IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 27, 2018, accepted June 4, 2018, date of publication June 13, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2846809

Forecasting Crude Palm Oil Prices Using Fuzzy Rule-Based Time Series Method

NUR FAZLIANA RAHIM[1], MAHMOD OTHMAN[1], RAJALINGAM SOKKALINGAM[1], AND EVIZAL ABDUL KADIR[2]

[1]Fundamental and Applied Science Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[2]Faculty of Engineering, Universitas Islam Riau, Pekanbaru 28284, Indonesia

Corresponding author: Nur Fazliana Rahim (rnfazliana@unimas.my)

This work was supported by Universiti Teknologi PETRONAS under Grant YUTP (Cost Centre: 0153AA-H27).

ABSTRACT Oil palm can be reflected as the main contributor to capital investment, technology, foreign workers' employment, and knowledge management. There are risks and uncertainties arising from instability of crude palm oil (CPO) prices. Therefore, this research develops a new CPO price forecasting method using weighted subsethood-based algorithm in order to generate fuzzy rules of forecasting. The concept of fuzzy rule-based systems was embedded in fuzzy time series application to generate fuzzy IF-THEN rules. This paper aims to enhance the efficacy of time-series forecasting, which would in turn increase the accuracy of the predictions. The CPO prices data set was used for validation purposes. The accuracy of forecasting of the proposed method was compared with previous methods. The numerical results are comparable with the previous methods. The outcomes of the proposed method have shown an increase in the accuracies of the CPO price forecasts. As such, the above-mentioned method can be utilized for the creation of a new set of fuzzy rules to better predict CPO prices.

INDEX TERMS Fuzzy time series, fuzzy rule-based systems, weighted subsethood based algorithm, forecasting CPO prices.

I. INTRODUCTION

Malaysia was, and still is, the world palm oil leader in terms of production and export in light of huge increases in its production and export capacities [1]. To date, the largest revenue of the agricultural sector in Malaysia is accounted for by palm oil products [2]. Besides, people from 150 countries are currently consuming it. This demonstrates the vital importance of this industry in Malaysia. From 2001 onwards, it has been known that, palm oil has surpassed timber and rubber in terms of contribution to the local gross domestic product (GDP) [3].

Besides, it is found in the previous literature that the price of palm oil is volatile and the instability of palm oil prices will affect many sectors especially the country's earnings indirectly [4]. CPO price volatility is a form of insecurity for the players in the palm oil industry, which includes the producers, marketers, and end users. Therefore, accuracy of price forecasting is required to ease decision making in the event of immense economic instability [5]. It was also mentioned in [6] that predictions which a high degree of accuracy are paramount, as the decisions devised with reference to the same will have an effect on the true market performance of this commodity. Thus, it is necessary to monitor the growth of Crude Palm Oil (CPO) prices in the past as well as in the future [7]. In order to improve CPO price forecasting accuracy, one should take daily CPO price data to measure whether there is any factor that would change the price [8].

Although there are many methods of CPO price forecasting, the forecasting error is still debated due to high percent errors. As all known that, the forecaster's decision making on forecasting is very important in order to maintain the production in the oil palm business. Furthermore, the CPO price instability due to factor uncertainties is the major issue. Thus, the fuzzy method is useful in handling the uncertainty in the data. In many cases, forecasting of future values of CPO price has the limitations in terms of accuracy. A forecaster also faces difficulties in determining the appropriate method to be used in forecasting. Although these methods are well suited for measuring the efficiency and accuracy of the forecasting, they do not intend to address the use of fuzzy approximate reasoning, which reflects human thinking, to allow decision-makers to make the best choice. There are numerous forecasting methods that are well established by multiple ways. However, the problem that arises in most forecasting methods is the inability of the method to reduce forecasting error. The methods are also unable to produce accurate prediction values. Therefore, certain approach has to be adopted in finding a more suitable method which can address these drawbacks. Thus, this research proposes a new forecasting method using fuzzy rule based time series.

The new approach can be expected to offer several contributions to enhance the forecasting method. The proposed WSBA produces new fuzzy rules, which carry new information, alongside with the previous rules obtained in the previous method. This may help to confirm or refute rules devised by humans or by other automated systems. The use of linguistic hedges in fuzzy approach allows more flexible judgment compared to numerical values, in particular when comparing forecasting performance. A fixed number of rules will be generated according to a number of classes which are defined in the methodology. This method is imperative to the field of forecasting as it can be used in different forecasting applications.

II. LITERATURE REVIEW

Many methods have been proposed in relation to CPO price forecasting. According to [9], many aspects should be taken into consideration for selection of the methods. This includes data, financial support, manpower and expertise, accuracy, as well as significance level. Nevertheless, most of the methods are relatively expensive and require various data types and expertise, whose availability could be scarce. Past methods have their respective advantages and disadvantages. The researcher in [10] used Box-Jenkins approach to forecast monthly CPO prices. Autoregressive Integrated Moving Average (ARIMA) was applied in finding the suitable time series to its past values but the results of the residuals were orthogonal and not normal. Upon observation, it is revealed that there were endured serial correlations in the series. Not only that, there were also problems in deciding the proper order of model identification stage of ARIMA, such as parameters and residuals from the fitted model. Re-identification of a model becomes necessary when an incorrectly identified model gives rise to inaccurate estimations. According to [11], copious data is required for the ARIMA model, notwithstanding the difficulty in updating its parameters in the presence of new data. These problems raised the risk of selecting models that consist of biggest error [12].

In another study, an Artificial Intelligence (AI) approach was used in order to forecast the CPO prices [13]. As such, this research suggested the utilization of the Artificial Neural Network (ANN) and Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to predict the aforementioned parameter with increased accuracy. Between the two models, the ANN provided superior results compared to ARFIMA's. Nevertheless, both models had a weakness that was more complicated underlying time series characteristics. As both of them had relatively more parameters, they required larger amounts of data. As reported by [14], this downside was also true for ARIMA. This issue was supported by [15], who stated that excessive problems occur more frequently in neural network models compared to other statistical models. Meanwhile, [16] in their study compared the Multivariate Autoregressive Moving Average (MARMA) model with an econometric model to forecast CPO prices. The results indicated that the efficiency of the predictions was higher when the MARMA model was used instead of the econometric one. However, the Root Mean Square Percentage Error (RMSPE) result showed that the model produced high percentage of errors.

In other cases, the researcher in [17] did their study for stock price forecasting, which the proposed model had significant performance than existing fuzzy time series models. From the RMSE value for the proposed model, the results indicated that it could reduce forecasting errors. Furthermore, [18] mentioned that their proposed model based on entropy discretization and a fast Fourier transform outperformed the conventional time series models and could effectively handle any kind of time series forecasting problems. There was also a study in [19] which proposed a high-order fuzzy time series model for internet stock trading and the results also outperformed the conventional fuzzy time series models. The researcher did mention that the study generated decision rules and effective high-order fuzzy time series model, but needed more research to further validate their performance. Besides that, according to [20] and [21], the application of fuzzy approach-based reasoning in forecasting was successful. In addition, fuzzy rule-based systems (FRBS) employ semantic models to reason with reference to intuition [22]. There were numerous applications of FRBS in real world problem classifications in the area of fuzzy systems. More and more studies over the past few years have demonstrated that FRBS is readily available and reliable in the sense that it produces results of higher precision as what supported by [22]. Apart from that, FRBS has been established as an alternative method to tackle various types of less precise data, the outcomes of which can affect the thought and decision-making processes [23]. Fuzzy rule determination is one of the most important things to be considered in fuzzy forecasting. The use of these rules implemented in the fuzzy time series increases the accuracy of the results produced in forecasting situations that involves subjective, vague, and imprecise information. Thus, in this research, Weighted Subsethood-Based Algorithm (WSBA) is proposed as it is suitable for generating fuzzy rules. WSBA is relatively an easier method and there are likelihood of the generation of highly-accurate outcomes in the presence of fewer rules vis-à-vis other models [24]. The development of WSBA comprises a modification of the Subsethood-Based Algorithm (SBA) and generation of fuzzy rules using SBA values as weights.

III. METHODOLOGY

The method used to achieve the objectives was discussed in this section. There are a total of 5 phases in the said method, which are described in detail as follows.

A. PHASE 1: DATA COLLECTION AND PRE-PROCESSING

Crude Palm Oil (CPO) prices data of Malaysia were collected based on the historical daily CPO prices data for the year 2012 until 2016. The data were obtained from the Malaysian Palm Oil Board (MPOB). A small dataset was used in order to show how this method works. The CPO dataset was divided into two subsets: CPO-1, which was used for training and CPO-2, which was used for testing. Each dataset contained 28 cases. The CPO price data comprised of three attributes: One day previous, Two days previous and Three days previous. The classification outcomes of the CPO price were ranked as Low, Medium, and High. The proposed fuzzy time series method is shown as follow referred in [25].

$$Y = \alpha X_{t-1} + \beta X_{t-2} + \delta X_{t-3} \tag{1}$$

where

X_{t-1} = one day previous,
X_{t-2} = two days previous, and
X_{t-3} = three days previous.

B. PHASE 2: FUZZY RULE GENERATION

In this phase, the process to produce the required fuzzy model based on WSBA was proposed. To improve the comprehensibility of the resultant system, WSBA employs an algorithm which generates rules with respect to the general fuzzy laws or the extension of a Mamdani-type FRBS. Five steps have been taken in this phase:

Step 1: The training dataset as in Table 1 was separated into three subgroups based on the classification outcomes. The classification outcomes are calculated based on measure of location, Q_k as follows.

$$Q_k = L_k + \left(\frac{\frac{k}{4}N - F_k}{f_k}\right) C_k \tag{2}$$

where $k = 1, 2, 3$,

L_k = lower boundary of the class where Q_k lies,
N = total number of observations,
F_k = cumulative frequency before the Q_k class,
f_k = frequency of the class where Q_k lies, and
C_k = size of the class where Q_k lies.

Step 2: The measure of locations Q_1, Q_2, and Q_3 was calculated from (2) and this was followed by construction of the fuzzy membership function for the CPO price dataset. The fuzzy partition is defined according to the constructed fuzzy membership function and it is used to transform crisp values into fuzzy values.

Step 3: For each linguistic term, the values of the fuzzy subsethood in each subgroup have been determined. These rules have been created to address the issues pertaining classification. The extent to which the fuzzy subsethood value A was a subset of B denoted by $S(B, A)$

TABLE 1. Training dataset (CPO-1).

Case	t-3	t-2	t-1	Y	Outcome
1	531.21	527.50	568.43	618.32	High
2	527.50	568.43	618.32	591.28	High
3	568.43	618.32	591.28	553.24	High
4	618.32	591.28	553.24	568.60	High
5	591.28	553.24	568.60	558.52	High
6	553.24	568.60	558.52	513.61	Medium
7	568.60	558.52	513.61	421.28	Low
8	558.52	513.61	421.28	431.05	Low
9	513.61	421.28	431.05	411.35	Low
10	421.28	431.05	411.35	439.21	Low
11	431.05	411.35	439.21	452.29	Low
12	411.35	439.21	452.29	456.42	Low
13	439.21	452.29	456.42	462.82	Low
14	452.29	456.42	462.82	468.54	Medium
15	456.42	462.82	468.54	497.86	Medium
16	462.82	468.54	497.86	501.36	Medium
17	468.54	497.86	501.36	524.21	Medium
18	497.86	501.36	524.21	536.65	Medium
19	501.36	524.21	536.65	528.92	Medium
20	524.21	536.65	528.92	566.86	High
21	536.65	528.92	566.86	552.88	High
22	528.92	566.86	552.88	535.10	Medium
23	566.86	552.88	535.10	560.73	High
24	552.88	535.10	560.73	592.24	High
25	535.10	560.73	592.24	565.61	High
26	560.73	592.24	565.61	541.45	High
27	592.24	565.61	541.45	521.66	Medium
28	565.61	541.45	521.66	531.41	Medium

where $S(B, A) \in [0, 1]$ [26], [27]:

$$S(B, A) = \frac{M(B, A)}{M(B)} = \frac{\sum_{x \in U} \nabla\left(\mu_B(x), \mu_A(x)\right)}{\sum_{x \in U} \mu_B(x)} \tag{3}$$

where $S(B, A)$ refer to fuzzy subsethood value of fuzzy set A to fuzzy set B, $\mu_A(x)$ and $\mu_B(x)$ refer to membership value of x in set A and B respectively.

Step 4: Based on the subsethood values in Step 3, weights for each linguistic term were calculated. The weightages in this study ranged between 0 and 1, whereby the former represented minimum weightage (i.e. of lowest importance) while the latter maximum weightage (i.e. of highest importance). Through an expansion of the subsethood explication, fuzzy sets could now be associated with various linguistic variables. With the assumption that the value of the subsethood for a particular linguistic term, *Aj*, of linguistic variable *A* with respect to classification *X* was *S(X, A)*, and that $A_1, A_2, \ldots, A_l$ were within the set of possible linguistic terms for *A*, then *Ai*'s relative weight, *W*, with respect *X* was:

$$W(X, A_i) = \frac{S(X, A_i)}{\max_{j=1,\ldots,l} S\left(X, A_j\right)} \tag{4}$$

where $W(X, A_i) \in [0, 1]$ and $i = 1, 2, \ldots, l$. Hence, every linguistic term in each conditional attribute now had a specific weightage. Also, the most and least important subsethood values of linguistic terms were the largest and smallest ones respectively. The weightages assigned to every antecedent via fuzzy rules were non-negative numbers, and these could be modified with respect to the learning datasets. Additionally, these weightages were incorporated into the linguistic terms, which were in turn associated with the conditional attributes. Thus, the following formula could be used to calculate the compound weightage, $T(A)$ of the weighted conjunction of linguistic terms for each A:

$$T(A) = \left(\frac{w_1}{w}(A_i)\nabla \ldots \nabla \frac{w_m}{w}(A_m)\right) \tag{5}$$

where A was the conditional attribute, ∇ the t-norm operator, A_i, $i = 1, 2, \ldots, m$ the linguistic terms of A which have been conjunctively combined, and W the largest one of m associated weights, $W(X, A_i)$. A t-norm (also called triangular norm) is a kind of binary operation used in the framework of probabilistic metric spaces and in multi-valued logic, specifically in fuzzy logic. A t-norm generalizes intersection in a lattice and conjunction in logic.

In a similar manner, the formula below could be used to determine the compound weight, $T(B)$, of the weighted disjunction of the linguistic terms related to variable B:

$$T(B) = \left(\frac{w_1}{w}(B_i)\Delta \ldots \Delta \frac{w_n}{w}(B_n)\right) \tag{6}$$

where Δ was the t-conorm operator and B_i, $i = 1, 2, \ldots, n$ the linguistic terms of variable B which have been combined disjunctively. T-conorms (also called S-norms) are dual to t-norms under the order-reversing operation which assigns $1 - x$ to x on [0, 1].

Step 5: This step generated the set of rules. The weighted conjunction $T(A)$ and weighted disjunction $T(B)$, which gave rise to the weightage in the Subsethood-Based Algorithm, were utilized to create the fuzzy IF-THEN rules. With reference to the initial SBA algorithm, "OR" and "AND" were deduced based on the t-conorm and t-norm operators respectively. In the process, the linguistic rules generated by the proposed WSBA could be expounded in terms of an amalgamation of Fuzzy General Rules as well as Fuzzy Quantifier Rules. Quantifiers like "some" or "all" could be used to describe the weightages of individual linguistic terms. To elaborate, "All" denoted a weightage of 1, while "some" otherwise. The interpretation of the degree of the latter depended on the weightages of each linguistic term. The execution of the FRBS utilizes these learned rules; the rule with the largest overall weightage would be regarded as the concluding classification.

C. PHASE 3: FINAL CLASSIFICATION OUTPUT

Once the rule set is obtained, the classification of CPO price rank can be performed. Then, the rules were calculated using the rule set generated and the transformed fuzzy values. In this part, the Min-Max Operator is used and the classification result is based on the highest truth-value.

D. PHASE 4: TESTING THE RULE SET FOR CLASSIFICATION TASKS

The rule set trained using CPO-1 dataset for classification of the CPO price rank was then tested using the CPO-2 dataset. The rules of forecasting were created and these rules were employed in the current phase to ascertain the forecast trend for individual CPO prices. Then, the CPO price distribution forecasting was commenced.

E. PHASE 5: TESTING FOR THE SIGNIFICANCE OF THE PROPOSED FUZZY TIME SERIES METHOD

Test statistic is tested for occurrence within either of the two critical regions on the two extremes of the distribution such as

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_1: \textit{at least one } \beta_i \neq 0$$

The test statistic used in this research is $F_0 = \frac{MS_R}{MS_E}$, where $MS_R = \frac{SSR}{k}$ and $MS_E = \frac{SSE}{n-p}$ with the rejection criteria is $F_0 = \frac{MS_R}{MS_E} > f_{\alpha,k,n-p}$.

IV. NUMERICAL EXAMPLE

The data collection was divided into two subset data, which is CPO-1 for training dataset and CPO-2 for testing dataset. Table 1 and Table 2 show the training and testing datasets respectively that have been randomly determined.

Table 3 below displays the phrases utilized to denote the linguistic terms in the entirety of the CPO price dataset. There are twelve Linguistics terms that have been defined based on the input and output that will be produced by the chosen CPO prices. Meanwhile, Table 4 shows the subgroups with respect to the classification outcomes calculated based on (2). According to classification outcomes, the training dataset was divided into three subgroups.

Figure 1 depicts the Fuzzy Membership Function of CPO price dataset. This function is used to transform CPO price data into fuzzy values. It is defined according to the criterion of CPO price ranking.

The subsethood values for each linguistic term in each subgroup were calculated and the subsethood values which were calculated according to each classification results are shown in Table 5, while Table 6 shows the weights for each linguistic term.

The three rules that have been generated are as follows:

Referring to the values in Table 1, row 1, given that one day previous, A_i: RM568.43, two days previous, B_i: RM527.50, and three days previous, C_i: RM531.21, where A_i, B_i, and C_i, $i = 1, 2, 3$ refer to the linguistic terms depicted in Table 3. These prices need to be transformed into fuzzy values, $\mu(p)$ based on the fuzzy membership function in Fig. 1. The calculations are as follows:

TABLE 2. Testing dataset (CPO-2).

Case	t-3	t-2	t-1	Y	Outcome
1	541.45	521.66	531.41	495.22	Medium
2	521.66	531.41	495.22	452.56	Low
3	531.41	495.22	452.56	450.77	Low
4	495.22	452.56	450.77	449.15	Low
5	452.56	450.77	449.15	428.88	Low
6	450.77	449.15	428.88	451.33	Low
7	449.15	428.88	451.33	456.11	Low
8	428.88	451.33	456.11	445.89	Low
9	451.33	456.11	445.89	437.60	Low
10	456.11	445.89	437.60	448.87	Low
11	445.89	437.60	448.87	455.39	Low
12	437.60	448.87	455.39	432.35	Low
13	448.87	455.39	432.35	387.01	Low
14	455.39	432.35	387.01	395.53	Low
15	432.35	387.01	395.53	417.23	Low
16	387.01	395.53	417.23	390.47	Low
17	395.53	417.23	390.47	403.56	Low
18	417.23	390.47	403.56	421.40	Low
19	390.47	403.56	421.40	475.80	Medium
20	403.56	421.40	475.80	506.72	Medium
21	421.40	475.80	506.72	560.89	High
22	475.80	506.72	560.89	542.42	High
23	506.72	560.89	542.42	518.39	Medium
24	560.89	542.42	518.39	509.16	Medium
25	542.42	518.39	509.16	548.57	High
26	518.39	509.16	548.57	571.88	High
27	509.16	548.57	571.88	530.11	Medium
28	548.57	571.88	530.11	530.96	Medium

(i) If the price, p falls in the **low region** ($p \leq 444.83$), the fuzzy value is equal to 1.

(ii) If the price, p falls in the **low region** ($444.83 < p \leq 502.66$), the fuzzy value is

$$\mu(p) = \frac{502.66 - p}{502.66 - 444.83}$$

(iii) If the price, p falls in the **medium region** ($444.83 < p \leq 502.66$), the fuzzy value is

$$\mu(p) = \frac{p - 444.83}{502.66 - 444.83}$$

(iv) If the price, p falls in the **medium region** ($502.66 < p \leq 560.49$), the fuzzy value is

$$\mu(p) = \frac{560.49 - p}{560.49 - 502.66}$$

TABLE 3. Labels used for each linguistic term in CPO price dataset.

Label	Linguistic Term
A1	One day previous is Low
A2	One day previous is Medium
A3	One day previous is High
B1	Two day previous is Low
B2	Two day previous is Medium
B3	Two day previous is High
C1	Three day previous is Low
C2	Three day previous is Medium
C3	Three day previous is High
X	Price rank is Low
Y	Price rank is Medium
Z	Price rank is High

TABLE 4. Subgroups of CPO prices with respect to the classification outcomes.

Subgroup	Cases	Outcome
Subgroup 1	7 - 13	Low
Subgroup 2	6, 14 - 19, 22, 27, 28	Medium
Subgroup 3	1 - 5, 20, 21, 23 - 26	High

(v) If the price, p falls in the **high region** ($502.66 < p \leq 560.49$), the fuzzy value is

$$\mu(p) = \frac{p - 502.66}{560.49 - 502.66}$$

(vi) If the price, p falls in the **high region** ($p \geq 560.49$), the fuzzy value is equal to 1.

(vii) Elsewhere, the fuzzy value is equal to 0.

Therefore, from the above calculations, the transformed fuzzy values are as follows:

$$\mu_{A1}(568.43) = 0, \quad \mu_{A2}(568.43) = 0, \ \mu_{A3}(568.43) = 1,$$
$$\mu_{B1}(527.50) = 0, \quad \mu_{B2}(527.50) = 0.57,$$
$$\mu_{B3}(527.50) = 0.43,$$
$$\mu_{C1}(531.21) = 0, \quad \mu_{C2}(531.21) = 0.51,$$
$$\mu_{C3}(531.21) = 0.50,$$

Then, the following rules were calculated using the rule set generated as shown in Fig. 2 and the above transformed fuzzy values. In this part, the Min-Max Operator is used.

Rule 1:

$$\begin{aligned} X &= \min[\max(0, 0, 0.04), \max(0, 0.1254, 0.1161), \\ &\qquad \max(0, 0.1275, 0.32)] \\ &= 0.04 \end{aligned}$$

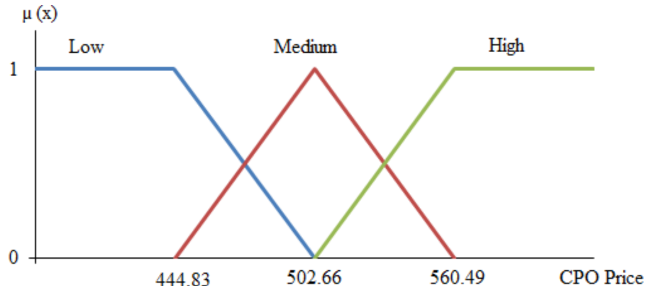

FIGURE 1. Fuzzy membership function for CPO price data.

TABLE 5. Subsethood values calculated from CPO price dataset.

Linguistic Term	Output		
	Low (X)	Medium (Y)	High (Z)
A1	0.84	0.18	0.00
A2	0.18	0.65	0.11
A3	0.03	0.50	0.89
B1	0.69	0.28	0.00
B2	0.15	0.45	0.21
B3	0.18	0.44	0.80
C1	0.53	0.41	0.00
C2	0.13	0.41	0.26
C3	0.34	0.38	0.75

TABLE 6. Weight for each linguistic term.

Linguistic Term	Output		
	Low (X)	Medium (Y)	High (Z)
A1	1.00	0.27	0.00
A2	0.21	1.00	0.12
A3	0.04	0.78	1.00
B1	1.00	0.63	0.00
B2	0.22	1.00	0.27
B3	0.27	0.99	1.00
C1	1.00	1.00	0.00
C2	0.25	1.00	0.34
C3	0.64	0.94	1.00

Rule 2:

$$Y = \min[\max(0, 0, 0.78), \max(0, 0.57, 0.4257), \max(0, 0.51, 0.47)] = 0.51$$

Rule 3:

$$Z = \min[\max(0, 0, 1), \max(0, 0.1539, 0.43), \max(0, 0.1734, 0.5)] = 0.43$$

Rule 1: The CPO Price Rank is Low (X)

IF one day previous is (A1 OR 0.21A2 OR 0.04A3) AND two days previous is (B1 OR 0.22B2 OR 0.27B3) AND three days previous is (C1 OR 0.25C2 OR 0.64C3), THEN the CPO Price Rank is Low.

Rule 2: The CPO Price Rank is Medium (Y)

IF one day previous is (0.27A1 OR A2 OR 0.78A3) AND two days previous is (0.63B1 OR B2 OR 0.99B3) AND three days previous is (C1 OR C2 OR 0.94C3), THEN the CPO Price Rank is Medium.

Rule 3: The CPO Price Rank is High (Z)

IF one day previous is (0.12A2 OR A3) AND two days previous is (0.27B2 OR B3) AND three days previous is (0.34C2 OR C3), THEN the CPO Price Rank is High.

FIGURE 2. Rule set generated.

TABLE 7. The rules of forecasting.

Type	Term and Condition
Rule 1	If the highest value of min[max(μ_{A1}×A1, μ_{A2}×0.21A2, μ_{A3}×0.04A3), max(μ_{B1}×B1, μ_{B2}×0.2B2, μ_{B3}×0.27B3), max(μ_{C1}×C1, μ_{C2}×0.25C2, μ_{C3}×0.64C3)], then the CPO prices forecasting go upward at the 0.25 point of the corresponding sub-interval.
Rule 2	If the highest value of min[max(μ_{A1}×0.27A1, μ_{A2}×A2, μ_{A3}×0.78A3), max(μ_{B1}×0.63B1, μ_{B2}×B2, μ_{B3}×0.99B3), max(μ_{C1}×C1, μ_{C2}×C2, μ_{C3}×0.94C3)], then the CPO prices forecasting be the middle of the corresponding sub-interval.
Rule 3	If the highest value of min[max(μ_{A1}×0A1, μ_{A2}×0.12A2, μ_{A3}×A3), max(μ_{B1}×0B1, μ_{B2}×0.27B2, μ_{B3}×B3), max(μ_{C1}×0C1, μ_{C2}×0.34C2, μ_{C3}×C3)], then the CPO prices forecasting go downward at the 0.75 point of the corresponding sub-interval.

Based on the rules calculated above, the classification result is Y, which is the CPO price rank is Medium because the highest truth-value is associated with Rule 2.

Next, Table 7 shows the rules of forecasting that have been generated and applied in order to determine CPO prices data forecasting trend which depends on the rules' terms and conditions.

The CPO price data trend was defined according to the classification of the rules and referring to the constructed sub-interval based on [28] as in Table 8. Then, the CPO prices was started to forecast.

V. ALGORITHM OF THE PROPOSED METHOD

By summarising the above discussion, an algorithm for the proposed method is listed as follows:

(i) Acquire and pre-process the data (divide data into two subsets: training and testing datasets).

TABLE 8. The sub-intervals.

Interval, U	Crude Palm Oil Prices			New Sub-interval, Sj
	2014	2015	2016	
u1,1	[415,424.75]	[352,365]	[409,448]	S1
u1,2	[424.75,434.5]	[365,378]	[448,467.5]	S2
u1,3	[434.5,444.25]	[378,391]	[467.5,487]	S3
u1,4	[444.25,454]	[391,400.75]	[487,496.75]	S4
u1,5	[454,473.5]	[400.75,410.5]	[496.75,506.5]	S5
u2,1	[473.5,493]	[410.5,420.25]	[506.5,516.25]	S6
u2,2	[493,506]	[420.25,430]	[516.25,526]	S7
u2,3	[506,519]	[430,437.8]	[526,533.8]	S8
u2,4	[519,532]	[437.8,445.6]	[533.8,541.6]	S9
u3,1	[532,539.8]	[445.6,453.4]	[541.6,549.4]	S10
u3,2	[539.8,547.6]	[453.4,461.2]	[549.4,557.2]	S11
u3,3	[547.6,555.4]	[461.2,469]	[557.2,565]	S12
u4,1	[555.4,563.2]	-	[565,578]	S13
u4,2	[563.2,571]	-	[578,591]	S14
u5	[571,610]	-	[591,604]	S15

(ii) Generate fuzzy rule for training dataset (CPO-1) based on Weighted Subsethood-Based Algorithm (WSBA).
(iii) Classify the output (CPO price rank) using the rule set obtained.
(iv) Test the rules set using the testing dataset (CPO-2) and determine the trend of forecasting.
(v) Use the trend to forecast CPO prices by referring to the sub-interval obtained.
(vi) The forecast values using the proposed method and Chen's Model are evaluated and the percentage of accuracy is compared between both methods.

VI. RESULTS AND DISCUSSION

In this section, the experimental results of the forecasted CPO price obtained are discussed. Table 9 illustrates the comparison of the outcomes obtained by using the previous method (Chen's Model) and those by the present approach. In order to test the rule set trained, the CPO-1 is tested for classification of CPO price dataset, whereas the CPO-2 dataset is used for testing. The differences between the forecasting methods are depicted in Fig. 3.

Referring to Fig. 3, the blue line refers to the actual price of the CPO. The dotted line refers to the forecasted value using WSBA and the red line refers to the previous forecasting method (Chen's Model). The plotted graph shows that the forecasted values of CPO prices using WSBA is very close to the actual prices of the CPO. This means that the forecasted values using fuzzy time series with the proposed WSBA is better compared to Chen's Model and performs more accurate forecasting of the CPO prices. According to the results in Table 9, it proves that by using present approach, the forecast values obtained are almost similar to the actual price of the CPO. Nevertheless, among 28 cases, there are 20 worst forecasts from Chen's Model and none from the proposed method. This indicates that the proposed method can consistently improve the forecasting results compare to Chen's Model.

TABLE 9. Comparison of CPO price forecasting between proposed method (WSBA) and Chen's model.

Cases	Actual Price	Forecast value (WSBA)	Forecast value (Chen's Model)
1	495.22	496.25	512.25
2	452.56	449.13	434.13
3	450.77	449.13	433.13
4	449.15	449.13	434.13
5	428.88	427.19	412.19
6	451.33	451.13	467.75
7	456.11	457.63	474.25
8	445.89	444.63	461.25
9	437.60	438.13	456.38
10	448.87	444.63	431.25
11	455.39	451.13	437.75
12	432.35	431.63	448.25
13	387.01	383.69	371.13
14	395.53	392.95	379.90
15	417.23	416.35	433.30
16	390.47	383.69	371.13
17	403.56	400.75	419.65
18	421.40	421.19	438.63
19	475.80	473.35	458.75
20	506.72	509.75	524.75
21	560.89	555.25	545.13
22	542.42	535.75	520.75
23	518.39	516.25	501.25
24	509.16	509.75	524.75
25	548.57	555.25	560.25
26	571.88	571.50	556.50
27	530.11	535.75	550.75
28	530.96	535.75	555.63

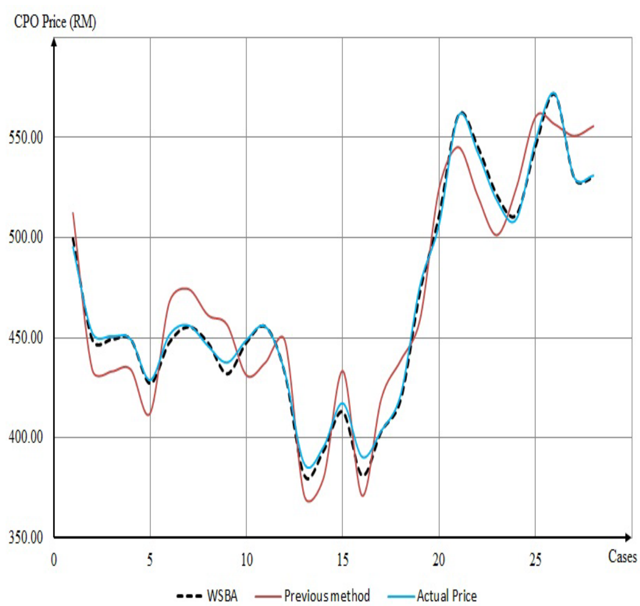

FIGURE 3. The differences in CPO price between actual and the forecasting methods.

TABLE 10. ANOVA table for the regression.

Model	Sum of Squares	*df*	Mean Square	*F*
Regression	61360.611	3	20453.537	28.517
Residual	17214.021	24	717.251	
Total	78574.632	27		

TABLE 11. Evaluation of methods.

Forecasting Method	MSE	RMSE	Accuracy of Forecasting
Proposed Method	0.02	3.28	98.43%
Chen's Model	0.65	17.39	96.28%

Table 11 below summarises the results of MSE, RMSE, and the percentage of accuracy for each forecasting method.

From Table 11, the results of MSE and RMSE for the proposed WSBA are found to be lower as compared to the Chen's Model. The MSE and RMSE results for the proposed WSBA are 0.02 and 3.28 respectively, while the MSE and RMSE results for the Chen's Model are 0.65 and 17.39 respectively. This shows that the proposed forecasting method is consistent. The accuracy of forecasting by using the proposed method is more accurate compared to the Chen's Model. Furthermore, Table 10 depicted the ANOVA table for the regression analysis to proven the efficiency of the proposed.

By taking the significance level, $\alpha = 0.025$, the value of $f_{0.025,3,24}$ from F table is 3.72. From Table 10, the value of F obtained *is* 28.517, which is greater than the value of F table. Thus, the hypothesis H_0 can be rejected. Therefore, it can be concluded that the proposed fuzzy time series method is fitted. Besides, from the Table 10 we can also determine the value of R^2, which is equal to 78.1 percent. Thus, the proposed fuzzy time series method is fitted and can be used for generating fuzzy rules in time series forecasting.

VII. CONCLUSION

This research presented the area of data driven Fuzzy Rule-Based System in the time series forecasting of CPO prices. This research showed that this new approach have several advantages that can be applied to strengthen the previous method. The development of the Weighted Subsethood-Based Algorithm (WSBA) – a preliminary data-driven FRBS – from the fuzzy subsethood values, provided ease of use because default fuzzy rules could be churned out without the requirement of threshold values. In light of the need for a system which is easily comprehensible by most people, particularly forecasters, the said algorithm could be highly practical. The actual CPO prices data were processed first by classifying the outcomes (CPO price rank). The data was taken based on historical daily dataset. Then, the number of class intervals and sub-intervals obtained from [28] in the previous work were used to test the developed WSBA. Utilising the rules created by using the WSBA, the CPO price forecasting was performed. The comparison of the CPO prices forecasting between the proposed method and Chen's Model also was conducted. This research focused on generating fuzzy rules using WSBA systematically rather than judgment by human expertise. Two types of statistical indicators – Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) – had been employed to assess the said model. As mentioned in the result, it can be seen that the MSE and RMSE results for the forecasting values of CPO prices using the proposed method were the smallest as compared to Chen's Model. The proposed method produced new fuzzy rules, which carried new information, alongside with the previous rules obtained in Chen's Model. As such, the rules generated by people or automated systems could be validated or rejected via this model. It can be summarized that the proposed method produced more effective and accurate forecasting compared to Chen's Model. Therefore, the application of this method could facilitate the development of systematic forms of forecasting application methods. Extensive research had been done in solving the algorithms of finding the best rule from training data. Embedding the rules in the proposed forecasting method and trying to produce good results were tedious work. To extend this research, an improve in rebuilding of membership functions from numerous training set data taken at random in generating the fuzzy rules are suggested.

REFERENCES

[1] Z. Mohamed and F. M. Arshad, "Price discovery through crude palm oil," *Int. Manag. Dev. Assoc. Univ. Sains Malaysia*, pp. 73–92, Jun. 1994.

[2] A. H. F. Fizari *et al.*, "A time series analysis of the relationship between total area planted, palm oil price and production of Malaysian palm oil," *World Appl. Sci. J.*, vol. 12(BES), pp. 34–40, 2011.

[3] M. K. Lam, K. T. Tan, K. T. Lee, and A. R. Mohamed, "Malaysian palm oil: Surviving the food versus fuel dispute for a sustainable future," *Renew. Sustain. Energy Rev.*, vol. 13, nos. 6–7, pp. 1456–1464, 2008.

[4] B. A. Talib and Z. Darawi, "An economic analysis of the malaysian palm oil market," *Oil Palm Ind. Econ. J.*, vol. 2, no. 1, pp. 19–27, 2014.

[5] Y. Basiron, "Palm oil and its global supply and demand prospects," *Oil Palm Ind. Econ. J.*, vol. 2, no. 1, pp. 1–10, 2002.

[6] D. Bigman, "Futures market efficiency and the time content of the information sets," *J. Futures Markets*, vol. 3, no. 3, pp. 321–334, 1983.

[7] R. Abdullah, "Short term and long term projection of Malaysian palm oil production," *Oil Palm Ind. Econ. J.*, vol. 3, no. 1, pp. 32–36, 2015.

[8] S. Ahmad and H. A. Latif, "Forecasting on the crude palm oil and kernel palm production: Seasonal ARIMA approach," in *Proc. IEEE Colloqium Humanities Sci. Eng. Res.*, Dec. 2011, pp. 939–944.

[9] S. D. Applanaidu, F. Mohamed Arshad, M. N. Shamsudin, and A. A. Abdel Hameed, "An econometric analysis of the link between biodiesel demand and malaysian palm oil market," *Int. J. Bus. Manage.*, vol. 6, no. 2, pp. 35–45, 2011.

[10] M. H. Ahmad, P. Y. Ping, and N. Mahamed, "Volatility modelling and forecasting of Malaysian crude palm oil prices," *Appl. Math. Sci.*, vol. 8, no. 124, pp. 6159–6169, 2014.

[11] M. Khashei, M. Bijari, and G. A. R. Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," *Neurocomputing*, vol. 72, nos. 4–6, pp. 956–967, 2009.

[12] O. Valenzuela *et al.*, "Hybridization of intelligent techniques and ARIMA models for time series prediction," *Fuzzy Sets Syst.*, vol. 159, no. 7, pp. 821–845, 2008.

[13] A. A. Karia, I. Bujang, and I. Ahmad, "Forecasting on crude palm oil prices using artificial intelligence approaches," *Amer. J. Oper. Res.*, vol. 3, no. 2, pp. 259–267, 2013.
[14] A. Aziz and I. Bujang, "Progress accuracy of CPO price prediction: Evidence from ARMA family and artificial neural network approach," *Int. Res. J. Financ. Econ.*, no. 64, pp. 66–79, 2011.
[15] G. Zhang, M. Y. Hu, B. E. Patuwo, and D. C. Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," *Eur. J. Oper. Res.*, vol. 116, no. 1, pp. 16–32, 1999.
[16] M. N. Shamsudin and F. M. Arshad, "Short term forecasting of Malaysian crude palm oil prices," Faculty Econ. Manage., UPM, Selangor, Malaysia, Tech. Rep., Jan 1999, pp. 1–12.
[17] M.-Y. Chen and B.-T. Chen, "A hybrid fuzzy time series model based on granular computing for stock price forecasting," *Inf. Sci.*, vol. 294, pp. 227–241, Feb. 2015.
[18] M.-Y. Chen and B.-T. Chen, "Online fuzzy time series analysis based on entropy discretization and a fast Fourier transform," *Appl. Soft Comput.*, vol. 14, pp. 156–166, Jan. 2014.
[19] M.-Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," *Futur. Gener. Comput. Syst.*, vol. 37, pp. 461–467, Jul. 2014.
[20] O. Cordon, F. Herrera, and I. Zwir, "Linguistic modeling by hierarchical systems of linguistic rules," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 1, pp. 2–20, Feb. 2002.
[21] A. Khaliq and A. Ahmad, "Fuzzy logic and approximate reasoning," Dept. Math. Sci., Blekinge Inst. Technol., Sweden, Tech. Rep., 2010, pp. 1–72.
[22] E. H. Ruspini, P. Bonissone, W. Pedrycz, *Handbook of fuzzy computation*, 1st ed. Oxford, U.K.: Inst. Phys. Publ. Oxford Univ. Press, Jan. 1998, pp. 1–24.
[23] K. A. Rasmani and Q. Shen, "Data-driven fuzzy rule generation and its application for student academic performance evaluation," *Appl. Intell.*, vol. 25, no. 3, pp. 305–319, 2006.
[24] S.-M. Chen and F.-M. Tsai, "Generating fuzzy rules from training instances for fuzzy classification systems," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 611–621, 2008.
[25] R. Mansor, M. M. Kasim, and M. Othman, "The framework of weighted subset-hood Mamdani fuzzy rule Based system rule extraction (MFRBS-WSBA) for forecasting electricity load demand," in *Proc. AIP Conf.*, 2016, vol. 1761, no. 1, p. 020061.
[26] A. González, R. Pérez, Y. Caises, and E. Leyva, "An efficient inductive genetic learning algorithm for fuzzy relational rules," *Int. J. Comput. Intell. Syst.*, vol. 5, no. 2, pp. 212–230, 2012.
[27] Y.-C. Tsai, C.-H. Cheng, and J.-R. Chang, "Entropy-based fuzzy rough classification approach for extracting classification rules," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 436–443, 2006.
[28] K. Huarng and T. H.-K. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 36, no. 2, pp. 328–340, Apr. 2006.

NUR FAZLIANA RAHIM was born in Kelantan, Malaysia, in 1987. She received the B.Sc. degree (Hons.) in management mathematics from Universiti Teknologi MARA, Perlis, Malaysia, in 2009, and the M.Ed. degree in mathematics from Universiti Pendidikan Sultan Idris, Perak, Malaysia, in 2011. She is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS, Perak, Malaysia. Her current research interests include forecasting crude palm oil prices, a field of fuzzy time series forecasting.

MAHMOD OTHMAN was an Associate Professor with Universiti Teknologi MARA, Perlis, Malaysia. He has a total of 23 years of working experience in education industries as an Academician, where 21 years with Universiti Teknologi MARA and another two years with Universiti Teknologi PETRONAS, Perak, until present. He has been an Associate Professor with Universiti Teknologi PETRONAS since 2016. His research interests include the fields of fuzzy mathematics, artificial intelligent, optimization, and decision making.

RAJALINGAM SOKKALINGAM received the Ph.D. degree in mathematics. He is currently a Senior Lecturer with Universiti Teknologi PETRONAS, Perak, Malaysia. He has total of 23 years of working experience, where 10 years in manufacturing industries in engineering field and another 13 years in education industries as an Academician. During his employment, he has made contributions to undergraduate and postgraduate teaching. The subjects that he has taught range from the first year undergraduates (mathematics and statistics) up to the postgraduate level (risk, project, and cost management). He managed to present his research in a number of national and international conferences, and published some results national and international journals. He possesses good experience in computer laboratory sessions and very familiar with some related mathematics and statistical software, such as MATLAB, Maple, and SPSS.

EVIZAL ABDUL KADIR received the M.Eng. and Ph.D. degrees in wireless communication from the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia, in 2008 and 2014, respectively. He is currently a Senior Lecturer with Universitas Islam Riau, Indonesia. He has been involved in several companies that provide system solution in telecommunication and radio frequency identification (RFID). He is currently continuing his research activity related to the wireless and mobile communication, computer networking and information system, RFID, and wireless sensor network (WSN). His research interest include in the field of antenna, RF, smart system, RFID, WSN, computer networking, and intelligent system.