Forecasting Palm Oil Production Using Fuzzy Time Forecasting Two-Factor Cross Associations with Frequency Density Partitions

(Peramalan Produksi Minyak Sawit Menggunakan Fuzzy Time Forecasting Two-Factor Cross Associations dengan Frequency Density Partitions)

R Wulandari[1*], L Aulia[1]

[1]*Department of Actuarial Sains, Faculty of Sains, Institut Teknologi Statistika dan Bisnis, Semarang, Indonesia*

E-mail: ratri.wulandari@itesa.ac.id, lathifatul.aulia@itesa.ac.id

ABSTRACT

Economic decisions have many determining factors based on estimates of macroeconomic variables. The accuracy of decision estimates can have an important impact. Forecasting is a method of reducing uncertainty about the future because economic decisions have multi-factor problems, the high order fuzzy time series forecast method is more suitable to overcome these problems. Predictions are made for main factors by taking influence from both factors. Fuzzy Logic Relationships (FLR) reflect the relationship between the premise and consequence. This paper will be discussed fuzzy time series forecasting multi-factor one order cross association based on frequency density partition as a forecasting method to forecast palm oil production influenced by large the area. The results of the estimates show that the proposed method has a high forecast performance, with an AFER value is 1.128% *according to the AFER criteria table* 1.128% *<10%, it can be concluded that the forecast has very good criteria.*

Keywords*: fuzzy forecasting, fuzzy logical relationship, two-factor one-order, cross association, frequency density partition.*

ABSTRAK

Keputusan ekonomi memiliki banyak faktor penentu berdasarkan perkiraan variabel makroekonomi. Keakuratan estimasi keputusan dapat memberikan dampak yang penting. Peramalan adalah metode untuk mengurangi ketidakpastian tentang masa depan karena keputusan ekonomi memiliki masalah multi-faktor, metode perkiraan deret waktu fuzzy order tinggi lebih cocok untuk mengatasi masalah ini. Prediksi dibuat untuk faktor-faktor utama dengan mengambil pengaruh dari kedua faktor tersebut. Fuzzy Logic Relationships (FLR) mencerminkan hubungan antara premis dan konsekuensinya. Tulisan ini akan dibahas fuzzy time series yang meramalkan multi-factor one order cross association berdasarkan partisi kerapatan frekuensi sebagai metode peramalan untuk meramalkan produksi minyak sawit yang dipengaruhi oleh area yang luas. Hasil estimasi menunjukkan bahwa metode yang diusulkan memiliki kinerja estimasi yang tinggi, dengan nilai AFER sebesar 1,128% sesuai dengan tabel kriteria AFER 1,128%<10%, dapat disimpulkan bahwa estimasi tersebut memiliki kriteria yang sangat baik.

Kata kunci: *fuzzy forecasting, fuzzy logical relationship, two-factor one-order, cross association, frequency density partition.*

INTRODUCTION

Known as a computer scientist at the University of California, Berkeley. He is Professor Lotfi A. Zadeh, who introduced a fuzzy set in 1965. A fuzzy set is a set that is represented by a membership function. The membership function is a curve that shows the mapping of data input points into their membership values which have intervals, the membership value can be called the degree of membership [4]. Over time, applications of fuzzy have developed into the forecasting field. The main purpose of forecasting is to reduce uncertainty about the future so that planning becomes more systematic and accountable. The time series methods that are often used are ARMA, ARIMA, ARIMAX, and so on. However, the ARIMA method has drawbacks such as this method requires large amounts of data, for the formation of a good model often requires time and other large resources. In addition, statistical forecasting methods such as ARIMA and ARIMAX cannot solve linguistic problems in forecasting. Therefore, the use of the fuzzy method has the advantage that the time series relation does not have to be long, this is because the use of the fuzzy method must first be changed in qualitative form. Another advantage of the Fuzzy Time Series (FTS) is that the calculation process does not require complicated systems such as genetic algorithms and neural networks, so it is certainly easier to develop. In addition, this method can also solve the problem of forecasting historical data in the form of linguistic values. Research using FTS has been expanded by various researchers, including the apply of the FTS method to overcome the problem of forecasting historical data in the form of linguistic values [6].

The application of several fuzzy forecasting methods has been implemented on time series forecasting on the University of Alabama registration data in 1971-1992, to get a method that has a small error value [5]–[8], these forecasts use one factor of time series data, namely the University of Alabama enrollment time series data from 1971-1992. Two or more factors forecasting is the development of one-factor forecasting. Forecasting of two or more factors is different from forecasting that uses only one factor, the expand of this forecasting allows researchers to predict time series data using the prime factor by considering the supporting factors which are then called high order factors. The basic model of time series fuzzy forecasting by Song and Chissom consists of 4 basic steps, the first step is to define the actual range of time series values in the interval and then define the fuzzy set of these intervals, the second step is to build the time series fuzzy from each data using the intervals at the previous step, step number three is to build fuzzy logic relations, the next step is to make predictions using defuzzification. For several decades the modification of fuzzy time series forecasting has developed rapidly to find the best method with a better error forecast. In the first basic step of defining the universe of discourse, Chen et all define how to find the range of intervals for the universe of discourse [8], still, at the same step, Jilani et al proposed dividing Chen's interval into several sub-intervals using frequency density partitions, like has been studied by B Irawanto et al with a modification of the metric approach for predicting fuzzy time series based on frequency density partitions [8]–[10]. The second basic step is to build a triangular fuzzy number using the definition of Kusuma et al [11]. The third step is to build fuzzy logic relations (FLR), FLR become so important to the forecasting method because it optimizes interval division and interval length as well as FLR, interval division greatly affects forecast accuracy [5], [12], [13].

Tri Oktarina Rasmila discussed the forecasting of crude palm oil production using the ARIMA method at PT Sampoerna Agro Tbk. The ARIMA method is expected to produce accurate forecasting of production results for the following year so that companies can make the right decisions if oil palm production declines and companies can continue to increase crude palm oil (CPO) production. Another discussion regarding forecasting oil palm fresh fruit bunches production using linear regression and genetic algorithms (Case Study: PT. Peputra Masterindo) by Fitri Insani, et al. Production is one of the expected results in a plantation, especially PT. Peputra Masterindo is engaged in palm oil processing. The amount of production of fresh fruit bunches (FFB) is very influential on oil palm production. The method used is the linear regression method to predict, while the genetic algorithm is used to optimize the variables that affect the prediction results.

In this paper, the author applies the fuzzy method for forecasting palm oil production. Where, FLR plays an important role in the time series fuzzy model, concerning FLR, factor and order are two related concepts that are important and influence the development of FLR, on the other hand, the development of forecasting is also carried out by S. Kumar et al by adding the time series factor into two factors (multi-factor) [14]. Predictions made for the main factor by considering the main factor itself are included in the one-factor forecasting problem, while in a multi-factor forecasting problem, predictions are made to predict the main factor by taking the influence of the second factor. Forecasting problems using multiple factors still use the relation defined by Chen, while in 2020 Li et al sparked a new relation definition called cross relation [3].

PRELIMINARIES

1.1 Model of Fuzzy Time Series

Definition 1

Suppose that two fuzzy time series $F(t)$ and $G(t)(t = \cdots,0,1,2,3,\dots)$, If $F(t)$ caused $F(t-1), G(t-1)$, then the FLR is mapping by :

$$\left(\boldsymbol{F}(\boldsymbol{t}-\mathbf{1})\boldsymbol{G}(\boldsymbol{t}-\mathbf{1})\right) \rightarrow \boldsymbol{F}(\boldsymbol{t}) \quad (1)$$

Equation (1) is called a two-factor FLR, where the prime factor is $F(t)$ and the second factor is $G(t)$ on $(t = \cdots,0,1,2,3,\dots)$.

The number of fuzzy sets $F(t)$ $(t = \cdots,0,1,2,3,\dots)$. For example, if $F(t-1) = A_t$, $G_2(t-1) = B_j^2$ dan A_r is a fuzzy set. Equation (1) can be written as:

$$\boldsymbol{A_t B_j^2} \rightarrow \boldsymbol{A_r} \quad (2)$$

where $A_t B_j^2$ is called the premise and A_r called consequences.

If $A_t B_j^2 \rightarrow A_r$ appear v times in the fuzzy time series, it will be written as $A_t B_j^2 \rightarrow A_r(v)$ where is a positive integer.

2.2 Short Cross Assosiation (SCAFLR and SAFLR)

Definition 2

Suppose that the forecast to forecasting the t moment from the main factor is F. Observations from F in the moment $t-1$ are $F(t-1)=A, G_i(t-1)=B_j^i (i=1,2,\ldots,n)$ respectively. SAFLR and SCAFLR are defined by [3] :

SAFLR, $\boldsymbol{A_u} \rightarrow \boldsymbol{A_k}$ is SAFLR if it meets $\boldsymbol{A_u} = \boldsymbol{A_k}$ (3)

SCAFLR, $\boldsymbol{B_j^i} \rightarrow \boldsymbol{A}$ is SCAFLR if it meets $\boldsymbol{B_f^i} = \boldsymbol{B_j^i}$ (4)

dan dipisahkan dengan tab berisi titik-titik (lihat Persamaan 1). Keterangan rumus dituliskan di bawah rumus, disusun rapi ke bawah sesuai dengan simbol rumus yang digunakan.

2.3 Short Cross Association Forecasting

The following figure is the flowchart of two-factor cross association based frequency density partitioning

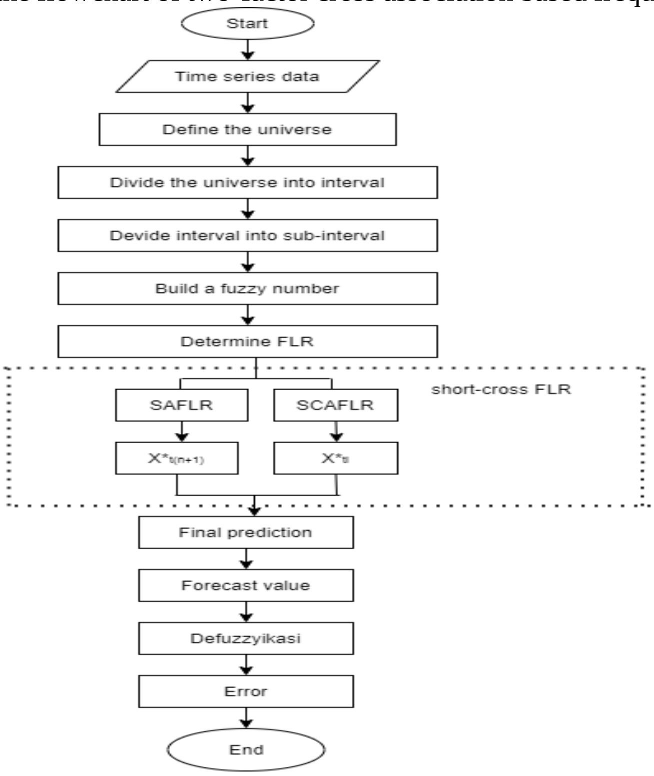

Figure 1 Short Cross Association Forecasting

SIMULATION

In this research simulation used data related to oil palm production. The data used are in the form of oil palm plantation area in hectares and data on oil palm production in tons from 2000 to 2019.

Step 1: Determined the universe for each factor, as follows:

$$X_{min} = 2991.3 \quad X_{max} = 8688.9 \quad D_{X1} = 91.3 \quad D_{X2} = 56.1$$

$$Y^1{}_{min} = 5094.86 \quad Y^1{}_{max} = 29637.5 \quad D_{Y^1{}_1} = 94.86 \quad D_{Y^1{}_2} = 62.5$$

For example, the prime factor of the universe (U)

$$U = [X_{min} - D_{X1}, X_{max} + D_{X2}] = [2900,8745]$$

Estimate the universe of supporting factors (W)

$$W = [Y^1{}_{min} - D_{Y^1{}_1}, Y^1{}_{max} + D_{Y^1{}_2}] = [5000,29700]$$

Define the extent of the interval for the prime factor

$$l_X = \frac{[(X_{max} + D_{X2}) - (X_{min} - D_{X1})]}{m} = 835$$

Define the extent of the interval for a second factor

$$l_{Y^1} = \frac{[(Y^1{}_{max} + D_{Y^1{}_2}) - (Y^1{}_{min} - D_{Y^1{}_1})]}{m} = 2470$$

Steps 2 and 3: Divide the universe into intervals [14] and divide the universe into some sub-intervals [13]

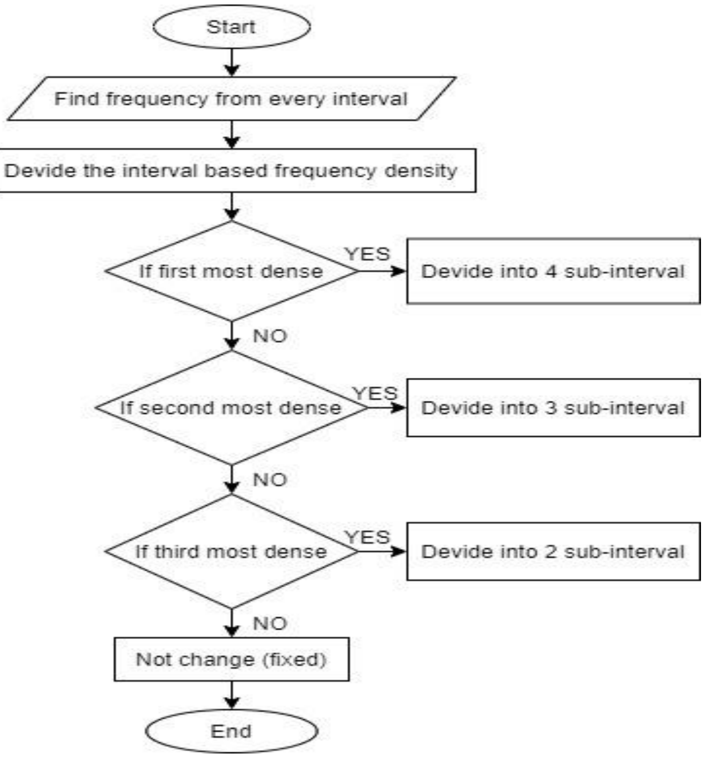

Figure 2 Sub-interval algorithm

Step 4: construct a fuzzy time series for each factor.

Definition 3 [3]

The triangle membership function can be made from the interval $u_1 = [5000,5617.5]$

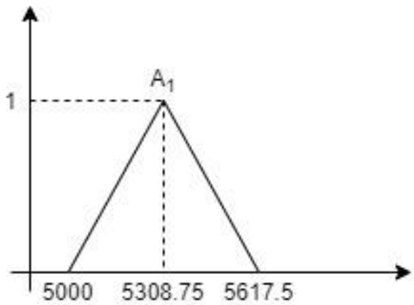

Figure 3 triangular membership function $\boldsymbol{A_1}$

$$\mu_{\tilde{a}}(x) = \begin{cases} \frac{x-5000}{5308.75-5000} & , 5000 \le x \le 5308.75 \\ \frac{5617.5-x}{5617.5-5308.75} & , 5308.75 \le x \le 5617.5 \\ 0 & , \quad other \end{cases} \tag{5}$$

Table 1 Divide second factor sub-interval

Large (second factor)	Sub-interval	FUZZY
[5000 ; 7470]	[5000; 5617.5]	A1
	[5617.5 ; 6235]	A2
	[6235; 6852.5]	A3
	[6852.5 ; 7470]	A4
[27230 ; 29700]	[27230 ; 28465]	A19
	[28465 ; 29700]	A20

Table 2 Divide the main factor sub-interval

Production (main factor)	Sub-interval	FUZZY
[2900 ; 3735]	[2900 ; 3108.75]	B1
	[3108.75 ; 3317.5]	B2
	[3317.5 ; 3526.25]	B3
	[3526.25 ; 3735]	B4
[7910 ; 8745]	[7910 ; 8745]	B13

Step 5: defines a fuzzy set group for each factor

5.1 Forecast of SCAFLR available at $R_i(i = 1,2,...,n)$

Suppose that the observation of the supporting factors $(i+1)$ when $(t-1)$ fuzzy by B_j^i. N_i is the number of available SCAFLRs found in R_i, according to the number N_i, the $x_{t_i}^*$ $(j \in 1,2,\dots,k_i+1),(i = 1,2,\dots,n)$ predicted time t can be calculated, for $N_i \neq 0$.

Several N_i FLRs are awarded, with:

$$A_1 \rightarrow B_1 \ (1)$$
$$A_1 \rightarrow B_2 \ (1)$$
$$A_2 \rightarrow B_2 \ (1)$$
$$A_4 \rightarrow B_3 \ (1)$$
$$A_5 \rightarrow B_3 \ (1)$$
$$A_6 \rightarrow B_4 \ (1)$$
$$A_7 \rightarrow B_5 \ (2)$$
$$A_9 \rightarrow B_6 \ (1)$$
$$A_{10} \rightarrow B_7 \ (1)$$
$$A_{10} \rightarrow B_8 \ (1)$$
$$A_{12} \rightarrow B_8 \ (1)$$
$$A_{13} \rightarrow B_9 \ (1)$$
$$A_{14} \rightarrow B_9 \ (1)$$
$$A_{15} \rightarrow B_{10} \ (1)$$
$$A_{16} \rightarrow B_{11} \ (1)$$
$$A_{16} \rightarrow B_{10} \ (1)$$
$$A_{18} \rightarrow B_{11} \ (1)$$
$$A_{19} \rightarrow B_{13} \ (1)$$
$$A_{20} \rightarrow B_{13} \ (1)$$

5.2 Forecast of SAFLR available at R

Suppose that the main factor when $(t-1)$ fuzzy by A_j, N is the number of available SAFLR found in R. The prediction $x_{t_{(n+1)}}^*$ can be calculated in the same way in step 5:

$B_1 \rightarrow B_2$ (1)	$B_4 \rightarrow B_5$ (1)	$B_8 \rightarrow B_8$ (1)	$B_{11} \rightarrow B_{10}$ (1)
$B_2 \rightarrow B_2$ (1)	$B_5 \rightarrow B_5$ (1)	$B_8 \rightarrow B_9$ (1)	$B_{11} \rightarrow B_{13}$ (1)
$B_2 \rightarrow B_3$ (1)	$B_5 \rightarrow B_6$ (1)	$B_9 \rightarrow B_9$ (1)	$B_{13} \rightarrow B_{13}$ (1)
$B_3 \rightarrow B_3$ (1)	$B_6 \rightarrow B_7$ (1)	$B_9 \rightarrow B_{10}$ (1)	$B_{23} \rightarrow \#$ (1)
$B_3 \rightarrow B_4$ (1)	$B_7 \rightarrow B_8$ (1)	$B_{10} \rightarrow B_{11}$ (2)	

Step 6: Forecasting, using the following definitions:

Definition 4 Forecasts from SCAFLR are available at.$R_i(i = 1,2,...,n)$

Then the prediction can be calculated by:

$$\boldsymbol{x'} = \frac{v_1 \times \bar{x}(l_1) + v_2 \times \bar{x}(l_2) + \dots + v_N \times \bar{x}\left(l_{N_i}\right)}{v_1 + v_2 + \cdots + v_{N_i}} \quad (6)$$

With the $\overline{\boldsymbol{x}}(\boldsymbol{l_r}) = \overline{\boldsymbol{x}}(\boldsymbol{a})$ defined as:

$$\bar{x}(a) = \begin{cases} \frac{0.5m(1)+Q(1)}{1.5} & a = 1 \\ \frac{0.5m(a-1)+Q(a)+0.5(a+1)}{2} & 2 \le a \le k \\ \frac{0.5m(a-1)+Q(a-1)+q}{1.5} & a = k+1 \end{cases} \quad (7)$$

The forecast from the SCAFLR is available R_i

The year 2001 data is fuzzy by A_1. By searching on R_i, it can be found that the available SCAFLRs for this data (whose premise is A_1) are as follows :

$$A_1 \to B_1 \ (1)$$
$$A_1 \to B_2 \ (1)$$

Because it uses case 1.$N_i \neq 0$

Using the FLR above, forecasts can be obtained as follows :

$$x' = \frac{v_1 \times \bar{x}(l_1) + v_2 \times \bar{x}(l_2) + \ldots + v_N \times \bar{x}(l_{N_i})}{v_1 + v_2 + \cdots + v_{N_i}} = \frac{1 \times \bar{x}(1) + 1 \times \bar{x}(2)}{1+1}$$

With,

$$\bar{x}(1) = \frac{0.5 \times m(1) + 1 \times Q(1)}{1 + 0.5} = \frac{0.5 \times (3004.38) + 1 \times (2900)}{1.5} = 2934.79$$

$$\bar{x}(2) = \frac{0.5 \times m(1) + 1 \times Q(2) + 0.5 \times m(3)}{0.5 + 1 + 0.5} = \frac{0.5 \times (3004.38) + 1 \times 3108.75 + 0.5 \times (33421.88)}{2} = 3160.64$$

Then obtained:

$$x' = \frac{1 \times (2934.79) + 1 \times (3160.64)}{2} = 3047.87$$

Definition 5 Forecasts from the SAFLR are available at R. [20]

Forecasts from the SAFLR are available R

The year 2001 data is fuzzy by B_1. By searching on R, it can be found that the SAFLR available for this data (whose premise is B_1) as follows :

$$B_1 \to B_2 \ (1)$$

Because $N_i \neq 0$ then using the FLR above, forecasts can be obtained as follows :

The prediction x_t^* is calculated by:

$$x_t^* = 0.5(\bar{x}(p) + x') \quad (8)$$

Then obtained:

$$x_{2001}^* = 0.5\big(\bar{x}(1) + \bar{x}(2)\big)$$
$$x_{2001}^* = 0.5\big((2934.79) + (3160.94)\big)$$
$$x_{2001}^* = 3047.87$$

Step 7: Defuzzification (Final prediction from two kinds of FLR).

The final prediction for the year 2001 is come from x_{2001}^* and $x_{2001} = \frac{x_{2001}^* + x_{2001}}{2} = 3047.87$

Step 8: Calculating the error.

Given the error calculation formula MSE (Mean Squared Error)

$$MSE = \frac{\sum_{i=1}^{n}(X_i - F_i)^2}{n} = \mathbf{304255.92}$$

Then, AFER (the Average Forecast Error Rate)

$$AFER = \frac{\frac{|X_i - F_i|}{X_i}}{n} \times 100\% = \mathbf{1.128}\%$$

With the following criteria:

Table 3 AFER Benchmark

AFER benchmark	
Excelent	<10%
Good	10% -20%
Normal	20% -50%
Bad	> 50%

RESULT AND DISCUSSION

In this section, the estimation result show that the proposed method has a higher forecasting performance. The method used in forecasting oil palm production is a database that has been grouped and influenced by plantation area then uses a modified method of forecasting fuzzy time series multi-factor one cross association with frequency density partitioning.

Table 4 Fuzzified historical data

Year	Actual Value X_i	Forecast Value F_i	$\frac{\lvert X_i - F_i \rvert}{X_i}$	$(X_i - F_i)^2$
2000	2991.3			
2001	3152.4	3047.86667	0.03316	10927.218
2002	3258.6	3130.49708	0.039312	16410.357
2003	3429.2	3239.22125	0.0554	36091.925
2004	3496.7	3451.23281	0.013003	2067.2651
2005	3593.4	3451.23281	0.039563	20211.509
2006	3748.5	3774.14188	0.006841	657.50575
2007	4101.7	4097.05188	0.001133	21.605066
2008	4451.8	4097.05188	0.079686	125846.23
2009	4888	4465.625	0.086411	178400.64
2010	5161.6	4909.21875	0.048896	63696.295
2011	5349.8	5083.17719	0.049838	71087.724
2012	5995.7	5304.97406	0.115204	477102.32
2013	6108.9	5885.56094	0.03656	49880.337
2014	6332.4	5885.56094	0.070564	199665.15
2015	6724.9	6396.5675	0.048823	107802.23
2016	6462.1	6558.92802	0.014984	9375.6656
2017	6685.2	6413.9625	0.040573	73569.781
2018	8507.4	6715.49052	0.21063	3210939.6

2019	8688.9	7492.50333	0.137693	1431365
		AFER	1.128273	6085118.3
			MSE	304255.92

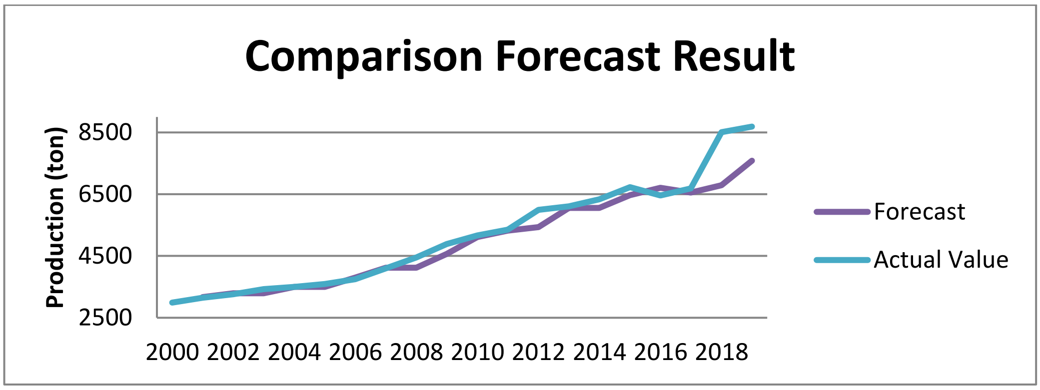

Figure 4 Chart of comparison forecast results

CONCLUTION

From the research conducted, the time series forecasting method has 4 basic steps. In the fuzzy time forecasting method, cross-relation based on the frequency density partition is modified from the basic step, among others, in the step before building the fuzzy number is a sub-interval division with the partition to the frequency, then the determination of fuzzy relationships using cross relationships, to minimize errors in the results of the forecasting value obtained. In this research, forecasting is done by using multi factors that can do forecasting better than using a single factor and order one because in forecasting using FLR built based on time series data several n-orders and using consideration of the predicted data factors. Implementation of fuzzy time forecasting factors many cross-relationships based on the frequency density partition imposed on the time run tun data related to the production and area of palm oil plantations. So it can be deduced that the forecasting performed has very good criteria with the obtained AFER value of 1.128% according to the AFER criteria table 1.128%.

REFERENCES

[1] J. B. and Serena 2005 Understanding and Comparing Factor-Based Forecasts *Comp. Factor-Based Forecast Underst* **vol. 1**, no. 3, pp. 117–151

[2] A. Hmadu 2017 A Novel Two – Factor High Order Fuzzy Time Series With Applications To Temperature And Futures Exchange Forecasting *Dept Engineering* **vol. 36**, no. 4, pp. 1124–1134

[3] F. Li and F. Yu 2020 Multi-factor one-order cross-association fuzzy logical relationships based forecasting models of time series *Inf. Sci.* **vol. 508** pp. 309–328

[4] L. A. Zadeh 1965 Fuzzy sets *inf. Control*, **vol. 8**, no. 3, pp. 338–353

[5] Q. Song and B. S. Chissom 1993 Forecasting enrollments with fuzzy time series—part I *Fuzzy sets Syst* **vol. 54**, no. 1, pp. 1–9

[6] Q. Song and B. S. Chissom 1994 Forecasting enrollments with fuzzy time series-Part II *Fuzzy sets Syst* **vol. 62**, no. 1, pp. 1–8

[7] S. Chen 1996 *Forecasting enrollments based on fuzzy time series* Fuzzy sets Syst vol. 81, pp. 311–319

[8] S.-M. Chen and C.-C. Hsu 2004 A new method to forecast enrollments using fuzzy time series *Int. J. Appl. Sci. Eng.* **vol. 2**, no. 3, pp. 234–244

[9] B. Irawanto, R. W. Ningrum, and B. Surarso 2019 An improved forecasting method of

frequency density partitioning (FDP) based on fuzzy time series (FTS) *Journal of Physics: Conference Series* **vol. 1321**, no. 2, p. 22082

[10] R. Wulandari, Farikhin, B. Surarso, and B. Irawanto 2020 First-Order Fuzzy Time Series based on Frequency Density Partitioning for Forecasting Production of Petroleum *IOP Conf. Ser. Mater. Sci. Eng.* **vol. 846**, no. 1,

[11] S. Kusumadewi and H. Purnomo 2010 *Aplikasi Logika Fuzzy untuk pendukung keputusan* (Yogyakarta:Graha Ilmu) pp. 33–34

[12] M. I. Hasan 2001 *Pokok-pokok Materi Statistik I (Statistik Deskriptif)* (Jakarta: Bumi Aksara)

[13] T. A. Jilani, S. M. A. Burney, and C. Ardil 2007 Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning *Int. J. Comput. Intell.* **vol. 4**, no. 1, pp. 112–117

[14] S. Kumar and N. Kumar 2015 Two factor fuzzy time series model for rice forecasting *Int. J. Comput. Math. Sci* **vol. 4**, no. 1, pp. 56–61