Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function

Yundari1,2*, U S Pasaribu2, U Mukhaiyar3, M N Heriawan3
1Department of Mathematics, Universitas Tanjungpura, Pontianak, Indonesia
2Department of Mathematics, Institut Teknologi Bandung, Indonesia
3Department of Mining Engineering, Institut Teknologi Bandung, Indonesia

*yundari@math.untan.ac.id

Abstract. The stochastic process models with the index parameters such as time and location were investigated in this paper. The model used was GSTAR (1;1), and it was applied to the Gamma ray log data. The important thing to be assessed in this model is the determination of the space weight matrix. Commonly, the spatial weight matrix was determined based on the Euclidean distance, but not based on data. In this work, we use the kernel function approach to determine the spatial weighting function whose domain was in the form of data observation. In addition, we also study the influence of this weight matrix to the stationary condition of GSTAR (1;1) model, and we use the inverse of auto-covariance matrices or IAcM methods. The results showed that the kernel weights matrix approach still being met influence on stationary of this model.

1. Introduction
In daily life, we frequently deal with the data that depend on time. The data that observed as a function of time was known as a time series. There were the monthly plantation production data, daily closing share price, the weekly number of criminal cases and many other examples of time series data. The data are observed only on one object in one time series and one location. If the observed data were in some locations with the assumption of location dependence, data can be analyzed in a space-time series.

Some models of spatial time series, which composed an autoregressive form were vector autoregressive (VAR) [1] which modeling multiple time series in a vector, and space-time autoregressive moving average (STARMA) [2] which modeling time series with the spatial dependence within a matrix and maintaining the three iterative of Box Jenkins time series. Further development was the forming rows of parameters of Autoregressive Model STARMA by [3], called generalized space-time autoregressive (GSTAR). This GSTAR model was perfect for heterogeneous geographic conditions, since the parameters were distinct for each location. It was the fundamental difference between the STAR and GSTAR models.

The geographical conditions were heterogeneous in Indonesia, therefore the GSTAR model was widely used and developed by researchers in Indonesia. Some application of GSTAR (1;1) model have been done by some researcher such as Wutsqa and Suhartono (2010) [4], examined the relationship between the number of foreign tourists who will come to Yogyakarta and Bali, Ruchjana(2012) [5]
applied GSTAR (1.1) model of petroleum production data, Yundari et al (2017) [6] in tea production data and Fadlilah et al (2015) [7] applied it to the red chili price data.

One of the main problems in composing GSTAR model was determining the appropriate spatial weight for space time series data that will be analyzed. Spatial weight generally, determined in the early development process model, therefore it’s role will affect the further modeling process. The weight was compiled into a matrix that called the weighted matrix. A weighted matrix represents the spatial relationship of the data, although it was the fully right of model user. The determination of weight matrix requires spatial lag definition which was useful for grouping nearby locations.

There were several methods to determine the spatial weight on GSTAR, that commonly used were uniform and binary weights. The uniform weights give equal value to each location (in same spatial lag). The next method was spatial weight that proposed to determine the weight through semivariogram spatial correlation [3]. Several methods in constructing the weight matrix were always determined by the distance. Some researchers have constructed spatial weight through it’s real observation. Suhartono and Subanar (2006) [8] used normalized cross correlation among locations on the corresponding lag to build spatial weight.

Subsequently, in 2015, Nugraha et al [9] defined that the weight matrix involving the observation value of a location, namely the determination of the space weight value through fuzzy set concept approach. The advantage of this fuzzy set was every object in the fuzzy set has a distinct degree of membership. To represent the degree of membership, usually used several functions that form a shape, such as a triangular shape, trapezoidal, Gaussian, or represented in the linear function. This paper examined the determination of spatial weight matrix by using the kernel function approach.

2. Spatial weight of GSTAR(1;1) model with kernel function approach

By using the idea of determining spatial weight with the Fuzzy set approach, we proposed a new method to determine the spatial weight matrix for GSTAR model using kernel function approach. Kernel function approach as a weight function has been used in density estimation and regression function. The working of kernel function was assumed some of kernel functions for each point with every point around it. With this idea we used this kernel approach to determine the weight matrix with location as a point used in the definition of the kernel function.

In general, kernel function for a point with its closest point was $k\left(\frac{x-y}{h}\right)$, where $h$ was the bandwidth that controls the smoothness level. By using the mean value of each location $\bar{Y}_i$, location weights $j$ to $i$ and matrix form can be written as follows:

$$W_{ij} = \frac{k\left(\frac{\bar{Y}_i - \bar{Y}_j}{h}\right)}{\sum_{\substack{\ell=1 \atop \ell \neq i}}^{n} k\left(\frac{\bar{Y}_i - \bar{Y}_\ell}{h}\right)}$$

(1)

and

$$W = \begin{bmatrix}
0 & W_{12} & \cdots & W_{1N} \\
W_{21} & 0 & \cdots & W_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
W_{N1} & W_{N2} & \cdots & 0
\end{bmatrix}$$

From the result of the weighted matrix $W$ with kernel function approach, the properties of weight matrix was still met. Selection of the mean value from observation of each location was intended to
determine the characteristics of overall data (centered data) by ignoring outliers of an observation data. The following uniform weight, Gaussian and Epanechnikov kernel function were used:

$$k(x) = \frac{3}{4}(1-x^2) \quad \text{for} \quad |x|<1.$$  

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\} \quad \text{for all} \quad x.$$  

In addition to the selection of kernel function, the most important thing was the selection of bandwidth of kernel functions. In this paper, we used normal scale rule. This rule has the same principle with the selection of optimal bandwidth through Asymptotic Mean Integrated Squared Error (AMISE) for normal density. Assuming that the density was normal and the properties of the normal kernel were changed variance and "roughness" into the following [10]:

$$h_N = \delta_K \left[ \frac{8\sqrt{\pi}}{3} \right]^{-\frac{1}{5}} \sigma n^{-\frac{1}{5}}$$  

with $\sigma$ was estimated variance and $\delta_k$ was the canonical bandwidth of kernel function shown as:

$$\delta_k = \left( \frac{R(K)}{\sigma_k^2} \right)^{\frac{1}{5}}.$$  

3. Structured generalized space-time autoregressive model

The approach combines the time-series analysis and space analysis that has been proposed, including Bayesian hierarchical model [11] and latent Gaussian Markov random field [12]. Another approach was the STAR (space-time autoregressive) [2], STAR moving average or STARMA [13], and Generalization STAR or GSTAR [3]. Among some of the space-time models, the concern of this study is GSTAR model. Research on GSTAR (1; 1) model pioneered by Ruchjana as a topic of the dissertation and was followed by several other researchers with different viewpoints.

The GSTAR model was a special form of the VAR (Vector Autoregressive) model. A process of $Y_i(t) = (Y_{i1}(t), Y_{i2}(t), ..., Y_{ip}(t))^\top$, follows the GSTAR process with time order $p$ and spatial order $\lambda_1, \lambda_2, ..., \lambda_p$, written as GSTAR $(p; \lambda_1, \lambda_2, ..., \lambda_p)$ if

$$Y(t) = \sum_{i=1}^{p} \Phi_{ki} Y(t-k) + \sum_{i=1}^{p} \Phi_{ki} \sum_{j=1}^{N} W_{ij} Y_j(t-k) + \epsilon(t), i = 1, 2, ..., N; t = 1, 2, ..., T$$

If the time order and spatial order used, respectively, were one, then the model can be written as

$$Y_i(t) = \phi_{0i} Y_i(t-1) + \phi_{1i} \sum_{j=1}^{N} W_{ij} Y_j(t-1) + \epsilon_i(t),$$

$t = 1, ..., T, i = 1, ..., N$.

with $Y_i(t)$ was the observation at time $t$ in a location $i$, $\phi_{0i}$ and $\phi_{1i}$ represents an autoregressive parameter for time and spatial respectively, and $\epsilon_i(t)$ was modeled error with normal distribution.

The GSTAR modeling follows three stages of the Box-Jenkins in time series analysis. These stages were the stage of model identification, parameter estimation, and diagnostic testing. Identifying a space-time model adapted the stage of identification carried out on the STARMA model by Pfeifer and Deutsch (1980)[13], which used the space-time autocorrelation function (STACF) and the space-time partial autocorrelation function (STPACF). At the time series analysis, both functions were similar to the ACF and PACF.
In modeling, parameter estimation was very important to be considered. There were several ways of obtaining estimation parameters such as Least Square (LS) method and likelihood maximum (LM). LS estimator obtained by minimizing the sum of squared errors, generally used in the regression model and proved to be the best linear unbiased estimator on GSTAR (1.1) model [14].

Another study about GSTAR (1:1) was the process of stationary testing. The stationary requirement of GSTAR process obtained if the modulus of all eigenvalues of the autoregressive parameter matrix was smaller than one. In 2012, Mukhaiyar [14] introduced a new method to test the GSTAR model using the Inverse Autocovariance Matrix (IAcM). In this paper, we used diagnostic test using IAcM but applied to GSTAR model (1: 1) with kernel spatial weight.

The IAcM form of the GSTAR model (1; 1) with the spatial weight of the kernel was

\[ M_1 = (N-1)I_N - (N-1)A_1'A_1 \text{, for } N > 1 \]  

(6)

with \( A_s = \Phi_s + \Phi_s'W \). The parameter requirements for the GSTAR model (1: 1) with the kernel weight as the diagnostic test stage are represented by the following Theorem 1.

**Theorem 1.**
A GSTAR process (1; 1) with a kernel weight matrix having IAcM (6) is said to be stationary if and only if the matrix \( M_1 \) is positive definite.

**Proof:**
A GSTAR process (1; 1) is known, with a kernel weight matrix having IAcM (6) said to be stationary. In this case, the stationary is weak. The weak stationary means the mean vector and the autocovariance matrix is constant for all time \( t \). The matrix \( M_1 \) is an inverse of an autocovariance matrix whose elements are independent of time \( t \). This matrix is also symmetric. Since it is constructed from the sum of squares of non-negative error values for any nonzero vector \( Y(t) \) applies \( Y(t)'M_1Y(t) > 0 \). So the matrix is a positive definite matrix.

Conversely, if known \( M_1 \) IAcM matrix is positive definite it will be shown process GSTAR (1: 1) with kernel weight is stationary. The \( M_1 \) matrix is a positive definite matrix and the sum of the error squares is worth up to then \( 0 < Y(t)'M_1Y(t) = \sum_{t=-\infty}^{1} \varepsilon(t)\varepsilon(t) < \infty \). Since \( M_1 \) being a function of autoregressive parameters and spatial weights that are independent of time, the GSTAR (1;1) process is stationary.

4. Implementation of space weight determination based on kernel function approach for GSTAR(1:1) model

In geometry, for six drill holes with each data \( T = 420 \) (depth \( m \) 0.2\( m \) to 84 drill holes). Typically, time series data used is data related to time such as an hour, day, month, year and so forth. In the field of geology, particularly stratigraphic sedimentation learnt about the layers of rock in which the layers stated time of formation of new stones. In other words, these rock layers can also be used as a time series data because each of the layers of rock is interdependent. This was reviewed by Davis [15]. The lowest rock layer states the oldest time (older rocks) and the highest layer states the youngest time (younger rocks). The purpose of modelling is to predict the previous rock layers (back-forecasting). In geology, this is important in terms of metal mineral drilling, to determine the metal reserves existed in a drill hole, so can be done the drilling that more effective and efficient in time, effort and cost.

The data used is the amount of Gamma Ray log data that can represent rock shape, so the time spent usually follow the direction of the horizontal axis. In this case, obtained data following the vertical axis from top to bottom. Because the data used for time series data must have a stationary assumption, the properties for backcasting equal to forecasting. Here is a summary of numerical and its graph (see table 1).
Table 1. Descriptive statistics of six drill holes (DH) data

| Drill Holes | Mean | 1st Qu. | Med | Mean | 3rd Qu. | Min | 3rd Qu. | Min |
|-------------|------|--------|-----|------|---------|-----|---------|-----|
| DH 21       | -2.41| 41.82  | 51.44| 47.18| 58.98   | 91.01|         |     |
| DH 25       | -1.64| 43.27  | 49.85| 47.84| 57.72   | 75.26|         |     |
| DH 31       | -1.29| 43.74  | 52.05| 49.28| 59.65   | 76.83|         |     |
| DH 35       | 4.34 | 43.75  | 51.97| 48.3 | 58.33   | 75.76|         |     |
| DH 41       | 0.34 | 35.34  | 52.63| 47.7 | 60.43   | 106.90|         |     |
| DH 49       | 0.46 | 46.85  | 53.18| 51.97| 59.43   | 98.73|         |     |

Figure 1. (a) Line-plot of six DH for GSTAR(1;1) model (b) box-plot of six DH. Visually the plot's data are not stationary so we transform to stationary through differencing once.

From the result of box plot seen that the Gamma ray data contain many outliers, but in this modeling, outliers are still ignored. Then will be done GSTAR (1; 1) model is used with uniform,
Gaussian, and Epanechnikov kernel weights. Uniform kernel represents the spatial weight matrix that commonly used by researchers, Gaussian kernel is the kernel that its asymptotic is simple and the last kernel, Epanechnikov, is the optimal kernel for any bandwidth.

All of the weight matrices produced satisfy properties of a spatial weight matrix. By using these three weights, will be estimated parameter for GSTAR (1; 1) model and the result can be seen in table 2. Parameter estimation using the LS method. After estimating the parameters, the parameter estimation results are validated using the IAcM method as described in Section 3.

The weighted matrix result obtained for each kernel is as follows:

\[
W_{\text{unif}} = \begin{pmatrix}
0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\
0.2 & 0 & 0.2 & 0.2 & 0.2 & 0.2 \\
0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.2 \\
0.2 & 0.2 & 0.2 & 0 & 0.2 & 0.2 \\
0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\
0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 \\
\end{pmatrix}
\]

\[
W_{\text{gauss}} = \begin{pmatrix}
0 & 0.11 & 0.25 & 0.25 & 0.25 & 0.14 \\
0.09 & 0 & 0.16 & 0.25 & 0.25 & 0.25 \\
0.25 & 0.17 & 0 & 0.21 & 0.25 & 0.12 \\
0.21 & 0.21 & 0.16 & 0 & 0.21 & 0.21 \\
0.20 & 0.20 & 0.20 & 0.20 & 0 & 0.20 \\
0.15 & 0.23 & 0.12 & 0.25 & 0.25 & 0 \\
\end{pmatrix}
\]

\[
W_{\text{Epan}} = \begin{pmatrix}
0 & 0.11 & 0.25 & 0.25 & 0.25 & 0.14 \\
0.09 & 0 & 0.16 & 0.25 & 0.25 & 0.25 \\
0.25 & 0.17 & 0 & 0.21 & 0.25 & 0.12 \\
0.21 & 0.21 & 0.16 & 0 & 0.21 & 0.21 \\
0.20 & 0.20 & 0.20 & 0.20 & 0 & 0.20 \\
0.15 & 0.23 & 0.12 & 0.25 & 0.25 & 0 \\
\end{pmatrix}
\]

The diagnostic test with IAcM yields valid model parameters because all Eigenvalues of the IAcM matrix are positive. In other words, the IAcM matrix is a positive definite matrix (see Table.2). Based on the smallest RMSE GSTAR (1; 1) model with Gaussian kernel weights is the best model for training data, but for prediction, Epanechnikov kernel is the best choice. Performance for prediction of 5 bottom layers can be seen in table 3.

Table 2. Result of parameter estimation and diagnostic test using Eigen value of parameter matrix and invers of covariance matrix (IAcM)

| Time Parameter | Spatial Parameter | Conclusion/diagnostic test (stationary test using IAcM) | RMSE  |
|----------------|-------------------|--------------------------------------------------------|-------|
| \( \phi_i \)   | \( \phi_i \)     | Valid                                                  | 12.3128 |
| Epanechnikov kernel |                      |                                                       |       |
| -0.3904        | 0.0343             | Valid                                                  |       |
| -0.3026        | 0.0200             | Valid                                                  |       |
| -0.2024        | -0.1306            | Valid                                                  |       |
| -0.3338        | 0.2596             | Valid                                                  |       |
| -0.1551        | 0.1537             | Valid                                                  |       |
| -0.1083        | 0.1643             | Valid                                                  |       |
| -0.3923        | 0.1402             | Valid                                                  |       |
| -0.3065        | 0.1020             | Valid                                                  |       |
| -0.2012        | -0.0862            | Valid                                                  |       |
| -0.3236        | 0.0925             | Valid                                                  |       |
| -0.1502        | 0.0856             | Valid                                                  |       |
| -0.1159        | 0.0661             | Valid                                                  |       |
| -0.3906        | 0.0754             | Valid                                                  |       |
| -0.3056        | 0.0716             | Valid                                                  |       |
| -0.2029        | -0.1412            | Valid                                                  |       |
| -0.3327        | 0.2432             | Valid                                                  |       |
| -0.1551        | 0.1537             | Valid                                                  |       |
| Gaussian kernel |                      |                                                       | 12.2983 |
| Uniform Weight |                   |                                                       | 12.3163 |
| -0.3906        | 0.0754             | Valid                                                  |       |
| -0.3056        | 0.0716             | Valid                                                  |       |
| -0.2029        | -0.1412            | Valid                                                  |       |
| -0.3327        | 0.2432             | Valid                                                  |       |
| -0.1551        | 0.1537             | Valid                                                  |       |
Table 3. The RMSE result of the 3 weights for prediction

| Type of weight     | RMSE   |
|--------------------|--------|
| Rectangular/ uniform kernel | 10.90838 |
| Gaussian kernel    | 9.52521 |
| Epanechnikov kernel| 9.518456 |

5. Conclusion

Based on the result of the previous section, it can be concluded that the determination of the spatial weight matrix in GSTAR model can also be determined based on observed data. From the result of the most minimum RMSE obtained weight matrix of the Gaussian kernel in which estimation result was quite accurate, with a small error. The resulting RMSE value was more or less the same. This was because the sampling of GR log data may not be as old (based on the rock layer). For the next research will be considered taking samples in accordance with lithostratigraphy appropriate to the relative age of the rock. For prediction purpose, the Epanechnikov kernel was the best spatial weight to this data.

Acknowledgments

This paper was supported by Research Grant “RISET DIKTI 2017” Ministry of Research, Technology, and Higher Education Indonesia.

References

[1] Melnyk I and Banerjee A 2016 Estimating Structured Vector Autoregressive Models, *Proceedings of the 33rd International Conference on Machine Learning*, New York, NY, USA

[2] Cliff A D and Ord J 1975 Space Time Modelling with an Application to Regional Forecasting, *Transactions of the Institute of British Geographers* 66 pp 119–128

[3] Ruchjana B N 2002 Generalized Space-time Autoregressive Model and its application on oil production (text in Indonesia). Ph.D Thesis. Institut Teknologi Bandung.(un-publish)

[4] Wutsqa D U and Suhartono 2010 Seasonal Multivariate Time Series Forecasting on Tourism Data by Using Var-Gstar Model. *Jurnal ILMU DASAR*, 11(1) pp 101-109

[5] Ruchjana B N, Borovkova S A and Lopuhaa H P 2012 Least Squares Estimation of Generalized Space Time Autoregressive (GSTAR) Model and Its Properties. *The 5th International Conference on Research and Education in Mathematics AIP Conf. Proc.* 1450 pp 61-64

[6] Yundari, Pasaribu U S and Mukhaiyair U 2017 Error Assumptions on Generalized STAR Model *Journal of Mathematical and Fundamental Sciences* 49 (2), pp 136-155

[7] Fadlilah N, Mukhaiyair U and Fahmi F 2015 The Generalized STAR (1;1) Modelling with Time Correlated Errors to Red-Chilli Weekly prices of some Traditional Markets in Bandung, West Java, *AIP Conference Proceeding* 1692, 020014

[8] Suhartono and Subanar 2006 The Optimal Determination of Space Weight in GSTAR Model by Using Cross-correlation Inference, *Journal of Quantitative Methods. Journal Devoted the Mathematical and Statistical Application in Various Field*, 2(2) pp 45-53

[9] Nugraha R F, Setyowati S, Mukhaiyair U and Apriliani Yulawati 2015 Prediction of oil palm production using the weighted average of fuzzy sets concept approach. *AIP Conf. Proc.* 1692, 020012

[10] Wand M and Jones M M C 1995 *Kernel Smoothing* (London : Chapman & Hall)

[11] Wikle C K, Milliff F, Nychka, D and Berliner L M 2001 Spatio-Temporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds, *Technometrics*, 96 pp 382–397

[12] Allcroft D J and Glasbey C A 2003 A Latent Gaussian Markov Random Field Model for Spatio-
Temporal Rainfall Disaggregation, *Applied Statistics*, **52** pp 487–498

[13] Pfeifer P E and Deutsch S J 1980 Identification and Interpretation of First Order Space-Time ARMA Models *Technometrics*, **22**(3) pp 397–408

[14] Mukhaiyar U and Pasaribu U S 2012 A New Procedure of Generalized STAR Modelling using IAcM Approach, *ITB Journal Sciences*, **44** A (2) pp 179–192.

[15] Davis J C 2002 *Statistics and Data Analysis in Geology* (New York : John Willey &Sons)