

## Research Article

# Long-Term Load Forecasting Based on a Time-Variant Ratio Multiobjective Optimization Fuzzy Time Series Model

Xiaojuan Liu<sup>1,2</sup> and Jian'an Fang<sup>1</sup>

<sup>1</sup> College of Information Science & Technology, Donghua University, Shanghai 201620, China

<sup>2</sup> School of Mathematics and Physics, Shanghai University of Electric Power, Shanghai 201300, China

Correspondence should be addressed to Jian'an Fang; [jafang@dhu.edu.cn](mailto:jafang@dhu.edu.cn)

Received 9 January 2013; Accepted 4 March 2013

Academic Editor: Engang Tian

Copyright © 2013 X. Liu and J. Fang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Load forecasting problem is a complex nonlinear problem linked with economic and weather factors. Long-term load forecasting provides useful information for maintenance scheduling, adequacy assessment, and limited energy resources for electrical power systems. Fuzzy time series forecasting models can be used for long-term load forecasting. However, the interval length has been chosen arbitrarily in the implementations of known fuzzy time series forecasting models, which has an important impact on the performance of these models. In this paper, a time-variant ratio multiobjective optimization fuzzy time series model (TV-RMOP) is proposed, and its performance is tested on the prediction of enrollment at the University of Alabama. Results clearly promote the forecasting accuracy as compared to the conventional models. A genetic algorithm is used to search for the length of intervals based on the training data while Pareto optimality theory provides the necessary conditions to identify an optimal one. The TV-RMOP model is applied for the long-term load forecasting in Shanghai of China.

## 1. Introduction

Electric power load forecast has been a research topic for many decades and the accuracy of load forecast is crucial to electricity power industry due to its direct influence on generating planning. During the past years, there have been numerous attempts to improve the accuracy of load forecasting methods [1, 2]. The forecasting model and methods can be divided into two categories: statistical methods and computational intelligence algorithms. The former method include regression analysis and time series [3, 4]. The latter include neural network, expert system, fuzzy logical, and so forth [5–7]. Now more and more scholar and researchers pay much attention to uncertainty hybrid method [8–10].

There are three different types of electric load forecasting depending on the time horizon and the operating decision that needs to be made: short-, medium-, and long-term forecasting. In general, long term forecasting includes prediction making for a range more than a year and is needed for power supply and delivery system. Therefore, the weather factor is not the main factor to affect the long-term load forecast, only using historical load data can predict the 1 year ahead load.

Fuzzy time series [11, 12] prediction method is likely to get a satisfactory result.

Fuzzy time series has been widely studied for recent years for the aim of forecasting. The basic process of conventional model is shown in Figure 1(a). Related studies mainly focus on three factors: the partition of discourse, the content of forecasting rules, and the methods of defuzzification, all of which influence the prediction accuracy of forecasting models. A number of fuzzy time series forecasting models have been presented since Song and Chissom [13, 14] first presented the concepts of fuzzy time series based on the fuzzy set theory. However, the length of intervals has been chosen arbitrarily in known models. This paper focuses on the partition of discourse in TV-RMOP. We propose a novel approach to determine the length of intervals based on multiobjective optimization method in order to obtain more accurate forecasts in fuzzy time series. The proposed approach is based on two-parameter constrained multiobjective optimization. A genetic algorithm is used to search for the two parameters based on the training data. The TV-RMOP model is shown in Figure 1(b). Its performance is tested on the prediction of enrollment at the University

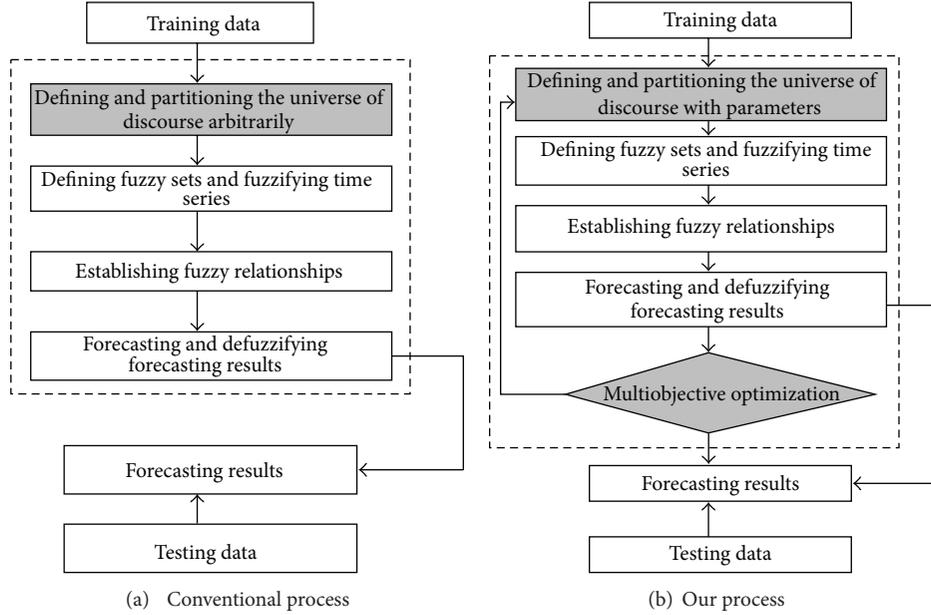


FIGURE 1: Two distinct prediction processes of fuzzy time series forecasting.

of Alabama and applied for the long-term load forecasting in Shanghai of China. Results investigated that TV-RMOP model considerably improves the forecasting performance.

The rest of this paper is organized as follows. Section 2 introduces the methodology background briefly. The presented model is described, and the forecasting results on the prediction of enrollment at the University of Alabama are presented and discussed in Section 3. In Section 4, TV-RMOP model is applied for the long-term load forecasting in Shanghai of china. Finally, conclusions are presented in Section 5.

## 2. Background

**2.1. Fussy Time Series.** The concepts of fuzzy time series are described as follows [13, 14].

Let  $U$  be the universe of discourse, where  $U = \{u_1, u_2, \dots, u_n\}$ . A fuzzy set  $A$  of  $U$  is defined as  $A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n$ , where  $f_A$  is the membership function of the fuzzy set  $A$ ,  $f_A : U \rightarrow [0, 1]$ .  $u_i$  is a generic elements of fuzzy set  $A$ .  $f_A(u_i)$  is the degree of belongingness of  $u_i$  to  $A$ ,  $f_A(u_i) \in [0, 1]$ , and  $1 \leq i \leq n$ .

**Definition 1.** Let  $Y(t)$  ( $t = \dots, 0, 1, 2, \dots$ ) be the universe of discourse and also a subset of real numbers. It is assumed that  $f_i(t)$  is defined on  $Y(t)$  and  $F(t)$  is the collection of  $f_i(t)$ , then  $F(t)$  is called a fuzzy time series on  $Y(t)$ .

**Definition 2.** It is assumed that  $F(t)$  is a fuzzy time series and  $F(t) = F(t-1) \times R(t, t-1)$ , where  $R(t, t-1)$  is a fuzzy relation and  $\times$  is an operator which is caused by  $F(t-1)$ . The relationship between  $F(t)$  and  $F(t-1)$  can be denoted by  $F(t-1) \rightarrow F(t)$  when  $F(t) = F(t-1) \times R(t, t-1)$  is the first-order fuzzy time series model of  $F(t)$ .

**Definition 3.** Let  $F(t)$  be a fuzzy time series. For any  $t$ ,  $F(t) = F(t-1)$  and  $F(t)$  have only finite elements and therefore  $F(t)$  is a time-invariant fuzzy time series; otherwise, it is a time-variant fuzzy time series.

**2.2. Multiobjective Optimization and Pareto Points.** It is common in decision-making problems to have a set of parameters that play a dominant role in the process. It is also desirable for the decision made to be the best possible with respect to the available data. In the context of mathematics, the best decision coincides with the optimization of a function of the decision parameters, known as cost function or objective.

On the one hand, if the goal is to optimize a single cost function, then the problem is classified as single objective optimization. On the other hand, the process of simultaneously optimizing two or more conflicting objectives, subject to certain constraints, is identified as multiobjective optimization [17–19].

Such problems are called multiobjective optimization problems and formulated as follows:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), f_k(\mathbf{x})] \\ \text{subject to} \quad & h_i(\mathbf{x}) \leq 0, \quad i = 0, 1, \dots, n_1 \\ & g_l(\mathbf{x}) = 0, \quad l = 0, 1, \dots, n_2, \end{aligned} \quad (1)$$

where  $k$  denotes the number of objective functions;  $n_1, n_2$  identify the number of inequality and equality constraints, respectively.  $\mathbf{x}$  is an  $n$ -dimensional vector in the solution space  $\mathbf{X}^n$ , and  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbf{X}^n$ .

Multiobjective optimization problems do not have a single global solution in contrast to the single objective problems. As a result, it is required to specify a number of criteria-conditions for which a set of solutions can be identified as

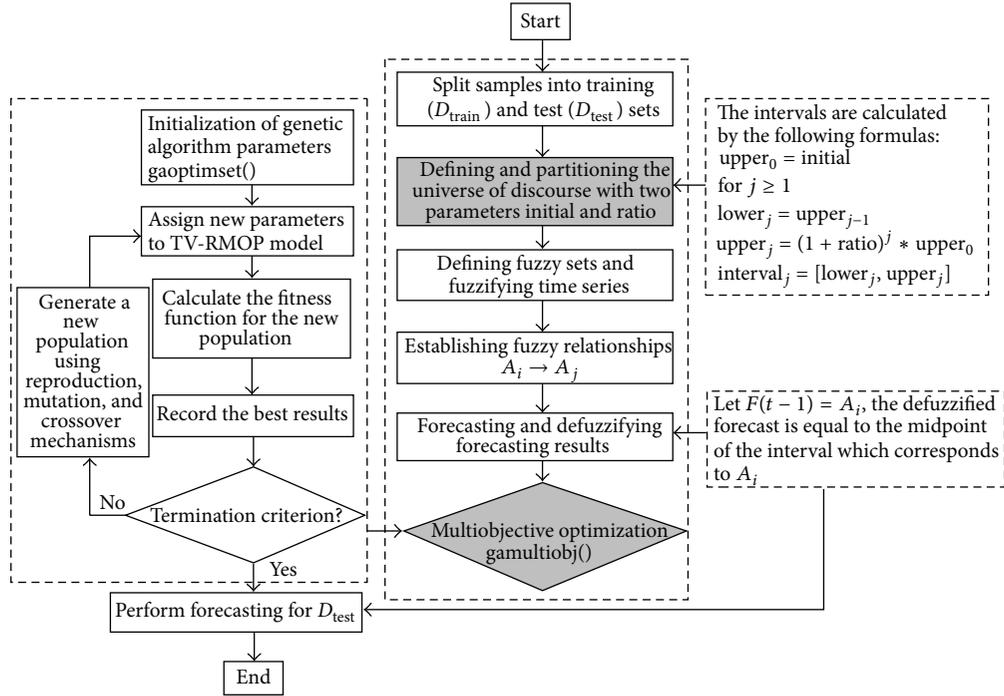


FIGURE 2: Proposed time-variant ratio multiobjective optimization fuzzy time series model.

optimal. In general, the optimal solutions for  $F(\mathbf{x})$  can be obtained based on the Pareto optimality theory. The set of optimal solutions, which is known as the Pareto optimal set  $\mathbf{X}^*$ , are those solutions such that each of these cannot improve any objective function in  $F(\mathbf{x})$  without the worsening of some other objective functions at the same time. The optimal solution set is defined as

$$\mathbf{X}^* = \{\mathbf{x}^* \in \mathbf{X} \mid \text{does not exist } \mathbf{x} \in \mathbf{X}, F(\mathbf{X}) \leq F(\mathbf{X}^*)\}. \quad (2)$$

The corresponding points, which are known as the nondominated points, form the Pareto front in the objective space. Since there exists different tradeoff solutions to minimize  $F(\mathbf{x})$ , one possible way to solve the multiobjective problem is to select one solution manually from the Pareto optimal set.

### 3. Modeling and Verification

**3.1. Proposed TV-RMOP Model.** Time series is divided into two parts. One part is used for training purpose while the other part is used for testing. The proposed model was applied to the training data to make estimation and then the predicted values were found for the test data. In the training phase, five measures are adopted to evaluate the model prediction accuracy with the intention to formulate the multiobjective problem. These measures are objective functions with parameters and defined as follows:

Root Mean Square Error (RMSE):

$$\text{RMSE}(\cdot) = \sqrt{\frac{1}{N} \sum_{t=1}^N (R_t - P_t)^2}$$

Mean Absolute Error (MAE):

$$\text{MAE}(\cdot) = \frac{1}{N} \sum_{t=1}^N |R_t - P_t|$$

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE}(\cdot) = \frac{100}{N} \sum_{t=1}^N \left| \frac{R_t - P_t}{R_t} \right|$$

Maximum Absolute Percentage Error (MAP):

$$\text{MAP}(\cdot) = \max_{t=1, \dots, N} \left( 100 \times \left| \frac{R_t - P_t}{R_t} \right| \right)$$

Theil Inequality Coefficient (THEIL):

$$\text{THEIL}(\cdot) = \frac{\text{RMSE}}{\sqrt{(\sum_{t=1}^N (R_t)^2) / N}}, \quad (3)$$

where  $R_t$  is the actual value and  $P_t$  is the predicted value at time instant  $t$ .  $N$  is the number of predictions made for a prespecified time interval. All measures should be minimized in the optimization procedure.

As shown in Figure 2, the TV-RMOP model can be summarized as follows.

*Step 1.* Define the universe of discourse and intervals using equations defined in the upper-right dashed box. Where *initial* and *ratio* are two unknown parameters.

TABLE 1: The enrollment data.

Year	Enrollments
1971	13055
1972	13563
1973	13867
1974	14696
1975	15460
1976	15311
1977	15603
1978	15861
1979	16807
1980	16919
1981	16388
1982	15433
1983	15497
1984	15145
1985	15163
1986	15984
1987	16859
1988	18150
1989	18970
1990	19328
1991	19337
1992	18876

*Step 2.* Define fuzzy sets based on the universe of discourse and fuzzify the historical data.

*Step 3.* Establish fuzzy relationships:  $A_i \rightarrow A_j$ .

*Step 4.* Forecast. Let  $F(t-1) = A_i$ ; the forecast is the midpoint of the interval which corresponds to  $A_j$ .

*Step 5.* Calculate RMSE, MAE, MAPE, MAP, and THEIL, respectively. Each of them is a function with variable *initial* and *ratio*.

In the training phase, the training samples ( $D_{\text{train}}$ ) are used to calculate the above five functions. The determination of variable *initial* and *ratio* is formulated as a multiobjective optimization problem:

$$\begin{aligned} \min \quad & F(\text{initial}, \text{ratio}) \\ & = [\text{RMSE}, \text{MAE}, \text{MAPE}, \text{MAP}, \text{THEIL}] \quad (4) \\ \text{subject to} \quad & \text{initial} \in (a, b), \text{ ratio} \in (A, B). \end{aligned}$$

In the test phase, an optimal solution is brought into the model, and the forecast value is determined in the same way as the training phase.

Genetic algorithm can be used to search for a solution of multiobjective optimization problems. The process is shown

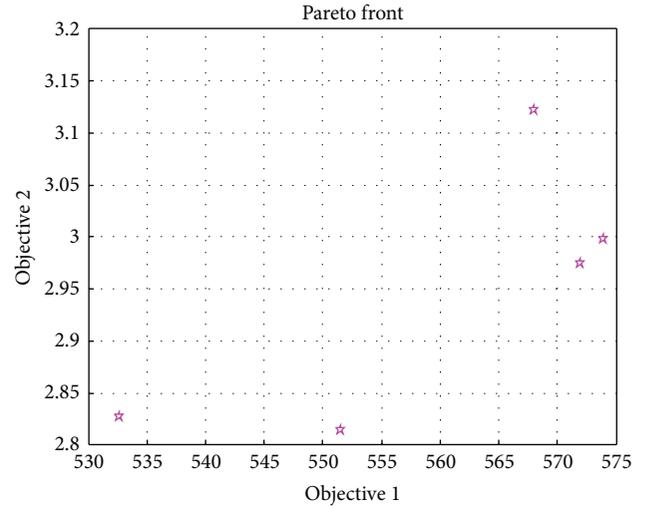


FIGURE 3: Pareto front of TV-RMOP model for enrollment.

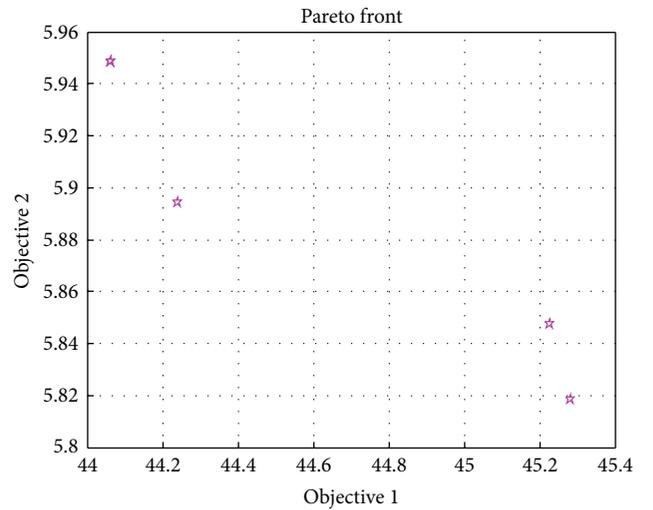


FIGURE 4: Pareto front of TV-RMOP model for load forecasting.

in the dashed box on the left in Figure 2. In this paper we use a MATLAB function “gamultiobj” to optimize the initial and ratio parameters.

**3.2. Verification of Proposed Model.** In order to show what we have achieved from the TV-RMOP model, we apply it to the enrollment of University of Alabama and then have compared the obtained results with those obtained from Chen’s method [20]. Table 1 lists the enrollment from 1971 to 1992. The observations from 1971 to 1989 are used for the training ( $D_{\text{train}}$ ), while the observations from 1990 to 1992 are used for test.

The minimal value of  $D_{\text{train}}$  is 13055, then [12000, 13000] is taken as constrains of variable *initial*. (0, 0.2] is taken as constrains of the other variable *ratio*, because when the ratio is greater than 0.2, the length of interval will be too large, and the number of intervals will decrease. It results in diminishing

TABLE 2: Pareto solutions of enrollment.

Initial	Ratio	RMSE	MAE	MAPE	MAP	THEIL
12266.7	0.0621	571.9	484.4	2.9745	5.8418	0.0359
12230.4	0.0530	573.9	486.0	2.9983	5.6954	0.0360
12135.5	0.0455	532.7	451.8	2.8276	6.6546	0.0334
12846.0	0.0676	568.0	501.2	3.1217	5.2926	0.0357
12266.7	0.0620	551.5	459.0	2.8142	5.8748	0.0346

TABLE 3: The forecasting results for  $D_{test}$ .

Year	Actual enrollment	[15]	[16]	TV-RMOP model
1990	19328	18685	18970	19271
1991	19337	19138	19306	19271
1992	18876	19176	19315	19271

TABLE 4: The measures for  $D_{test}$ .

Initial	Ratio	RMSE	MAE	MAPE	MAP	THEIL
12135.5	0.0455	233.5	172.7	0.9098	2.79	0.01212

TABLE 5: Load in Shanghai region.

Year	Load (hundred million Kwh)
1990	264.74
1991	288.78
1992	317.38
1993	345.86
1994	337.3
1995	403.27
1996	430.4
1997	454.26
1998	482.94
1999	501.2
2000	559.42
2001	592.99
2002	645.71
2003	745.97
2004	821.44
2005	921.97
2006	990.15
2007	1072.38
2008	1138.22
2009	1153.38
2010	1295.87

the fluctuation in fuzzy time series. Then the problem can be formulated as

$$\begin{aligned}
 &\min F(\text{initial, ratio}) \\
 &= [\text{RMSE, MAE, MAPE, MAP, THEIL}] \quad (5) \\
 &\text{subject to } \text{initial} \in [12000, 13000], \text{ ratio} \in (0, 0.2].
 \end{aligned}$$

We use a MATLAB function “gamultiobj” to solve (5). The initialization of genetic algorithm parameters is set as follows. The fraction of population on non-dominated front is 0.1, the number of individuals is 50, maximum number of generations allowed is 100, and termination tolerance on fitness function value is  $1e - 100$ . The Pareto solutions of training phase are provided in Table 2 (seen in Figure 3).

We select 12135.5 and 0.0455 as optimal parameters to forecast the enrollment from 1990 to 1992. For this initial and ratio, the universe of discourse is defined as [12135.5, 19694.1]. The intervals are finally defined as:  $u_1 = [12135.5, 12681.6]$ ,  $u_2 = [12681.6, 13252.3]$ ,  $u_3 = [13252.3, 13848.6]$ ,  $u_4 = [13848.6, 14471.8]$ ,  $u_5 = [14471.8, 15123]$ ,  $u_6 = [15123, 15803.6]$ ,  $u_7 = [15803.6, 16514.7]$ ,  $u_8 = [16514.7, 17257.9]$ ,  $u_9 = [17257.9, 18034.5]$ ,  $u_{10} = [18034.5, 18846]$ ,  $u_{11} = [18846, 19694.1]$ . Tables 3 and 4 list the forecast results and corresponding measures, respectively. The best RMSE value is 500 with the interval length of 500 for Chen’s method [20]. The RMSE value is 425.46 for Kuo’s model [15] and 327.54 for Wong’s model [16]. The proposed model produces 233.5 for RMSE. Obviously, the TV-RMOP model gives more accurate forecast than [15, 16, 20].

#### 4. Applying TV-RMOP Model on Long Term Load Forecasting

Long term load forecasting is needed for power supply and delivery system. We use twenty years (1990~2010) load of Shanghai region in China for forecast (data from the 2011 Shanghai statistical yearbook). The load is shown in Table 5. Where loads from 1990 to 2007 are used for the training ( $D_{train}$ ), and loads from 2008 to 2010 are used for test ( $D_{test}$ ).

The minimal value of  $D_{train}$  is 264.74, [240, 260] is taken as constrains of variable initial. (0, 0.4] is taken as constrains

TABLE 6: Pareto solutions of load.

Initial	Ratio	RMSE	MAE	MAPE	MAP	THEIL
257.68	0.2317	45.28	34.47	5.8185	12.184	0.0719
250.80	0.2400	44.06	34.49	5.9488	13.627	0.0699
256.51	0.2329	45.22	34.54	5.8476	12.448	0.0718
250.80	0.2400	44.06	34.49	5.9488	13.627	0.0699
252.48	0.2380	44.24	34.30	5.8942	13.273	0.0702

TABLE 7: The forecasting results for  $D_{\text{test}}$ .

Year	Actual load	TV-RMOP model
2008	1138.22	1021.1
2009	1153.38	1266.2
2010	1295.87	1266.2

TABLE 8: The measures for  $D_{\text{test}}$ .

Initial	Ratio	RMSE	MAE	MAPE	MAP	THEIL
250.8	0.24	95.43	86.53	7.45	10.29	0.07966

of the other variable ratio. Then the problem can be formulated as

$$\begin{aligned} \min \quad & F(\text{initial}, \text{ratio}) \\ & = [\text{RMSE}, \text{MAE}, \text{MAPE}, \text{MAP}, \text{THEIL}] \quad (6) \\ \text{subject to} \quad & \text{initial} \in [240, 260], \text{ratio} \in (0, 0.4]. \end{aligned}$$

The initialization of genetic algorithm parameters is set as follows. The fraction of population on non-dominated front is 0.1, the number of individuals is 50, maximum number of generations allowed is 100, and termination tolerance on fitness function value is  $1e - 100$ . The Pareto solutions of training phase are provided in Table 6, also seen in Figure 4.

Select 250.8 and 0.24 as optimal parameters to forecast the loads from 2008 to 2010. Tables 7 and 8 list the forecast results and corresponding measures, respectively. It is seen that the MAPE and MAP measures are 7.45% and 10.29%, which meet the actual demand.

## 5. Conclusions

Fuzzy time series forecasting method are getting quite popular in recent years. Most studies focus on the forecasting rules and methods of defuzzification. How to determine the lengths of intervals is less studied. The decision on what the lengths will be is important for forecasting accuracy.

In this study, a new model TV-RMOP is proposed, which formulates the forecasting as a multiobjective optimization problem and solve it by the use of genetic algorithms. The optimal lengths of intervals can be determined in the proposed model. The model is tested for forecasting the enrollment of University of Alabama. The experimental results

show that the TV-RMOP model is more accurate than existing models. Finally, we apply this model for long-term load forecasting of Shanghai Region with satisfactory result.

## Acknowledgments

The authors would like to thank the anonymous reviewers in MPE for helpful suggestions and correction. This work was conducted by using of the MATLAB software and the authors declared that it has no conflict of interests to this work. The work was financially supported by Shanghai Municipal Nature Science Foundation under Grant 10ZR1401400.

## References

- [1] E. Almeshai and H. Saltan, "A methodology for electric power load forecasting," *Alexandria Engineering Journal*, vol. 50, pp. 137–144, 2011.
- [2] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: tools for decision making," *European Journal of Operational Research*, vol. 199, no. 3, pp. 902–907, 2009.
- [3] Y. M. Wi, S. K. Joo, and K. B. Song, "Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 596–603, 2012.
- [4] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 134–141, 2012.
- [5] J. Taylor, "Short-term load forecasting with exponentially weighted methods," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 458–464, 2012.
- [6] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Interval type-2 fuzzy logic systems for load forecasting: a comparative study," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1274–1282, 2012.

- [7] M. Xia, Y. C. Zhang, L. G. Wen, and X. L. Ye, "Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs," *Knowledge-Based Systems*, vol. 36, pp. 253–259, 2012.
- [8] N. Amjady and A. Daraeepour, "Midterm demand prediction of electrical power systems using a new hybrid forecast technique," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 755–765, 2011.
- [9] R. A. Hooshmand, H. Amooshahi, and M. Parastegari, "A hybrid intelligent algorithm based short-term load forecasting approach," *International Journal of Electrical Power & Energy Systems*, vol. 45, no. 1, pp. 313–324, 2013.
- [10] S. Zhaowei, Z. Lingfeng, M. Shangjun, F. Bin, and Z. Taiping, "Incomplete time series prediction using max-margin classification of data with absent features," *Mathematical Problems in Engineering*, vol. 2010, Article ID 513810, 14 pages, 2010.
- [11] E. Tian, D. Yue, and Z. Gu, "Robust  $H_\infty$  control for nonlinear systems over network: a piecewise analysis method," *Fuzzy Sets and Systems*, vol. 161, no. 21, pp. 2731–2745, 2010.
- [12] E. G. Tian and D. Yue, "Reliable  $H_\infty$  filter design for T-S fuzzy model-based networked control systems with random sensor failure," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 1, pp. 15–32, 2013.
- [13] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—part I," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1–9, 1993.
- [14] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—part II," *Fuzzy Sets and Systems*, vol. 62, no. 1, pp. 1–8, 1994.
- [15] I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin, C. L. Lee, and Y. Pan, "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6108–6117, 2009.
- [16] W. K. Wong, E. Bai, and A. W. C. Chu, "Adaptive time-variant models for fuzzy-time-series forecasting," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 40, no. 6, pp. 1531–1542, 2010.
- [17] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, 2004.
- [18] Y. Tang, Z. Wang, H. Gao, S. Swift, and J. Kurths, "A constrained evolutionary computation method for detecting controlling regions of cortical networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 6, pp. 1569–1581, 2012.
- [19] Y. Tang, Z. Wang, W. K. Wong, J. Kurths, and J.-A. Fang, "Multiobjective synchronization of coupled systems," *Chaos*, vol. 21, no. 2, Article ID 025114, 12 pages, 2011.
- [20] S. M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311–319, 1996.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

