A New Procedure for Generalized STAR Modeling using IAcM Approach

Utriweni Mukhaiyar & Udjianna S. Pasaribu

Statistics Research Division, Institut Teknologi Bandung,
Jalan Ganesha 10 Bandung, Jawa Barat 40132, Indonesia
Email: utriweni@math.itb.ac.id

Abstract. A new procedure of space-time modeling through the Invers of Autocovariance Matrix (IAcM) is proposed. By evaluating the IAcM behaviors on behalf of the Generalized Space-Time Autoregressive (GSTAR) process stationarity, we may find an appropriate model to space-time data series. This method can complete the Space-Time ACF and PACF methods for identifying space-time models. For study case, we apply the GSTAR models to the monthly tea production of some plantations in West Java, Indonesia.

1 Introduction

An observation at a certain spatial location and time may be commonly influenced by previous-time observations in that location and its neighboring locations. For example, the oil production of a certain well during this month can be affected by the production during the previous months at that same location and at the neighboring wells; the number of today's criminal actions in a city may be influenced by the number of criminal actions that happened during the previous days in that city and the nearest cities; or, then the number of vehicles at a certain intersection within a certain time range is influenced by the number of vehicles at that location and related intersections during the previous time ranges. These examples are the portrait of space-time analysis problems.

Also, we would like to be able to forecast the oil production of that particular well during the coming months; how many criminal actions will happen in the city; and how many vehicles can be found at that particular intersection during the next range times. Knowing these forecasts, people in charge may take the necessary actions. We can do forecasting after obtaining the appropriate model. One way of doing this is space-time modeling.

In this paper we proffer a new procedure for space-time modeling, especially for Generalized STAR (GSTAR) modeling. For this purpose, we modify the Pfeifer and Deutsch three stage-iterative procedure for space-time modeling

(see [1]), which consist of identification, parameter estimation and diagnostic checking. We improve model selection by checking the stationarity of possible models using the IAcM. The IACM has been studied for autoregressive models of time-series by Rahardjo in 2005 [2]. We will explain these processes briefly in Section 3. Before that, we will describe the GSTAR model and spatial weight matrices in Section 2. In Section 4, we employ real data of the monthly tea production in some plantations in West Java. Finally Section 5 contains remarks and conclusions.

2   Generalized STAR Model

The observation at site i and time t, $Y_{i,t}$, which is a realization of a stochastic process, follows the GSTAR model if it can be declared as a linier combination of past observations for both time and spatial indices. A ($N \times 1$)-dimensional centered process vector $\mathbf{Y}_t = \left(Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t} \right)'$ follows the GSTAR($p; \lambda_1, \lambda_2, \ldots, \lambda_p$) model if $\mathbf{Y}_t$ can be presented as,

$$\mathbf{Y}_t = \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \mathbf{\Phi}_{k\ell} \mathbf{W}^{(\ell)} \mathbf{Y}_{t-k} + \mathbf{e}_t \tag{1}$$

with k is time autoregressive order ($k = 1, 2, \ldots, p$), ℓ is spatial autoregressive order ($\ell = 0, 1, \ldots, \lambda_k$), and $\mathbf{e}_t$ is ($N \times 1$)-dimensional vector of errors that assumed have i.i.d normal distribution. The $\mathbf{W}^{(\ell)}$ is ℓ-th order weight matrix which the main diagonal is zero and the sum of each row is one, and $\mathbf{\Phi}_{k\ell}$ is ($N \times N$) diagonal matrix which presents autoregressive parameter of k-th time order and ℓ-th spatial order for each location $i = 1, 2, \ldots, N$. We write $\mathbf{\Phi}_{k\ell} = \mathrm{diag} \left(\phi_{k\ell}^{(1)}, \phi_{k\ell}^{(2)}, \ldots, \phi_{k\ell}^{(N)} \right)$.
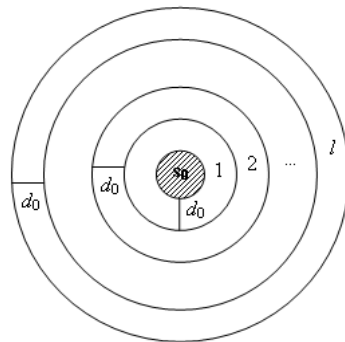

Figure 1 Spatial order/lag in two dimensional radius system. Sites in same 'category of distance' are in the same spatial order. The category of distance is classified by a fixed radius value d_0.

The ℓ-th spatial order shows the ordered neighbors of a number of spatial locations, while spatial weight $w_{ij}^{(\ell)}$ shows the influence of location j on location i in spatial lag ℓ-th. The locations which are closer to the reference location (s_0) will have a smaller spatial lag, which is distinguished by fixed distance d_0 (Figure 1). Since the configuration in each spatial lag order is distinct, different spatial lags may give different weight matrices.

2.1   Spatial Weights

There are three types of spatial weight: binary, uniform, and non-uniform. Binary weight matrix has values 0 and 1 in off-diagonal entries; uniform weight is determined by the number of sites surrounding a certain site in ℓ-th spatial order; and non-uniform weight gives unequal weight for different sites. The element of the uniform weight matrix is formulated as,

$$w_{ij}^{(\ell)} = \begin{cases} \dfrac{1}{n_i^{(\ell)}} & , j \text{ is neighbor of } i \text{ in } \ell\text{-th order} \\ 0 & , \text{ others} \end{cases}$$

$n_i^{(\ell)}$ is the number of neighbor locations with site-i in ℓ-th order.

The non-uniform weight may become uniform weight when some conditions are met. One method in building non-uniform weight is based on inverse distance, which is used by Borovkova in 2008 [3]. The weight matrix of spatial lag k is based on the inverse weights $\dfrac{1}{1+d_{ij}}$ for sites i and j whose Euclidean distance d_{ij} lies within a fixed distance range, and otherwise is weight zero.

2.2   Stationarity

GSTAR models are invertible since the vector of observations $\mathbf{Y}_t$ could be expressed as a weighted linear combination of past observations with weights that converge to zero. In order for a GSTAR model to represent a stationary process, one in which the covariance structure of $\mathbf{Y}_t$ does not change with time, certain conditions should be met. We obtain the stationary conditions of the first and second order of GSTAR models using its IAcM by the following propositions.

Proposition 1 Suppose that ($N{\times}N$)-dimensional matrix $\mathbf{A} = \mathbf{\Phi_{10}} + \mathbf{\Phi_{11}W}$ and IAcM, $\mathbf{M_1} = \mathbf{I}_N - \mathbf{A'A}$, is invers of autocovariance matrix whose elements are the parameters of GSTAR(1;1) process. If the determinants of all the leading principal submatrices of IAcM are positive, then GSTAR(1;1) model is stationary.

Proof. The IAcM is built from the sum square of errors from all past observations until the first observation. For $(N \times 1)$-dimensional vector of observations, $\mathbf{Y}_1 = (Y_{1,1}, Y_{2,1}, \dots, Y_{N,1})'$, we can write $\mathbf{Y}_1'\mathbf{M}_1\mathbf{Y}_1 = \sum_{t=-\infty}^{1} \mathbf{e}_t'\mathbf{e}_t$ which has positive and finite values. The $\mathbf{M}_1$ is symmetric and since $\mathbf{Y}_1'\mathbf{M}_1\mathbf{Y}_1 > 0$ then $\mathbf{M}_1$ is positive definite matrix [4]. It implies the determinants of all leading principal submatrices of $\mathbf{M}_1$ are positive. Furthermore, since $\mathbf{M}_1$ and the covariance structure is do not change with the time, the GSTAR(1;1) process is stationary. *Q.E.D.*

The IAcM is a symmetric positive definite matrix. For GSTAR(1; λ_1) models with $\lambda_1 \geq 0$, we may define $\mathbf{A} = \sum_{\ell=0}^{\lambda_1} \mathbf{\Phi}_{1\ell}\mathbf{W}^{(\ell)}$, and the IAcM is,

$$\mathbf{M}_1 = \mathbf{I}_N - \mathbf{A}'\mathbf{A} \qquad (2)$$

Proposition 2 GSTAR(1; λ_1) models are stationary if the determinants of all the leading principal submatrices of $\mathbf{M}_1$ are positive.

Proof. Similar to Proposition 1, suppose $\mathbf{Y}_1$ is a $(N \times 1)$-dimensional vector of observations, $\mathbf{Y}_1 = (Y_{1,1}, Y_{2,1}, \dots, Y_{N,1})'$. The $\mathbf{Y}_1'\mathbf{M}_1\mathbf{Y}_1$ is the collection of squared errors from all past observations until the first observation. This means that $\mathbf{Y}_1'\mathbf{M}_1\mathbf{Y}_1$ always has positive value and is only equal to null if $\mathbf{Y}_1$ is a null vector. We write,

$$\mathbf{Y}_1'(\mathbf{I} - \mathbf{A}'\mathbf{A})\mathbf{Y}_1 > 0$$

$$\|\mathbf{Y}_1\|^2 - \|\mathbf{A}\mathbf{Y}_1\|^2 > 0$$

$$\|\mathbf{A}\mathbf{Y}_1\|^2 < \|\mathbf{Y}_1\|^2$$

The absolute eigen value of $\mathbf{A}$ must be less than one, which means that the GSTAR(1; λ_1) process is stationary. *Q.E.D.*

The stationarity condition in *Proposition* 1 and 2 meets the stationarity condition that was formulated by Wei [5] for Vector AR models. Furthermore for GSTAR(2; λ_1, λ_2) models, we define $(N \times N)$-dimensional matrix $\mathbf{A}_k = \sum_{\ell=0}^{\lambda_k} \mathbf{\Phi}_{k\ell}\mathbf{W}^{(\ell)}$, $k = 1, 2$ and the $(2N \times 2N)$-dimensional IAcM is,

$$\mathbf{M}_2 = \begin{pmatrix} \mathbf{I}_N - \mathbf{A}_2'\mathbf{A}_2 & -\mathbf{A}_1' - \mathbf{A}_2'\mathbf{A}_1 \\ -\mathbf{A}_1 - \mathbf{A}_1'\mathbf{A}_2 & \mathbf{I}_N - \mathbf{A}_2'\mathbf{A}_2 \end{pmatrix} \qquad (3)$$

Proposition 3 GSTAR(2; λ_1, λ_2) models are stationary if the determinants of all the leading principal submatrices of its IAcM are positive.

Proof. Using a similar procedure as with Proposition 2 and by letting $\mathbf{Y}_2 = (Y_{1,1}, \dots, Y_{N,1}, Y_{1,2}, \dots, Y_{N,2})'$, we may consider Proposition 3 as proved. Q.E.D

These stationarity conditions can provide a significance contribution to the process of identifying possible GSTAR models.

3   Identification Process GSTAR Model

A GSTAR model is a generalization of a STAR model with respect to its parameters. In a STAR model, the parameters are assumed to be the same for every location, while a GSTAR model assumes that different locations have different parameter values. Although both have different assumptions concerning the parameters, they still have the same procedures in doing the identification process of space-time models.

3.1   Space Time ACF and PACF

The STACF and STPACF had been employed by Pfeifer and Deutsch [1] in identifying Space-Time Autoregressive Moving Average (STARMA) models. Since STARMA and GSTAR models have many things in common, STACF and STPACF may be applied in identifying GSTAR models. The STACF is obtained by standardizing the autocovariance function of lag time s for observations with lag spatial k-th and ℓ-th, $\gamma_{\ell k}(s)$. Autocovariance function in matrix notation is

$$\gamma_{\ell k}(s) = \frac{1}{N} E\left[\left(\mathbf{W}^{(\ell)} \mathbf{Y}_s \right)' \left(\mathbf{W}^{(k)} \mathbf{Y}_{t+s} \right) \right] \tag{4}$$

Since equation (4) is a scalar then we write

$$\gamma_{\ell k}(s) = \frac{1}{N} tr \left\{ \mathbf{W}^{(\ell)} \mathbf{W}^{(k)} E\left[\mathbf{Y}_t' \mathbf{Y}_{t+s} \right] \right\} = \frac{1}{N} tr \left\{ \mathbf{W}^{(\ell)} \mathbf{W}^{(k)} \mathbf{\Gamma}(s) \right\} \tag{5}$$

The estimator for $\mathbf{\Gamma}(s)$ is $\hat{\mathbf{\Gamma}}(s) = \sum_{t=1}^{T-s} \frac{\mathbf{Y}_t \mathbf{Y}_{t-s}'}{T-s}$ and by substituting it to (5) then estimated covariance, $\hat{\gamma}_{lk}(s)$ holds. Furthermore the STACF, which is defined as,

$$\rho_{\ell k}(s) = \frac{\gamma_{\ell k}(s)}{\left[\gamma_{\ell \ell}(0) \gamma_{kk}(0) \right]^{1/2}}$$

can be estimated by using the sample covariance,

$$\hat{\rho}_{\ell k}(s) = \frac{\hat{\gamma}_{\ell k}(s)}{\left[\hat{\gamma}_{\ell\ell}(0)\hat{\gamma}_{kk}(0)\right]^{1/2}} = \frac{\sum_{i=1}^{N}\sum_{i=1}^{T-s} L^{(\ell)}Y_{i,t}L^{(k)}Y_{i,t+s}}{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(L^{(\ell)}Y_{i,t}\right)^2 \sum_{i=1}^{N}\sum_{t=1}^{T}\left(L^{(k)}Y_{i,t}\right)^2\right]^{1/2}} \tag{6}$$

with $L^{(\ell)}$ as the spatial lag operator of spatial order ℓ-th, so that

$$L^{(\ell)}Y_{i,t} = \begin{cases} Y_{i,t}, & \ell = 0 \\ \sum_{j=1}^{N} w_{ij}^{(\ell)}Y_{j,t}, & \ell = 1,2,...,\lambda_p \end{cases}$$

the STPACF for spatial order λ is the last coefficient of $\phi_{h\ell}$ ($\ell = 0,1,\ldots,\lambda$ and $h = 1,2,\ldots$) in space-time Yule Walker equation as defined in [1].

$$\begin{bmatrix} \gamma_{00}(1) \\ \gamma_{10}(1) \\ \vdots \\ \gamma_{\lambda 0}(1) \\ \hline \gamma_{00}(2) \\ \gamma_{10}(2) \\ \vdots \\ \gamma_{\lambda 0}(2) \\ \hline \vdots \\ \hline \gamma_{00}(h) \\ \gamma_{10}(h) \\ \vdots \\ \gamma_{\lambda 0}(h) \end{bmatrix} = \begin{bmatrix} \begin{array}{cccc|cccc|c|c} \gamma_{00}(0) & \gamma_{01}(0) & \cdots & \gamma_{0\lambda}(0) & \gamma_{00}(-1) & \gamma_{01}(-1) & \cdots & \gamma_{0\lambda}(-1) & & \\ \gamma_{10}(0) & \gamma_{11}(0) & \cdots & \gamma_{1\lambda}(0) & \gamma_{10}(-1) & \gamma_{11}(-1) & \cdots & \gamma_{1\lambda}(-1) & & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \cdots & (1-h) \\ \gamma_{\lambda 0}(0) & \gamma_{\lambda 1}(0) & \cdots & \gamma_{\lambda\lambda}(0) & \gamma_{\lambda 0}(-1) & \gamma_{\lambda 1}(-1) & \cdots & \gamma_{\lambda\lambda}(-1) & & \\ \hline & & (1) & & & & (0) & & & (2-h) \\ \hline & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\ \hline & & (h-1) & & & & (h-2) & & \cdots & (0) \end{array} \end{bmatrix} \begin{bmatrix} \phi_{10} \\ \phi_{11} \\ \vdots \\ \phi_{1\lambda} \\ \hline \phi_{20} \\ \phi_{21} \\ \vdots \\ \phi_{2\lambda} \\ \hline \vdots \\ \hline \phi_{p0} \\ \phi_{p1} \\ \vdots \\ \phi_{h\lambda} \end{bmatrix} \tag{7}$$

Space time model selection may be obtained by observing autocorrelation behavior among time lags and spatial lags which are represented by the cut off and tail off pattern of the STACF and STPACF.

3.2   IAcM as a Part of GSTAR Modeling

In this paper we propose a new procedure for GSTAR modeling by involving stationarity investigation through the IAcM characteristics as presented in *Proposition* 1 to 3. This stage is executed after identifying the model through the STACF and STPACF, and estimating the parameters. We see it as a first stage of model validation in addition to checking the mean square of residuals as the second stage. The residuals are obtained from the differences between real and estimated observations. The model with the minimum mean square of residuals will be recommended as the fitting model. The procedure of GSTAR modeling is shown in the following flow chart (Figure 2).

This procedure combines and advances the three-stage iterative procedure for space-time modeling of Pfeifer & Deutsch [1] and time-series modeling of Box & Jenkins [6]. The new aspect of this procedure is the process stationarity check with the IAcM [*].
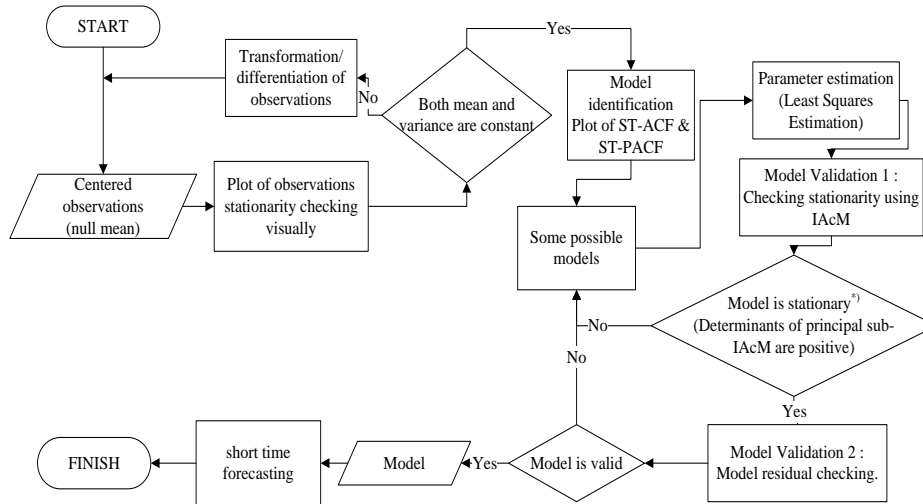

Figure 2 The new procedure of space time autoregressive modeling.

4   Case Study

For our case study we used the data of the monthly tea production of six plantation sites ($N = 6$) in West Java, Indonesia (Purbasari, Santosa, Sedep, Papandayan, Cisaruni, and Dayeuh Manggung), from January 1992 to December 2010 ($T = 228$). We executed space-time modeling by using Matlab and started by centering the processes in order to obtain the null mean of the processes. The coordinates of the observed plantations and the distances between them are displayed in a symmetric ($N \times N$)-dimensional matrix of which the main diagonal is null (see Figure 3).

From Figure 3 we can draft the numbers of spatial lag order and its members, and also the weight matrices for each spatial lag. We can build the range of spatial lag based on the fixed radius system in Figure 1 with a fixed d_0 of 0.4 km. The interval distance (0, 0.4) is categorized as spatial lag 1, (0.4, 0.8] as spatial lag 2, and so on until (0.4(-1), 0.4] as spatial lag ℓ. For this case we got four spatial lags for eachweight matrix (uniform, euclidean-uniform and binary). The appropriate weight matrices for the first and second spatial lag are shown in Table 1.
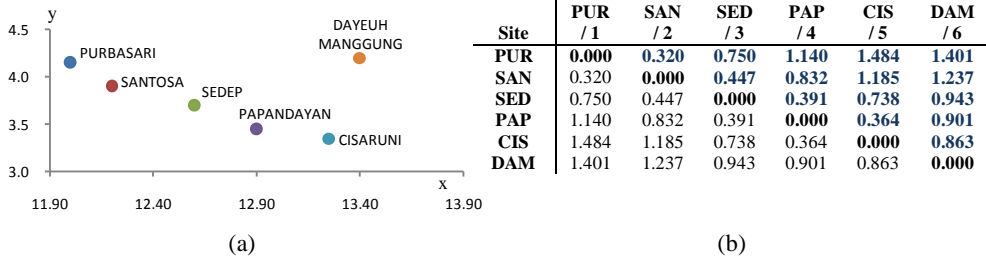
Site	PUR / 1	SAN / 2	SED / 3	PAP / 4	CIS / 5	DAM / 6
PUR	0.000	0.320	0.750	1.140	1.484	1.401
SAN	0.320	0.000	0.447	0.832	1.185	1.237
SED	0.750	0.447	0.000	0.391	0.738	0.943
PAP	1.140	0.832	0.391	0.000	0.364	0.901
CIS	1.484	1.185	0.738	0.364	0.000	0.863
DAM	1.401	1.237	0.943	0.901	0.863	0.000

(a)   (b)

Figure 3 (a) The coordinate of six observed plantations, and (b) the distances (in kilometre, km). PUR is Purbasari, SAN is Santosa, SED is Sedep, PAP is Papandayan, CIA is Cisaruni, and DAM is Dayeuh Manggung.

Table 1 The weight matrices for spatial lag ℓ-th, $\ell = 1,2,3,4$. For $\ell = 0$, the weight matrix $\mathbf{W}^{(0)} = \mathbf{I}_N$. The sum of each row must be equal to one with a null main diagonal.

	Uniform	Euclidean - Uniform	Biner
$\mathbf{W}^{(1)}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.495 & 0 & 0.505 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$
$\mathbf{W}^{(2)}$	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0.311 & 0.376 & 0 & 0 & 0.313 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$
$\mathbf{W}^{(3)}$	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.544 & 0.456 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0.304 & 0.355 & 0 & 0 & 0 & 0.342 \\ 0 & 0.460 & 0 & 0 & 0 & 0.540 \\ 0 & 0 & 0.326 & 0.334 & 0.340 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$
$\mathbf{W}^{(4)}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0.491 & 0.509 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0.482 & 0.518 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$

The appropriate weight of the matrices in Table 1 has similar values. This is acceptable because they have the same criterion for determining the spatial lag interval. This happens because there are only six observed sites, which should be categorized into four levels of spatial authority ($\lambda_p = 4$). Since there is subjective judgment involved in building the weight matrices, we should pay more attention to the rules, especially when the ratio between the number of locations and the number of spatial lags is close to 1:1. In these cases, it will be more difficult to obtain appropriate and obedient weight matrices.

The first stage in GSTAR modeling is to identify possible models by computing the STACF and STPACF values as in equation (6) and (7). These values arepresented in Table 2.

Table 2   The STACF and STPACF values for several time and spatial lag.

	STACF						STPACF				
spatial lag time lag	0	1	2	3	4	spatial lag time lag	0	1	2	3	4
1	0.4697	0.3102	0.3315	0.1989	0.3329	1	**0.3578**	-0.0086	0.1975	-0.0984	0.0634
2	0.1554	0.006	0.0105	-0.0339	0.0401	2	0.0191	-0.0385	**-0.2131**	0.1136	-0.0886
3	0.0699	-0.0779	-0.0912	-0.1194	-0.0532	3	0.1688	0.076	-0.088	-0.0345	-0.0845
4	-0.0102	-0.1232	-0.0859	-0.0821	-0.0374	4	-0.1734	-0.0166	-0.0854	0.0221	0.1226
5	0.0651	-0.0633	0.0083	0.0136	0.0473	5	0.0486	-0.0135	0.1478	-0.0569	-0.0015
6	0.0408	-0.0868	-0.004	0.0331	0.0372	6	-0.0401	-0.0529	-0.0472	0.0905	-0.045
7	0.0032	-0.1322	-0.0478	0.0236	-0.0063	7	0.0488	-0.0038	-0.0297	0.1858	-0.1577
8	-0.0343	-0.1714	-0.1305	-0.0862	-0.0808	8	0.036	-0.0311	-0.0982	0.0292	-0.0511

From Table 2, the STACF values for all spatial lags already have cut off after first time lag. This alert us to consider Moving Average (MA) factors in the models. But if we plot those values for larger time lags, it seems like sinusoidal graphs are coming out. This means that we can temporarily neglect the MA factors and more focus on the STPACF values that represent autoregressive terms. This is also because this paper discusses about the GSTAR modeling. The STPACF values are cut off after the first and second lag time for zero, second, and third spatial lag. From these lags (both spatial and time) we obtain some possible models, namely: GSTAR(1;0), GSTAR(1;1), GSTAR(1;3), GSTAR(2;1,1), GSTAR(2;1,3), etc.

The next step is parameter estimation with the least squares method. To execute this step we should rearrange the GSTAR model to be a linear structure, $\mathbf{Y}=\mathbf{X\Phi}+\mathbf{e}$, with $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', ..., \mathbf{Y}_N')'$, $\mathbf{X} = \text{diag}(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N)$, $\mathbf{\Phi} = (\mathbf{\Phi}_1', \mathbf{\Phi}_2', ..., \mathbf{\Phi}_N')$, and $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N)'$. For each location we have linear structure $\mathbf{Y}_i = \mathbf{X}_i \mathbf{\Phi}_i + \mathbf{e}_i$, $i = 1,2, ..., N$, with $\mathbf{Y}_i = (Y_{i,p}, Y_{i,p+1}, ..., Y_{i,T})'$, $\mathbf{e}_i = (e_{i,p}, e_{i,p+1}, ..., e_{i,T})'$, $\mathbf{\Phi}_i = (\phi_{10}^{(i)}, ..., \phi_{1\lambda_1}^{(i)}, \phi_{20}^{(i)}, ..., \phi_{2\lambda_2}^{(i)}, ..., \phi_{p0}^{(i)}, ..., \phi_{p\lambda_p}^{(i)})'$, and

$$\mathbf{X}_i = \begin{pmatrix} V_{i,p-1}^{(0)} & V_{i,p-1}^{(1)} & \cdots & V_{i,p-1}^{(\lambda_1)} & \cdots & V_{i,0}^{(0)} & V_{i,0}^{(1)} & \cdots & V_{i,0}^{(\lambda_p)} \\ V_{i,p}^{(0)} & V_{i,p}^{(1)} & \cdots & V_{i,p}^{(\lambda_1)} & \cdots & V_{i,1}^{(0)} & V_{i,1}^{(1)} & \cdots & V_{i,1}^{(\lambda_p)} \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ V_{i,T-1}^{(0)} & V_{i,T-1}^{(1)} & \cdots & V_{i,T-1}^{(\lambda_1)} & \cdots & V_{i,T-p}^{(0)} & V_{i,T-p}^{(1)} & \cdots & V_{i,T-p}^{(\lambda_p)} \end{pmatrix}$$

with $V_{i,t}^{(k)} = \sum_{j=1}^N w_{ij}^{(k)} Y_{j,t}$ for $k \geq 1$ and $V_{i,t}^{(0)} = Y_{i,t}$ (see [3]).

For $(\mathbf{X}'\mathbf{X})$ nonsingular, the least square (LS) estimator may be obtained by

$$\widehat{\mathbf{\Phi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad (8)$$

The LS estimator is unbiased with minimum variance.

After obtaining the LS estimators, we can construct the matrix of parameters ($\mathbf{A}$) and the IAcM for checking the process stationarity. The IAcM determinant values of possible models with 1st, 2nd time lags and 0th, 1st, 2nd, 3rd spatial lag are presented in Figure 4.

Based on *Proposition* 1 to 3, the model that produce at least one negative value of the leading principal submatrix determinant value in the IAcM, will be eliminated from appropriate model candidates. Looking at Figure 4, we notice that models of which at least one of its leading principal matrix has negative determinant are: GSTAR(2;0,2), GSTAR(2;0,3), GSTAR(2;1,3), GSTAR(2;3,1), GSTAR(2;3,2), and GSTAR(2;3,3). We may leave these models and move forward to the next stage, diagnostic checking (Model Validation 2).

In diagnostic checking stage, the residuals of rest possible models are compared. Models with least mean square of residuals (MSR) is recommended as the appropriate models. MSR is formulated as,

$$\text{MSR} = \frac{1}{NT-p}\big(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t\big)'\big(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t\big) \qquad (9)$$

$\widehat{\mathbf{Y}}_t$ is a ($NT \times 1$)-dimensional vector of estimated observations that follows GSTAR($p; \lambda_1, \dots, \lambda_p$) process.

Table 3 MSR of possible GSTAR($p; \lambda_1, \dots, \lambda_p$) models, $p = 1, 2$. There are no patterns in MSR values as the order becomes larger or smaller.

GSTAR($p; \lambda_1, \dots, \lambda_p$) Models					
Order	MSR	Order	MSR	Order	MSR
(1;0)	98.6716	(2;0,1)	96.6197	(2;1,2)	89.9513
(1;1)	89.8531	(2;1,0)	**87.3273**	(2;2,1)	88.3205
(1;2)	**88.0149**	(2;2,0)	**86.7508**	(2;2,2)	88.9287
(1;3)	**82.5222**	(2;3,0)	**81.9143**	(2;2,3)	90.6263

The model with the smallest MSR value, GSTAR(2;3,0), can be proposed as the most appropriate model for this case and be employed to execute the short forecasting. Since there are some models that have close MSR values, we will perform an additional step. We let the forecasting stage serve as an advanced checking diagnostic stage. For efficiency, we choose the five possible models

that produced the lowest MSR from Table 3. Those models are: GSTAR(2;3,0), GSTAR(1;3), GSTAR(2;2,0), GSTAR(2;1,0), and GSTAR(1;2).
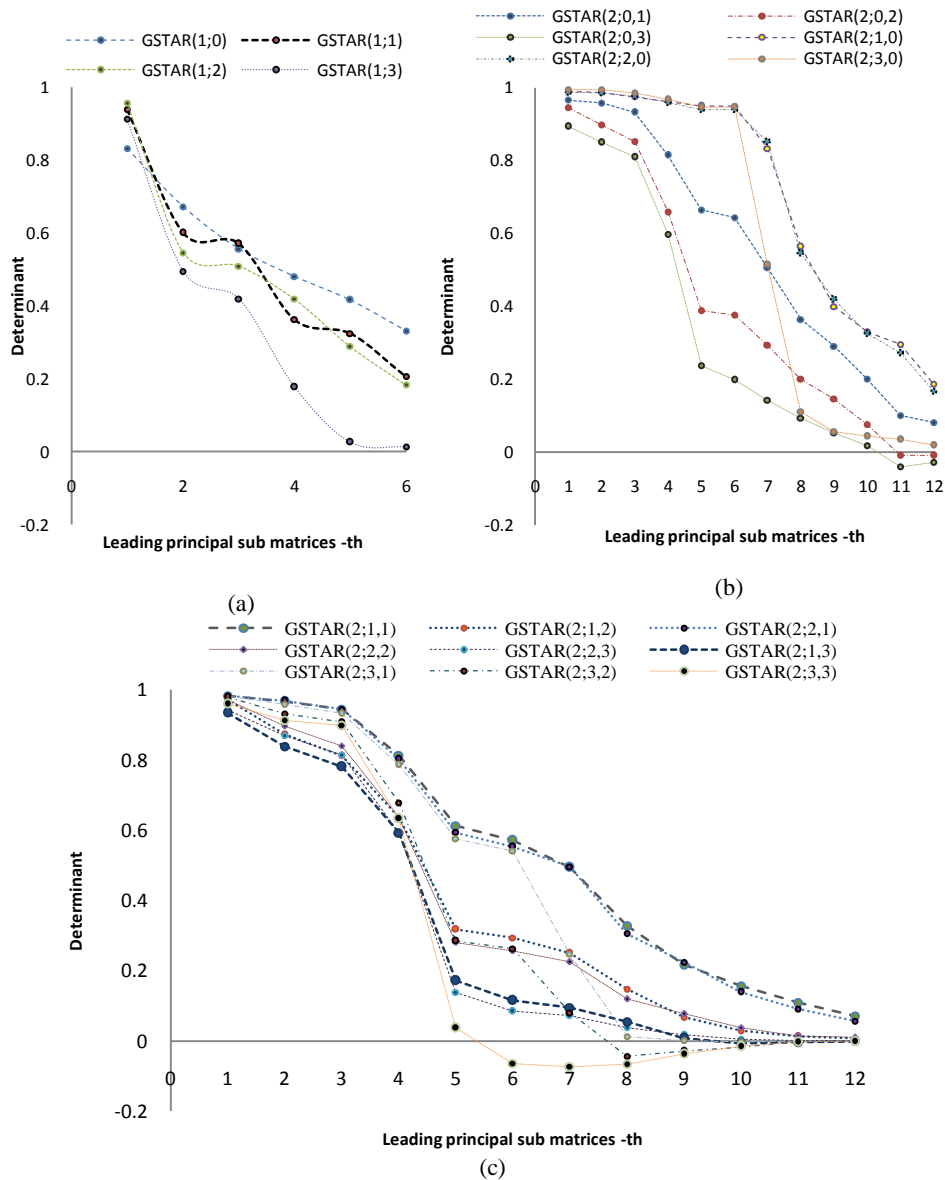

(a)

(b)

(c)

Figure 4 The determinant values of all the leading principal matrices of the IAcM tend to be smaller as the dimensions of the leading principle matrices become larger, (a) possible models with 1st time lag, (b) possible models with 2nd time lag and one of spatial lag is 0, and (c) others possible models with 2nd time lag.

For the forecasting stage, we prepared the ten latest data, collected from January to October 2011. In the advanced checking diagnostic stage, we estimate these data using the five selected models in order to obtain what we call *forecast-estimated observations*. These forecast-estimated observations will be compared with real observations using (9), so that we have the MSR of the forecast-estimated observations. The results of the MSR can be seen in Table 4.

Table 4 MSR of forecast-estimated of five possible appropriate GSTAR models. The smallest MSR is given by GSTAR(2;3,0) model.

GSTAR	MSR
(1;2)	175.3843
(1;3)	139.4374
(2;1,0)	155.788
(2;2,0)	141.6173
(2;3,0)	**108.9293**
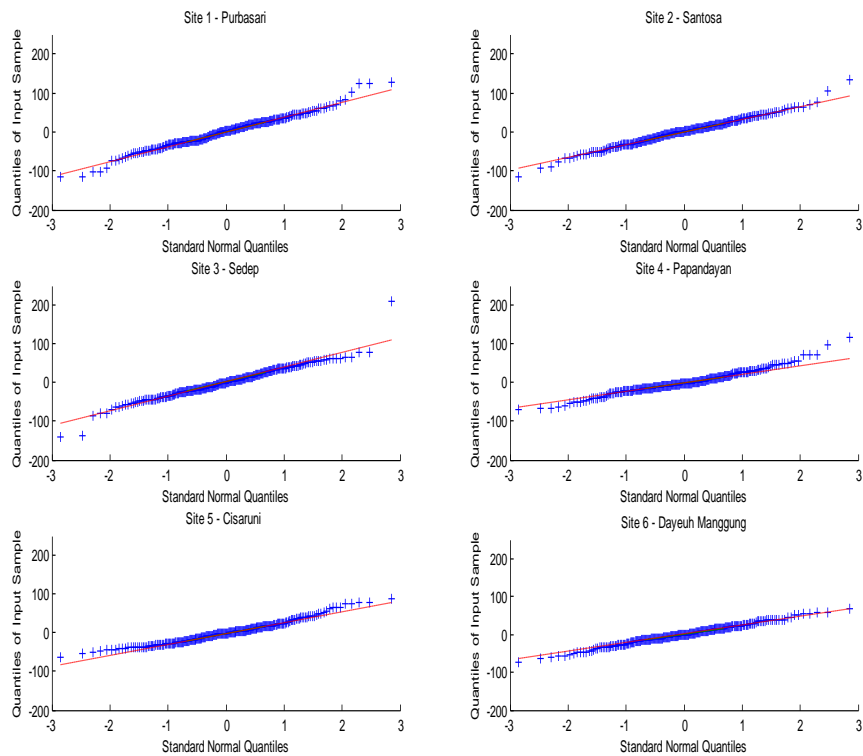
Figure 5 QQ Plot of Sample Data versus Standard Normal Data for each site *i*, *i* = 1,2,...,6 (from top left to bottom right). Quantile-quantile plot of the sample quantiles of residuals versus theoretical quantiles from a normal distribution.

Based on the smallest MSR value, we have GSTAR(2;3,0) as the most appropriate model. For convenience, we display a quantile-quantile plot of residuals data versus normal distribution data in Figure 5. Since the plot close to linear, we may say that the residuals have normal distribution.

In order to do forecasting of $(t+1)^{\text{th}}$ observations, the GSTAR(2;3,0) may be presented as the following equation.

$$\widehat{\mathbf{Y}}_{t+1} = \left(\widehat{\boldsymbol{\Phi}}_{10} + \widehat{\boldsymbol{\Phi}}_{11}\mathbf{W}^{(1)} + \widehat{\boldsymbol{\Phi}}_{12}\mathbf{W}^{(2)} + \widehat{\boldsymbol{\Phi}}_{13}\mathbf{W}^{(3)}\right)\mathbf{Y}_t + \widehat{\boldsymbol{\Phi}}_{20}\mathbf{Y}_{t-1} \quad (10)$$

where $\widehat{\mathbf{Y}}_{t+1}$ is a estimated $(N{\times}1)$-dimensional vector of observations at time $t{+}1$, $\widehat{\mathbf{Y}}_{t+1} = \left(\widehat{Y}_{1,t+1}, \widehat{Y}_{2,t+1}, \dots, \widehat{Y}_{N,t+1}\right)'$, $\widehat{\boldsymbol{\Phi}}_{1\ell}$ is a estimated $(N{\times}N)$-dimensional matrix of autoregressive parameters for the first time lag, with spatial lag $\ell = 0,1,2,3$. The $\ell = 0$ shows the influence of a location on itself.
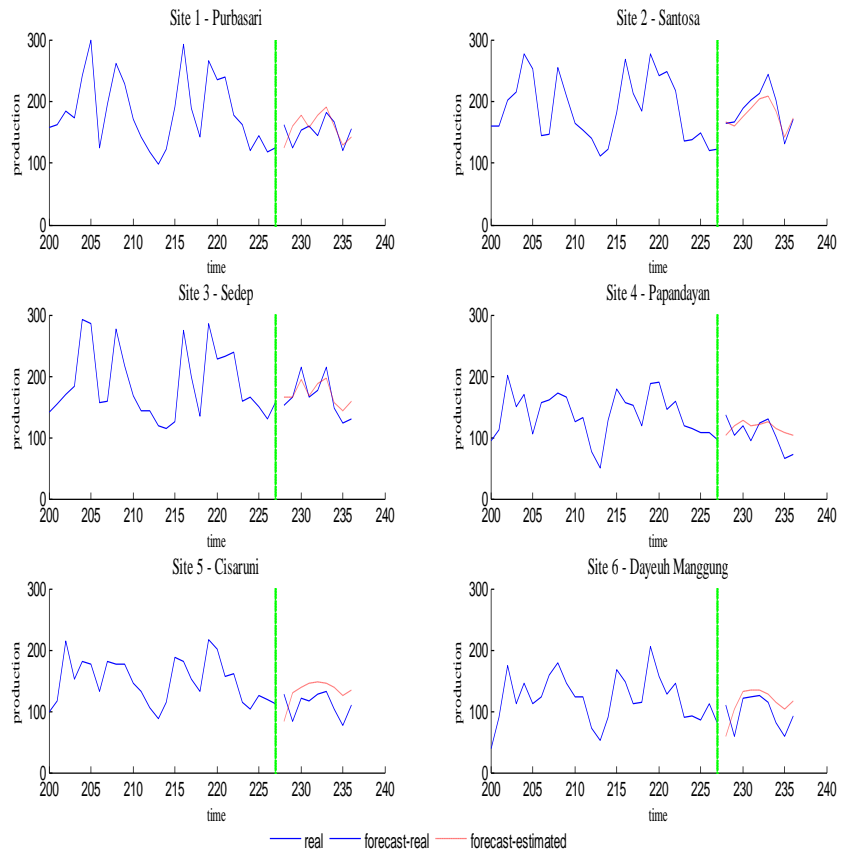

Figure 6 Plot of forecast-estimated observations using GSTAR(2;3,0) model. for each site i, $i = 1,2,\dots,6$ (from top left to bottom right).

The GSTAR(2;3,0) model tells us that a location will be influenced by itself until the second time lag, while its neighbors can exert influence during the first time lag within three spatial lags.

The plots of the 28 latest observations and the 10 next forecasted observations, the total MSR of which is 108.9293, are shown in Figure 6. The forecast-estimated plots display similar patterns and are close enough to the real observations.

5   Conclusions

An IAcM is suggested to be used in selecting the approriate space-time models in completion of STACF and STPACF values. Using an IAcM decreases the subjectivity of the judgment about which models are appropriate. In the case study of the monthly tea production of six plantations in West Java, we found that GSTAR(2;3,0) is the most appropriate model. We also observed that three types of matrices (uniform, euclidean, and binary) will have the same influence on the modeling process, since the ratio between the number of spatial lags and observed locations is close to one. They are supposed to have different results if the number of observed locationsis much higher than the number of observed spatial lags.

Acknowledgement

References

[1]   Pfeifer, P.E., & Deutsch, S.J., *A Three-Stage Iterative Approach for Space-Time Modeling*, Technometrics, **22**(1), pp. 35-47, 1980.
[2]   Rahardjo, S. *Identifikasi Kestasioneran Proses Autoregresif Secara Analitis*, PhD Dissertation, Mathematics, Inst. Teknol. Bandung, 2005.
[3]   Borovkova, S.A., Lopuhaä, H.P., & Nurani, B., *Consistency and Asymptotic Normality of Least Squares Estimators in Generalized Space-Time Models*, Statistica Neerlandica, **62**, pp. 482-508, 2008.
[4]   Laub, A.J., *Matrix Analysis for Scientists & Engineers*, SIAM, Philadelphia, 2005.
[5]   Wei, W.W.S., *Time Series Analysis: Univariate and Multivariate Methods* 2nd. Ed., Pearson Addison Wesley, Boston, 2006.
[6]   Box, G.E.P., Jenkins, G.M. & Reinsel, G., *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, New Jersey, 1994.