n_oil_production_and_oil_consumption_using_fuzzy_time_series.pdf

by

Forecasting of Malaysian Oil Production and Oil Consumption Using Fuzzy Time Series

Riswan Efendi[1,2](✉) and Mustafa Mat Deris[1]

[1] Faculty of Computer Science, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
riswan.efendi@uin-suska.ac.id, mmustafa@uthm.edu.my
[2] Mathematics Department, Faculty of Science and Technology, State Islamic University of Sultan Syarif Kasim Riau, Panam, 28294 Pekanbaru, Indonesia

Abstract. Many statistical models have been implemented in the energy sectors, especially in the oil production and oil consumption. However, these models required some assumptions regarding the data size and the normality of data set. These assumptions give impact to the forecasting accuracy. In this paper, the fuzzy time series (FTS) model is suggested to solve both problems, with no assumption be considered. The forecasting accuracy is improved through modification of the interval numbers of data set. The yearly oil production and oil consumption of Malaysia from 1965 to 2012 are examined in evaluating the performance of FTS and regression time series (RTS) models, respectively. The result indicates that FTS model is better than RTS model in terms of the forecasting accuracy.

Keywords: Fuzzy time series · Regression time series · Oil production · Oil consumption · Interval number

1 Introduction

The oil, gas and energy are very essential sectors to force the modern economy in the world, especially in Malaysia [1]. The decision makers and researchers should pay attention seriously to enhance the studies in organizing and managing the oil production and oil consumption, respectively. This is due to the fact that, both components are very determinative sectors for the future of this country. Additionally, the forecasting models and the good planning have to be considered in maintaining both components continuously.

Some of the previous studies have been focused on the forecasting of the oil price and the oil consumption by using the specific models. For example, the forecasting of crude oil prices by using econometric model, neural network (ANN), fuzzy regression (FR), support vector machine (SVM), fuzzy neural network (FNN), Markov Chain, and wavelet analysis [2], energy price and consumption for ASEAN countries [3], statewide fuel consumption forecast models [4], short-term energy outlook [5], peak models oil forecast China's supply, demand by Hubbert, generalized Weng, and HCZ models [6]. Recently, the oil production consumption from the middle east countries already

T. Herawan et al. (eds.), *Recent Advances on Soft Computing and Data Mining*,
Advances in Intelligent Systems and Computing 549, DOI 10.1007/978-3-319-51281-5_4

modelled by using Artificial Bee Colony algorithm (ABC-LH) and Levenberg-Marquardt Neural Network (LMNN) [7]. From these studies, there are some perspectives can be identified such as, the model types, the time intervals, the forecasting accuracy, the non-statistical approaches, and the gap between theoretical models and their implementations in real situation.

In the forecasting area, the accuracy of models are still in issue and very important, both components are very challenging task to achieve by the researchers and academicians in the world, especially to forecast the sensitive data and the dynamic variables. Because, not easy to get the historical data accurately and many factors may influence these data. The conditions above occurred in the oil production and oil consumption data set. On the others hand, the prediction for the future of both data should be provided by government seriously. The studies are still going on to find out the systematic procedure and models in the oil, gas and energy sectors as mentioned the second paragraph.

In this paper, we present the univariate-FTS and multiple-RTS models. For the first model, there are some merits such as, no assumptions should be provided regarding data set, no explanatory variables needed, and it can be used to forecast the linguistic values [8]. This model has been frequently implemented to forecast some real data sectors such as, education [9–14], economics [15–18], energy [19–22], others. In this model, the forecasting accuracy is improved by using modification of interval numbers of data set and also out-sample model for linguistic time series. On the other hands, multiple-RTS model is used to evaluate and verify the performance of FTS model by using the historical data of yearly oil production and consumption of Malaysia from 1965 to 2012. From the second model, the causal relationship between oil production, oil consumption and time (year) may be investigated. Through this paper, we want to show how the mathematical and statistical approaches may be implemented in modelling the real time series data.

The rest of paper is organized as follows: In Sect. 2, the theories of FTS and RTS are described. The proposed method in modification of interval number of data set is presented in Sect. 3. In Sect. 4, the empirical analysis of oil production and consumption are discussed. In the end of this paper, the conclusion is mentioned briefly.

2 The Basic Theories of FTS and RTS

2.1 FTS Theory and Its Procedure

Fuzzy time series (FTS) is an implementation of the fuzzy theory to the time series data which the historical data are the linguistic values. From literature, no the conventional time series methods can be used to forecast this data type [8]. There are some definitions related to FTS as follows:

Definition 1. Fuzzy time series [8]
Let $Y(t)$ $(t = 0, 1, 2,\ldots)$, a subset of real numbers, be the universe of discourse on which fuzzy sets $f_i(t)$ $(i = 1, 2,\ldots)$ are defined in the universe of discourse. $Y(t)$ and $F(t)$ is a collection of $f_i(t)$ $(i = 1, 2,\ldots)$. Then $F(t)$ is called a fuzzy time series defined on $Y(t)$ $(t = 0, 1, 2,\ldots)$. Therefore, $F(t)$ can be understood as a linguistic time series variable, where $f_i(t)$ $(i = 1, 2,\ldots)$, are possible linguistics values of $F(t)$.

Definition 2. Fuzzy relations [8]
If there exists a fuzzy relationship R $(t-1, t)$, such that $F(t) = F(t-1) \circ R(t-1, t)$, then $F(t)$ is said to be caused by $F(t-1)$ as denoted as

$$F(t-1) \rightarrow F(t). \quad (1)$$

Definition 3. Fuzzy logical relationship [15]
Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive data (called a FLR), i.e., $F(t)$ and $F(t-1)$, can be denoted as $A_i \rightarrow A_j$, $i, j = 1, 2, \ldots, p$ is called the LHS, and A_j is the RHS of the FLR.

Definition 4. Fuzzy logical group [15]
Let $A_i \rightarrow A_{j1}$, $A_i \rightarrow A_{j2}, \ldots$, $A_i \rightarrow A_{jn}$ are FLRs with the same LHS which can be grouped into an ordered FLG (called a fuzzy logical group) by putting all their RHS together as on the RHS of the FLG. It can be written as:

$$A_i \rightarrow A_{j1,}A_i \rightarrow A_{j2}, \ldots, A_{i,} \rightarrow A_{jn}\ ;\ i, j, \ldots, p = 1, 2, \ldots, n. \quad (2)$$

FTS Procedure
The basic of FTS forecasting algorithm can be derived by using steps as follows [9, 10, 15]:

Step 1: Define the universe of discourse (U) and divide it into several equal length interval.
Step 2: Fuzzify each interval into linguistic time series values (A_i, $i = 1, 2, \ldots, p$, p is the partition number).
Step 3: Establish fuzzy logical relationships (FLRs) among linguistic time series values ($A_i \rightarrow A_j$, $i, j = 1, 2, \ldots, p$).
Step 4: Establish forecasting rule. Actually, there are two rule for forecasting as follows:

Rule 1: IF no relationship occurred among linguistic THEN the final forecast is equal with the midpoint interval value of $A_{i.}$
Rule 2: otherwise, the final forecast is determined by Step 5.

Step 5: Determine the forecast value. Basically, there are three models in determining the final forecast by Song and Chissom [9], Chen [10] and Yu [15] as follows:
Song and Chissom's model:

$$F(t) = F(t-1)\,.\,R(t, t-1), \quad (3)$$

where $F(t)$ is the forecasted data of year t represented by fuzzy sets, $F(t-1)$ is the fuzzified of data year t-1, "." is max-min composition operator, R is union of fuzzy relations.
Chen's model:

$$F(t+1) = \text{Average}\left(m_1, m_2, \ldots, m_p\right), \quad (4)$$

where m_1, $m_2, \ldots$, m_p are the midpoint interval values from fuzzy relationships.
Yu's model:

$$F(t) = \mathbf{M}(t) \times \mathbf{W}(t)^T, \tag{5}$$

where $\mathbf{M}(t)$ is the midpoint matrix ($1 \times n$) and $\mathbf{W}(t)$ is the weight matrix ($n \times 1$).

2.2 RTS Theory and Its Procedure

Let we have time series data available on two variables, say y and x, where y_t and x_t are dated contemporaneously. A static model relating y and x is

$$y_t = a + bx_t + u_t, \ \ t = 1, 2, \ldots, n. \tag{6}$$

Equation (6) is known as a "static model" which comes from the fact that we are modelling a contemporaneous relationship between y and x. Actually, a static model is postulated when a change in x at time t is believed to have an immediate effect on y. Static regression models are also used when we are interested in knowing the tradeoff between y and x. If we want to know the effect of the series of time to y, then it can be written as:

$$y_t = a + bt + u_t, \ \ t = 1, 2, \ldots, n. \tag{7}$$

Equations (6) and (7) may be combined, if any a time trend in a regression model as written as:

$$y_t = a + bx_t + ct + u_t, \ \ t = 1, 2, \ldots, n. \tag{8}$$

Equation (8) is called as a regression time series model with a linear trend (time trend). Three equations above may be applied to forecast the real data [23]. The forecasting algorithm can be calculated by following steps as:

Step 1: Identify the visual trend and relationship of x, y to t by scatter plot.
Step 2: Check the correlations among x, y and t.
Step 3: Estimate parameters a, b and c by using ordinary least square method (OLS).
Step 4: Check the validity of parameters a, b and c by t-test and f-test.

3 Proposed Interval Adjustment and Algorithm in FTS

In FTS forecasting, the interval length and partition number of data are very important to be considered, because their contribution in reducing the forecasting error is very significant. In previous studies, many rules and approaches have been discussed

[9, 10, 15], but they are still not standard to be followed. In 2013, Ismail *et al.* [18] proposed a new approach by using inter-quartile range approach, this approach more compatible if we compared with the existing approaches, however this approach has a problem when the ranges between quartiles are too long. In this section, we suggest to modify it as follows:

Let X_t (t = 1, 2, 3,…, n) be a time series data and ($X_{\min,}$ $X_{\max}$) element of X_t. The effective interval length and partition number can be determined as:

Step 1: Determine the quartiles Q_1 Q_2 and Q_3 from data set, X_t.
Step 2: Divide data set into four-intervals and determine their frequency respectively.
Step 3: Re-divide each interval from Table 1 by using each frequency respectively.

Thus, the total of interval number is equal with total of frequency also. Mathematically, the comparison of our proposed interval (b) with Ismail's *et al.* (a) can be written as:

$$\text{Total of interval number}_{(\mathrm{a})} < \text{Total of interval number}_{(\mathrm{b})}. \tag{9}$$

Thus,

$$l_a > l_b, \tag{10}$$

$$R_a/k_a > R_b/k_b. \tag{11}$$

By taking $a = b$, thus

$$R_a/(1 + 3.3\ \log(f_a)) > R_a/f_a, \tag{12}$$

$$1/(1 + 3.3\ \log(f_a)) > 1/f_a, \tag{13}$$

which l_a, k_a, R_a, F_a are interval length, number of interval, range, and frequency data from Ismail's *et al.* approach, otherwise from our proposed approach. Equation (13) shows that the interval length from proposed approach is smaller than Ismail's approach. The final forecast can be derived by using an algorithm as follows:

Step 1: Follow Steps 1–3 as mentioned in the first paragraph of Sect. 3.
Step 2: Transform the actual data to be linguistic time series values.
Step 3: Forecast the series of linguistic by using statistical time series methods.

Table 1. Quartile, range, and frequency of data set.

Interval	Range between quartiles	Frequency
$[X_{\min}, Q_1]$	R_1	f_1
$[Q_2, Q_2]$	R_2	f_2
$[Q_2, Q_3]$	R_3	f_3
$[Q_3, X_{\max}]$	R_4	f_4

Step 4: Forecast the numerical time series by using midpoint interval values.
Step 5: Verify the forecasting accuracy by using mean square error (MSE) and compare with regression time series model.

$$MSE = \sum (\text{actual}_t - \text{forecast}_t)^2 / n, \tag{14}$$

where actual_t is actual data at time t, forecast_t is a forecast value at time t, and n is a number of observation/testing data.

4 Empirical Analysis

In this section, the FTS and RTS models are implemented to forecast the yearly oil production and oil consumption of Malaysia, the period 1965 to 2012 which are used as model building. By using the proposed algorithm given in Sect. 3, the forecast values can be calculated as follows:

Step 1: Calculate the quartiles of oil production and oil consumption data sets (Table 2).
Step 2: Divide each data set into four intervals, and determine the ranges and frequencies (Table 3).
Step 3: Re-divide into sub-interval and determine the midpoint interval values (Table 4).
Step 4: Transform the actual data to be linguistic time series values (Table 5).
Step 5: Forecast the series of each linguistic by using statistical models. In this paper, regression time series model is considered, because there is strong linear correlation between oil production, oil consumption and year, respectively.

Table 6 shows the forecasted indexes of oil production and oil consumption are obtained using regression time series model as follows:

$$\text{Index OP} = 0.002 + 0.994(\text{Year}). \tag{15}$$

$$\text{Index OC} = -0.602 + 1.059(\text{Year}). \tag{16}$$

Step 6: Forecast the numerical time series value for each linguistic time series by using the midpoint interval values (Table 7).
Step 7: Verify and compare the forecasting results from FTS and RTS models.

Table 2. The quartiles of oil production and consumption

Data	Q_1	Q_2	Q_3
Oil production (barrels)	171	564.5	703.75
Oil consumption (barrels)	99.25	214.50	519.50

Table 3. The intervals, ranges, and frequencies of oil production and consumption

Oil Production		
Interval	Range between quartiles	Frequency
[1.00, 171.5]	170	12
[171.5, 564.5]	393.5	12
[564.5, 703.7]	139.5	12
[703.7, 776.0]	72.25	12
Oil Consumption		
[46, 99.25]	52.25	12
[99.25, 214.50]	115.25	12
[214.50, 519.50]	61	12
[519.50, 718.00]	39.7	12

Table 4. The sub-interval and the linguistic time series values

Oil Production		
Sub-interval	Midpoint interval value	Linguistic value
[1, 14.16]	7.58	A_1
[14.16, 28.32]	21.24	A_2
[28.32, 42.68]	35.40	A_3
…	…	…
[763.90, 769.90]	766.90	A_{47}
[769.90, 776.00]	772.90	A_{48}
Oil Consumption		
[46, 50.35]	48.175	A_1
[50.35, 54.70]	52.52	A_2
[54.70, 59.05]	56.875	A_3
…	…	…
[701.44, 718.00]	709.72	A_{48}

Table 5. The transformation actual data to be linguistic time series

Year	Actual oil production	Linguistic time series	Year	Actual oil consumption	Linguistic time series
1965	1	A_1	1965	46	A_1
1966	1	A_1	1966	54	A_2
1967	1	A_1	1967	54	A_2
…	…	…	…	…	…
2012	670	A_{34}	2012	283	A_{48}

Table 6. The actual and forecasted indexes of linguistic time series value

Year	Oil production		Oil consumption	
	Actual index	Forecasted index	Actual index	Forecasted index
1965	1	1	1	1
1966	1	2	2	2
1967	1	3	2	3
…	…	…	…	…
2012	34	48	48	49

Table 7. The forecasted real time series for each linguistic

Year	Oil production		Oil consumption	
	Forecasted index	Forecasted real time series	Forecasted index	Forecasted real time series
1965	A_1	8	A_1	48
1966	A_2	21	A_2	53
1967	A_3	35	A_3	57
…	…	…	…	…
2012	A_{48}	773	A_{49}	79

Table 8 shows three different models are used to forecast the yearly oil production and consumption which derived by regression time series model. For the actual data, the oil consumption and oil production models can be written as:

$$(\text{Oil Consumption})_t = 21.962(\text{Year}) - 0.30(\text{Oil Production})_t - 43213.97. \tag{17}$$

$$(\text{Oil Production})_t = 33.299(\text{Year}) - 0.891(\text{Oil Consumption})_t - 65478.573. \tag{18}$$

Table 8. The forecasting results using RTS and FTS models

Year	Oil production			
	Actual	RTS	FTS-1	FTS-2
1965	1	9	0	8
1966	1	28	18	21
…	…	…	…	…
2012	670	897	754	773
Year	Oil consumption			
	Actual	RTS	FTS-1	FTS-2
1965	46	-73	0	48
1966	54	-53	51	53
…	…	…	….	….
2012	712	755	738	721

Table 9. The comparison of forecasting accuracy from three different models

MSE (mean square error)					
	RTS-(OC, T)	RTS-(OP, T)	RTS-(T)	FTS-1	FTS-2
Oil Consumption		34014.199	6855.874	521.5045	232.6588
Oil Production	6295.419		8666.020	2328.861	1881.947
Rank	3	6	4, 5	2	1

In Eqs. (16) and (17), the interval numbers is 48. Additionally, the comparison between these models in term of forecasting accuracy can be shown in Table 9.

Table 9 indicates that MSE from FTS-1 and FTS-2 are smaller than RTS models, but FTS-2 has smaller MSE if compared with FTS-1.

5 Conclusion

In achieving the higher forecasting accuracy, we adjusted the interval length based on Ismail's *et al.* approach. The performance of adjusting interval length can be shown through Eq. (13) and the comparison errors with regression time series and existing FTS models. From MSE of data training and testing, the contribution of proposed interval length is very significant in reducing of the forecasting error. Therefore, the optimization of the partition number is still in issue to be considered in FTS forecasting procedure.

Acknowledgment. The authors are grateful to Research and Innovation Fund, UTHM for their financial support.

References

1. Economic Transformation Programme Chap. 1. http://www.etp.pemandu.gov.my
2. Gabralla, L.A., Abraham, A.: Computational modeling of crude oil price forecasting: a review of two decades of research. Int. J. Comput. Inf. Syst. Ind. Manag. Appl. **5**, 729–740 (2013)
3. Kimura, S.: The 2nd ASEAN Energy Outlook. The Energy Data and Modeling Centre, Japan (2009)
4. Washington State Department of Transportation – Economic Analysis: Statewide Fuel Consumption Forecast Models (2010)
5. US Energy Information Administration: Short-Term Energy Outlook. Independent Statistics & Analysis (2015). Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid information services for distributed resource sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181–184. IEEE Press, New York (2001)
6. Liangyong, F., Junchen, L., Xiongqi, P., Xu, T., Lin, Z.: Peak oil models forecast China's oil supply, demand. Oil Gas J. 43–47 (2008)

7. Chiroma, H., et al.: An intelligent modeling of oil consumption. In: El-Alfy, El-Sayed M., Thampi, S.M., Takagi, H., Piramuthu, S., Hanne, T. (eds.). AISC, vol. 320, pp. 557–568 Springer, Cham (2015). doi:10.1007/978-3-319-11218-3_50
8. Song, Q., Chissom, B.S.: Fuzzy time series and its models. Fuzzy Sets Syst. **54**, 269–277 (1993)
9. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series – Part 1. Fuzzy Sets Syst. **54**, 1–9 (1993)
10. Chen, S.M.: Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst. **81**, 311–319 (1996)
11. Singh, S.R.: A robust method for forecasting based on fuzzy time series. Int. J. Comput. Math. **188**, 472–484 (2007)
12. Kuo, I.: An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. Expert Syst. Appl. **36**, 6108–6117 (2009)
13. Ismail, Z., Efendi, R.: Enrollment forecasting based on modified weight fuzzy time series. J. Artif. Intell. **4**, 110–118 (2011)
14. Ismal, Z., Efendi, R., Deris, M.M.: Inter-quartile range approach to length – interval adjustment of enrollment data. Int. J. Comput. Intell. Appl. **3**, 10 p. (2013)
15. Yu, H.K.: Weighted fuzzy time series models for TAIEX forecasting. Phys. A **349**, 609–624 (2005)
16. Yu, H.K., Huarng, K.H.: A bivariate fuzzy time series model to forecast the TAIEX. Expert Syst. Appl. **34**, 2945–2952 (2008)
17. Lee, H.L., Liu, A., Chen, W.S.: Pattern discovery of fuzzy time series for financial prediction. IEEE Trans. Syst. Man Cybern. Part B **18**, 613–625 (2006)
18. Efendi, R., Ismail, Z., Deris, M.M.: Improved weight fuzzy time series used in the exchange rates forecasting US Dollar to Ringgit Malaysia. Int. J. Comput. Intell. Appl. **12**, 19 p. (2013)
19. Bolturuk, E., Oztayzi, B, Sari, I.U.: Electricity consumption forecasting using fuzzy time series. In: IEEE 13th International Symposium on Computer Intelligence and Informatics, 20–22 November 2012, Istanbul, Turkey, pp. 245–249
20. Alpaslan, F., Cagcag, O.: Seasonal fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering. J. Soc. Econ. Stat. **2**, 1–13 (2012)
21. Azadeh, A., Saberi, M., Gitiforouz, A.: An integrated simulation-based fuzzy regression-time series algorithm for electricity consumption estimation with non-stationary data. J. Chin. Inst. Eng. **34**, 1047–1066 (2012)
22. Efendi, R., Ismail, Z., Deris, M.M.: New linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. Appl. Soft Comput. **28**, 422–430 (2015)
23. Wooldridge, J.M.: Introductory Econometrics A Modern Approach, 3rd edn. Thomson South Western, Mason (2006)

n_oil_production_and_oil_consumption_using_fuzzy_time_se...

ORIGINALITY REPORT

19%	%	19%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Computational Intelligence in Information Systems", Springer Nature, 2017 Publication	5%
2	Efendi, Riswan, Zuhaimy Ismail, and Mustafa Mat Deris. "A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand", Applied Soft Computing, 2015. Publication	5%
3	Riswan Efendi, Mustafa Mat Deris, Zuhaimy Ismail. "Implementation of Fuzzy Time Series in Forecasting of the Non-Stationary Data", International Journal of Computational Intelligence and Applications, 2016 Publication	3%
4	Studies in Fuzziness and Soft Computing, 2016. Publication	2%
5	Riswan Efendi, Mustafa Mat Deris. "Non-Probabilistic Inverse Fuzzy Model in Time Series Forecasting", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2018 Publication	2%
6	"Recent Advances on Soft Computing and Data Mining", Springer Science and Business Media LLC, 2018 Publication	2%

7	Riswan Efendi, Zuhaimy Ismail, Nor Haniza Sarmin, Mustafa Mat Deris. "A reversal model of fuzzy time series in regional load forecasting", International Journal of Energy and Statistics, 2015 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		