Modeling of red onion production in Central Java using hybrid ARIMA-ANFIS

I H Diarsih¹, Tarno¹ and A Rusgiyono¹

¹Department of Statistics, Diponegoro University
Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia
E-mail: inashusnad.26@gmail.com

Abstract. Red onion is one of the strategic horticulture commodities in Indonesia considering its function as the main ingredients of the basic ingredients of Indonesian food. For increasing production to supply national necessary, Central Java as the main center of red onion production should be able to predict the production of several periods to maintain the balance of national production. The purpose of this research is to get the best model to forecast the production of red onion in Central Java by ARIMA, ANFIS, and hybrid ARIMA-ANFIS method. The smallest RMSE and AIC values measure model accuracy. The results show that the best model for modeling red onion production in Central Java is obtained by hybrid ARIMA - ANFIS model which is a combination between SARIMA ([2], 1, [12]) and residual ARIMA using ANFIS model with input $\epsilon_{t,1}$, $\epsilon_{t,2}$ on the grid partition technique, gbell membership function, and membership number of 2 that produce RMSE 12033 and AIC 21.6634. While ARIMA model yield RMSE 13301.24 and AIC 21.89807 with violation of assumption. And the ANFIS model produces RMSE 14832 and AIC 22.0777. It shows that ARIMA-ANFIS hybrid method is better than ARIMA and ANFIS.

1. Introduction

The agriculture sector has an important role in Indonesia’s economic growth. As a sector that provides food and industrial raw material, the public’s demand in this factor will continue to increase in line with the increase of population, needs, and public consumption. The increasing demand causes the government to plan a special effort to increase food production.

One of the seven national strategic commodities in special effort to increase the production is red onion, due to its function as the basic ingredients of Indonesian food. Four provinces in Indonesia are the central of red onion production; one of them is Central Java. Thus, Central Java must be able to predict the production for several periods ahead to have a clear image during the planning of special effort. Which is needed to keep a balance in the production whether in fulfilling its own needs or other regions in Indonesia?

One of the commonly used traditional prediction methods is Autoregressive Integrated Moving Average (ARIMA). The downsides of this method are a stationary assumption and the difficulty on order determination (p,d,q) [1]. However, along with the development of science, there is one method from an expert system called Adaptive Neuro-Fuzzy Inference System (ANFIS), which is a compilation between Artificial Neural Network (ANN) and Fuzzy Inference System (FIS) [2]. The use of ANFIS method does not need an independent assumption, homoscedasticity, and residual distributed normal does not often seen in the data. Thus, this method is suitable to predict the data that has extreme value [1].
In this research, the hybrid ARIMA-ANFIS method will be used to create modeling of red onion production in Central Java. The purpose of this research is to determine the best prediction method between ARIMA, ANFIS, and hybrid ARIMA-ANFIS in red onion’s production modeling in Central Java.

2. Literature review

2.1 Autoregressive Integrated Moving Average (ARIMA)
ARIMA model \((p,d,q)\) is the result of a merging between non-stationary processes that have been stationed using \(p\) and \(q\) states that the order from autoregressive and moving average and \(d\) are the order from differencing. The common form from ARIMA model \((p,d,q)\) can be defined as follow [3]

\[
\phi_p(B)(1-B)^dZ_t = \theta_q(B)a_t
\]

where \(\phi_p(B) = 1 - \phi_1B - \phi_2B^2 - \ldots - \phi_pB^p,\)
\[
\theta_q(B) = 1 - \theta_1B - \theta_2B^2 - \ldots - \theta_qB^q.
\]

The model that will be used if the data contains the seasonal patterns is Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Generally, the form of Seasonal Autoregressive Integrated Moving Average (SARIMA \((p,d,q)(P,D,Q)^S\)) model is

\[
\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^DZ_t = \theta_q(B)\Theta_Q(B^S)a_t
\]

where \(\Phi_P(B^S) = 1 - \Phi_1B^S - \Phi_2B^{2S} - \ldots - \Phi_PB^{pS},\)
\[
\Theta_Q(B^S) = 1 - \Theta_1B^S - \Theta_2B^{2S} - \ldots - \Theta_QB^{qS}.
\]

According to the equation (2), the subset ARIMA model can also be formulated. Subset ARIMA model is a part of the generalized ARIMA model. Thus it cannot be stated in general form. The example of subset ARIMA model \(([1,5][0,1],[12])\) can be written as follows

\[
(1 - \phi_1B - \phi_2B^5)Z_t = (1 - \theta_1B - \theta_2B^{12})a_t
\]

Therefore the subset ARIMA model is an ARIMA model with some of its parameters equal with zero [4].

During the determination of ARIMA Box-Jenkins model that consists of seven steps is used, which are identification, parameter estimation, diagnostic checking, and forecasting.

2.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)
ANFIS is an architecture that functionally the same with fuzzy rule base Sugeno model and is a method that during rules setting is using learning algorithm towards a set of data and also enable the rules to adapt [5].

For simplicity, if two inputs consist of \(Z_{t,1}\) is \(Z\) at \(t-1\) the time and \(Z_{t,2}\) is \(Z\) at \(t-2\) the time and output \(\hat{Z}_t\), therefore according to the assumption on the rule-base of Sugeno order-one there are two rules, as follows

If \(Z_{t,1}\) is \(A_1\) dan \(Z_{t,2}\) is \(B_1\) then \(Z_t^{(1)} = p_1Z_{t,1} + q_1Z_{t,2} + r_1\)

If \(Z_{t,1}\) is \(A_2\) dan \(Z_{t,2}\) is \(B_2\) then \(Z_t^{(2)} = p_2Z_{t,1} + q_2Z_{t,2} + r_2\)

Where \(Z_{t,1}\) is \(A_1\) and \(Z_{t,2}\) is \(B_1\); \(Z_{t,1}\) is \(A_2\) and \(Z_{t,2}\) is \(B_2\) being called as the premise (nonlinear) section, while \(Z_t^{(1)} = p_1Z_{t,1} + q_1Z_{t,2} + r_1\) and \(Z_t^{(2)} = p_2Z_{t,1} + q_2Z_{t,2} + r_2\) are called consequent (linear) section. \(p_1, p_2, q_1, q_2, r_1, r_2\) are linear parameters and \(A_1, B_1, A_2, B_2\) are called non-linear.
If the firing strength of each rule is \( w_1 \) and \( w_2 \) for two values \( Z_t^{(1)} \) and \( Z_t^{(2)} \), then the output \( \hat{Z}_t \) is computed as weighted mean

\[
\hat{Z}_t = \frac{w_1 Z_t^{(1)} + w_2 Z_t^{(2)}}{w_1 + w_2} = \overline{w}_1 Z_t^{(1)} + \overline{w}_2 Z_t^{(2)}
\]

(4)

The architecture of ANFIS with the base rule above could be seen in Figure 1 [6].

ANFIS model as proposed by Jang [2] has five layers feed forward, which each layer is described as bellow.

Layer 1, every node in the first layer is adaptive with one parametric activation function. The output is membership degree of given inputs which satisfy membership function \( \mu_{A_i}(Z_{t,1}) \) and \( \mu_{B_i}(Z_{t,2}) \). One example of the fuzzy membership function is generalized bell (bell),

\[
\mu(Z_{t,i}) = \frac{1}{1 + \left| \frac{Z - c_{i}}{a} \right|^{2b}}, \quad i = 1,2
\]

Where \( Z \) is input and \( \{a_i, b_i, c_i\} \) is the parameter set of this node, parameters in this layer are called premise parameters.

Layer 2, every node in the second layer is fixed node labeled \( \Pi \). The output of each node represents the firing strength \( w_i \) of \( i \)-th rule.

\[
w_i = \mu_{A_i}(Z_{t,1}) \mu_{B_i}(Z_{t,2}), \quad i = 1,2
\]

(5)

Layer 3, every node in the third layer is fixed node labeled \( N \). The \( i \)-th node calculates the ratio of the \( i \)-th rule’s firing strength to the sum of all rules’ firing strengths,

\[
\overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2
\]

(6)

The output of this layer is called normalized firing strengths.

Layer 4, every node in the fourth layer is an adaptive node, the output of each node is

\[
\overline{w}_i Z_t^{(i)} = \overline{w}_i \left(p_i Z_{t,1} + q_i Z_{t,2} + r_i \right)
\]

(7)
\( \bar{w}_i \) is a normalized firing strength from layer 3 and \( \{p_i, q_i, r_i\} \) is the parameter set of this node. Parameters in this layer are called consequent parameters.

Layer 5, every node in the fifth layer is a fixed node labeled \( \Sigma \), which computes the overall output as the summation of all incoming signal. The output of the fifth layer is the output of the whole network.

\[
\hat{Z}_t = \sum_{i=1}^{3} \bar{w}_i Z_t^{(i)} \\
\hat{Z}_t = \sum_{i=1}^{3} \bar{w}_i (p_i Z_{t,1} + q_i Z_{t,2} + r_i) \\
\hat{Z}_t = \bar{w}_1 (p_1 Z_{t,1} + q_1 Z_{t,2} + r_1) + \bar{w}_2 (p_2 Z_{t,1} + q_2 Z_{t,2} + r_2) \\
\hat{Z}_t = p_1 (\bar{w}_1 Z_{t,1}) + q_1 (\bar{w}_2 Z_{t,2}) + r_1 \bar{w}_1 + p_2 (\bar{w}_2 Z_{t,2}) + q_2 (\bar{w}_2 Z_{t,2}) + r_2 \bar{w}_2
\]

(8)

ANFIS apply the hybrid learning algorithms in the system to identify parameters. In the forward pass of the hybrid learning algorithm, node outputs go forward until layer 4, and the consequent parameters are identified by the least squares method. In the backward pass, the error signals propagate backward, and the premise parameters are updated by gradient descent.

2.3 Hybrid ARIMA-ANFIS

The hybrid model is a combined method of one or more model in function of a system. According to Zhang [7], generally, time series of the combined model that possess linear and non-linear structure can be written as follows

\[
Z_t = L_t + N_t
\]

(9)

Where \( L_t \) shows linear component and \( N_t \) shows the non-linear component. ARIMA model is used to solve linear cases, where residual from linear model still contains information on the non-linear connection. Mathematically, it could be written as follows

\[
e_t = Z_t - \hat{L}_t
\]

(10)

Where \( \hat{L}_t \) is the prediction value of ARIMA at \( t \) time and \( Z_t \) is data at \( t \) time. Then, to model residuals from ARIMA model using ANFIS, the non-linear connection could be resolved. ANFIS prediction result then combined with prediction result using ARIMA method. Mathematically, the overall prediction result obtained are as follows

\[
\hat{Z}_t = \hat{L}_t + \hat{N}_t
\]

(11)

where \( \hat{N}_t \) is a prediction from equation (10) and \( \hat{Z}_t \) is a prediction from the combined value of prediction from ARIMA model and ANFIS model.

2.4 Accuracy measurement

Many statistical measures can be used as an accurate measurement of forecasting, and this research performance of the model can be evaluated by using Akaike’s information criterion (AIC) and root mean square error (RMSE). AIC is used to select the best model by considering the amount of parameter in a model where the smaller the AIC value, the better the model is, and it is adequate to use. AIC could be defined as follows

\[
AIC(M) = -2 \ln(\text{max likelihood}) + 2M
\]

(12)

with

\[
\ln \hat{L} = -\frac{n}{2} \ln \hat{\sigma}_a^2 - \frac{n}{2} (1 + \ln 2\pi)
\]

(13)
Where $M$ is a number of the parameter from the ARIMA model and $n$ is a number of observation \[8\]. While RMSE is used to selecting the best model in prediction is also considered by errors produced.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} = \sqrt{\frac{\sum_{t=1}^{n} (Z_t - \hat{Z}_t)^2}{n}} \quad (14)$$

with $e_t$ is residual on period t, $Z_t$ is actual data on period t, and $\hat{Z}_t$ is prediction data result on period t.

3. Methodology
The data used in this research is secondary data regarding red onion production in Central Java on 2006-2016 collected from Central Java’s Department of Agriculture and Plantation. Data variable used in this research is red onion production data in Central Java with the monthly period from January 2006 until December 2016, as much as 132 data.

Modeling of red onion production is performed with Eviews8 and supported with Minitab14 and Matlab R2009a. The steps of data analysis on prediction with hybrid ARIMA-ANFIS model begin with linear modeling and forecasting using ARIMA to get the best ARIMA model. Then non-linear modeling using ANFIS that used input variable is based on AR lag from the best ARIMA model residual. The next step is determining the membership degree which is 2 and 3, also the membership function what will be used; Gauss, Generalized Bell, Triangle, and Trapezoidal, and clustering method that be used; grid partition and Fuzzy Cluster Mean (FCM), and determining a number of the epoch. And then running the fuzzy inference system parameter training. After we got ANFIS model then combine it with ARIMA model that we used, to acquire hybrid ARIMA-ANFIS model. The performance of the model can be viewed based on the value of AIC and RMSE.

4. Results and discussion
4.1 Data analysis using Autoregressive Integrated Moving Average (ARIMA)
Model identification is the first step during the analysis using ARIMA, which is by creating a time series plot, ACF, and PACF.

Refer to the Figure 2, it is shown that the red onion’s production data in Central Java during January 2006 until December 2016 is not yet stationer in mean, due to the upwards trend patterned and it also can be seen from the unstable plot distribution and very fluctuate.
After the visual and formal assumption has been fulfilled, thus, according to Figure 5, it is assumed that it is the time series model by observing ACF and PACF plot. ACF plot shows cut off on lag 1, 11 and seasonal lag that is lag 12. Meanwhile, PACF plot shows cut off on lag 1, 2, 3, 11, and 23. From
the expected model, it is obtained that the best model is SARIMA ([1,2,3],[12]) with the smallest rate of RMSE and AIC. It can be seen in Equation (15).

$$Z_t = 0.385Z_{t-4} + 0.079Z_{t-2} + 0.165Z_{t-3} + 0.371Z_{t-4} + a_t + 0.884a_{t-12} \quad (15)$$

### Table 1. Summary of the best arima model analysis

| Testing                              | Result                  |
|--------------------------------------|-------------------------|
| Parameter significance test          | $\phi_1, \phi_2, \phi_3, \theta_{12}$ significant |
| Residual independent test            | Fulfilled               |
| Residual normality test              | Fulfilled               |
| Residual homoscedasticity test       | Unfulfilled             |
| RMSE                                 | 13301.24                |
| AIC                                  | 21.89807                |

Table 1 represents that SARIMA ([1,2,3],[1,12]) cannot be used to describe the red onion production in Central Java because it cannot fulfill the residual homoscedasticity assumption. However, it also can be caused as the effect of ARCH/GARCH on the residual.

4.2 Data Analysis with Hybrid ARIMA-ANFIS

According to ARIMA analysis, the best model obtained is SARIMA ([1,2,3],[1,12]) because the model has sufficient parameter. Thus, it can be formed based on parsimony principal which mentions that the simpler model is more preferred rather than the one with high parameter [3], that is SARIMA ([2,1],[12]). Thus, the SARIMA model is formed as below

$$Z_t = Z_{t-4} + \phi_1 Z_{t-2} - \phi_2 Z_{t-3} + a_t - \theta_{12} a_{t-12}$$

PACF plot of residual as a result of SARIMA ([2,1],[12]) is used to observe the cut off lag. According to Figure 6, it can be seen that there are cut off lag on lag 1, 2, and 3 in PACF plot. Therefore, several combinations are applied from lag 1, 2, and 3 as an input in the ANFIS model to predict the residual of SARIMA ([2,1],[12]).

![Figure 6. PACF Plot of The Residual of SARIMA ([2,1][12])](image)

After had examined the input that will be used, then the functions that will be applied are Gauss, generalize bell (bell), triangle, and trapezoidal with membership degree as much as 2 or 3 with two clustering technique, there are Fuzzy C-Means (FCM) and grid partition. The result of both techniques can be seen as below.

In FCM, the membership function used is the default function, Gauss. The number of rules in this technique is equal to the membership degree. Therefore, there is no combination in forming the rules.

### Table 2. RMSE and AIC of hybrid ARIMA-ANFIS with FCM

| Input | MF | Membership Degree of 2 | Membership Degree of 3 |
|-------|----|------------------------|------------------------|
|       |    | RMSE  | AIC  | RMSE  | AIC  |
|       |    |       |     |       |     |
According to Table 2 and parsimony principle, it can be concluded that the best ARIMA-ANFIS hybrid model contains of membership degree of 2, membership function of Gauss with input $e_{t,1}$, $e_{t,2}$, and $e_{t,3}$.

In grid partition, the membership function applied is Gauss, generalized bell (gbell), triangle and trapezoidal. Each rule formed in this technique is combination of many partitions in each input.

### Table 3. RMSE and AIC of Hybrid ARIMA-ANFIS with Grid Partition

| Input   | MF    | Membership Degree of 2 | Membership Degree of 3 |
|---------|-------|------------------------|------------------------|
|         |       | RMSE                   | AIC                    | RMSE                   | AIC                    |
| $e_{t,1}$ |       | 13625                  | 21,8944                | 13309                  | 21,8475                |
| $e_{t,2}$ |       | 16065                  | 22,2240                | 15418                  | 22,1419                |
| $e_{t,3}$ |       | 16001                  | 22,2162                | 15695                  | 22,1776                |
| $e_{t,1}, e_{t,2}$ | Gauss | 12104                  | 21,6752                | 11709                  | 21,6088                |
| $e_{t,1}, e_{t,3}$ |       | 13435                  | 21,8842                | 12380                  | 21,7206                |
| $e_{t,2}, e_{t,3}$ |       | 15035                  | 22,1093                | 14180                  | 21,9922                |
| $e_{t,1}, e_{t,2}, e_{t,3}$ |       | 12921                  | 21,8238                | 35491                  | 23,8446                |
| $e_{t,1}$ | Gbell  | 13659                  | 21,8995                | 13207                  | 21,8322                |
| $e_{t,2}$ |       | 15929                  | 22,2071                | 15352                  | 22,1332                |
| $e_{t,3}$ |       | 15992                  | 22,2151                | 15629                  | 22,1692                |
| $e_{t,1}, e_{t,2}$ | Gbell | 12033*                 | 21,6634*               | 11846                  | 21,6322                |
| $e_{t,1}, e_{t,3}$ |       | 13414                  | 21,8810                | 12290                  | 21,7060                |
| $e_{t,2}, e_{t,3}$ |       | 14879                  | 22,0884                | 14059                  | 21,9749                |
| $e_{t,1}, e_{t,2}, e_{t,3}$ |       | 11545                  | 21,5985                | 40142                  | 24,0909                |
| $e_{t,1}$ | Triangle | 13602                  | 21,8910                | 13461                  | 21,8702                |
| $e_{t,2}$ |       | 16073                  | 22,2251                | 15564                  | 22,1607                |
| $e_{t,3}$ |       | 16004                  | 22,2166                | 15752                  | 22,1849                |
| $e_{t,1}, e_{t,2}$ | Triangle | 12542                  | 21,7464                | 25161                  | 23,1388                |
| $e_{t,1}, e_{t,3}$ |       | 13081                  | 21,8308                | 16577                  | 22,3045                |
| $e_{t,2}, e_{t,3}$ |       | 15626                  | 22,1864                | 16116                  | 22,2481                |
| $e_{t,1}, e_{t,2}, e_{t,3}$ |       | 24668                  | 23,1170                | 80321                  | 25,4781                |
| $e_{t,1}$ | Trapezoidal | 13777                  | 21,9166                | 13191                  | 21,8297                |
| $e_{t,2}$ |       | 16072                  | 22,2249                | 15325                  | 22,1297                |
| $e_{t,3}$ |       | 15943                  | 22,2089                | 15651                  | 22,1720                |
| $e_{t,1}, e_{t,2}$ | Trapezoidal | 12696                  | 21,7707                | 12056                  | 21,6673                |
| $e_{t,1}, e_{t,3}$ |       | 13459                  | 21,8878                | 12709                  | 21,7730                |
| $e_{t,2}, e_{t,3}$ |       | 15315                  | 22,1461                | 14729                  | 22,0682                |
| $e_{t,1}, e_{t,2}, e_{t,3}$ |       | 11646                  | 21,6160                | 10204                  | 21,3516                |

Based on Table 3 and parsimony principle, it can be concluded that the best ARIMA-ANFIS hybrid model contains of membership degree of 2, membership function of gbell with input $e_{t,1}$, $e_{t,2}$.

According to Table 2 and 3, a comparison is made between the hybrid ARIMA-ANFIS with FCM and grid partition technique.

### Table 4. Comparison between RMSE and AIC of Hybrid ARIMA-ANFIS
Based on Table 4, it can be concluded that the best ARIMA-ANFIS hybrid model in this research is by applying grid partition technique with the input $e_{t,1}, e_{t,2}$ of gbell membership function and the membership degree is two.

According to the best hybrid ARIMA-ANFIS model, it can be seen the comparison between the target data and output data in Figure 7.

![Figure 7. Comparison between the target and output on the best hybrid ARIMA-ANFIS](image)

The best hybrid ARIMA-ANFIS model applies a grid partition technique with input as much as 2 and the membership degree as much as 2. Therefore, it is formed 4 rules as seen below.

If $e_{t,1}$ is $A_1$ dan $e_{t,2}$ is $B_1$ then $e_t^{(1)} = p_1 e_{t,1} + q_1 e_{t,2} + r_1$

If $e_{t,1}$ is $A_1$ dan $e_{t,2}$ is $B_2$ then $e_t^{(2)} = p_2 e_{t,1} + q_2 e_{t,2} + r_2$

If $e_{t,1}$ is $A_2$ dan $e_{t,2}$ is $B_1$ then $e_t^{(3)} = p_3 e_{t,1} + q_3 e_{t,2} + r_3$

If $e_{t,1}$ is $A_1$ dan $e_{t,2}$ is $B_2$ then $e_t^{(4)} = p_4 e_{t,1} + q_4 e_{t,2} + r_4$

Where $A_1, B_1, A_2, B_2$ are called premise or non-linear parameters. Meanwhile $p_1, p_2, ..., p_4, q_1, q_2, ..., q_4$ and $r_1, r_2, ..., r_4$ are called consequent or linear parameters.

On the layer 1 in ANFIS architecture, there are 4 groups of nonlinear parameter. The early value of the parameter will be used for learning until the expected iteration. In this research, the iteration used is as much as 50. The output resulted in layer 1 is membership function in every input in the whole fuzzy set: $\mu_{A_1}(e_{t,1}), \mu_{A_2}(e_{t,1}), \mu_{B_1}(e_{t,2})$ and $\mu_{B_2}(e_{t,2})$. The value of membership function will be used as an input in layer 2. It is also used to produce firing strength in each rule. On the best ANFIS model, it has 4 rules. Therefore output in layer 2 is $w_1, w_2, ..., w_4$. The value of $w_i$ earned on layer 2 will be applied as an input on layer 3 in order to create normalized firing strength of fuzzy rule and output that will be obtained as $\bar{w}_1, \bar{w}_2, ..., \bar{w}_4$. The output on layer 3 is functioned in defuzzification process on layer 4. However, it will occur when the premise or nonlinear parameter has been repaired backward from layer 5 to layer 1 with gradient descent in the best ANFIS model learning. Besides the premise or nonlinear parameter, it will also be obtained the consequent or linear parameter as a result of Least Square Estimator (LSE) Recursive. Consequent parameters can be observed in linear form in each rule as below:

$$e_t^{(i)} = -45.86 e_{t,1} - 1.837 e_{t,2} - 935400$$
\[ e_t^{(2)} = 4.686e_{t,1} + 5.88e_{t,2} - 30800 \]
\[ e_t^{(3)} = 4.27e_{t,1} - 0.8446e_{t,2} + 229600 \]
\[ e_t^{(4)} = -1.86e_{t,1} - 3.29e_{t,2} + 35170 \]

According to the formed consequent parameter, the output on layer 4 is \( \overline{w}_i e_{t}^{(i)}, \overline{w}_2 e_{t}^{(2)}, \ldots, \overline{w}_4 e_{t}^{(4)} \). On layer 5, the output from layer 4 will be added up and will be obtained this value as below:

\[
\hat{e}_t = \sum_{i=1}^{4} \overline{w}_i e_{t}^{(i)}
\]
\[
\hat{e}_t = \sum_{i=1}^{4} \overline{w}_i \left( p_i e_{t,1} + q_i e_{t,2} + r_i \right)
\]
\[
\hat{e}_t = \overline{w}_1 (-45.86e_{t,1} - 1.837e_{t,2} - 935400) + \overline{w}_2 (4.686e_{t,1} + 5.88e_{t,2} - 30800) + \overline{w}_3 (-4.27e_{t,1} - 0.8446e_{t,2} + 229600) + \overline{w}_4 (-1.86e_{t,1} - 3.29e_{t,2} + 35170)
\]

with

\[
\overline{w}_i = \frac{w_i}{\sum_{i=1}^{4} w_i}
\]

So that, the hybrid ARIMA-ANFIS that is formed is a is a combination of the SARIMA model ([2],[1],[12]) and ARIMA residual by using ANFIS model. It can be written mathematically as below.

\[
\hat{Z}_t = \hat{L}_t + \hat{N}_t
\]

with

\[
\hat{L}_t = Z_{t-1} - 0.202Z_{t-2} + 0.202Z_{t-3} + a_t + 0.887a_{t-12}
\]
\[
\hat{N}_t = \overline{w}_1 (-45.86e_{t,1} - 1.837e_{t,2} - 935400) + \overline{w}_2 (4.686e_{t,1} + 5.88e_{t,2} - 30800) + \overline{w}_3 (-4.27e_{t,1} - 0.8446e_{t,2} + 229600) + \overline{w}_4 (-1.86e_{t,1} - 3.29e_{t,2} + 35170)
\]

4.3 The Selection of the Best Model
The selection of the best model between ARIMA, ANFIS, and ARIMA-ANFIS hybrid is obtained based on the smallest AIC and RMSE. The comparison of AIC and RMSE value on ARIMA, ANFIS, and ARIMA-ANFIS hybrid can be seen in Table 5.

| Model          | AIC    | RMSE  |
|----------------|--------|-------|
| ARIMA          | 21,89807 | 13301,24 |
| ANFIS          | 22,0777  | 14832 |
| Hybrid ARIMA-ANFIS | 21,6634 | 12033 |

Based on Table 5, it is obtained that the RMSE and AIC value of ARIMA-ANFIS hybrid is smaller than ARIMA and ANFIS. Thus, it can be concluded that the analysis using ARIMA-ANFIS hybrid is better than ARIMA and ANFIS in this research.

5. Conclusion

Based on the results and discussion, the best model to represent red onion production in Central Java is ARIMA-ANFIS hybrid. This model is a combination of SARIMA model ([2],1,[12]) and ARIMA residual by using ANFIS model with grid partition technique, the membership function of gbell, input $e_{t,1}$, $e_{t,2}$ and the membership degree as much as 2. The equation can be seen as below:

$$\hat{Z}_t = \hat{L}_t + \hat{N}_t$$

with

$$\hat{L}_t = Z_{t-1} - 0.202Z_{t-2} + 0.202Z_{t-3} + a_t + 0.887a_{t-12}$$

$$\hat{N}_t = \bar{w}_1(-45.86e_{t,1} - 1.837e_{t,2} - 935400) + \bar{w}_2(4.686e_{t,1} + 5.88e_{t,2} - 30800) + \bar{w}_3(-4.27e_{t,1} - 0.8446e_{t,2} + 229600) + \bar{w}_4(-1.86e_{t,1} - 3.29e_{t,2} + 35170)$$

References

[1] Faulina R and Suhartono 2013 International Journal of Science and Research (IJSR) 2 159.
[2] Jang JSR, Sun CT, and Mizutani E 1997 Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence (London: Prantice-Hall, Inc).
[3] Soejoeti Z 1987 Materi Pokok Analisis Runtun Waktu (Jakarta: Karunika).
[4] Tarno 2013 Kombinasi Prosedur Permodelan Subset ARIMA dan Deteksi Outlier untuk Prediksi Data Runtun Waktu Prosiding Seminar Nasional Statistika UNDIP Semarang ISBN 978-602-14387-0-1.
[5] Kusumadewi S and Hartati S 2006 Neuro Fuzzy: Integrasi Sistem Fuzzy & Jaringan Syaraf (Yogyakarta: Graha Ilmu).
[6] Tarno, Subanar, Rosadi D and Suhartono 2013 IJCSI 10 491.
[7] Zhang GP 2003 Neurocomputing 50 159.
[8] Wei WWS 2006 Time Series Analysis: Univariate and Multivariate Methods Second Edition (Pearson Education Inc).