GSTARX-SUR Modeling Using Inverse Distance Weighted Matrix and Queen Contiguity Weighted Matrix for Forecasting Cocoa Black Pod Attack in Trenggalek Regency

Arif Ashari[1], Achmad Efendi[2], Henny Pramoedyo[3]
{arifashari@student.ub.ac.id[1]}

University of Brawijaya, Indonesia[1, 2, 3]

Abstract. Generalized Space Time Autoregressive (GSTAR) is one of the multivariate time series models considering heterogenic location. One of the GSTAR model developments is GSTARX model with additional exogenous variables. Parameters of GSTAR model could be estimated using Seemingly Unrelated Regression (SUR) approach to cope with the residual model between locations generally related to each other. The model is commonly called GSTARX-SUR. It is applicable in various fields such as agricultural sector. In this research, GSTARX-SUR model was applied to predict cocoa black pod attack in Trenggalek Regency. Rainfall was used as the exogenous variable. One of the characteristics of GSTARX model is the spatial weights. The correct spatial weights in GSTARX model is expected to improve the accuracy result. The research aimed to obtain the best GSTARX model to predict cocoa black pod attack in Trenggalek Regency. The research findings showed that GSTARX-SUR model (1,[1,12])(0,0,0) using inverse distance weighted matrix was the best model. The prediction result was highly accurate, indicated by a small MAPE value less than 15%.

Keyword: GSTAR, GSTARX, Forecasting, Cocoa, Black Pod.

1 Introduction

Generalized Space Time Autoregressive (GSTAR) is one of the space-time models. It is a distinctive form of Vector Autoregressive (VAR) model. Additionally it takes into account the location where data are collected. GSTAR model was also a general form of Space Time Autoregressive (STAR) model that allows different parameter value between locations. Therefore, GSTAR model is more flexible than STAR model [1]. GSTAR model was introduced by Ruchjana in 2002 using the Ordinary Least Square (OLS) method as a method of parameter estimation. However, researchers found weaknesses in OLS method. They argued that when the residual between locations were correlated, the result of parameter estimation was ineffective. Following the issue, Iriany *et al.* introduced Seemingly Unrelated Regression (SUR) approach using Generalized Least Square (GLS) as a method of parameter estimation [2]. Their findings showed that even though the residual between locations were correlated, SUR approach was able to provide more efficient estimator than OLS method.

There had been many researches on GSTAR model, either on its model development or on application. One of the GSTAR model developments is exogenous variable added into the

model. Suhartono *et al.* developed GSTARX-GLS model applied in the prediction of inflation rate in four big cities in East Java with increased fuel oil as the non-metrix exogenous variable [3]. They found that GSTARX was better than VARIMAX. Another researchers, Andayani et al. compared GSTARIMA to GSTARIMAX in approaching the data of rice price in six provinces in Java [4]. The findings showed that GSTARIMAX was better than GSTARIMA. Based on both research, the addition of exogenous variable in the GSTAR model was able to increase the prediction accuracy.

Another development of GSTAR model was the determination of spatial weighting matrix, a distinctive feature of GSTAR model. Several types of spatial weighting were developed i.e. uniform weight, queen contiguity weight, inverse distance weight, and non-geographical weight i.e. using inter-correlation weight, spatial clustering weight, etc. Another development and application of GSTAR model was the model applied in seasonal data. The study was conducted by Setiawan *et al.* in 2016. They developed S-GSTAR-SUR model and did a case study on the number of tourists visiting four tourism objects in Indonesia [5].

GSTAR and GSTARX models have been widely used in various fields of life, including in agriculture. In this research GSTAR model was used to predict cocoa black pod attack. The disease is categorized as a serious problem in cocoa cultivation as it can decline cocoa production by 26-50% [6]. It certainly caused tremendous loss for farmers. Cocoa black pod disease might attack anytime. It might infect the fruit when it was still small to when it was already ripe. The infection was marked by color changing, started from the tip of the fruit or near to the stamp and quickly extended to all parts of the fruit. The fruit was rotten within 12-22, making the fruit color turn to black [6]. Cocoa black pod attack was caused by *Phytophthora* sp. Severity of the black pod attack was affected by many factors i.e. rainfall, air humidity, planting method, number of fruit on a tree, and plant type. High air humidity facilitated spore formation and increased infection. The infection spread on the fruit surface the same exposed by any water i.e. rain water. However, it also could be caused by water produced by condensation of water vapor on the surface of the fruit. Rain also facilitated the spread of spore in addition to increase field humidity. Fluctuation of disease intensity tended to be in line with that of daily rainfall. The peak of disease intensity occurred in 1-3 weeks following the peak of rainfall [6].

Referring to this condition and previous research on development and application of GSTAR model, we analyzed the application of GSTARX-SUR model using inverse distance weighted matrix and queen contiguity weighted matrix to predict cocoa black pod attack. We conducted our research at one of the cocoa centers in East Java that was Trenggalek Regency. The GSTARX model obtained is expected to provide accurate forecasts as a reference for cocoa farmers and stakeholders in order to anticipate prevent and suppress the development of cocoa pod rot disease in Trenggalek Regency.

2   Research Methods

2.1 Data

Data used in this research were secondary data, namely data of cocoa black pod disease and monthly rainfall in 14 sub-districts in Trenggalek, East Java. Data on cocoa black pod disease was obtained from BBPPTP Surabaya, while rainfall data was obtained from NASA's website. The data were divided into two, namely in-sample data (January 2013 - December

2017) to be used in building the model and out-sample data (January 2018 - December 2018) for use in model validation.

2.2 Method of Analysis

Stages of GSTARX-SUR modeling in this research consisted of: (1) Testing data stationarity; (2) Identifying model order; (3) Calculating spatial weight; (4) Estimating and testing significance of the model parameter; (5) Examining model diagnostic; (6) Forecasting; and (7) Selecting the best model.

2.2.1 GSTARX Model

GSTARX model with autoregressive order (p), spatial order (λ), and transfer function order (b, r, s) can be formulated as follows:

$$\boldsymbol{z}_{(t)} = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \boldsymbol{\phi}_{kl} \boldsymbol{W}^{(l)} \boldsymbol{z}_{(t-k)} + \boldsymbol{\omega}(\boldsymbol{B})[\boldsymbol{\delta}(\boldsymbol{B})]^{-1} \boldsymbol{W}^{(l)} \boldsymbol{X}_{(t-b)} + \boldsymbol{\varepsilon}_{(t)} \quad (2.1)$$

where:

$\boldsymbol{z}_{(t)}$: (Nx1) observation vector in t-time

$\boldsymbol{\lambda}_k$: spatial order from AR to k

$\boldsymbol{\phi}_{kl}$: diagonal matrix with diagonal elements as AR and space-time on each location ($\Phi_{kl(1)}, ..., \Phi_{kl(N)}$)

$\boldsymbol{X}_{(t-b)}$: variable X vector with a size of $m \times 1$ in $t - b$ time

$\boldsymbol{\omega}(\boldsymbol{B})[\boldsymbol{\delta}(\boldsymbol{B})]^{-1}$: diagonal matrix of transfer function parameter with a size of $m \times m$, with :
$$\boldsymbol{\omega}(\boldsymbol{B}) = (\boldsymbol{\omega}_0 - \boldsymbol{\omega}_1 \boldsymbol{B} - \cdots - \boldsymbol{\omega}_s \boldsymbol{B}^s)$$
$$\boldsymbol{\delta}(\boldsymbol{B}) = (\boldsymbol{I} - \boldsymbol{\delta}_1 \boldsymbol{B} - \cdots - \boldsymbol{\delta}_r \boldsymbol{B}^r)$$

$\boldsymbol{\varepsilon}_{(t)}$: white noise with 0 mean vector and covariance matrix $\boldsymbol{\sigma}^2 \boldsymbol{I}$

2.2.2 Order Identification on GSTARX Model

Identification on GSTARX model is divided into two. Firstly identification of space time order covering identification of time order (p) and spatial order (λ). Secondly identification of transfer function orders (b, r, s). Spatial order is generally limited to order $\lambda = 1$. Time order (p) can be determined using MACF plot (autocorrelation function matrix) and MPACF (partial autocorrelation function matrix) [7]. Meanwhile transfer function order (b, r, s) can be determined using CCF plot (Cross Correlation Function) between input sequence (exogenous variable) and output sequence (endogenous variable) after going through the pre-whitening process [8].

2.2.3 Seemingly Unrelated Regression (SUR)

Seemingly Unrelated Regression (SUR) is an equation system consisting of multiple regression equations where the residual is not correlated to the observation in one equation, but correlated to one equation to another. To test the residual correlation between equations Lagrange Multiplier (LM) a statistic test should be conducted [9].

SUR model with equation m is as follows [10]:

$$Z = X\beta + \varepsilon \qquad (2.2)$$

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Assumption in SUR model is $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma_{ij} I_T$ where $i,j = 1,2, \dots, m$. Covariance matrix is stated by Ω as follows:

$$\Omega = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1m} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & \cdots & \sigma_{2m} I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} I_T & \sigma_{m2} I_T & \cdots & \sigma_{mm} I_T \end{bmatrix} = \Sigma \otimes I_T \qquad (2.3)$$

where matrix Ω had a size of $(NxT) x (NxT)$

Parameter estimation in GSTAR-SUR and GSTAR-X-SUR models are performed by applying Generalized Least Square (GLS), or minimizing the total generalized square $\varepsilon'\Omega^{-1}\varepsilon$. The result of GLS estimation for GSTAR-SUR and GSTAR-X-SUR models is obtained by the following formula [5]:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z \qquad (2.4)$$

2.2.4 Spatial Weight Matrix

One of the distinctive features in space time model is spatial weight matrix. Spatial weight matrix is a tool to model relation between locations. Before spatial weight matrix is made, it is necessary to formulate spatial-neighbours matrix. The spatial-neighbours matrix can be illustrated as follows:

$$c_{ij} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{in} & c_{in} & \cdots & c_{nn} \end{bmatrix}$$

a. Inverse Distance Weighted Matrix

Distance between locations can be determined using Euclid distance. Euclid distance is defined as the shortest line between two points. Weighting using inverse distance method is performed based on latitude and longitude coordinates from the central point of location being observed [11]. If u and v represent latitude and longitude coordinates of the location and d_{ij} represent the location distance of the $-i^{th}$ on the location of the $-j^{th}$, so it is found:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \qquad (2.5)$$

and an element of the spatial-neighbors matrix:

$$c_{ij} = 1/d_{ij} \qquad (2.6)$$

Therefore, the inverse distance weighted matrix is as follows:

$$w_{ij} = \begin{cases} \frac{c_{ij}}{\Sigma_j c_{ij}} & , untuk\ i \neq j \\ 0 & , untuk\ i = j \end{cases} \qquad (2.7)$$

b. Queen Contiguity Weighted Matrix

Queen contiguity weight is one of the spatial weight types using contiguity concept where edge and corner contiguity is used. *Contiguity Matrix* is a matrix describing a near relation between locations [11]. If location i is adjacent and directly tangent to the location j, the matrix element (i, j) has a value of 1. Meanwhile, if the location i is not tangent to the location j, the matrix element (i, j) has a value of 0. The form of this weighting matrix is as follows:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{in} & w_{in} & \cdots & w_{nn} \end{bmatrix}$$

where $w_{ij} = \frac{c_{ij}}{c_{i.}}$, with c_{ij} is the neighbours value of location i with location j, and $c_{i.}$ was the total neighbours value in line (location) i. The $\boldsymbol{W}$ matrix is the normalized queen contiguity weighted matrix.

2.2.5 Criteria for Selecting the Best Model

The best model was selected based on the accuracy of prediction result from the out-sample data. Criteria used in this research were Root Mean Square Error (RMSE), which meant that the best model was the model having the smallest RMSE value. RMSE was the more precise measurement when the residual followed the normal distribution [7]. RMSE value can be calculated using the formula as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \qquad (2.8)$$

With $e_t^2 = \left(Z_t - \hat{Z}_t\right)^2$ where n is the amount of time (t) being predicted; e_t is the model residual at the t-time; Z_t is the observation value at the t-time; and $\hat{Z}_t$ is the prediction value at the t-time.

3 Result and Discussion

3.1 Testing Data Stationarity

Testing data stationarity was to collect data with stable mean and variance. If stationary, the data were stable for forecasting process. It was because stable mean and variance were required to obtain model able to make an accurate prediction. Based on the data plot, the variance was still unstable, so it was necessary to perform Box-Cox transformation. After the data had been transformed and the variance was already constant, the mean was tested by stationary test using Dickey Fuller test. The testing result showed that the data had represented stationary mean, so differencing could be ignored.

3.2 Identifying Model Order

Identification of autoregressive order (p) was performed by referring to MPCCF plot, AICC value, and data plot. The AICC value is presented in Table 1.

Table 1. AICC Value

Minimum Information Criterion Based on AICC			
Lag	MA0	MA1	MA2
AR0	-71.78840	-68.80602	-60.16277
AR1	-78.93838	-70.49176	-49.08230
AR2	-69.85757	-43.31916	

Based on MPCCF plot, the smallest AICC value, and the data plot of percentage of cocoa black pod attack showing the 12-month seasonal pattern; cocoa black pod attack in a month was closely related to that occurring in the previous 1 month and 12 months. Therefore, autoregressive order in GSTARX model in this research was [1,12].

Next, transfer function order (b, r, s) was determined using CCF plot between input and output lines obtained after the pre-whitening process. There are 14 CCF plots, according to the number of locations. Figure 1 is a CCF plot at one location (Durenan Sub-district).
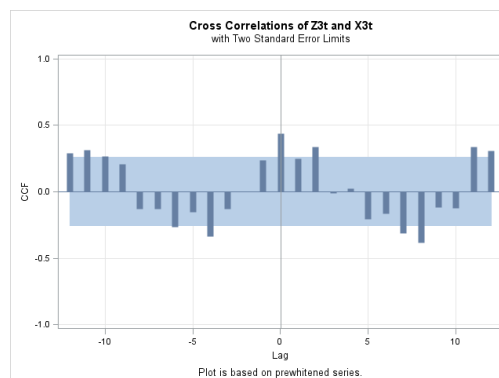

Fig. 1. CCF Plot at One Location.

Based on CCF plot in Figure 1, order *b* was 0 because the first significance lag was the 0th lag; order *r* was 0 because the CCF plot had several significant lag; while order s was 0 because the correlation value for lag 1 (lag after the first significance lag) was less than lag 0. Because most CCF plots had almost the same pattern, order *b, r, s* could be determined that was (0,0,0). By limiting the spatial order with spatial $\lambda = 1$, GSTARX-SUR model $(\lambda,p)(b,r,s)$ formed was GSTARX-SUR $(1,[1,12])(0,0,0)$ with the equation model as follows:

$$Z_{(t)} = \phi_{10}Z_{t-1} + \phi_{11}WZ_{t-1} + \phi_{120}Z_{t-12} + \phi_{121}WZ_{t-12} + \omega_0 X_t + \varepsilon_t$$

3.3 Calculating the Spatial Weight Matrix

3.3.1 Inverse Distance Weighted Matrix

The calculation of inverse distance weighted matrix is based on the distance between observation locations. The assumption was that cocoa black pod attack more highly occurred in a location closely related to other nearby location also attacked by the disease than that occurred in other further location. Here was the inverse distance weighted matrix (ID) among the 14 sub-districts in Trenggalek Regency:

$$W_{ij}(ID) = \begin{pmatrix}
0.00 & 0.04 & 0.07 & 0.08 & 0.05 & 0.10 & 0.04 & 0.04 & 0.10 & 0.05 & 0.07 & 0.19 & 0.13 & 0.04 \\
0.04 & 0.00 & 0.04 & 0.06 & 0.10 & 0.06 & 0.12 & 0.12 & 0.05 & 0.15 & 0.10 & 0.04 & 0.06 & 0.07 \\
0.07 & 0.04 & 0.00 & 0.11 & 0.06 & 0.09 & 0.04 & 0.03 & 0.22 & 0.04 & 0.06 & 0.11 & 0.07 & 0.06 \\
0.05 & 0.05 & 0.08 & 0.00 & 0.08 & 0.18 & 0.04 & 0.03 & 0.13 & 0.05 & 0.09 & 0.08 & 0.09 & 0.06 \\
0.04 & 0.09 & 0.06 & 0.10 & 0.00 & 0.08 & 0.11 & 0.05 & 0.06 & 0.07 & 0.09 & 0.05 & 0.06 & 0.13 \\
0.06 & 0.04 & 0.06 & 0.16 & 0.06 & 0.00 & 0.03 & 0.03 & 0.11 & 0.05 & 0.10 & 0.09 & 0.16 & 0.04 \\
0.04 & 0.14 & 0.05 & 0.07 & 0.14 & 0.06 & 0.00 & 0.08 & 0.05 & 0.08 & 0.07 & 0.04 & 0.05 & 0.12 \\
0.05 & 0.17 & 0.04 & 0.06 & 0.08 & 0.06 & 0.10 & 0.00 & 0.05 & 0.14 & 0.09 & 0.05 & 0.06 & 0.06 \\
0.07 & 0.04 & 0.17 & 0.13 & 0.05 & 0.12 & 0.03 & 0.03 & 0.00 & 0.04 & 0.06 & 0.13 & 0.09 & 0.04 \\
0.05 & 0.15 & 0.04 & 0.07 & 0.08 & 0.07 & 0.07 & 0.10 & 0.05 & 0.00 & 0.15 & 0.05 & 0.07 & 0.05 \\
0.05 & 0.08 & 0.05 & 0.10 & 0.08 & 0.12 & 0.05 & 0.05 & 0.06 & 0.12 & 0.00 & 0.06 & 0.11 & 0.05 \\
0.15 & 0.04 & 0.09 & 0.09 & 0.05 & 0.12 & 0.03 & 0.03 & 0.15 & 0.04 & 0.06 & 0.00 & 0.13 & 0.04 \\
0.09 & 0.04 & 0.06 & 0.09 & 0.05 & 0.19 & 0.04 & 0.03 & 0.09 & 0.06 & 0.11 & 0.12 & 0.00 & 0.04 \\
0.04 & 0.08 & 0.07 & 0.09 & 0.17 & 0.07 & 0.12 & 0.05 & 0.07 & 0.06 & 0.07 & 0.05 & 0.06 & 0.00
\end{pmatrix}$$

3.3.2 Queen Contiguity Weight Matrix

The calculation of queen contiguity weight matrix was based on the concept of neighbors or intersection between locations, or the intersection of the edge and corner. The assumption was that cocoa black pod attack in a certain location was affected by attack occurring in another location nearby (neighbor location) and not by attack from any far location. Here was the queen contiguity (QC) weight matrix among 14 sub-districts in Trenggalek Regency:

$$W_{ij}(QC) = \begin{pmatrix}
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.50 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.20 & 0.00 & 0.20 & 0.20 & 0.00 & 0.20 & 0.20 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.00 & 0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.17 & 0.00 & 0.17 & 0.00 & 0.17 & 0.17 & 0.00 & 0.00 & 0.00 & 0.17 & 0.00 & 0.00 & 0.17 \\
0.00 & 0.00 & 0.00 & 0.17 & 0.17 & 0.00 & 0.00 & 0.00 & 0.00 & 0.17 & 0.17 & 0.17 & 0.17 & 0.00 \\
0.00 & 0.25 & 0.00 & 0.00 & 0.25 & 0.00 & 0.00 & 0.25 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.25 \\
0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 & 0.33 & 0.00 & 0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.33 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.33 & 0.00 & 0.00 \\
0.00 & 0.20 & 0.00 & 0.00 & 0.00 & 0.20 & 0.00 & 0.20 & 0.00 & 0.00 & 0.20 & 0.00 & 0.20 & 0.00 \\
0.00 & 0.25 & 0.00 & 0.00 & 0.25 & 0.25 & 0.00 & 0.00 & 0.00 & 0.25 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.20 & 0.00 & 0.00 & 0.20 & 0.00 & 0.20 & 0.00 & 0.00 & 0.20 & 0.00 & 0.00 & 0.00 & 0.20 & 0.00 \\
0.25 & 0.00 & 0.00 & 0.20 & 0.00 & 0.25 & 0.00 & 0.00 & 0.00 & 0.25 & 0.00 & 0.25 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00
\end{pmatrix}$$

3.4 Estimation of GSTARX Model Parameter

Estimation of GSTARX model parameter was carried out to produce residual. Next, from the residual obtained using OLS method, the residual variance covariance was obtained to be used to estimate model parameter with SUR approach. Briefly, the result of estimation of GSTARX-SUR model parameter using inverse distance weighted matrix is presented in Table 2.

Table 2. The Result of Estimation of GSTARX-SUR Model Parameter Using Inverse Distance Weighted Matrix (ID)

Location	β	$\hat{\beta}$	p value	Sig.	Eff. (%)	Location	β	$\hat{\beta}$	p value	Sig.	Eff. (%)
Bendungan	$\varphi_{10}^{(1)}$	0.243	0.018	*	77.18	Panggul	$\varphi_{10}^{(8)}$	0.506	0.000	**	71.57
(Z_1)	$\varphi_{120}^{(1)}$	0.048	0.257	NS	67.78	(Z_8)	$\varphi_{120}^{(8)}$	-0.136	0.207	NS	68.86
	$\varphi_{11}^{(1)}$	0.518	0.000	**	74.87		$\varphi_{11}^{(8)}$	0.146	0.315	NS	65.96
	$\varphi_{121}^{(1)}$	0.138	0.052	NS	50.88		$\varphi_{121}^{(8)}$	0.375	0.014	*	61.15
	$\omega_0^{(1)}$	0.006	0.025	*	40.59		$\omega_0^{(8)}$	0.008	0.045	*	41.27
Dongko	$\varphi_{10}^{(2)}$	0.419	0.001	**	71.87	Pogalan	$\varphi_{10}^{(9)}$	0.487	0.000	**	66.71
(Z_2)	$\varphi_{120}^{(2)}$	0.121	0.044	*	65.06	(Z_9)	$\varphi_{120}^{(9)}$	0.124	0.139	NS	63.57
	$\varphi_{11}^{(2)}$	0.320	0.027	*	68.69		$\varphi_{11}^{(9)}$	0.362	0.169	NS	62.63
	$\varphi_{121}^{(2)}$	0.102	0.308	NS	52.98		$\varphi_{121}^{(9)}$	0.060	0.796	NS	58.72
	$\omega_0^{(2)}$	0.008	0.017	*	38.25		$\omega_0^{(9)}$	0.012	0.088	NS	35.86
Durenan	$\varphi_{10}^{(3)}$	0.572	0.000	**	71.69	Pule	$\varphi_{10}^{(10)}$	0.849	0.000	**	56.67
(Z_3)	$\varphi_{120}^{(3)}$	0.241	0.002	**	72.56	(Z_{10})	$\varphi_{120}^{(10)}$	0.149	0.005	**	56.81
	$\varphi_{11}^{(3)}$	0.156	0.122	NS	68.32		$\varphi_{11}^{(10)}$	-0.053	0.234	NS	40.42
	$\varphi_{121}^{(3)}$	-0.059	0.510	NS	66.45		$\varphi_{121}^{(10)}$	0.022	0.624	NS	41.16
	$\omega_0^{(3)}$	0.009	0.011	*	41.61		$\omega_0^{(10)}$	0.003	0.082	NS	33.20
Gandusari	$\varphi_{10}^{(4)}$	0.427	0.000	**	78.74	Suruh	$\varphi_{10}^{(11)}$	0.522	0.000	**	76.16
(Z_4)	$\varphi_{120}^{(4)}$	0.155	0.009	**	80.14	(Z_{11})	$\varphi_{120}^{(11)}$	0.042	0.557	NS	77.63
	$\varphi_{11}^{(4)}$	0.375	0.001	**	72.92		$\varphi_{11}^{(11)}$	0.196	0.057	NS	68.96

Location	β	β̂	p value	Sig.	Eff. (%)	Location	β	β̂	p value	Sig.	Eff. (%)
	$\varphi_{121}^{(4)}$	-0.014	0.897	NS	70.04		$\varphi_{121}^{(11)}$	0.124	0.238	NS	69.61
	$\omega_0^{(4)}$	0.010	0.006	**	42.01		$\omega_0^{(11)}$	0.008	0.008	**	41.46
Kampak	$\varphi_{10}^{(5)}$	0.844	0.000	**	68.71	Trenggalek	$\varphi_{10}^{(12)}$	0.232	0.062	NS	39.64
(Z_5)	$\varphi_{120}^{(5)}$	0.208	0.000	**	68.52	(Z_{12})	$\varphi_{120}^{(12)}$	-0.091	0.122	NS	39.05
	$\varphi_{11}^{(5)}$	-0.066	0.079	NS	56.44		$\varphi_{11}^{(12)}$	0.322	0.018	*	29.54
	$\varphi_{121}^{(5)}$	-0.016	0.675	NS	57.90		$\varphi_{121}^{(12)}$	0.347	0.009	**	25.38
	$\omega_0^{(5)}$	0.004	0.009	**	39.52		$\omega_0^{(12)}$	0.007	0.186	NS	20.78
Karangan	$\varphi_{10}^{(6)}$	0.771	0.000	**	71.05	Tugu	$\varphi_{10}^{(13)}$	0.734	0.000	**	69.18
(Z_6)	$\varphi_{120}^{(6)}$	0.222	0.000	**	70.88	(Z_{13})	$\varphi_{120}^{(13)}$	0.301	0.000	**	69.43
	$\varphi_{11}^{(6)}$	-0.020	0.657	NS	58.56		$\varphi_{11}^{(13)}$	-0.006	0.909	NS	58.13
	$\varphi_{121}^{(6)}$	-0.017	0.732	NS	57.36		$\varphi_{121}^{(13)}$	-0.069	0.203	NS	58.42
	$\omega_0^{(6)}$	0.004	0.025	*	41.14		$\omega_0^{(13)}$	0.004	0.012	*	40.09
Munjungan	$\varphi_{10}^{(7)}$	0.797	0.000	**	46.17	Watulimo	$\varphi_{10}^{(14)}$	0.723	0.000	**	65.50
(Z_7)	$\varphi_{120}^{(7)}$	-0.055	0.447	NS	47.73	(Z_{14})	$\varphi_{120}^{(14)}$	0.331	0.000	**	65.85
	$\varphi_{11}^{(7)}$	-0.325	0.161	NS	33.22		$\varphi_{11}^{(14)}$	0.005	0.916	NS	52.87
	$\varphi_{121}^{(7)}$	0.597	0.025	*	33.20		$\varphi_{121}^{(14)}$	-0.105	0.049	*	55.92
	$\omega_0^{(7)}$	0.010	0.258	NS	29.81		$\omega_0^{(14)}$	0.005	0.009	**	37.51

Meanwhile the result of estimation of GSTARX-SUR model parameter using queen contiguity weighted matrix is presented in Table 3.

Table 3. The Result of Estimation of GSTARX-SUR Model Parameter
Using Queen Contiguity Weighted Matrix (QC)

Location	β	β̂	p value	Sig.	Eff. (%)	Location	β	β̂	p value	Sig.	Eff. (%)
Bendungan	$\varphi_{10}(1)$	0.417	0.000	**	67.41	Panggul	$\varphi_{10}(1)$	0.581	0.000	**	67.85
(Z1)	$\varphi_{120}(1)$	0.069	0.135	NS	54.88	(Z8)	$\varphi_{120}(1)$	0.042	0.738	NS	66.48
	$\varphi_{11}(1)$	0.373	0.000	**	71.03		$\varphi_{11}(1)$	0.079	0.411	NS	67.54
	$\varphi_{121}(1)$	0.051	0.175	NS	62.33		$\varphi_{121}(1)$	0.108	0.291	NS	67.55
	$\omega_0(1)$	0.012	0.000	**	39.24		$\omega_0(1)$	0.013	0.002	**	43.55
Dongko	$\varphi_{10}(2)$	0.706	0.000	**	64.82	Pogalan	$\varphi_{10}(1)$	0.464	0.000	**	62.42
(Z2)	$\varphi_{120}(2)$	0.163	0.032	*	57.34	(Z9)	$\varphi_{120}(1)$	0.174	0.020	*	55.43
	$\varphi_{11}(2)$	-0.042	0.784	NS	64.08		$\varphi_{11}(1)$	0.356	0.089	NS	62.04
	$\varphi_{121}(2)$	0.032	0.714	NS	57.19		$\varphi_{121}(1)$	-0.136	0.293	NS	60.44
	$\omega_0(2)$	0.014	0.000	**	34.68		$\omega_0(1)$	0.024	0.002	**	34.69
Durenan	$\varphi_{10}(3)$	0.666	0.000	**	69.12	Pule	$\varphi_{10}(1)$	0.887	0.000	**	57.85
(Z3)	$\varphi_{120}(3)$	0.176	0.004	**	67.92	(Z10)	$\varphi_{120}(1)$	0.115	0.083	NS	57.77
	$\varphi_{11}(3)$	0.017	0.628	NS	71.77		$\varphi_{11}(1)$	-0.085	0.122	NS	53.40

Location	Parameter	Value	p-value	Sig.		Location	Parameter	Value	p-value	Sig.	
	$\varphi 121(3)$	-0.018	0.526	NS	69.69		$\varphi 121(1)$	0.014	0.790	NS	54.57
	$\omega 0(3)$	0.014	0.000	**	39.94		$\omega 0(1)$	0.005	0.003	**	33.40
Gandusari (Z4)	$\varphi 10(4)$	0.502	0.000	**	74.87	Suruh (Z11)	$\varphi 10(11)$	0.660	0.000	**	69.04
	$\varphi 120(4)$	0.102	0.067	NS	74.01		$\varphi 120(11)$	0.010	0.865	NS	72.74
	$\varphi 11(4)$	0.237	0.003	**	74.34		$\varphi 11(11)$	0.000	0.998	NS	68.06
	$\varphi 121(4)$	0.010	0.882	NS	70.20		$\varphi 121(11)$	0.173	0.043	*	71.00
	$\omega 0(4)$	0.016	0.000	**	39.86		$\omega 0(11)$	0.012	0.000	**	37.93
Kampak (Z5)	$\varphi 10(5)$	0.860	0.000	**	65.96	Trenggalek (Z12)	$\varphi 10(12)$	0.300	0.017	*	40.15
	$\varphi 120(5)$	0.172	0.001	**	66.90		$\varphi 120(12)$	-0.117	0.051	NS	37.93
	$\varphi 11(5)$	-0.079	0.003	**	67.38		$\varphi 11(12)$	0.243	0.050	NS	34.11
	$\varphi 121(5)$	-0.013	0.624	NS	69.87		$\varphi 121(12)$	0.309	0.009	**	31.55
	$\omega 0(5)$	0.006	0.000	**	38.43		$\omega 0(12)$	0.012	0.022	*	20.96
Karangan (Z6)	$\varphi 10(6)$	0.824	0.000	**	64.29	Tugu (Z13)	$\varphi 10(13)$	0.802	0.000	**	62.35
	$\varphi 120(6)$	0.164	0.014	*	68.20		$\varphi 120(13)$	0.249	0.000	**	65.05
	$\varphi 11(6)$	-0.064	0.257	NS	64.11		$\varphi 11(13)$	-0.069	0.218	NS	63.27
	$\varphi 121(6)$	-0.004	0.936	NS	68.83		$\varphi 121(13)$	-0.057	0.249	NS	66.65
	$\omega 0(6)$	0.006	0.001	**	40.18		$\omega 0(13)$	0.006	0.001	**	38.52
Munjungan (Z7)	$\varphi 10(7)$	0.758	0.000	**	44.41	Watulimo (Z14)	$\varphi 10(14)$	0.759	0.000	**	57.65
	$\varphi 120(7)$	-0.101	0.159	NS	46.60		$\varphi 120(14)$	0.265	0.000	**	59.22
	$\varphi 11(7)$	-0.414	0.046	*	37.09		$\varphi 11(14)$	-0.029	0.216	NS	60.73
	$\varphi 121(7)$	0.679	0.007	**	38.94		$\varphi 121(14)$	-0.061	0.020	*	63.60
	$\omega 0(7)$	0.020	0.029	*	28.15		$\omega 0(14)$	0.007	0.000	**	35.73

Based on the result of estimation of GSTARX-SUR model parameter, SUR approach produced more efficient parameter predictor than OLS method. Besides, from the result of estimation of model parameter, cocoa black pod attack in one sub-district in Trenggalek Regency was commonly affected by high and low rainfall intensity during that time and attack occurring in the previous one and twelve months in either sub-district or surrounding sub-districts.

3.5 Model Diagnostics

Model diagnostic was to ensure that the model residual was white noise and normally distributed. Diagnostic of white noise assumption was performed to see residual MPCCF plot; while normality test used QQ plot. The test result showed that residual of GSTARX-SUR model with inverse distance weighted matrix was white noise and normally distributed. Meanwhile, GSTARX-SUR model with queen contiguity weighted matrix was white noise, but there were two locations where the residuals were not normally distributed.

3.6 Forecasting

Based on the result of estimation of model parameter on either GSTARX-SUR model with inverse distance weighted matrix or GSTARX-SUR model with queen contiguity weighted matrix, the forecasting of cocoa black pod attack was processed for the next one month in every i^{th} location using the following equation:

$$\hat{Z}_{t+1}^{(i)} = \hat{\phi}_{10}^{(i)} Z_{it} + \hat{\phi}_{11}^{(i)} W Z_t + \hat{\phi}_{120}^{(i)} Z_{t-11} + \hat{\phi}_{121}^{(i)} W Z_{t-11} + \hat{\omega}_0^{(i)} X_{t+1}$$

Forecasting was performed using one step ahead method for the next 12 months using out-sample data. The following figure is the result of forecasting the GSTARX-SUR model with both weights in one sub-district Kampak.
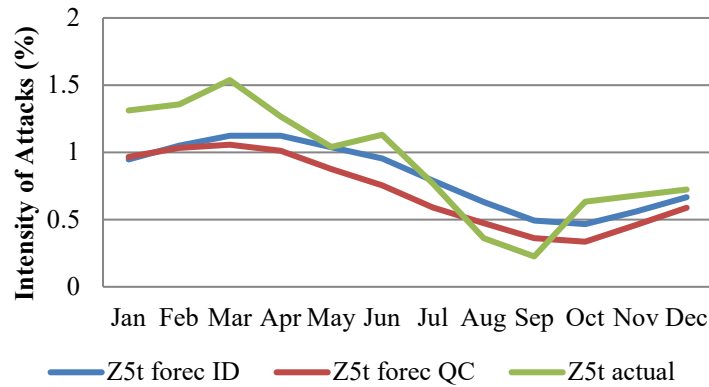

Fig. 1. GSTARX-SUR Forecasting Results at One Location

Based on Figure 1, it can be seen that the plot of the forecasting of cocoa pod rot black pod disease in Kampak formed a sine wave with the peak of the attack occurring in March 2018. The result of prediction showed the value of overall MAPE that was 12.97% and 13.46% for GSTARX-SUR model with inverse distance weighted matrix and GSTARX-SUR model with queen contiguity weighted matrix respectively. It indicated that GSTARX-SUR had a high accuracy.

3.7 Selecting the Best Model

To select a better GSTARX-SUR model between GSTARX-SUR model with inverse distance weighted matrix and queen contiguity weighted matrix, RMSE value from both prediction results were compared. The comparison of RMSE value from both models is presented in Table 4. Based on the comparison between RMSE values in Table 4, RMSE values of GSTARX-SUR model using inverse distance weighted matrix was smaller than RMSE values of GSTARX-SUR model using queen contiguity weighted matrix at 11 subdistricts. Generally, based on overall RMSE value, GSTARX-SUR model using inverse distance weighted matrix provided higher accuracy in forecasting cocoa black pod attack in Trenggalek than GSTARX-SUR model using queen contiguity weighted matrix.

Table 4. Comparison between RMSE Value of GSTARX-SUR (ID) Model and RMSE Value of GSTARX-SUR (QC) Model

| Location | RMSE | | Location | RMSE | |
	GSTARX-SUR (ID)	GSTARX-SUR (QC)		GSTARX-SUR (ID)	GSTARX-SUR (QC)
Bendungan (Z_1)	0.1810	0.1881	Panggul (Z_8)	0.2243	0.2343
Dongko (Z_2)	0.1681	0.1642	Pogalan (Z_9)	0.2751	0.2984
Durenan (Z_3)	0.2154	0.1731	Pule (Z_{10})	0.0657	0.0678
Gandusari (Z_4)	0.1830	0.1805	Suruh (Z_{11})	0.1711	0.1725
Kampak (Z_5)	0.0693	0.0716	Trenggalek (Z_{12})	0.4365	0.4479
Karangan (Z_6)	0.0750	0.0767	Tugu (Z_{13})	0.0778	0.0795
Munjungan (Z_7)	0.4573	0.4659	Watulimo (Z_{14})	0.0866	0.0895
Overall RMSE of GSTARX-SUR (ID) = 0.1919					
Overall RMSE of GSTARX-SUR (QC) = 0.1936					

4 Conclusion

GSTARX-SUR (1,[1,120])(0,0,0) using inverse distance weighted matrix was the best model with a high accuracy, indicated by a small MAPE value less than 15%. It means that this model can be relied for forecasting the cocoa black pod attack in Trenggalek Regency. For further research, the GSTARX-SUR model can be tried to be applied for forecasting pest and disease attacks on other plants that have similar characteristics with cocoa black pod disease. Recommendation for farmers are to maintain the humidity of the cocoa plantation by improving water sanitation before the rainy season.

References

[1] B. N. Ruchjana, S. A. Borovkova, and H. P. Lopuhaa, "Least squares estimation of Generalized Space Time AutoRegressive (GSTAR) model and its properties," in *AIP Conference Proceedings*, 2012, vol. 1450, no. 1, pp. 61–64.

[2] A. Iriany and B. N. Suhariningsih, "Ruchjana and Setiawan,' Prediction of Precipitation data at Batu Town using the GSTAR (1, p)-SUR Model,'" *J. Basic Appl. Sci. Res.*, vol. 3, no. 6, pp. 860–865, 2013.

[3] Suhartono, S. R. Wahyuningrum, Setiawan, and M. S. Akbar, "GSTARX-GLS model for spatio-temporal data forecasting," *Malaysian J. Math. Sci.*, vol. 10, pp. 91–103, 2016.

[4] N. Andayani, I. M. Sumertajaya, B. N. Ruchjana, and M. N. Aidi, "Comparison of GSTARIMA and GSTARIMA-X Model by using Transfer Function Model Approach to Rice Price Data," in *IOP Conference Series: Earth and Environmental Science*, 2018, vol. 187, no. 1, p. 12052.

[5] S. Setiawan and M. Prastuti, "S-GSTAR-SUR model for seasonal spatio temporal data forecasting," *Malaysian J. Math. Sci.*, vol. 10, pp. 53–65, 2016.

[6] H. Semangun, "Penyakit-penyakit tanaman perkebunan Indonesia (p. 835)." Yogyakarta: Gadjah Mada University Press, 2000.

[7] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–

Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[8]   W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate (2nd Edition)*. USA: Pearson Education, Inc, 2006.

[9]   W. H. Greene, *Econometrics Analysis (7th Edition)*. Boston: Prentice Hall, 2002.

[10]   S. C. Nisak, "Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts," *CAUCHY*, vol. 4, no. 2, pp. 57–64, 2016.

[11]   H. Pramoedyo, *Analisis Spasial Dasar*. Malang: Universitas Negeri Malang Press, 2017.