Oil_Production_and_Consumtion_Using_Fuzzy_Time_Series_Model.pdf

by

Submission date: 14-Apr-2023 10:33AM (UTC+0700)
Submission ID: 2064085717
File name: Oil_Production_and_Consumtion_Using_Fuzzy_Time_Series_Model.pdf (815.81K)
Word count: 6139
Character count: 30263

Advances in Data Science and Adaptive Analysis
Vol. 9, No. 1 (2017) 1750001 (17 pages)
© World Scientific Publishing Company
DOI: 10.1142/S2424922X17500012

World Scientific
www.worldscientific.com

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Prediction of Malaysian–Indonesian Oil Production and Consumption Using Fuzzy Time Series Model

Riswan Efendi[*,†,‡] and Mustafa Mat Deris[*,§]

[*]*Faculty of Computer Science and Information Technology*
Universiti Tun Hussein Onn Malaysia
86400 Batu Pahat, Johor, Malaysia

[†]*Mathematics Department, Faculty of Science and Technology*
State Islamic University of Sultan Syarif Kasim Riau
28294 Panam, Pekanbaru, Indonesia
[‡]*riswan@uthm.edu.my*
[§]*mmustafa@uthm.edu.my*

Received 13 October 2016
Revised 20 November 2016
Accepted 4 December 2016
Published 13 April 2017

Fuzzy time series has been implemented for data prediction in the various sectors, such as education, finance-economic, energy, traffic accident, others. Moreover, many proposed models have been presented to improve the forecasting accuracy. However, the interval-length adjustment and the out-sample forecast procedure are still issues in fuzzy time series forecasting, where both issues are yet clearly investigated in the previous studies. In this paper, a new adjustment of the interval-length and the partition number of the data set is proposed. Additionally, the determining of the out-sample forecast is also discussed. The yearly oil production (OP) and oil consumption (OC) of Malaysia and Indonesia from 1965 to 2012 are examined to evaluate the performance of fuzzy time series and the probabilistic time series models. The result indicates that the fuzzy time series model is better than the probabilistic models, such as regression time series, exponential smoothing in terms of the forecasting accuracy. This paper thus highlights the effect of the proposed interval length in reducing the forecasting error significantly, as well as the main differences between the fuzzy and probabilistic time series models.

Keywords: Fuzzy time series; index of linguistic; oil production–consumption; interval-length; forecasting accuracy.

1. Introduction

Oil, gas, and energy sectors are essential to the modern economy throughout the world, especially in Malaysia [Economic Transformation Programme Chapter 1, http://www.etp.pemandu.gov.my.]. Synergistic efforts between the decision-makers and industrial researchers to enhance research in oil production (OP) and oil

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

consumption (OC) will not only improve the efficiency in both sides, but may also determine the future of this country. Additionally, the development of accurate forecasting models and strategic planning are crucial in maintaining both aspects continuously.

Some of the previous studies have focused on the forecasting of oil price and OC using specific models such as the econometric model, artificial neural network (ANN), fuzzy regression (FR), support vector machine (SVM), fuzzy neural network (FNN), Markov Chain, wavelet analysis [Gabralla and Abraham (2013)], forecasting models for energy price and consumption in ASEAN countries [Kimura (2009)], statewide fuel consumption forecast models [Washington State Department of Transportation (2010)], short-term energy outlook [U.S. Energy Information Administration (2015); Czajkowski *et al.* (2001)], peak load forecasting models for oil supply in China, Hubbert's demand model, generalized Weng and HCZ models [Liangyong *et al.* (2008)]. More recently, the oil production–consumption in the Middle East countries was modeled using Artificial Bee Colony algorithm (ABC-LH) and Levenberg–Marquardt Neural Network (LMNN) [Chiroma *et al.* (2015)]. From these studies, several components can be observed, namely the model type, time interval, forecasting accuracy, nonstatistical approach, and gap between theoretical models and their implementations in a real situation.

In the field of forecasting, researchers and academics face challenges in improving the accuracy of the forecasting models, especially when it involves sensitive data and dynamic variables. This is because of the difficulty of obtaining accurate historical data in addition to the influence of many other factors on the data; these conditions commonly occur in the OP and OC data sets. Studies on the systematic procedures and prediction models in oil, gas, and energy sectors such as the ones mentioned in the previous paragraph must be pursued by researchers. The government's role in providing resources for data prediction is also essential to ensure the future of these two sectors.

In this paper, we present the univariate-FTS and the probabilistic-time series models. The advantage of the first model, besides not needing assumptions and explanatory variables for the data set, is that it can be used to forecast linguistic values [Song and Chissom (1993a)]. The univariate-FTS model has been frequently implemented to forecast real data in sectors such as education [Song and Chissom (1993b); Chen (1996); Singh (2007); Kuo (2009); Ismail and Efendi (2011)], economy [Yu (2005); Yu and Huarng (2008); Lee *et al.* (2006); Efendi *et al.* (2013); Sun *et al.* (2015); Gholizade and Chafi (2015)], and energy [Bolturuk *et al.* (2012); Alpaslan and Cagcag (2012); Azadeh *et al.* (2012); Efendi *et al.* (2015); Ismail *et al.* (2015); Efendi and Deris (2017)]. The forecasting accuracy of this model is improved by modifying the interval numbers of the data set and using out-sample model for linguistic time series. Meanwhile, the multiple-RTS model has been used to evaluate and verify the performance of the FTS model by using the historical data of Malaysia's yearly oil production and consumption from 1965 to 2012. Through this model, the causal relationship between OP, OC, and time (year) may

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

be investigated. This paper shows how the mathematical and statistical approaches may be implemented in modeling the real time series data.

The next section describes the theories of FTS and probabilistic-TS. The proposed method for the modification of the interval number of the data set is presented, followed by the discussion on the empirical analysis of oil production and consumption.

2. Fundamental Theories of Fuzzy and Regression Time Series

2.1. *Fundamental fuzzy time series and its procedure*

Fuzzy time series (FTS) is the implementation of the fuzzy theory on the time series data in which the historical data are linguistic values. From the literature, there are no conventional time series methods that can be used to forecast this type of data [Song and Chissom (1993a)]. The following are some recurring terms relevant to FTS.

Definition 1 Fuzzy time series [Song and Chissom (1993a)]. Let $Y(t)$ $(t = 0, 1, 2, \ldots)$, a subset of real numbers, be the universe of discourse in which fuzzy sets $f_i(t)$ $(i = 1, 2, \ldots)$ are defined in the universe of discourse. $Y(t)$ and $F(t)$ are 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 linguistics time series variable, where $f_i(t)$ $(i = 1, 2, \ldots)$ are possible linguistics values of $F(t)$.

Definition 2 Fuzzy relations [Song and Chissom (1993a)]. 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 $F(t-1)$, denoted as

$$F(t-1) \rightarrow F(t). \tag{1}$$

Definition 3 Fuzzy logical relationship (FLR) [Yu and Huarng (2008)]. Let $F(t-1) = A_i$ and $F(t) = A_j$. The FLR, which is the relationship between two consecutive data, i.e. $F(t)$ and $F(t-1)$, can be denoted as $A_i \rightarrow A_j$, $i, j = 1, 2, \ldots, p$, where A_i is the left-hand side (LHS) and A_j is the right-hand side (RHS) of the FLR.

Definition 4 Fuzzy logical group (FLG) [Yu (2005)]. Let $A_i \rightarrow A_{j1}$, $A_i \rightarrow A_{j2}, \ldots, A_i \rightarrow A_{jn}$ be FLRs with the same LHS which can be grouped into an ordered FLG by putting all their RHS together as the RHS of the FLG. This can be written as

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

The basics of FTS forecasting can be derived from the following steps [Song and Chissom (1993b); Chen (1996); Yu (2005)]:

Step 1: Define the universe of discourse, U, and divide it into several intervals of equal length.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

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 rules. There are generally two rules for forecasting:

Rule 1: IF no relationship occurred among linguistic values, THEN the final forecast is equal to the midpoint value of interval A_i.

Rule 2: Otherwise, the final forecast is determined by Step 5.

Step 5: Determine the forecast value. There are three models in determining the final forecast, respectively, developed by Song and Chissom [1993b], Chen [1996] and Yu [2005].

In Song and Chissom's model,

$$F(t) = F(t-1) \cdot R(t, t-1), \tag{3}$$

where $F(t)$ is the forecasted data of year t represented by fuzzy sets, $F(t-1)$ is the fuzzified data year $t-1$, "$\cdot$" is the max–min composition operator, and R is the union of the fuzzy relations.

In Chen's model,

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

where m_1, $m_2, \ldots, m_p$ are the midpoint values of the interval from the fuzzy, relationships.

In 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 $(n \times 1)$. However, the flaw of these models is that they are not able to forecast the series of each linguistic values. Therefore, the final forecasted values could not be obtained clearly and logically.

2.2. *Regression time series model and its procedure*

In this section, regression time series (RTS) model was chosen as one of the probabilistic time series (PTS) group models. Let us say we have a time series data of two variables, y and x, where y_t and x_t are dated contemporaneously. A static model relating y to x can be written as

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

Equation (6) is known as a "static model" which comes from the fact that a contemporaneous relationship between y and x is being modeled. 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 one is interested in knowing the

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

trade-off between y and x. To find out the effect of the series of time to y, the relationship can be written as

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

Equations (6) and (7) may be combined if any time trends exist in a regression model and are written as

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

Equation (8) is called a regression time series model with a linear trend (time trend). The three equations above may be applied to forecast real data. The forecasting algorithm can be calculated using the following steps [Wooldridge (2006)]:

Step 1: Identify the visual trend and relationship of (x, y) to t through a 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 using t-test and f-test.

3. Proposed Interval Length and Forecasting Algorithm

In FTS forecasting, the interval length and partition number of data are very important because their contribution in reducing the forecasting error is significant. Although many rules and approaches have been applied in previous studies [Song and Chissom (1993b); Chen (1996); Yu (2005)], the findings have yet to produce a guiding standard that can be followed. In 2013, Ismail *et al.* [2013] proposed the inter-quartile range approach that is more compatible than the existing approaches. However, this approach is not suitable when the range between quartiles is too big. In this section, the modifying of the interval length is suggested as follows.

Let X_t $(t = 1, 2, 3, \ldots, n)$ be a time series data and the range $(X_{\min}, X_{\max})$ be the elements of X_t. The effective interval length and partition number can be determined by following this procedure:

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

Thus, the total of interval numbers is equal to the total of frequency as well. On the other hand, Ismail *et al.* [2013] proposed that the sub-interval number is equal to the range at interval i divided by $1+3.3 \log(f_i)$, $(i = 1, 2, 3, 4)$. The modification of our proposed interval can be mathematically compared with the interval proposed in Efendi *et al.* [2013] as shown below.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

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

Given,

$$\text{Total of interval number}_{(\text{Ismail } et\ al.)} < \text{Total of interval number}_{(\text{proposed model})}, \tag{9}$$

thus, the length of interval from both models is

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

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

By assuming $a = b$, therefore

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

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

because

$$(1 + 3.3\log(f_a)) < f_a, \tag{14}$$

in which l_a, k_a, R_a, and f_a are interval length, number of interval, range, and frequency of data, respectively, from the former approach by Ismail *et al.*, while the other parameters are from our proposed interval. Equation (13) shows that the proposed interval length is smaller than the interval proposed by Ismail *et al.* [2013]. The rest of the steps in determining the final forecast can be shown systematically:

Step 4: Transform the actual data into linguistic time series values.
Step 5: Forecast the series of linguistic values by using statistical time series models (forecasting type-1) as shown in Fig. 1 [Efendi *et al.* (2015)].
Step 6: Forecast the numerical time series by using midpoint values of the interval (forecasting type-2) as shown in Fig. 2.
Step 7: Measure the forecasting accuracy of FTS and RTS models by looking at the mean square error (MSE).

The computation steps of the proposed forecasting algorithm are summarized in the form of a flow diagram in Fig. 3.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

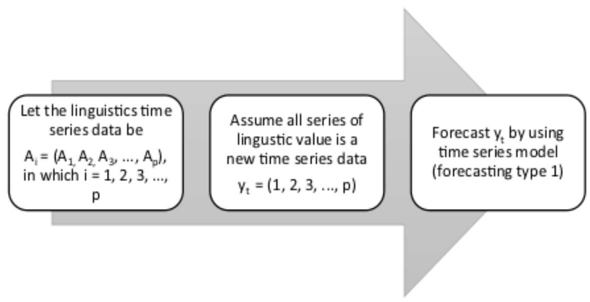

Fig. 1. Forecasting type-1.

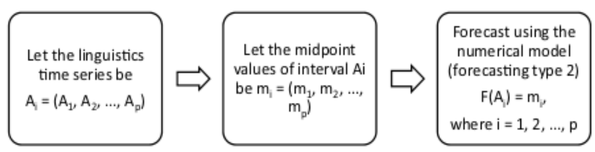

Fig. 2. Forecasting type-2.

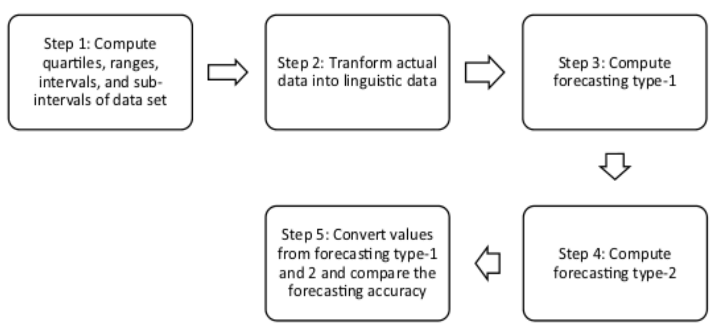

Fig. 3. Forecasting steps based on the proposed algorithm.

4. Empirical Analysis

In this section, the FTS and RTS models implemented to forecast the yearly OP and OC of Malaysia and Indonesia data from 1965 to 2012 are used to examine both the models separately. Using the proposed algorithm delineated in Sec. 3 and Malaysia's yearly oil data, the computation of the forecasted values is presented as follows:

Step 1: The quartiles of the OP and OC data sets were calculated in Table 2.
Step 2: Each data set was divided into four intervals, and the ranges and frequencies were determined in Table 3.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Table 2. Quartiles of oil production and consumption.

Data	Q_1	Q_2	Q_3
OP (barrel)	171	564.5	703.75
OC (barrel)	99.25	214.50	519.50

Table 3. Interval, range, and frequency of oil production and consumption.

Interval	Range between quartiles	Frequency
	OP	
[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
	OC	
[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

Step 3: The intervals were re-divided into sub-intervals and the midpoint values of the intervals were determined in Table 4.

Step 4: The actual data were transformed into linguistic time series values in Table 5.

Step 5: The index series of each linguistic value can be forecasted by using the RTS model, as presented in Table 7. While, the descriptive statistics between the actual OP, OC and their linguistic indexes are presented in Table 6.

With reference to Table 6, the ranges of the actual OC–OP are very huge when compared to the ranges of OC–OP indexes, observed by the large difference between the maximum and minimum data. At the same time, the variance of the actual data

Table 4. Sub-intervals and their linguistic time series values.

Sub-interval	Midpoint value of interval	Linguistic value
	OP	
[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}
	OC	
[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}

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Table 5. Transformation of actual data into linguistic time series.

Year	Actual OP	Linguistic time series	Year	Actual OC	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
1968	4	A_1	1968	54	A_2
1969	9	A_1	1969	56	A_3
1970	18	A_2	1970	63	A_4
1971	69	A_5	1971	68	A_6
1972	93	A_7	1972	76	A_7
—	—	—	—	—	—
2012	670	A_{34}	2012	283	A_{48}

Table 6. Descriptive statistics.

	N	Range	Minimum	Maximum	Mean	Std. deviation	Variance
OP-MY (actual)	48	775	1	776	452.04	280.757	78824.594
OC-MY (actual)	48	672	46	718	319.21	233.011	54293.913
Index OC	48	47	1	48	25.35	14.902	222.063
Index OC 20	48	19	1	20	10.79	6.202	38.466
Index OP 20	48	19	1	20	10.40	6.066	36.797
Index OP 48	48	47	1	48	24.33	14.781	218.482
Valid N (listwise)	48						

(real data) is also higher than that of the index data (linguistics data). The index data is a key point to consider in forecasting type-1. Furthermore, by using the index data, the forecasting error may be reduced significantly because the digit number of the data will become smaller when they are transformed into linguistic values (index data type).

Table 7 shows that the forecasted index of each linguistic from OP and OC is derived by RTS models. Moreover, the causal relationships between index-OP and

Table 7. Actual and forecasted indexes of linguistic time series values.

	OP		OC	
Year	Actual index	Forecasted index	Actual index	Forecasted index
1965	1	1	1	1
1966	1	2	2	2
1967	1	3	2	3
1968	1	4	2	2
1969	1	5	3	3
1970	2	6	4	4
1971	5	7	6	6
1972	7	8	7	8
—	—	—	—	—
2012	34	48	48	49

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Table 8. Forecasted real time series for each linguistic value.

	OP		OC	
Year	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
1968	A_4	50	A_2	53
1969	A_5	64	A_3	57
1970	A_6	78	A_4	61
1971	A_7	92	A_6	61
1972	A_8	106	A_8	69
—	—	—	—	—
2012	A_{48}	773	A_{49}	79

year (time), and also between index-OC and year (time) can be written mathematically as

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

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

Step 6: The numerical time series value (forecasting type-2) for each linguistic time series is forecast using each midpoint interval as presented in Table 8. While all forecasted results obtained by the FTS and RTS models can be presented in Table 9.

Table 8 shows the transformation process from the forecasted index (forecasting type-1) into the forecasted real value (forecasting type-2) for OC and OP, respectively. For example, if the forecasted index is A_8, then the forecasted numerical value is 108 (in barrels). From this procedure, we can describe the clear way in determining the forecasted linguistic index and numerical value. Both forecasting types are significantly different from the previous models in FTS forecasting. Thus, forecasting type-1 is our main contribution to improve the performance of FTS model as compared to the existing studies.

Table 9 shows three different models used to forecast the yearly oil production and consumption. For the RTS model, the correlation coefficients between OP, OC, and year (time) indicate strong linear relationships among these variables. Therefore, the regression time series model for both data sets can be written as

$$(\text{OC})_{\text{Year}} = 21.962\ \text{Year} - 0.307(\text{OP})_{\text{Year}} - 43213.97, \quad (17)$$

$$(\text{OP})_{\text{Year}} = 33.299\ \text{Year} - 0.891(\text{OC})_{\text{Year}} - 65478.573. \quad (18)$$

In Eqs. (17) and (18), the number of intervals used was 48 while the forecasting results from FTS-1 and FTS-2 were derived using the model proposed by Efendi *et al.* [2015].

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Table 9. Forecasted results using RTS and FTS models.

Year	Actual	RTS	FTS-1	FTS-2
		OP		
1965	1	9	0	8
1966	1	28	18	21
1967	1	47	18	35
1968	4	66	51	50
1969	9	84	51	64
1970	18	103	51	78
—	—	—	—	—
2012	670	897	754	773
		OC		
Year	Actual	RTS	FTS-1	FTS-2
1965	46	−73	0	48
1966	54	−53	51	53
1967	54	−31	51	57
1968	54	−9	62	53
1969	56	12	62	57
1970	63	32	73	61
—	—	—	—	—
2012	712	755	738	721

Table 10. MSE comparison between proposed and the existing models (In-sample forecast).

	MSE	
Model	Malaysia's OP	Indonesia's OP
RTS-1 (OC, T)	6295.419	53019.63
RTS-2 (T)	8666.020	—
ARIMA	1235.066	1788.996
FTS-1 [Yu (2005)]	4556.098	6450.987
FTS-2 [Cheng *et al.* (2008)]	3456.865	3214.779
FTS-3 [Ismail *et al.* (2013)]	2328.861	560.187
FTS-4 (Proposed model)	1881.947**	109.466**
	MSE	
Model	Malaysia's OC	Indonesia's consumption
RTS-1 (OP, T)	34014.199	3091.468
RTS-2 (T)	6855.874	—
DES-TS		5181.182
ARIMA	1558.987	2998.765
FTS-1 [Yu (2005)]	2887.898	4578.786
FTS-2 [Cheng *et al.* (2008)]	1775.633	3788.665
FTS-3 [Ismail *et al.* (2013)	521.504	9496.788
FTS-4 (Proposed model)	232.658**	2801.686**

Note: **: smallest MSE.

Table 11. MSE comparison between proposed and the existing models (Out-sample forecast).

Model	MSE Malaysia's OP	MSE Indonesia's OP
RTS-1 (OC, T)	6467.323	69117.67
RTS-2 (T)	9200.112	—
ARIMA	1834.122	2388.221
FTS-1 [Yu (2005)]	5556.698	6959.587
FTS-2 [Cheng *et al.* (2008)]	4456.653	4214.865
FTS-3 [Ismail *et al.* (2013)]	2922.786	765.989
FTS-4 (Proposed model)	2478.543**	259.678**

Model	MSE Malaysia's OC	MSE Indonesia's consumption
RTS-1 (OP, T)	34014.199	4145.661
RTS-2 (T)	6855.874	—
DES-TS		5181.182
ARIMA	1728.113	3003.225
FTS-1 [Yu (2005)]	2997.199	4600.110
FTS-2 [Cheng *et al.* (2008)]	1909.201	3811.667
FTS-3 [Ismail *et al.* (2013)]	777.234	9811.667
FTS-4 (Proposed model)	411.334**	3001.145**

Note: **: smallest MSE.

Step 7: The accuracy of the forecasting performance of FTS and RTS models is evaluated by using training (in-sample forecast) and testing (out-sample forecast), respectively. The comparison of MSE between proposed and the existing models is shown in Table 10.

Tables 10 and 11 indicate that the MSE (in-out-sample forecast) of Malaysian–Indonesian OP from FTS-4 model are smaller than the MSE which derived by others models. Both MSEs show the effect of different interval numbers is very significant

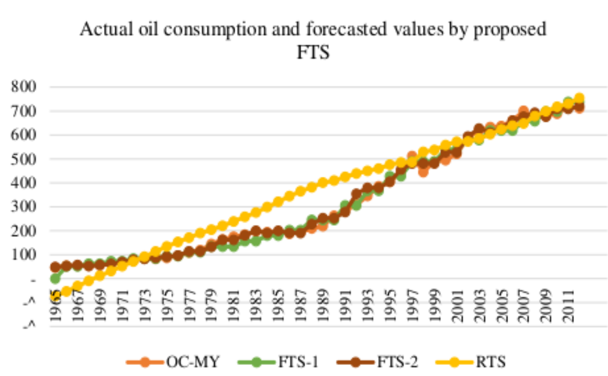

Fig. 4. OC forecast and actual data using FTS and RTS models.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

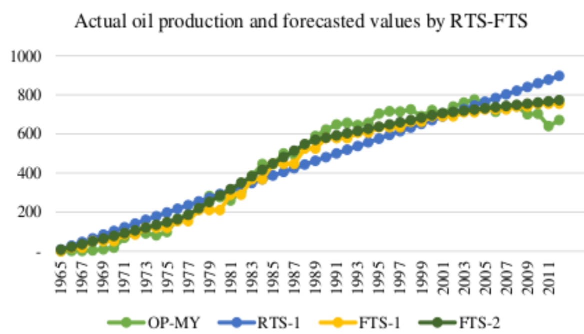

Fig. 5. OP forecast and actual data using FTS and RTS models.

in increasing the forecasting accuracy. The performance of FTS and RTS models in forecasting the OP and OC (Malaysia's oil data) is shown graphically in Figs. 4 and 5.

Figure 4 illustrates that the actual OC does not always increase linearly with year, although both variables have strong linear positive relationship. The yellow plot is the forecasted values obtained by the RTS model. From 1976 to 1996, the actual OC was not precisely predicted by this model. However, a more accurate estimation was obtained by the FTS model, especially FTS-2 with 48 intervals.

In Fig. 5, the estimated values of OP are more precise using FTS as compared to the RTS models, especially FTS-2 model with 48 intervals. Many actual data of

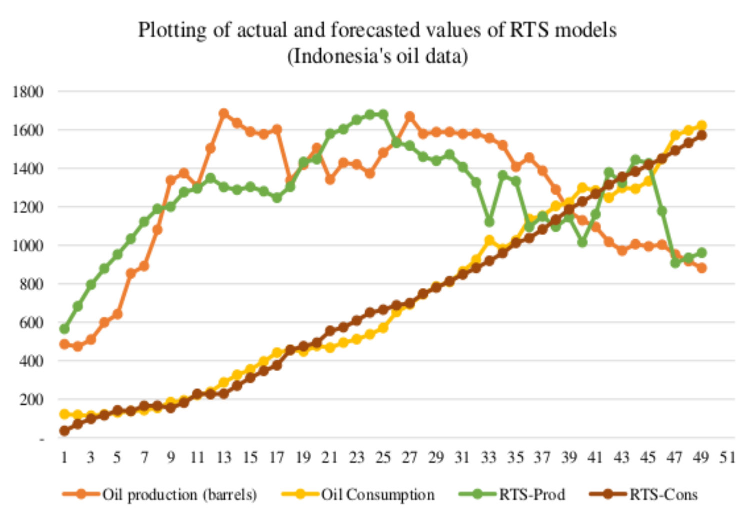

Fig. 6. Actual and forecasted values of Indonesia's oil production and consumption using RTS model.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com
by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

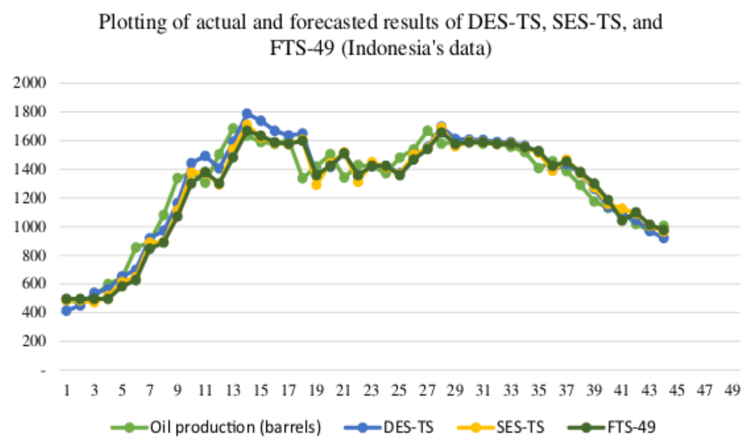

Fig. 7. Indonesia's yearly OP using DES-TS, SES-TS, and FTS models.

OP could not be accurately forecasted by the RTS model especially from year 1985 to 1998.

Next, the plotting of the actual and forecasted results of Indonesia's OP and OC is illustrated in Figs. 6 and 7.

Table 12. Characteristic of FTS and RTS models in forecasting.

Component	FTS forecasting model	RTS forecasting model
Data type	Real data and linguistic data are needed but without any assumption, e.g. number of data can be less than 30.	Real data will full assumption are needed.
Model type	Linguistic index: Probabilistic model Numerical value: Deterministic model.	Probabilistic model.
Forecasting algorithm	Straightforward steps. Simple in computation because index data uses digit numbers.	Some steps need looping. The computation steps are complex when the digit of the data is very big.
Ability	Suitable for predicting fluctuating data with an upward or downward trend (see Fig. 7).	Suitable for predicting trend data with no fluctuation (for upward or downward trend).
Accuracy	Depends on the partition number and interval length. More accurate when data fluctuates.	Depends on the number of data and the correlation between the explanatory and response variables. Not accurate when data fluctuates.
Process	(1) To forecast a series of linguistic values using statistical models. (2) To forecast numerical values of linguistic using the midpoint of intervals.	To forecast numerical values using the model obtained from historical data.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Figure 6 shows the performance of the RTS model in predicting both Indonesia's yearly OP and OC. For the OP graph, it was difficult to estimate the actual data using this model because of the fluctuating data. Nevertheless, the same model was able to forecast the yearly OC more precisely.

Figure 7 illustrates the performance of three different models, namely double exponential smoothing time series (DES-TS), seasonal exponential smoothing time series (SES-TS), and FTS models, in predicting Indonesia's yearly OP data. These models are generally able to cope with the actual data. Based on this empirical study, the characteristic of FTS and RTS models can be distinguished and is listed in Table 12.

5. Conclusion

In this paper, the FTS and RTS models have been applied to forecast the yearly OP and OC data. The performance of the FTS model was better than the RTS model in terms of the MSE for both Indonesia's and Malaysia's OC. Moreover, the FTS model performed well in terms of the MSE for Indonesia's OP. From this performance, some merits of FTS model could be identified. The model does not require a scope for the number of data or any assumption for the data type, and also no statistical testing in terms of the model parameters, plus it is easier to be understood with simple calculations. The optimization of the partition number, however, is very important in the FTS model in order to achieve high forecasting accuracy. Apart from that, forecasting type-1 is also very helpful in obtaining better forecast values.

Acknowledgment

This study is supported by Research, Innovation, Commercialization, and Consultancy Management Office (ORICC) at UTHM, Malaysia.

References

Alpaslan, F. and Cagcag, O. (2012). Seasonal fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering. *J. Soc. Eco. Stat.*, **2**: 1–13.

Azadeh, A., Saberi, M. and Gitiforouz, A. (2012). An integrated simulation-based fuzzy regression-time series algorithm for electricity consumption estimation with non-stationary data. *J. Chinese Inst. Eng.*, **34**: 1047–1066.

Bolturuk, E., Oztayzi, B. and Sari, I. U. (2012). Electricity consumption forecasting using fuzzy time series. *IEEE 13th Int. Symp. Computer Intelligence and Informatics*, 20–22 November 2012, Istanbul, Turkey, pp. 245–249.

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets Syst.*, **81**: 311–319.

Cheng, C. H., Chen, T. L., Teoh, H. J. and Chiang, C. H. (2008). Fuzzy time series based on adaptive expectation model for TAIEX forecasting. *Expert Syst. Appl.*, **34**: 1126–1132.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Chiroma, H., Abdulkareen, S., Muaz, S. A., Abubakar, A. I., Sutoyo, E., Mungad, M., Saadi, Y., Sari, E. N. and Herawan, T. (2015). An intelligent modeling of oil consumption. *Advances in Intelligent Systems and Computing*, Vol. 320, Springer International Publication, Switzerland, pp. 557–568.

Economic Transformation Programme Chapter 1, http://www.etp.pemandu.gov.my.

Efendi, R., Ismail, Z. and Deris, M. M. (2013). Improved weight fuzzy time series used in the exchange rates forecasting US dollar to ringgit Malaysia. *Int. J. Comput. Intell. Appl.*, **12**: 19.

Efendi, R., Ismail, Z. and Deris, M. M. (2015). New linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. *App. Soft. Comput.*, **28**: 422–430.

Efendi, R., Deris, M. M. and Ismail, Z. (2016). Implementation of fuzzy time series in forecasting of the non-stationary data. *Int. J. Compt. Intel. Appl.*, **15**: 50–60.

Efendi, R and Deris, M. M (2017). Forecasting of Malaysian oil production and oil consumption using fuzzy time series. Herawan, T., *et al.* (eds.), *Advances Intel. Syst. Compt.*, **549**: 31-40.

Gabralla, L. A. and Abraham, A. (2013). Computational modeling of crude oil price forecasting: A review of two decades of research. *Int. J. Comp. Inf. Syst. Ind. Manag. App.*, **5**: 729–740.

Gholizade, H. and Chafi, M. R. S. (2015). Using repetitive fuzzy method for chaotic time series prediction. *J. Intell. Fuzzy Syst.*, **28**: 1937–1946.

Ismail, Z. and Efendi, R. (2011). Enrollment forecasting based on modified weight fuzzy time series. *J. Artif. Intell.*, **4**: 110–118.

Ismail, Z., Efendi, R. and Deris, M. M. (2013). Inter-quartile range approach to length — Interval adjustment of enrollment data. *Int. J. Comput. Intell. Appl.*, **3**: 10.

Ismail, Z, Efendi, R and Deris, M. M. (2015). Application of fuzzy time series approach in electric load forecasting. *New Mathematics and Natural Computation*, **11**: 229–248.

Kimura, S. (2009). The 2nd ASEAN Energy Outlook. The Energy Data and Modeling Centre, Japan.

Kuo, I. (2009). An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Syst. Appl.*, **36**: 6108–6117.

Liangyong, F., Junchen, L., Xiongqi, P., Xu, T. and Lin, Z. (2008). Peak oil models forecast China's oil supply, demand. *Oil Gas J.*, **14**: 43–47.

Lee, H. L., Liu, A. and Chen, W. S. (2006). Pattern discovery of fuzzy time series for financial prediction. *IEEE Trans. Syst. Man. Cyber-Part B.*, **18**: 613–625.

Singh, S. R. (2007). A robust method for forecasting based on fuzzy time series. *Int. J. Comput. Math.*, **188**: 472–484.

Song, Q. and Chissom, B. S. (1993a). Fuzzy time series and its models. *Fuzzy Sets. Syst.*, **54**: 269–277.

Song, Q. and Chissom, B. S. (1993b). Forecasting enrollments with fuzzy time series — Part 1. *Fuzzy Sets Syst.*, **54**: 1–9.

Sun, B. Q., Haifeng, G., Karimi, H. R., Ge, Y. and Xiong, S. (2015). Prediction of stock index futures pried based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing*, **151**: 1528–1536.

U.S. Energy Information Administration: Short-Term Energy Outlook. Independent Statistics & Analysis (2015); Czajkowski, K., Fitzgerald, S., Foster, I. and Kesselman, C. (2001). Grid information services for distributed resource sharing. *10th IEEE Int. Symp. High Performance Distributed Computing*, IEEE Press, New York, pp. 181–184.

Washington State Department of Transportation — Economic Analysis: Statewide Fuel Consumption Forecast Models (2010).

Wooldridge, J. M. (2006). *Introductory Econometrics A Modern Approach*. 3rd edition, Thomson, South Western, USA.

Yu, H. K. (2005). Weighted fuzzy time series models for TAIEX forecasting. *Physica A*, **349**: 609–624.

Yu, H. K and Huarng, K. H. (2008). A bivariate fuzzy time series model to forecast the TAIEX. *Expert Syst. Appl.*, **34**: 2945–2952.

Advances in Data Science and Adaptive Analysis 2017.09. Downloaded from www.worldscientific.com by NANYANG TECHNOLOGICAL UNIVERSITY on 09/20/17. For personal use only.

Oil_Production_and_Consumtion_Using_Fuzzy_Time_Series_...

ORIGINALITY REPORT

9%	%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	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	3%
2	C. Sasikala, A. Senthilkumar. "Secure disjoined multi-hop communication in wireless sensor network using trust nodes based data transmission technique", Materials Today: Proceedings, 2021 Publication	2%
3	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	2%
4	Riswan Efendi, Adhe N. Imandari, Yusnita Rahmadhani, Suhartono, Noor A. Samsudin, Nureize Arbai, Mustafa M. Deris. "Fuzzy Autoregressive Time Series Model Based on Symmetry Triangular Fuzzy Numbers", New Mathematics and Natural Computation, 2021 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		