First-Order Fuzzy Time Series based on Frequency Density Partitioning for Forecasting Production of Petroleum

R. Wulandari, Fariikhin, B Surarso, B Irawanto
Universitas Diponegoro, Semarang, Indonesia

Abstract. Forecasting method based on fuzzy time series has been widely developed in recent years. In this paper, we propose a new improvement at determining universe of discourse, variation historical data and partitioning stage. At early stage, we define the universe of discourse then calculate the basis value to find out how much interval should be used with variatin historical data. Secondly, we are partitioning the main intervals into several numbers of sub-intervals. The empirical analysis shows that sub-interval caused the fuzzy number getting closer to crisp value. It causes the better forecasting value. We use the data of yearly production petroleum Indonesia for simulation. We compare the forecasting results and error value of the method with previous existing methods. The modifications give better forecasting results than previous methods indicated with smaller The Means Squared Error (MSE) and Average Forecasting Error (AFER).

Keywords: First-Order Fuzzy time series, Frequency density, Partitioning, Sub-intervals.

1. Introduction
Mathematics as a fundamental science in the knowledge of mathematical causes can be applied in various disciplines in the world both science, technology, Economics and other sciences (Masykur 2008:43). The set of fuzzy was introduced by Zadeh (1965:338-353) and is expanding in various applications in the field of science one of which is in the field of forecasting, the forecasting using the fuzzy is commonly referred to as fuzzy forecasting, in different Forecasting using statistical methods such as ARMA, regression and others that can directly be in the process with existing values, in fuzzy forecasting in large lines there are 3 steps that is step Fuzzyfikasi, must change the existing values or Commonly called a Crips value to a linguistic value first, then fuzzy relation seeks the right relation to the value of linguistic values, subsequent defuzzyfication then by using a method to return the linguistic value to the Crips value, the advantage of using linguistic values is that it can be applied even if some data is lost (Klir, 1998). The Fuzzy time series was first introduced by song and Chissom in 1993 and grew rapidly in recent years, later (Chen,1996), (Ismail,2011), (Chissom,1993), (Chissom,1994) introducing several forecasting methods To foresee one object i.e. enrollment at the University of Alabama, Chen brought about an easier and simpler method. (Malike and Konstantin,2005) developed the first-order fuzzy time series to complete the university's forecasting of Alabama, then (Jilani TA,2007) writes that forecasting based on Frequencys density has a smaller error value Compared to the previous metoe. From the above, we modified the method belonging to Jilani with Malike and Constantine's methods became a new method which was later used to find
forecasting of data Petroleum Indonesia in 1996-2017. The result of each forecast method will be evaluated by comparing the value of MSE and AFER.

2. Methodology

2.1 Fuzzy Time Series

The main concepts of fuzzy time series is a fuzzy set \( A \) of the universe of discourse \( U \), \( U = \{u_1, u_2, ..., u_n\} \), is defined as follows: \( A = \{f_A(u_i)\}, i = 1, 2, ..., n \) where \( f_A \) is the membership function of the fuzzy set \( A \), \( f_A: U \to [0, 1], f_A(u_i) \) denotes the grade of membership of \( u_i \) in the fuzzy set \( A \), and \( 1 \leq i \leq n \).

Definition 1

Imprecise data at equally spaced discrete time points are modeled as fuzzy variables. The set of this discrete fuzzy data forms a fuzzy time series.

Definition 2

Chronological sequences of imprecise data are considered as time series with fuzzy data. A time series with fuzzy data is referred to as fuzzy time series.

Definition 3

Let \( Y(t), (t = \cdots, 0, 1, 2, \cdots) \) be universe of discourse and \( Y(t) \subseteq R \). Assume that \( f_i(t), i = 1, 2, \cdots \) is defined in the universe of discourse \( Y(t) \) and \( F(t) \) is collection of \( f(t_i), (i = \cdots, 0, 1, 2, \cdots) \). Then \( F(t) \) is called a fuzzy time series of \( (t), i = 1, 2, \cdots \). Using fuzzy relation, we define \( F(t) = F(t - 1) \circ R(t, t - 1) \) where \( R(t, t - 1) \) is a fuzzy relation and “\( \circ \)” is a max-min composition operator, then \( F(t) \) is caused by \( F(t - 1) \) where \( F(t) \) and \( F(t - 1) \) are fuzzy set.

Definition 4

The membership function of the triangle is marked by the presence of three (3) parameters fuzzy \( \tilde{a} = (\alpha_1, \alpha_2, \alpha_3) \), with the \((\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3 \) that will determine x coordinate from three angles. This curve is essentially a combination of the two lines. As for the triangular form of the equation for this is:

\[
\mu_\tilde{a}(x) = \begin{cases} 
  \frac{x - \alpha_1}{\alpha_2 - \alpha_1}, & \alpha_1 \leq x \leq \alpha_2 \\
  \frac{\alpha_3 - x}{\alpha_3 - \alpha_2}, & \alpha_2 \leq x \leq \alpha_3 \\
  0, & \text{other}
\end{cases}
\]
2.2 First order method

![Flowchart of first order forecasting](image)

**Figure 1.** Flowchart of first order forecasting

2.3 Simulation

**Step 1 Determine the universe of discourse**
Determine the universe of discourse\(^2\). From the table, we knew that the value of \(X_{min} = 548648.30\) dan \(X_{max} = 286814.20\)

\[
U = [X_{min} - D_1, X_{max} + D_2]
\]

\[
U = [250000 ; 550000]
\]

**Step 2 Find the number of intervals**
Find the number of intervals with the mean\(^3\).

\[
|\bar{X}| = 16536.045
\]

\[
J = X_{max} - X_{min} = 81719.6
\]

\[
|\bar{X}| / 2 = 8268.0225
\]

\[
I = \frac{J}{basis} = \frac{81719.6}{8000} = 10.21495
\]

Since the data is single data and single median data, the number of intervals is an odd number. 10.21495 are rounded to the odd 11 number of the nearest number.

**Step 3 Divide into sub-interval**
Divide into sub-interval with partitioning based on mean\(^4\). The division returned the average base intervals according to their respective frequency, if there is no data that is distributed in the interval (frequency is zero) then the interval will be discarded.

| Table 1. Fuzzy sub-interval using mean based partitioning |
| Sub-Interval | |
| U1 | [250000 ; 277272.73] |
| U2 | [277272.73 ; 304545.45] |
| U3 | [304545.45 ; 331818.18] |
| \vdots | \vdots |
| U11 | [522727.27 ; 550000] |
We have assumed 11 partitions of the universe of discourse of the main factor fuzzy time series.

**Step 4 define a fuzzy triangle based on table**
Let \( A_1, A_2, \ldots, A_k \) be the fuzzy sets which are linguistic values of the linguistic values of the linguistic variable IHSG data. In this paper, there are 11 linguistic values, as formula.

From \( u_1 \) triangular membership functions can be made from the interval [250000 ; 277272.73]

\[
\mu_d(x) = \begin{cases} 
\frac{x - 25000}{277272.73 - 263636.3} & , \quad 4194 \leq x \leq 4242,5 \\
\frac{277272.73 - x}{277272.73 - 263636.3} & , \quad 4242,5 \leq x \leq 4291 \\
0 & , \quad \text{other}
\end{cases}
\]

And further by replacing the as linguistic. Based on the fuzzify historical enrollments obtained in step 4, we can get the fuzzy logical relationship group (FLRG).

| Year | Actual Data | Fuzzification |
|------|-------------|---------------|
| 1996 | 548648.30   | A1            |
| 1997 | 548648.30   | A1            |
| 2015 | 286814.20   | A2            |
| 2017 | 292373.80   | A2            |

**Step 5 Determine first-order relation (FLR &FLRG)**
Logic Relationship between two consequent variations is as show Table 2 we assume that the fuzzy logical relationship " \( A_p \rightarrow A_q \) " denotes that “ if the fuzzified enrollments of year q are \( A_q \) respectively, then the fuzzified enrollments of year (p) is \( A_p \)."

FLR for \( A_1 \) is \( A_1 \rightarrow A_1, A_1 A_0 \)
Then FR left groupstarting from identical left –hand sides
The same rules for FLRG\( A_2 \), For \( A_2 \) FLRGnya \( A_2 \rightarrow A_2, A_2 A_2 \)

**Step 6 Defuzzyfikasi**
For the determination of the value of forecasting by determining the value prediction by defuzzification of linguistic forms through the Song and Chissom, its used principles that following:
\[
\begin{align*}
\text{if} \quad & \begin{cases}
0 & \forall \mu = 0 \\
\text{mid point} & \forall \mu \text{ has one maximum} \\
\text{mid point of corresponding conjunct interval} & \forall \mu \text{ has more than one maximum}
\end{cases}
\end{align*}
\]

Step 7 fault finding forecasting

Jilani, Burney, and Ardil (2007) using the AFER (Average Forecast Error Rate) and MSE (Mean Squared Error) methods to find the magnitude of the irregularities which occurred in the data results of the forecasting actual data. As for the calculation of AFER:

\[
AFER = \frac{|X_i - F_i|}{X_i} \times 100\%
\]

\[
MSE = \frac{\sum_{i=1}^{n}(X_i - F_i)^2}{n}
\]

We use MSE and AFER to compare the forecasting result of different forecasting methods, where \(A_i\) denotes the actual enrollment and \(F_i\) denotes the forecasting enrollment of year \(i\).

| AFER | Criteria |
|------|----------|
| <10% | Very good|
| 10%-20% | Good |
| 20%-50% | Good enough |
| >50% | Bad |

3 Results and Discussion

In this paper, we apply a new algorithm method to forecast the petroleum data from the year 1996-2017 which obtained from bps.co.id. Using the first-order fuzzy time series based on frequency density partitioning method, we compare the result of first-order fuzzy time series based on frequency density partitioning with error 0.033 % and Error of the first-order at the (Melike Sah and Konstantin 2005) give about 2% until 3%.

| Year | Actual Data | Proposed method |
|------|-------------|-----------------|
| 1996 | 548648.30   | 336363.6364     |
| 1997 | 543752.60   | 336363.6364     |
| 2010 | 344888.00   | 340000          |
| 2017 | 292373.80   | 290909.0909     |
| MSE  | 6382409916  |                 |
| AFER | 0.0330083%  |                 |
Conclusion
Using the first-order fuzzy time series based on frequency density partitioning method, we compare the result of first-order fuzzy time series based on frequency density partitioning with error 0.033% and Error of the first-order at the (Melike Sah and Konstantin 2005) give about 2% until 3%.

References
[1] M. Sah and Konstantin, “Freecasting Enrollment Model Based and First-Order Fuzzy Time Series,” Proceedings of World Academy of Science Engineering and Technology., vol. 1, pp. 1307-6884, 2003.
[2] T.A. Jilani, S.M.A. Burney and C. Ardil “Fuzzy Metric Approach for Fuzzy time series Forecasting based on Frequency Density based Partitioning,” International Journal of Computer and Information Engineering, vol. 4, pp. 1194 – 1199, 2007.
[3] L. Abdullah, “Performance of exchange rate forecast using distance-based fuzzy time series,”International Journal of Engineering and Technology (IJET), vol.5, pp. 452 – 459, 2007.
[4] W. Wu, “Fuzzy reasoning and fuzzy relational equations,”, Fuzzy Sets and Systems, vol. 20, pp. 67 – 78.
[5] S.M. Chen, “Forecasting Enrollments Based on Fuzzy Time Series,”, Fuzzy Sets and Systems, vol. 81, pp. 311–319, 1986.
[6] S.M. Chen and K. Tanuwijaya,”Fuzzy Forecasting Based on High-order Fuzzy Logical Relationships and Automatic Clustering Techniques,” Expert Systems with Applications, vol.3, pp. 15425–15437, 2011.
[7] R.C. Tsaur, “A fuzzy time series-markov chain model with an application to forecast the exchange rate between the Taiwan and US Dollar,”International Journal of Innovative Computing. Information and Control, vol.8, pp. 4931 – 492, 2011.
[8] K. Huarr, “Heuristic models of fuzzy time series for forecasting,” Fuzzy Sets and Systems, vol.123, pp. 369 – 386, 2001.
[9] S.M. Chen, “Forecasting Enrollment on Fuzzy Time Series,”Fuzzy Sets and System, vol.81, pp. 311-319,1996.
[10] Q. Song and B,S. Chissom, “Forecasting Enrollment with Fuzzy Time Series Part-1,”Fuzzy Sets and System, vol.pp.54 1-9,1993.
[11] Q. Song and B,S. Chissom, “Forecasting Enrollment with Fuzzy Time Series Part-2”, Fuzzy Sets and System, vol.62,pp. 1-8 1994.
[12] A.M. Masykur dan F.A. Halim, “Mathematical Intelegence,” AR-Ruzz Media 2008.
[13] L. A. Zadeh, “Fuzzy Sets”, Information and Control, vol.8,pp.338-353,1965.