IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Forecasting The Broad Proportion Attack of Rice Blast Disease in Indonesia

[1] Iman Setiawan; [2] I Made Sumertajaya; [3] Farit Mochammad Afendi

[1] Department of Statistics, Bogor Agricultural University
Bogor, 16680, Indonesia

[2] Department of Statistics, Bogor Agricultural University
Bogor, 16680, Indonesia

[3] Department of Statistics, Bogor Agricultural University
Bogor, 16680, Indonesia

Abstract - Classical regression analysis is a statistical technique for modeling, forecasting and investigating the relationship between response variable and explanatory variables. However, there are model adequacy must be checked on residual model i.e. autocorrelation. The autocorrelation problem can be solved by modeling the residual of regression model into model that specifically incorporates the autocorrelation structure. Autocorrelation can be caused by residual of regression model increasing over time. The time series regression model is one of the analyzes used to accommodate the model residual which increasing over time. This study used data on the broad proportion of rice blast (*Pyricularia grisea*) attacks. The purpose of this study is to forecast the broad proportion of rice blast attacks used classical regression model and time series regression model. Evaluate forecast values used mean absolute percentage error (MAPE). The comparison results showed that the forecast of time series regression model better than classical regression model.

Keywords – *forecasting, MAPE, pyricularia grisea, regression, time series regression model*

1. Introduction

Forecasting the broad proportion of plant disease attacks is very important problem. Considering the impact is the decrease in the quantity and quality of the rice plant production. There are three main factors that can cause plant diseases: (1) host plant in a vulnerable state, (2) virulent pathogens (high infection power) with sufficient quantities, and (3) supportive environment. Environmental factors on plant diseases divided into physical environment (abiotic) e.g. rain, temperature, humidity, light and wind and the biotic environment e.g. natural enemies and organisms competitors. [1]

Blast (*Pyricularia grisea*) disease is caused by *Pyricularia grisea* fungus. Blast disease becomes an important issue in Indonesia. The rapid development of the Pyricularia grisea and the changing of populations become an obstacle in controlling blast disease. Blast disease is able to break the resistance of varieties quickly, so that the superior varieties resistant of Blast disease will turn into sensitive after extensive planting for 2-3 consecutive planting seasons [2]. One of the causes of blast disease is the physical environment factor i.e. climate factor. For example, the role of wind velocity is essential for the release of spores or the intensity of light that affects the penetration process of infection [3].

Classical regression analysis is a statistical technique for modeling, forecasting and investigating the relationship between response variables and explanatory variables. However, there are model adequacy must be checked on residual model i.e. normality, constant variance and autocorrelation. The assumption problems of normality and constant variance problems can be solved by transformation. The autocorrelation problem can be solved by modeling the residual of regression model into model that specifically incorporates the autocorrelation structure [4]. Autocorrelation can be caused by residual of regression model increasing over time. The time series regression model is one of the analyzes used to accommodate the model residual which increasing over time by performed time series model on residual of regression model [5]. Therefore, the purpose of this study is to forecast the broad proportion of rice blast attacks in

Indonesia using the classical regression model and time series regression model.

2. Time series regression model

The time series regression model is one of the analyzes used to accommodate the model residual which increasing over time by performed time series model on residual of regression model. For example, given the linear regression model with the residual of model follows Autoregressive (1).

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t \\ \varepsilon_t &= \phi\varepsilon_{t-1} + w_t \end{aligned} \quad (1)$$

where y_t and x_t are the observations on the response and predictor variables at time period *t*, ε_t is the error term in model at time period *t*, ω_t is an IIDN $(0, \sigma_w^2)$, β_0 is an intercept, β_1 is a slope and ϕ is parameter that defines the relationship between successive values of the model errors ε_t and ε_{t-1}. [5]

3. Materials and Methods

3.1 Materials

The data used in this study is the broad proportion of rice blast attacks, flood and drought with rice planting area obtained from the Director General of Food Crops. The explanatory variables used are physical environmental factors (climate factors) sourced from the Meteorological, Climatological and Geophysical Agency. The time period of data is the monthly period from January 2010 - December 2016. Data divided into two parts, namely January 2010 - December 2015 as training data and January 2016 - December 2016 as testing data. This study used variables given as follows.

Table 1: Variables used in this study

Variables	Description
Y	Broad proportion of rice blast attacks
X_1	Humidity (%)
X_2	Average temperature (°C)
X_3	Average wind velocity (knot)
X_4	Length exposed to light (hours)
X_5	Minimum temperature (°C)
X_6	Maximum temperature (°C)
X_7	Broad proportion of floods
X_8	Broad proportion of drought

3.2 Methods

The method of analysis in this study divided into several steps given as follows:

1. Exploration of data by plot to see pattern of broad proportion of rice blast attacks.
2. Divided data into training data to do modeling while evaluation of forecast results used testing data.
3. Performed classical regression analysis and assumption test of regression model.
4. Predict and forecast for the next 12 months and evaluate the forecast by using Eq. (2) [4].

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} x\,100 \right|}{n} \quad (2)$$

5. Conducted modeling and forecasting time series regression model. The steps to do the modeling are as follows :

   a. Identify stationary assumption on variance using Bartlett test. If the stationary assumptions on variance was not satisfied then performed a logit transformation [6].

   $$y_t = \ln\left(\frac{y_t}{1-y_t}\right) \quad (3)$$

   b. Identify stationary assumption on mean with Augmented Dickey-Fuller (ADF) test. If the assumption is not satisfied then performed differencing (d and D).
   c. Conducted regression modeling using response variables (Y) and explanatory variables (X) in Table 1 with residual of model follow ARIMA (2, d, 0) for non-seasonal data and SARIMA (2, d, 0) $(1, D, 0)_m$ for seasonal data.
   d. Performed ARIMA or SARIMA modeling on the residuals of the regression model. The steps given as follows :

      i. Specified the initial model based on the smallest Akaike's Information Criterion Bias Corrected (AICc) values of the model:

- Non-seasonal data : ARIMA (2,d,2), ARIMA (0,d,0), ARIMA (1,d,0) and ARIMA (0,d,1).
- Seasonal data : SARIMA (2,d,2) $(1,D,1)_m$, SARIMA (0,d,0) $(0,D,0)_m$, SARIMA (1,d,0) $(1,D,0)_m$ and SARIMA (0,d,1) $(0,D,1)_m$.

ii. Determined the new model based on the smallest AICc value by performed simulation on the initial model order.
- Simulation 1: added one of the order *p*,*q*,*P* and *Q* with ± 1.
- Simulation 2: added *p* and *q* order with ± 1.
- Simulation 3: added *P* and *Q* order with ± 1.
- Simulation 4: with and without constant.

iii. Repeat point (ii) on the new model. This process stops if the new model obtained has an AICc value greater than the previous new model.

e. Conducted back regression modeling with residual of model follow the order of ARIMA or SARIMA model from point d.

f. Checked the residual assumption of model and performed evaluation of prediction results of broad proportion of rice blast attacks in 2010-2015 with training data using MAPE.

g. Perform forecasting for the next 12 months and evaluate forecasting results with testing data using MAPE.

6. The best model determined by the smallest MAPE value on the predictions and forecasts of the classical regression model and the time series regression model.

4. Results

4.1 Data Exploration

The data exploration conducted using plot of data to obtain an overview of the pattern of broad proportion of rice blast attacks in 2010-2016 as shown in Fig. 1.

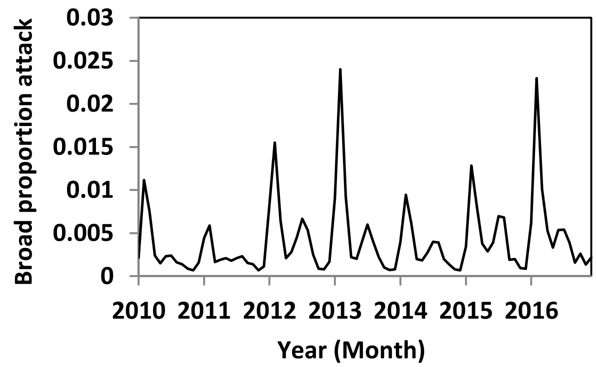

Fig. 1 Broad proportion of rice blast attacks

The largest broad proportions of rice blast attacks are always occurring at the beginning of the year (February) and mid-year (July-August) as shown in Fig. 1. The similarity of patterns over a period of time suggests that the broad proportion of attacks follows a seasonal pattern. Fig. 1 also shows that there are highly enough proportion of attacks at a certain time (outliers). The highly enough proportion of attacks make the precision of the prediction decreases so that transformation data performed to solved this problem.

4.2 Classical regression model

Forecasting the broad proportion of rice blast attacks performed using a classical regression model. Modeling is done on training data of 2010-2015 using physical environment factor (X) by performing logit transformation on response variable (Y) as shown in Table 1. The structure model of broad proportion of rice blast attacks is given as follows.

$$\ln\left(\frac{Y_t}{1-Y_t}\right) = -13.984 + 0.21X_1 - 0.271X_2 + 0.553X_3 + 0.382X_4 - 0.527X_5 + 0.218X_6 + 12.875X_7 + 2.905X_8 + \varepsilon_t \quad (4)$$

The coefficient of model can be interpreted as how much linear relationship of physical environmental factors to the ratio of the proportion of attack area and the proportion of area not affected. For example, the interpretation of the physical environmental factors of length exposed to light (X_4). if the length of exposed to light increases by one hour and other physical environmental factors is assumed constant then the ratio of the proportion of rice blast attack area and the proportion of area not affected increases by exp (0.382) =

1.465 times. The evaluation of residual assumptions is shown in Table 2.

Table 2: Model adequacy checking for classical regression model

Assumptions	Tests	Description
Autocorrelation	Durbin Watson	Positively autocorrelated
Normality	Shapiro Wilks	Normal
Constant variance	Breusch Pagan	Constant variance

Table 2 shows that there is autocorrelation problem on residual of model. Autocorrelation on residual of model causes the estimation of regression coefficients to be inefficient, thus potentially reducing the precision of prediction and forecast [4].

Prediction and forecast of the broad proportion of rice blast attacks are shown in Fig. 2. Figure 2 shows that the prediction in 2010-2015 and the forecast in 2016 of the broad proportion of rice blast attacks generally follows pattern of observation data with lower values than observation data.

This shows that prediction and forecast obtained only serve as an indicator of the arrival of blast disease population causing attacks on rice commodities so that, predictions and forecast can’t explain the actual value of broad proportion of rice blast attack.

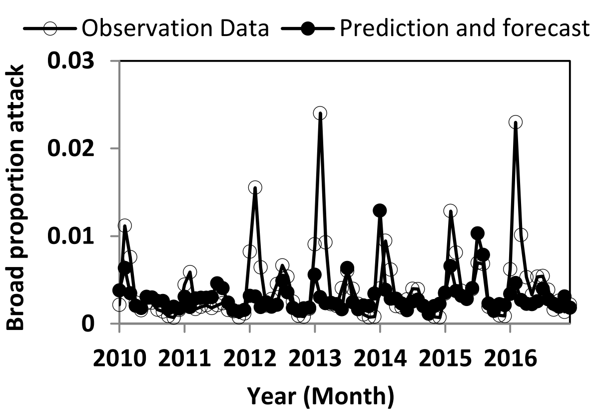

Fig. 2 Observation data of broad proportion of rice blast attack with the results of prediction and forecast using classical regression model

Evaluation of prediction results using training data and prediction results using testing data. MAPE values for each prediction and forecast were 60,819% and 51.2%. This shows that the average percentage of prediction and forecast errors compared to the actual value are 60,819% and 51.2% so that the prediction and forecast results is still not good enough.

4.3 Time series regression model

The time series regression model performed on the training data by using physical environment factor (X) by performing logit transformation on the response variable (Y) as shown in Table 1. The model structure obtained is the regression model with residual of model follows SARIMA (0,1,2) $(0,1,1)_{12}$.

$$\begin{aligned}\ln\left(\frac{Y_t}{1-Y_t}\right) &= -0.032\,X_1 + 0.605\,X_2 - 0.711\,X_3 - 0.251\,X_4 \\ &- 0.507\,X_5 - 0.176\,X_6 + 3.502\,X_7 + 1.146\,X_8 \\ &+ \varepsilon_{t-1} + \varepsilon_{t-12} - \varepsilon_{t-13} + w_t + 0.420\,w_{t-1} \\ &+ 0.492\,w_{t-2} + 0.634\,w_{t-12} + 0.266\,w_{t-13} \\ &+ 0.311\,w_{t-14} \end{aligned} \tag{5}$$

The coefficients model can be interpreted as how much linear relationship of physical environmental factors to the ratio of the proportion of attack area and the proportion of area not affected.

For example, the interpretation of the physical environmental factors of average wind velocity (X_3). if the average wind speed increased by one knot and other physical environmental factors influences is assumed constant then the ratio of the proportion of attack area and the broad proportion of not affected increases exp(-0.711) = 0.491 times. Model evaluation is shown in Table 3.

Table 3: Model adequacy checking for time series regression model

Assumptions	Tests	Description
Autocorrelation	Ljung-Box	Negatively autocorrelated
Normality	Shapiro Wilks	Normal
Constant variance	McLeod-Li	Constant variance

Table 3 shows that no model assumptions are violated so as to predict the broad proportion of rice blast attacks in 2010-2015 and forecasting in 2016 as shown in Figure 3. Predicted results evaluation is done using training data and forecast results using data testing.

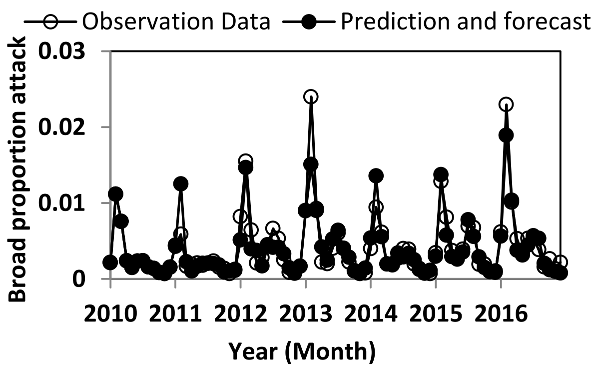

Fig. 3 Observation data of broad proportion of rice blast attack with the results of prediction and forecast using time series regression

Figure 3 shows that the prediction and forecast of broad proportion of rice blast attack generally follows pattern of observation data. The prediction and forecast value is not significantly different from the observation data. This shows that predictions and forecast obtained not only serve as indicators of the arrival of blast disease population causing attacks on rice commodities but predictions and forecast can also explain the actual value of broad proportion of rice blast attack.

Evaluation of predicted results is done using training data and forecast results using testing data. The MAPE value for each prediction and forecast is 21.242% and 24.533%. This shows that the average percentage of predictions and forecast of error compared with the actual value are 21.242% and 24.533% so that the prediction and forecast results looks good.

4.4 The best model

The best model determined using MAPE value comparison. The MAPE value comparison performed to find out the best model based on prediction and forecast obtained as shown in Table 4.

Table 4: MAPE value comparison

Model	Prediction (%)	Forecast (%)
Classical regression	60.819	51.2
Time series regression	21.242	24.533

Table 4 shows that prediction and forecast in the time series regression model are better than the classical regression models. The results of the prediction evaluation that is not significantly different from the forecast evaluation indicates that the model obtained is good enough in doing the forecasting. In addition to predictive values and better forecasts, the estimation of regression coefficients is more efficient because no assumptions are violated.

5. Conclusions

Forecasting using the time series regression model is better than the classical regression model with the average percentage of forecast error compared to the actual value are 24.533%. The results of the forecast obtained not only serve as an indicator of the arrival of blast disease population causing attacks on rice commodities but the forecast can also explain the actual value of broad proportion of rice blast attack.

References

[1] 1. M. S. Sinaga, Dasar-dasar ilmu penyakit tumbuhan, Jakarta, Indonesia : Penebar swadaya, 2003.

[2] Trisnaningsih, A. Nasution, “Respons ketahanan berbagai galur padi rawa terhadap wereng coklat, penyakit blas dan hawar daun bakteri”, Biodiversitas, Vol. 2, 2016, pp. 85-92.

[3] Sudir, A. Nasution, Santoso, B. Nutryanto, Penyakit Blas Pyricularia grisea pada tanaman padi dan strategi pengendaliannya. Jakarta, Indonesia : Buletin Iptek Tanaman Pangan KEMENTAN, 2014.

[4] D. C. Montgomery, L. C. Jennings and M. Kulahci, Introduction to Time series Analysis and Forecasting. New Jersey, US : John Wiley & Sons Inc, 2008.

[5] H.K Michael, J. N. Christopher, N. John, W. Li, Applied Linear Statistical Models Fifth Edition, NewYork, US: McGraw-Hill, 2005.

[6] B. McCune, J. B. Grace, Analysis of Ecological Communities, Oregon, US : MjM Software Design, 2002.

Authors -

I. Setiawan master student in Department of Statistics, Bogor Agricultural University. His main interests is on Statistical Modelling.

I.M. Sumertajaya Currently worked as a lecture in Department of Statistics, Bogor Agricultural University. His main interests is on Statistical Modelling Design of Experimental and Sampling Methodology.

F.M. Afendi Second Author Currently worked as a lecture in Department of Statistics, Bogor Agricultural University. His main interests is on Geoinformatics.