., Intell. Syst. 21st Century, vol. 5. Oct. 1995, pp. 4004–4008.
- [7] C. Chatfield, "Calculating interval forecasts," J. Business Econ. Stat., vol. 11, no. 2, pp. 121–135, Apr. 1993.
- [8] H. Quan, D. Srinivasan, A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of neural network-based prediction intervals for short term electrical load forecasting," in Proc. IEEE Symp. CIASG, Apr. 2013, pp. 66–72.
- [9] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network

based prediction intervals," IEEE Trans. Neural Netw., vol. 22, no. 3, pp. 337–346, Mar. 2011.

- [10] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, Forecasting -Methods and Applications, 3rd ed., John Wiley & Sons, New York, 1998.
- [11] D. Pelusi, M. Tivegna and P. Ippoliti, "Improving the profitability of Technical Analysis thorough intelligent algorithms," Journal of Interdisciplinary Mathematics, vol. 16, no. 2&3, pp. 203-215, 2013.
- [12] D. Pelusi and M. Tivegna, "Optimal trading rules at hourly frequency in the foreign exchange markets," Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer: Milan, 2012 pp. 341-348.
- [13] E. Yesil, M. F. Dodurka and L. Urbas, "Triangular fuzzy number representation of relations in Fuzzy Cognitive Maps," Proceedings of the FUZZ-IEEE 2014 – International Conference on Fuzzy Systems, July 6-11, Beijing, China, 2014.
- [14] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 2, pp. 303–315, Feb. 2014.
- [15] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, Time Series Analysis Forecasting and Control, 4th ed., John Wiley & Sons, New Jersey, 2008.
- [16] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 2nd ed., Springer-Verlag, New York, 2002.
- [17] D. Frechtling, Forecasting Tourism Demand: Methods and Strategies, 1st ed., Butterworth Heinemann, Oxford, 2001.