International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Forecasting the Production of Thai Pepper and Chili Pepper in Garut Indonesia Based on Fuzzy Time Series and X-13 ARIMA-SEATS Methods

Mamlakatul Fardaniyah[a], Kusman Sadik[b*], Anang Kurnia[c]

[a,b,c]*IPB University, Jl. Raya Dramaga, Babakan, Dramaga District, Bogor City, West Java, Indonesia*

[a]*Email: mamlakatulfardaniyah@apps.ipb.ac.id*

[b]*Email: kusmans@apps.ipb.ac.id*

[c]*Email: anangk@apps.ipb.ac.id*

Abstract

Time series data is a series of data measured over a certain period of time based on fixed time intervals divided into seasonal and non-seasonal time series data. Anomaly observation that can affect the consistency of time series data are known as outliers. Outliers are often found in data that is influenced by weather, one of which is the production data of thai pepper and chili pepper that is one of the important types of vegetables that are cultivated commercially in tropical countries, Indonesia. A robust forecasting method against outliers is required in this case. The computational method developed at this time is the Fuzzy Time Series (FTS) method with the Fuzzy K-Medoids (FKM) clustering algorithm that can handle data that contains outliers. In addition, the X-13 ARIMA-SEATS method is a seasonal adjustment method that automatically detects and solves the problem of outliers and overcomes the effects of moving holiday. This study aims to forecast the monthly thai pepper and chili pepper production using both non-linear and linear methods. The result showed that FTS – FKM method give the best accuracy with smallest value of RMSE and MAPE in short forecasting period, otherwise X-13 ARIMA-SEATS give its best performance in long forecasting period.

Keywords: Fuzzy Time Series; Fuzzy K-Medoids; X-13 ARIMA-SEATS; ARIMA; Seasonal Time Series Data; Outliers; Thai and Chili Pepper Production.

----

* Corresponding author.

1. Introduction

Time series data is the value of observations measured over a certain period of time based on fixed time intervals divided into seasonal and non-seasonal time series data. Unpredictable natural phenomena such as floods, earthquakes, volcanic eruptions and other similar phenomena can affect the consistency of time series data. Observations of these disturbing events are often known as outliers. Outliers are often found in agricultural sector data in Indonesia due to the dependence of crop yields on the weather. Thai pepper (Capsicum frutescens) and chili pepper (Capsicum annum L.) are one of the agricultural sector commodities that have high value and are always needed by the people of Indonesia. The consumption pattern of the Indonesian people for thai pepper and chili pepper has made the demand for these two commodities, but their availability is uncertain. The stock uncertainty causes the prices of thai pepper and chili pepper to fluctuate in the market and become a benchmark for the prices of other basic necessities. To minimize these price fluctuations, it is necessary to know how to forecast the amount of thai pepper and chili pepper production so that the Ministry of Agriculture as the authority holder can determine the right steps in overcoming the scarcity of thai pepper and chili pepper in the coming period. Therefore, we need an appropriate method in predicting the production data of thai pepper and chili pepper in Garut Regency, Indonesia as a center for producing that two commodities in West Java, Indonesia.

In some studies classical methods such as the Box-Jenkins method or known as the ARIMA model are not effective enough in handling cases of outlier data. Therefore, forecasting time series data containing outliers must be done with the right forecasting method and can overcome the case of outliers in order to obtain good forecasting results. The method of handling outliers until now is quite fast, both from the linear method and the non-linear method. X-13 ARIMA-SEATS is one of the latest linear methods that can solve outliers in seasonal time series. The basic idea of the X-13 ARIMA-SEATS method is to combine the regARIMA method which can eliminate the moving holiday effect and the seasonal adjustment method or X-12 ARIMA so that it can help improve forecasting accuracy and identify time series data patterns better [1]. This method is a linear method that still has to fulfill the assumptions of stationary, white noise, normality, and homoscedasticity. However, in reality many time series data do not meet this assumption. So that non-linear methods that can ignore these assumptions began to be developed. The non-linear Fuzzy Time Series (FTS) method is the choice for forecasting seasonal time series data containing outliers because it has advantages in overcoming uncertainty. FTS using the Fuzzy K-Medoids (FKM) clustering algorithm in the fuzzification step can minimize the negative effect of outliers and abnormal observations on forecasting performance with the FTS model [2].

A study on the X-13 ARIMA-SEATS method was conducted by [1] and [3] in overcoming the case of outlier data in time series data, but the method has not worked optimally because the data used does not meet the characteristics of the data according to the X-13 ARIMA-SEATS method. Meanwhile, the FTS method with the FKM clustering algorithm has been carried out by [2] in predicting air pollution at 65 observation stations in Turkey. Forecasting the production of chili pepper was done by [4] using Singular Spectrum Analysis (SSA) method, but the result give unsatisfactory result based on its MAPE value is 18.23% which means the result still have significant risk. Based on the background above, this study aims to forecast the production of thai pepper and chili pepper in Garut Regency, Indonesia. The forecast result is comparing between FTS – FKM method, X-

13 ARIMA-SEATS method, and Multiplicative Seasonal ARIMA Method.

2. Methodology

2.1 Thai Pepper and Chili Pepper

Chili (Capsicum sp.) is one of the important types of vegetables that are cultivated commercially in tropical countries. Various species of chili have been domesticated, but only thai pepper (Capsicum frutescens L) and chili pepper or (Capsicum annum L.) have economic potential [5]. Chili is one of the important vegetable commodities that have prospective business opportunities and is one of the important vegetable commodities that have prospective business opportunities.

Thai pepper is one of the horticultural plants of a type of vegetable that has small fruit with a spicy taste. This type of chili is cultivated by farmers because it is widely needed by the community, not only on a household scale but also used on an industrial scale and exported abroad. In general, thai pepper can be planted in almost all parts of Indonesia. Currently, thai pepper cultivation is generally still carried out on a small scale due to relatively narrow land ownership by farmers.

Chili pepper is one of the important horticultural crops that is cultivated commercially, this is because in addition to chili having a fairly complete nutritional content, it also has a high economic value which is widely used both for household consumption and for the food industry. According to [6] chili pepper provide color and taste that can arouse appetite, contain lots of vitamins and can also be used as medicines, food mixtures and animal husbandry.

2.2 Outliers in Time Series Data

Outliers are observations of disturbing events. The effects that arise from these events can be processed with an intervention model if the time and cause are known. According to [7] there are several types of outliers, namely:

a. Additive Outlier (AO)

Additive Outlier (AO) is an event that has an effect on time series data only in one period. The general form of an AO is described as follows:

$$Z_t = \begin{cases} X_t & , t \neq T \\ X_t + \omega, & t = T \end{cases} = X_t + \omega_{AO} I_t^{(T)} \tag{1}$$

with X_t is the observation data at time t, ω_{AO} is the size of AO, and $I_t^{(T)} = \begin{cases} 1, t \neq T \\ 0, t = T \end{cases}$ is an indicator variable that represents the presence or absence of outliers at time T.

b. Innovational Outlier (IO)

Innovational Outlier (IO) is an event whose effect follows the process that occurs in X_t. Suppose X_t s assumed to follow the ARMA process then the general form of an IO is defined as follows:

$$Z_t = X_t + \frac{\theta(B)}{\phi(B)} \omega_{IO} I_t^{(T)} = \frac{\theta(B)}{\phi(B)} (a_t + \omega_{IO} I_t^{(T)}) \tag{2}$$

with ω_{IO} is the size of IO.

From equations (1) and (2) it can be concluded that AO only affects the T-th observation, whereas IO affects all observations $(Z_t, Z_{t+1}, \dots)$ exceeds the time T throughout the system memory described by $\frac{\theta(B)}{\phi(B)}$ (Assuming data follows ARMA process).

c. Temporary Change (TC)

Temporary change (TC) is an event where the outlier produces an initial effect of ω at time t, then slowly decreases according to the magnitude of δ. The TC model can be written as follows:

$$Z_t = X_t + \frac{1}{(1 - \delta B)} \omega_{TC} I_t^{(T)} \tag{3}$$

with ω_{TC} is the size of TC. When $\delta = 0$ then TC will be case of AO.

d. Level Shift (LS)

Level Shift (LS) is an event that affects the series at a certain time which gives a sudden and permanent change. LS is a special case of the TC outlier model when $\delta = 1$. The LS outlier model is stated as follows:

$$Z_t = X_t + \frac{1}{(1 - B)} \omega_{LS} I_t^{(T)} = X_t + \frac{1}{(1 - B)} \omega_{LS} I_t^{(T)} = X_t + \omega_{LS} S_t^{(T)} \tag{4}$$

with ω_{LS} is the size of LS and $S_t^{(T)} \begin{cases} 1, t \geq T \\ 0, t < T \end{cases}$.

2.3 Fuzzy Time Series with Fuzzy K-Medoids Clustering Algorithm Method

Fuzzy time series (FTS) is a forecasting technique based on fuzzy logic where the results obtained can be discussed [8]. Forecasting and modeling techniques based on fuzzy logic are widely used in many fields for a large number of applications because of their advantages, such as the ability to model complex, non-linear and uncertain systems; to process linguistic variables; does not require statistical assumptions; and is suitable for data with a small number of observations. The FTS model can be considered as a version of time series statistical analysis applied to fuzzy sets.

The definition and properties of FTS forecasting are summarized as follows:

Definition 1. Suppose $X(t)$ $(t = \cdots, -2, -1, 0, 1, 2, \dots)$ is part of a real number (R). Then the sample set of fuzzy is (t) $(i = 1,2, \ldots)$. If $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$ then $F(t)$ called fuzzy time series on $Y(t)$.

Definition 2. If there is a fuzzy relationship $R(t, t-1)$, then:

$$F(t) = F(t-1) \circ R(t, t-1) \tag{5}$$

where $\circ$ is an arithmetic operator, then it can be said $F(t)$ can happen because $F(t-1)$. Relationship between $F(t)$ and $F(t-1)$ can be written as $F(t-1) \rightarrow F(t)$. If $F(t)$ calculated only by $F(t-1)$, and $F(t) = F(t-1) \circ R(t, t-1)$. For all t, if $R(t, t-1)$ independent of t then $F(t)$ is time invariant fuzzy time series. Otherwise, F(t) is time variant.

Definition 3. If $F(t-1) = A_i$ and $F(t) = A_j$, Fuzzy Logical Relationship (FLR) can be written as $A_i \rightarrow A_j$, where A_i and A_j named Left-Hand Side (LHS) and Right-Hand Side (RHS) from FLR.

Definisi 4. Suppose $F(t)$ is fuzzy time series which has a seasonal pattern with a period of m, then the fuzzy relationship can be written as $F(t-m) \rightarrow F(t)$ [9].

Clustering Analysis (CA) is a data mining technique that is widely used to divide data sets $X = (x_1, x_2, \ldots, x_n)$ into c group. CA aims to allocate data points (x_k) into the cluster in such a way that points in the same cluster are points near each other and points in different clusters are points far from each other. The CA technique is divided into two groups, namely crisp clustering and fuzzy clustering. The basic logic in both clustering techniques is to allocate data points to clusters with a certain degree of membership. Crisp clustering and fuzzy clustering are defined by the following equation:

$$\mu_H = \{U \in \mu : u_{ik} = 0 \ atau \ 1 \ \forall \ i \ dan \ k\} \tag{6}$$

$$\mu_F = \{U \in \mu : \sum_{i=1}^{c} u_{ik} = 1;\ \forall\ k; \sum_{k=1}^{n} u_{ik} > 0\ \forall\ i\} \tag{7}$$

Equation (6) is crisp clustering if and only if each data point is a member of one cluster and the degree of membership of each data point is 1. Equation (7) is a fuzzy clustering, one data point can be a member of two or more clusters simultaneously with different degrees of membership i.e. between 0 and 1. Most of the fuzzy clustering algorithms are based on the concept of minimizing the objective function which is defined as follows:

$$J(U, X, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - v_j \right\|^2, \ \ 1 \le m < \infty \tag{8}$$

where U shows a matrix consisting of degrees of membership$(u_{ij},\ i = 1,2, \ldots, n,\ j = 1,2, \ldots, c)$, X show data matrix$(x_i,\ i = 1,2, \ldots, n)$, V denotes a matrix containing the center of the cluster $(v_j,\ i = 1,2, \ldots, c)$, m show fuzziness index, and || || show euclidian distance.

In this algorithm, the cluster prototype represents each cluster. The cluster prototype includes some information such as the center, shape, and size of the cluster. The degree of membership is obtained by calculating the distance between the data point and the cluster center and indicating the data point assigned to the corresponding cluster. The shape of the cluster prototype can be a weighted mean, medoid, or linear. K-Medoid Fuzzy Clustering (FKM) is a fuzzy clustering algorithm whose cluster form is medoid and is widely used, especially in cases of data collection with outliers and noise data points. A medoid is defined as $V_j = \{x_1, x_2, \dots, x_{n_j}\}$ with j showing cluster and v_j is medoid from V_j. v_j is an element of V_j with the sum of the distances from v_j to another V_j's element is minimum. When minimizing the objective function is given in equation (8), the new equations for membership degree and medoid are obtained as follows [2]:

$$u_{ij} = \left[\sum_{k=1}^{c} \frac{\left\|x_i - v_j\right\|^2}{\left\|x_i - v_k\right\|^2}\right]^{\frac{1}{m-1}} \quad (9)$$

$$v_j = argmin_{1\leq z\leq n} \sum_{k=1}^{n} u_{kj}^{m} \left\|x_z - v_k\right\|^2 \quad (10)$$

2.4 ARIMA Method

ARIMA or Autoregressive Integrated Moving Average is a statistical method that is often used to forecast time series data that is not stationary in the mean value and is known as the Box-Jenkins model which has become a very popular time series forecasting model in various fields [10]. The ARIMA model is denoted by ARIMA(p,d,q), where p indicates the order of autoregressive (AR), d indicates the order of differencing and q indicates the order of moving average (MA). A series Y_t is said to follow the ARIMA(p,d,q) process if it has the form:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\, a_t \quad (11)$$

where in time series analysis a_t assumed to be white noise process, $a_t \sim WN(0, \sigma^2)$.

The ARIMA method can also be used for seasonal models. The seasonal model is a model that has recurring properties after a certain period of time. For example weekly for daily or annual time series data and 6 months for monthly time series data. The seasonal ARIMA model is expressed in the following form:

$$\Phi_p(B^s)(1-B^s)^D Y_t = \Theta_Q(B^s) a_t \quad (12)$$

Meanwhile, the combination of the non-seasonal ARIMA model and the seasonal ARIMA model is called multiplicative seasonal ARIMA and is written with ARIMA notation(p,d,q)(P,D,Q)[s] with p, d, q indicating the non-seasonal part and P, D, Q, and s represent the seasonal portion. A series Y_t is said to follow the ARIMA(p,d,q)(P,D,Q)[s] process if it has the form of equation [11]:

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s) a_t \quad (13)$$

2.5 X-13 ARIMA-SEATS Method

X-13 ARIMA-SEATS is a seasonal adjustment method that is based on the automatic fitting of the ARIMA model and includes detection of additive outliers (AO) and level shift (LS) [12]. The X-13 ARIMA-SEATS method is a popular technique in the parametric approach to estimating Level Shift (LS). The X-13 ARIMA-SEATS is an improved version of the X-12 ARIMA developed by the United States Census Bureau [13]. The difference between X-13 ARIMA-SEATS and X-12 ARIMA is in the expansion data. In the X-13 ARIMA-SEATS method the expansion data was obtained from the regARIMA model, while the X-12 ARIMA method the expansion data was obtained from the ARIMA model.

The regARIMA model is a form of calendar variation model that can be used to forecast data based on seasonal patterns with varying period lengths [14]. The use of the regARIMA model aims to eliminate the effect of the calendar in this case, namely moving holiday. According to [14] time series data Z_t which contains the effect of calendar variations written in the equation $Z_t = f(X_t; \xi) + N_t$ where $N_t = \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D}\alpha_t$. So that the regARIMA model is obtained as follows:

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D\big(Z_t - f(X_t;\xi)\big) = \theta_q(B)\Theta_Q(B^s)a_t \qquad (14)$$

There are two kinds of decomposition in the X-13 ARIMA-SEATS method, namely the X-12 ARIMA and TRAMO/SEATS decomposition which aims to estimate seasonal factors. The X-12 ARIMA decomposition method developed by David F. Findley and his colleagues from the United States Census Bureau in 1995 is one of the methods used to decompose the factors in the data pattern [16]. This method consists of expanding the time series data provided by forecasting and backcasting N_t from the regARIMA model in equation (14). The stages in the ARIMA X-12 decomposition method consist of 3 stages, namely initial estimation, final estimation of seasonal components and seasonal adjustment data, and final estimation of trend-cycle and irregular components.

In general, the X-13 ARIMA-SEATS process is as follows [1]:

a. Eliminating the moving holiday effect contained in the Z_t data by using the regARIMA model in eq. (14).
b. Expanding observation data by forecasting and backcasting based on the regARIMA model in eq. (14).
c. Decompose the expansion data in the previous step using the X-12 ARIMA decomposition assuming the decomposition model used is a multiplicative model.

2.6 Accuracy

The measure of forecasting accuracy will be measured using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). RMSE is an alternative method in evaluating a forecasting technique. RMSE serves to combine the magnitude of the error in predictions for various data points into a single measure of predictive power [17]. RMSE is positive and is said to be getting better if it is close to zero. So a model that has a lower RMSE can be said to be better than one that has a higher value. The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{t}(Y_t - X_t)^2}{t}} \tag{15}$$

where Y_t is the forecast value in period t for each model, while X_t is the actual value.

Meanwhile, MAPE is calculated by finding the absolute error for each period by dividing it with the observed value in that period and then presenting it. This method provides an indication of how much the forecast error is compared to the true value. A forecast is said to have very good performance if it has a MAPE value below 10% and has good performance if the MAPE value ranges from 10% to 30% [18]. The formula for MAPE is as follows:

$$MAPE = \frac{100}{n}\sum\left|X_t - \frac{Y_t}{X_t}\right| \tag{16}$$

where n is the value of the time period, X_t is the actual value in the t-t period, and Y_t the forecast value in the t-period.

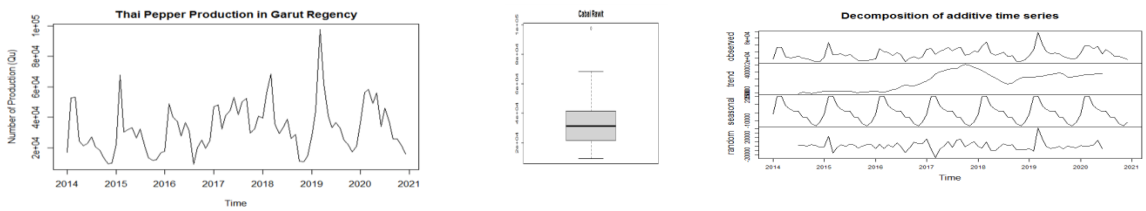
2.7 Data

The time series data used in this study is secondary data. It is monthly production of thai pepper and chili pepper data in Garut Regency with quintal units for January 2014 - December 2020 period so that there are 84 observations of time series data sourced from the Directorate General of Horticulture, Ministry of Agriculture. As a validation of model consistency, modeling was carried out using three splitting data to training and testing data. Training data is first 81, 78, and 72 data and the rest as testing data.

3. Result and Discussion

3.1 Data Exploration and Visualization

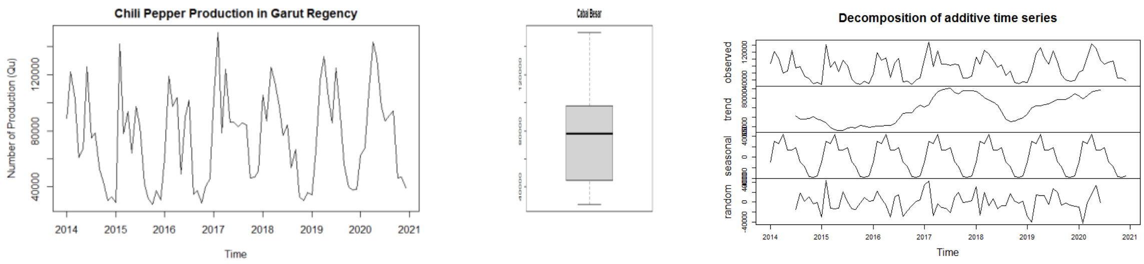
Data characteristics of thai pepper and chili pepper can be known by conducting data exploration, such as data plots, boxplots, and decomposition. Each data exploration is shown in Figure 1 and Figure 2.



(a) Data Plot (b) Boxplot (c) Decomposition

Figure 1: Exploration of Thai Pepper Production Data in Garut Regency

In Figure 1 (a) it can be seen that the production data of thai pepper has a seasonal pattern which can be seen from the presence of the same peak production but the differences in production patterns each year are not too extreme. Figure 1 (b) shows the distribution of thai pepper production data. In this plot, outliers were detected and the upper whisker was longer, indicating that the data distribution was skewed to the right (positive skewness). Figure 1 (c) shows the decomposition of thai pepper production data consisting of observed, trend, seasonal, and random components. In accordance with the reality that is happening in the community, the trend component shows that the production data of thai pepper fluctuates from 2014 to 2020.



(a) Data Plot (b) Boxplot (c) Decomposition

Figure 2: Exploration of Chili Pepper Production Data in Garut Regency

In Figure 2 (a) it can be seen that the chili pepper production data has a seasonal pattern, it can be seen from the pattern that repeats every year, but there is a slight difference in the pattern every year. Figure 2 (b) shows the distribution of chili production data. In this plot no outliers were detected, but the median line which tends to be upwards and the longer upper whisker indicates that the data distribution is skewed to the right (positive skewness). Figure 2 (c) shows the decomposition of chili pepper production data consisting of observed, trend, seasonal, and random components. In accordance with the reality that is happening in the community, the trend component shows that the chili pepper production data fluctuated from 2014 to 2020. Furthermore, to be clearer about the presence or absence of outliers in the data, outliers were checked with the help of the 'tsoutliers' package in the Rstudio software.

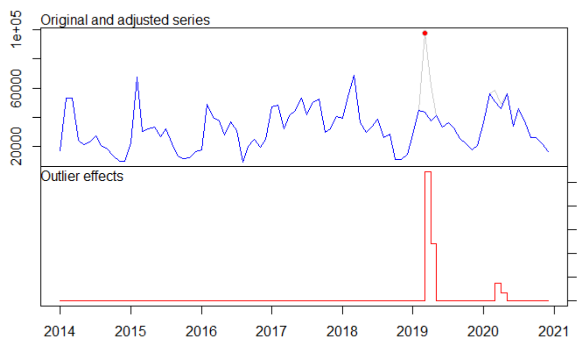


Figure 3: Plot of Outliers Detection of Thai Pepper Production Data

Table 1: Types and Sizes Outliers of Thai Pepper Production Data

Type of Outliers	Period	Time	Coefhat ($\widehat{\omega}$)
IO	63	2019:03	54200

Figure 3 and Table 1 show the results of detecting outliers in thai pepper production data. Outliers in Figure 3 are marked with a red dot, while the blue line shows the adjusted series plot and the gray line represents the original data series. The description of the outliers is presented in Table 1 and one outlier was detected, namely the Innovational Outlier (IO) type.

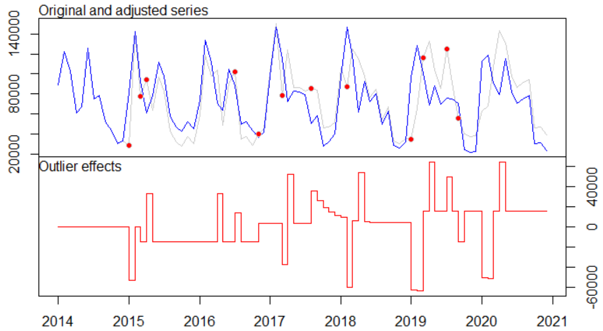


Figure 4: Plot of Outliers Detection of Chili Pepper Production Data

Table 2: Types and Sizes Outliers of Chili Pepper Production Data

Type of Outliers	Period	Time	Coefhat ($\widehat{\omega}$)	Type of Outliers	Period	Time	Coefhat ($\widehat{\omega}$)
AO	13	2015:01	-52459	TC	44	2017:08	31318
LS	15	2015:03	-14636	SLS	50	2018:02	-67022
SLS	16	2015:04	47836	IO	61	2019:01	-66003
AO	31	2016:07	28469	LS	63	2019:03	11985
LS	35	2016:11	18457	AO	67	2019:07	33342
AO	39	2017:03	-40949	AO	69	2019:09	-30535

Figure 4 and Table 2 show the results of outlier detection in chili pepper production data. The outliers in Figure 4 are marked with a red dot, while the blue line is the adjusted series plot and the gray line is the original data series plot. The description of the outliers is presented in Table 2. 12 outliers of various types were detected, namely Additional Outlier (AO), Innovational Outlier (IO), Temporary Change (TC), Level Shift (LS), and Seasonal Level Shift (SLS).

Based on the exploration of the data above, it can be seen that the characteristics of the production data of thai pepper and chili pepper are seasonal time series data that contain outlier data, but meet the stationarity requirements in the mean value. Therefore, a seasonal time series data forecasting method is needed that can overcome outliers. The best method for forecasting seasonal time series data in overcoming outlier data based on simulation results is the FTS - FKM method. So the FTS - FKM method will be applied to predict the production data of thai pepper and chili pepper containing outlier data. In addition, to validate the forecasting accuracy of the FTS – FKM method, two other forecasting methods will also be carried out, namely the X-13 ARIMA-SEATS method and the Multiplicative Seasonal ARIMA. In predicting scenarios, 3, 6, and 12 test data were made based on the planting and harvesting period of thai and chili pepper themselves. The determination of this scenario is based on the harvest period of thai pepper, which is on the 65th to 75th days after planting [19] and chili pepper, which is on days 70 to 120 after planting [20].

3.2 Comparison of Thai Pepper and Chili Pepper Production Forecasting Accuracy

The application of the Fuzzy Time Series (FTS) method uses the Chen [20] and Song [8] algorithms with the Fuzzy K-Medoids (FKM) clustering algorithm to predict cayenne pepper production data which is seasonal time series data with outlier data. The first step in forecasting using the FTS - FKM method is to determine the universe of discourse, which is symbolized by U where $\boldsymbol{U} = [\boldsymbol{X_{min}}; \boldsymbol{X_{max}}]$. $\boldsymbol{X_{min}}$ is the lowest data and $\boldsymbol{X_{max}}$ is the highest data. The second step is the formation of intervals and sub intervals. The interval is formed based on the theory of Fuzzy K-Medoids (FKM) which is strong against outliers. This interval calculation uses the Partitioning Around Medoids (PAM) algorithm with the 'pam' function available in the RStudio software.

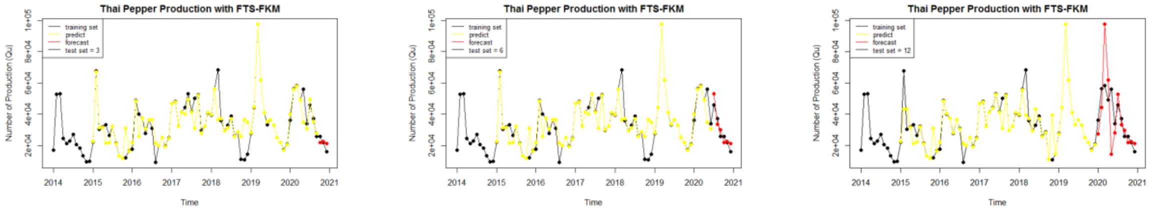
The third step is the formation of a fuzzy set which aims to simplify by converting numerical data into linguistic data based on the sub-intervals that have been formed. The next step is to perform fuzzification based on the obtained sub-intervals and the linguistic value can be determined according to the number of sub-intervals that have been formed. The fourth step is the formation of a Fuzzy Logical Relationship (FLR) and a Fuzzy Logical Relationship Group (FLRG). FLR is identified based on historical data which was fuzzified in the previous step. FLR is defined as $A_i \to A_j$. A_i is Left Hand Side (LHS) and A_j is Right Hand Side (RHS). The definition proposed by Song in 1999 is that time series data has a seasonal pattern with a period of m, so the fuzzy relationship can be written as $F(t-m) \to F(t)$, so that LHS is the observation of the previous 12 months and RHS is the observation of this month.

The fifth step is defuzzification using the median value obtained in the first step of each FLRG so that the predicted value is obtained. Forecasting calculations using FTS Chen. Then all the mean values are calculated on average so that the average value becomes the forecast value ($F(t)$). In general, the formula is as follows:

$$F(t) = \frac{me_1 + me_2 + \cdots + me_n}{n}$$

Forecasting values for all FLRGs are calculated based on the above formula so that the forecast values are obtained.

The forecast value obtained is then extracted to all of the cayenne pepper production data based on the fuzzification of the previous data so that it will produce predictive data from the original data. Forecasting data according to the amount of test data obtained from FLR information from the previous 12 months data, so that the results of prediction and forecasting data for each scenario are as follows:

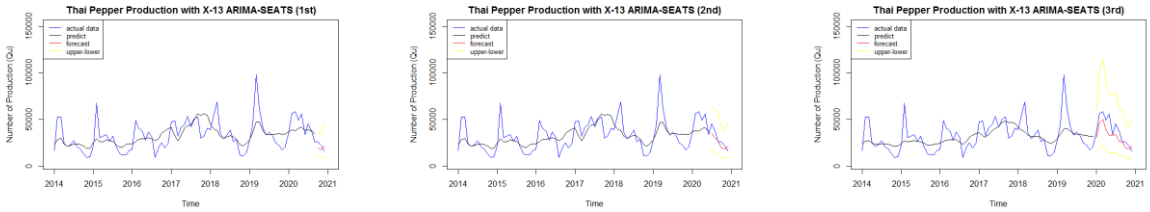


(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 5: Plot of Forecasting Thai Pepper Production Data with FTS - FKM Method

Based on the results of forecasting using the FTS - FKM method in Figure 5, the predicted data values are close to the actual data values, as well as the forecasting data are close to the test data. Furthermore, forecasting will be carried out using the X-13 ARIMA-SEATS method on thai pepper production data.

Modeling and forecasting the X-13 ARIMA-SEATS method using the RStudio software with the 'seas' and 'series' packages. Based on the exploration of the thai pepper production data in Figure 44, it is known that the thai pepper production data has a seasonal pattern. The 'seas' package goes through a logarithmic transformation process so that the prediction and forecasting results are obtained in Figure 6.

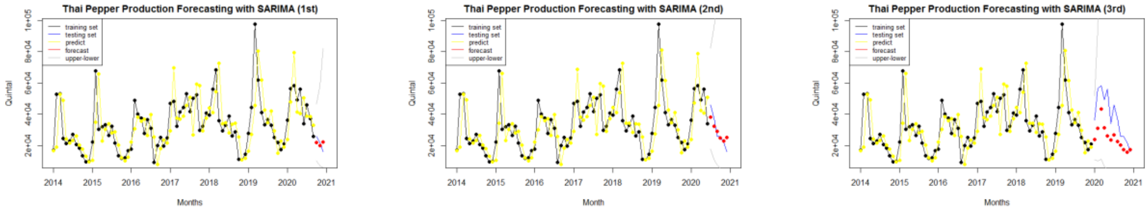


(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 6: Plot of Forecasting Thai Pepper Production Data with X-13 ARIMA-SEATS Method

Based on the results of forecasting using the X-13 ARIMA-SEATS method in Figure 6, it is obtained that predictive data has followed the training data pattern, and the forecast data has followed the pattern and is close to the value of the test data. Forecasting thai pepper production data is continued with the Multiplicative Seasonal ARIMA method with the first step being to test the fulfillment of stationary assumptions in the variance and the mean value. After the data is stationary, modeling can be continued by identifying candidates for the Multiplicative Seasonal ARIMA model and forecasting based on the best model in Figure 7.

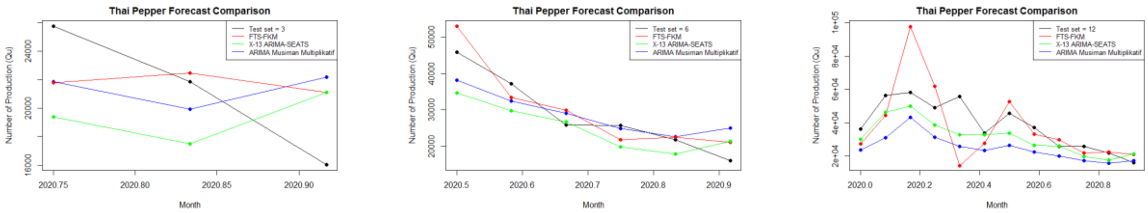
Based on the forecasting results using the Multiplicative Seasonal ARIMA method, prediction data is obtained that is close to the actual data value, as well as forecast data that follows the appropriate test data pattern. However, this forecasting result is obtained from the ARIMA model which does not meet the residual assumptions.



(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 7: Plot of Forecasting Thai Pepper Production Data with Multiplicative Seasonal ARIMA Method

After forecasting using the three methods, the next step is to compare the results and forecasting accuracy of thai pepper production. Forecasting accuracy is measured based on the RMSE and MAPE values.



(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 8: Result Comparison of Thai Pepper Production Forecasting from the Three Methods

Figure 8 shows that the forecasting results using the FTS - FKM method are closer to the test data than the other two methods. In order to be clear about the forecasting performance for the three methods, the next step is to calculate the RMSE and MAPE values for the forecasting of the three methods.

Table 3: Thai Pepper Production Forecasting Accuracy Comparison in Three Methods

Scenarios	Methods	RMSE	MAPE
	FTS – FKM	3732.36	16.58
1	X-13 ARIMA-SEATS	5314.35	25.32
	Multiplicative Seasonal ARIMA	4333.32	20.67
	FTS – FKM	4510.66	15.12
2	X-13 ARIMA-SEATS	6552.52	20.19
	Multiplicative Seasonal ARIMA	5396.79	17.51
	FTS – FKM	17812.73	26.73
3	X-13 ARIMA-SEATS	9920.50	20.25
	Multiplicative Seasonal ARIMA	15985.13	33.15

The comparison of the forecasting accuracy of cayenne pepper production for the three methods in each scenario shown in Table 3 concludes that the test data for 3 periods of the FTS - FKM method is better than the other two methods. It is based on the smallest RMSE and MAPE values.

3.3 Thai Pepper and Chili Pepper Production Forecasting with Best Method

It is better to forecast the production of thai pepper using the FTS - FKM method with forecasting for 3 periods based on the comparison of forecasting accuracy in the previous step. The next step is to forecast the next 3 periods, namely January 2021 to March 2021 using the FTS - FKM method.

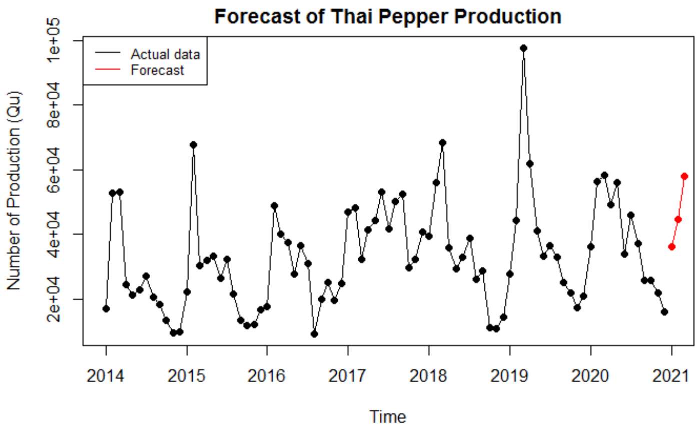
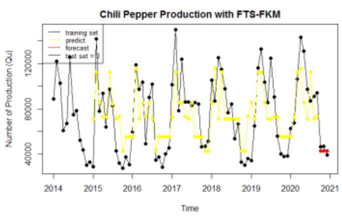


Figure 9: Forecasting the Future Period of Thai Pepper Production

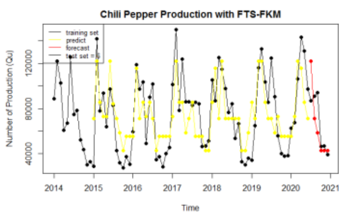
Based on Figure 9, the production of thai pepper is predicted to increase in January 2021. However, from February 2021 to March 2021 it will experience a decrease compared to February 2020 to March 2020.

3.4 Comparison of Chili Pepper Production Forecasting Accuracy

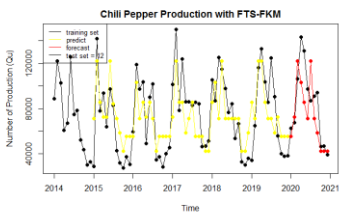
The characteristics of chili pepper production data are not much different from the characteristics of thai pepper production data, so the forecasting process with the three proposed methods will be the same as the forecasting process for thai pepper production data. The results of forecasting using the FTS – FKM method are shown in Figure 10, while the results of forecasting using the X-13 ARIMA-SEATS and Multiplicative Seasonal ARIMA methods are shown in Figure 11 and Figure 12.



(a) 1st Scenario



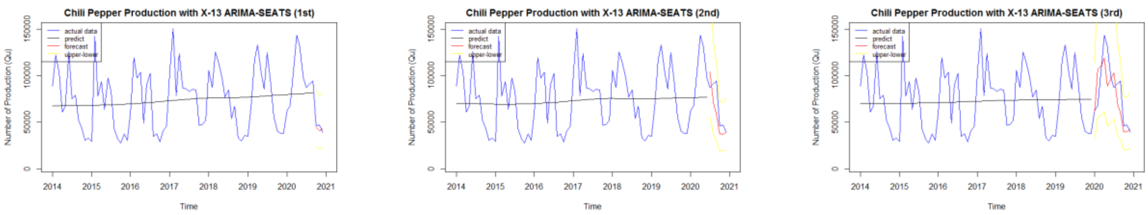
(b) 2nd Scenario



(c) 3rd Scenario

Figure 10: Plot of Forecasting Chili Pepper Production Data with FTS - FKM Method

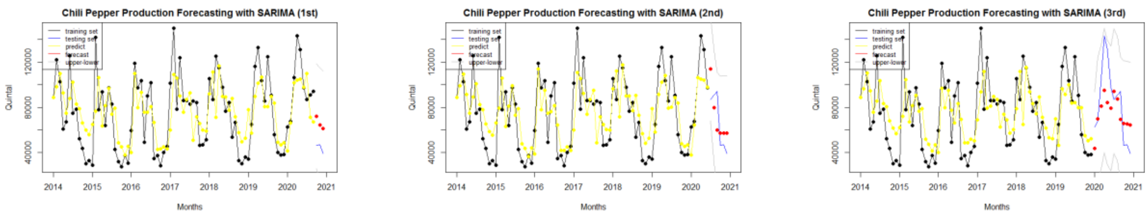
Based on the results of forecasting using the FTS - FKM method in Figure 10, it is obtained that the predicted data value is close to the actual data value, as well as the forecast data is close to the test data.



(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 11: Plot of Forecasting Chili Pepper Production Data with X-13 ARIMA-SEATS Method

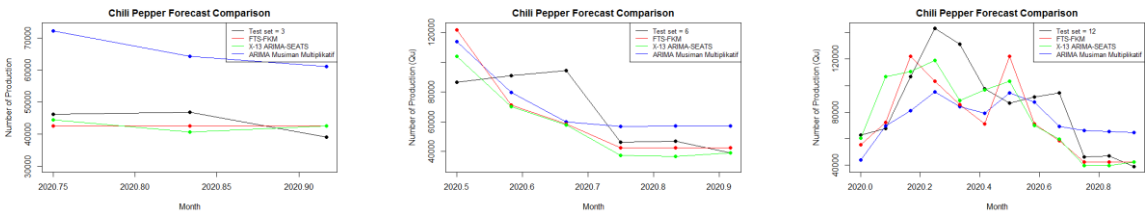
Based on the results of forecasting using the X-13 ARIMA-SEATS method in Figure 11, it is obtained that predictive data is constant, but the forecast data has followed the pattern and is close to the value of the test data.



(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 12: Plot of Forecasting Chili Pepper Production Data with Multiplicative Seasonal ARIMA Method

Based on the results of forecasting using the Multiplicative Seasonal ARIMA method in Figure 12, prediction data is obtained that is close to the actual data value, as well as forecast data that follows the test data pattern. After forecasting using the three methods, the next step is to compare the results and forecasting accuracy of chili pepper production. Forecasting accuracy is measured based on the value of RMSE and MAPE.



(a) 1st Scenario (b) 2nd Scenario (c) 3rd Scenario

Figure 13: Result Comparison of Chili Pepper Production Forecasting from the Three Methods

Based on Figure 13, it shows that the forecasting results using the FTS - FKM method are closer to the test data than the other two methods. In order to be clear about the forecasting performance for the three methods, the next step is to calculate the RMSE and MAPE values for the forecasting of the three methods..

Table 4: Chili Pepper Production Forecasting Accuracy Comparison in Three Methods

Scenarios	Methods	RMSE	MAPE
1	FTS – FKM	3789.31	8.58
	X-13 ARIMA-SEATS	4255.27	8.65
	Multiplicative Seasonal ARIMA	22050.71	49.81
2	FTS – FKM	22225.54	21.01
	X-13 ARIMA-SEATS	19333.14	20.43
	Multiplicative Seasonal ARIMA	20841.71	28.74
3	FTS – FKM	25266.74	20.81
	X-13 ARIMA-SEATS	22274.62	19.25
	Multiplicative Seasonal ARIMA	25685.71	27.65

The comparison of the forecasting accuracy of chili pepper production for the three methods in each scenario shown in Table 4 concludes that the test data for 3 periods of the FTS - FKM method shown by the red line in Figure 13 is better than the other two methods. It is based on the smallest RMSE and MAPE values.

3.5 Chili Pepper Production Forecasting with Best Method

It is better to forecast chili pepper production using the FTS - FKM method with forecasting for 3 periods based on the comparison of forecasting accuracy in the previous step. The next step is to forecast the next 3 periods, namely January 2021 to March 2021 using the FTS - FKM method.

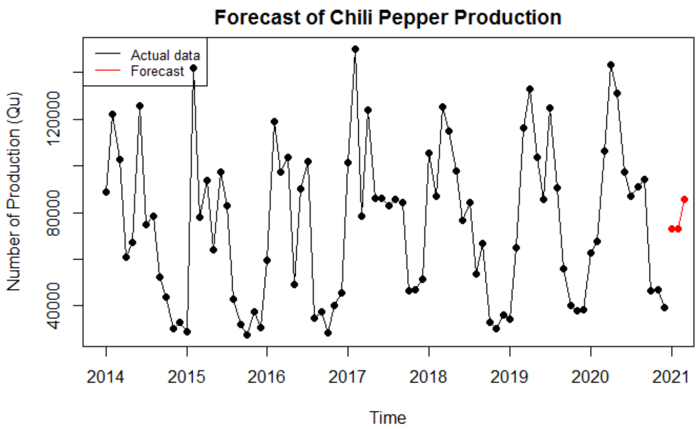


Figure 14: Forecasting the Future Period of Chili Pepper Production

Based on Figure 14, it is predicted that large chili production will increase in production in January 2021 to February 2021 compared to the same month in the previous year, namely January 2020 to February 2020. However, in March 2021 it will experience a decrease compared to March 2020.

4. Conclusions and Suggestions

Production of thai pepper and chili pepper data belongs to seasonal time series data type with outliers, so forecasting that data requires a robust method. Outliers can cause forecast using classical method, namely Multiplicative Seasonal ARIMA, produce not accurate enough forecast and the result model does not meet the assumptions. This assumption can be ignored by using non-linear methods that can handled outliers, namely the FTS-FKM and X-13 ARIMA-SEATS methods proposed in this study. Based on the result comparison of the production of thai pepper and chili pepper data forecasting, the FTS – FKM method produces better forecasting accuracy than the X-13 ARIMA-SEATS method in short forecasting period. However, the X-13 ARIMA-SEATS method gives the best performance in long forecasting period. FTS – FKM method is able to predict thai pepper production data with RMSE and MAPE values of 3732.36 and 16.58% and RMSE and MAPE values in forecasting chili pepper production data of 3789.31 and 8.58% in the 3 period data test scenarios.

References

[1] E. Lestari, T. Widiharih, R. Rahmawati. “Forecasting non-oil and gas exports with islamic calendar variations using X-13 ARIMA SEATS”. Gaussian, vol. 7, pp. 236-247, 2018.

[2] N.G. Dincer and O. Akkus. “A new fuzzy time series model based on robust clustering for forecasting of air pollution”. Ecological Informatics, vol. 43, pp. 157-164, 2018.

[3] A.M. Putri, K. Sadik, F.M. Afendi. 2021. “Study of X-13 ARIMA SEATS modeling for rice price index data in Indonesia”. Journal of Physics: Conference Series, vol. 1863, p. 012055, 2021.

[4] A. Asrof, R. Ischak, and G. Darmawan, “Forecasting Chili Pepper Production in West Java Using the Singular Spectrum Analysis (SSA) Method”. Statistika, vol. 17, pp. 77-87, 2017.

[5] S. Sulandari. “Biological characterization, serology and DNA fingerprint analysis of viruses that cause chili yellow curly leaf disease”. Perlindungan Tanaman Indonesia, vol. 12, pp. 1-12, 2006.

[6] A. Marliah, M. Nasution, A. Armi. “Growth and yield of several red chili varieties on different growing media”. Floratek, vol. 6, pp. 84-91, 2011.

[7] W. Wei. Time Series Analysis, Univariate and Multivariate Methods. New York: Pearson Education, 2006.

[8] L.A. Zadeh. “Fuzzy Sets”. Information and Control, vol. 8, pp. 338-353, 1965.

[9] Q. Song. “Seasonal forecasting in fuzzy time series”. Fuzzy Sets and System, vol. 107, pp. 235-236, 1999.

[10] G. Box and G. Jenkins. Time series analysis: forecasting and control. (Revised ed). San Fransisco: Holden day, 1976.

[11] J.D. Cryer and K.S. Chan. Time Series Analysis with Application in R. New York: Springer, 2008.

[12] P. Rousseeuw, D. Perrotta, M. Riani, M. Hubert. “Robust monitoring of time series with application to fraud detection”. Elsevier, vol. 9, pp. 108-121, 2018.

[13] U.S. Census Bureau. (2017, January 18). X-13 ARIMA-SEATS Reference Manual. (Version 1.1). [On-line]. Available: http://www.census.gov/srd/www/x13as [February 22, 2022].

[14] J.L. Lin and T.S. Liu. “Modeling lunar calendar holiday effect in Taiwan”. Taiwan, 2022.

[15] W.R. Bell and S.C. Hilmer. “Modelling time series with calendar variation”. Journal of Bussiness and Economics Statistics, vol. 78, pp. 526-534, 1983.

[16] N. Statistics. “Guide to seasonal adjustment with X-12 ARIMA”. ONS Methodology and Statistical Development, 2017.

[17] R.J. Hyndman and A.B. Koehler. “Another look at measures of forecast accuracy”. International Journal of Forecasting, vol. 22, pp. 679-688, 2006.

[18] D.C. Montgomery, et.al. Forecasting Time Series Analysis 2nd. New York: McGraw-Hill, 2008.

[19] BPTP of Maluku. 2020. "Harvesting divine varieties of chili at tagrimart garden /OPAL". Internet: https://maluku.litbang.pertanian.go.id/?p=6797, 2022 [Mei 4, 2022].

[20] BPTP of West Sumatera. 2017. "Cultivation of chili in pots/polybags". Internet: http://sumbar.litbang.pertanian.go.id/index.php/info-tek/1022-budidaya-cabai-dalam-pot-polybag, 2017 [Mei 4, 2022].

[21] S.M. Chen. "Forecasting enrollments based on fuzzy time series". Fuzzy Sets and Systems, vol. 81, pp. 311-319, 1995.