

## THE APPLICATION OF THE SEMIPARAMETRIC GSTAR MODEL IN DETERMINING GAMMA-RAY LOG DATA ON SOIL LAYERS

Yundari<sup>1</sup>, Shantika Martha<sup>2</sup>

<sup>1</sup> Mathematics Department, FMIPA Universitas Tanjungpura

<sup>2</sup> Statistics Department, FMIPA Universitas Tanjungpura

E-mail: [yundari@math.untan.ac.id](mailto:yundari@math.untan.ac.id)

DOI: 10.14710/medstat.14.2.108-117

### Article Info:

Received: 22 September 2020

Accepted: 28 December 2021

Available Online: 11 Januari 2022

### Keywords:

*space-time model; spatial  
weighted; superposition of rock-  
layer*

**Abstract:** This research examines the semiparametric Generalized Space-Time Autoregressive (GSTAR) spacetime modeling and determines its spatial weight. In general, the spatial weights used are uniform, binary weights, and based on the distance, the result is a fixed weight. The GSTAR model is a stochastic model that takes into account its random variables. Thus, it is necessary to study the random spatial weights. This study introduced a new method to estimate the observed value of the GSTAR model semiparametric with a uniform kernel. The data involved the Gamma Ray (GR) log data on four coal drill holes. The semiparametric GSTAR modeling aimed to predict the amount of log GR in the unobserved soil layer based on the observation data information on the layer above it and its surrounding location. The results revealed that semiparametric GSTAR modeling could predict the presence of coal seams and their thickness of drill holes. The results also highlight the validity test on the out-sample data that the error in each borehole results in a small error. In addition, the error tends to approach the actual observed value at a depth of 1 meter down.

## 1. INTRODUCTION

Coal exploration involves several initial stages, such as regional evaluation, basin study, coal seam distribution study, and resource evaluation. The purpose of the regional evaluation stage is to determine sedimentary basins and the tectonic control that controls them. The part of the basin that potentially carries coal seams is analyzed at the basin study stage. Further, the study of coal seam distribution aims to determine the depth of coal, the direction of coal seam continuity, coal thickness, and coal quality. The last stage is called resource evaluation, which estimates the potential amount of coal resources (Thomas, 2013).

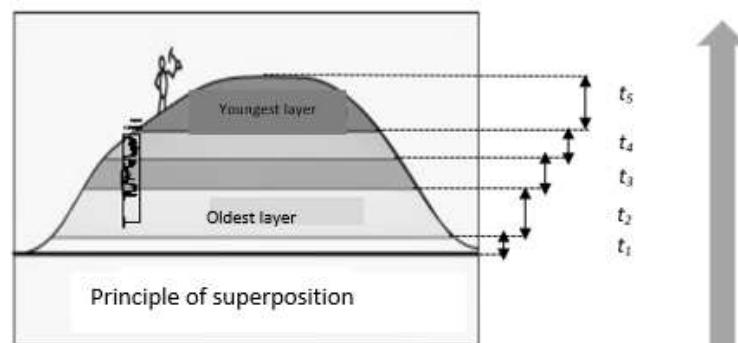
This initial stage known as the coal mapping stage, aims to obtain field data, such as the distribution of coal and non-coal seams and developing geological structures. The distribution of coal and non-coal seams includes strike, dip, and thickness sections. Geological structure defines whether there are faults or folds in the subsurface. The next stage is the drilling program which is a continuation of surface survey activities. This stage is carried out if the initial information of strike direction, dip, thickness, and coal quality gives economically promising results. The stage attempts to produce data in the form of

depth, thickness, and quality (from the sample), produce more detailed information on exposed coal (knowing the presence of undisclosed coal) and provide detailed information on the characteristics of flanking rocks. This drilling process results in a description and sampling of the cuttings or cores. The results of these samples are sometimes incomplete. Thus, they are usually equipped with geophysical logging data (Matuszak, 1972).

Geophysical logging is also known under different names, such as wireline logging, downhole logging, and well logging. It is commonly used in the petroleum industry and has various types. In coal exploration, the most-used types of geophysical logging include Gamma-Ray logs (GR), density logs, and caliper logs. These three logs have different roles. The GR log, for instance, plays a role in determining the seam correlation. Meanwhile, the density logs help identify coal, and the caliper logs are used to determine the borehole profile. With geophysical logging, the definition of coal thickness, parting thickness, and coal contact area can be known more convincingly. This data logging is also effective if there are samples lost during coring drilling, or what is generally called core-lost. Thus, the results of geophysical logging are used to correct the thickness of the coal in the drill log (reconciliation) (Thomas, 2013).

Stochastic modeling on logging data has been carried out by (Kyriakidis & Journal, 1999; Sahu, 1982). In the present research, the quantitative analysis used space-time modeling. One of the space-time models suitable for Indonesia's geographical conditions is the Generalized Space-Time Autoregressive (GSTAR) model (Yundari et al., 2017). The GSTAR model is the extension of the Autoregressive (AR) stochastic time series model. While the AR model is a stochastic model with time-dependent data, GSTAR data is both time- and location-dependent. This location parameter index is represented by a spatial weight matrix. In practice, the GSTAR modeling procedure still refers to the time series modeling developed by (Box et al., 1994). It is characterized by a spatial weight matrix that interprets the spatial dependencies between locations.

In this study, the researchers introduced a new method for estimating the observed value of the GSTAR model through a semiparametric method. This method combines parametric analysis with the smallest squares method and the non-parametric method with the kernel function approach. The role of the kernel function as a weight function can also be applied to the estimation of the GSTAR model, especially its semiparametric method, which serves as a smoothing of the parametric method results that have been previously obtained.



**Figure 1.** The Interpretation of Time Parameter Index in the Form of Rock Layer Depth in Time Series.

The application of the GSTAR model has been carried out for Gamma-Ray log data on soil layers (Yundari et al., 2018). Rock layers are represented as the index of time parameters in rock layers and can be used according to the law of superposition in rock stratigraphy (Figure 1). Stratigraphy is a geological science that explicitly discusses the lateral (space) and vertical (time) relationships per rock layer. The age of the rocks can be measured from the stratigraphic sequence, where the upper layers are formed later and can be influenced by the layers below them, which are formed earlier (Nichols, 1999). In this study, the researchers used GR log data on soil layers in four coal boreholes using the GSTAR Semiparametric model and kernel random spatial weights. More accurate modeling can be obtained with this model based on observational data in determining the spatial weight matrix.

## 2. LITERATURE REVIEW

GR log served as case study data in this study. The principle of the GR log measurement involves the recording of the earth's natural radioactivity, which originates from radioactive elements in Uranium (U), Thorium (Th), and Potassium (K) rocks. These elements continuously emit Gamma rays in the form of pulses of high radiation energy. Then, this GR log can penetrate rock and be detected by the GR sensor. Each GR detected by the sensor will cause an electric pulse on the detector. The parameter is recorded based on the number of pulses that occurred per unit of time (Harsono, 1993).

The unit of log GR is API (American Petroleum Institute) unit. The intensity of radioactivity emitted from rocks varies. When the rocks have more radioactive elements, the GR content will be higher. It also indicates that the layer is a non-coal layer. In contrast, if the number of radioactive elements is small, the GR content will be lower, implying the presence of coal seams. GR log data can be analyzed qualitatively or quantitatively. Qualitative analysis with log data can determine the type of lithology and the type of fluid in the formation that is penetrated by the borehole. Meanwhile, the quantitative analysis aims to determine thickness, porosity, permeability, fluid saturation, and hydrocarbon density.

**Table 1.** Rock Classification Based on GR Range Value.

Very Low Radioactive	Low Radioactive (32.5-60 API)	Medium Radioactive (60-100 API)	Very High Radioactive (>100 API)
Anhydrite; Salt; <b>Coal</b>	<b>Sandstone;</b> Limestone Dolomite	Arkose; Granite Stone; <b>Claystone;</b> Sand; Limestone	<b>Shale;</b> Volcanic Ash; Bentonite

Source: <http://teknik-perminyakan-Indonesia>

Geophysical log data is interpreted to determine the lithology at each depth below the earth's surface. Each rock has a unique response to the log curve to determine the type of lithology. Each lithology has different characteristics with different GR range values and density logs. Based on the value of the GR range, lithology types can be classified as in Table 1.

Research on identifying the thickness of rock or soil layers typically uses the geoelectric resistivity method (Muliadi et al., 2019). This method utilizes a resistivity

measuring instrument which is verified using drilling data. This research, then, provides a new contribution in determining the thickness of the soil or rock layers. In addition, the research uses quantitative analysis to predict GR log data in unobserved soil layers. Hence, it undoubtedly provides benefits to mining companies in making considerations prior to drilling. It depicts the predicted magnitude of the GR log in the undrilled soil layer.

One of the models that can predict data with time and location dependence is the GSTAR space-time model. Several studies have been conducted to investigate GSTAR modeling, and its development. For example, the GSTAR(1;1) model was carried out with modification of spatial weights through railroad passenger's mobility to research Covid-19 growth data in Java Island (Pasaribu et al., 2021). Further, the GSTAR-Kriging model was used in rainfall data (Abdullah et al., 2018) while the GSTAR-SUR Kriging model was applied to the data of coffee borer beetle attacks in East Java (Pramoedyo et al., 2020). The developed model is parametric because the model parameters must be estimated. The models developed are parametric because the model parameters must be estimated. Models that combine parametric and non-parametric GSTAR models are still rare. Little information can be gathered since no research has been carried out in this field by far. To date, semiparametric modeling and its statistical properties are primarily used for regression models (e.g. (Kuzairi et al., 2021; Nurcahyani et al., 2021; Wu et al., 2021)). In this study, the semiparametric GSTAR model was developed for the GR log data used in the Semiparametric GSTAR modeling as a state space. In contrast, the constant depth difference was used as a time index for the stochastic process. The time used is discretized continuous time. GSTAR modeling must have a time parameter index that has a constant interval. Through the correlation of rocks with the same relative age, the same depth interval is obtained. Therefore, in this study, the same depth of a 0.1-meter interval was used.

The GSTAR Semiparametric model was developed for models with one-time and spatial order. It also will be developed for higher-order models in the subsequent research. It is also adjusted to the GR log data on coal drilled wells whose research locations are still quite close/homogeneous based on the distance. The representation of the first-order GSTAR model and first-order spatial order was written GSTAR(1;1). It serves as an individual observation (single observation), namely observations at locations with index  $i$  at time  $t$ , can be written as:

$$Y_i(t) = \Phi_{0i}Y_i(t - 1) + \Phi_{1i} \sum_{j=1}^N w_{ij}Y_j(t - 1) + \varepsilon_i(t) \quad (1)$$

where  $N$  defines the quantity of research location,  $\Phi_{0i}$  dan  $\Phi_{1i}$  are called time AR parameter and spatial AR parameter for the location indexed  $i$ , respectively. The process of  $\{Y_i(t)\}$  on Equation (1) is the observation data at location- $i$  and at time- $t$ . The notation of  $w_{ij}$  defines weight matrices at location- $j$  towards location- $i$ . Spatial weights are zero to one. Furthermore, the  $\varepsilon_i(t)$  notation is the error at the  $i$ -location when  $t$  is normally distributed identically and independently with zero mean and constant variance.

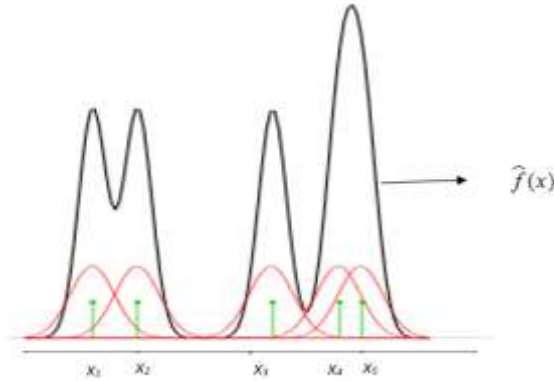
### 3. MATERIAL AND METHOD

The characteristic of the space-time model is the dependence on location and time. In the GSTAR(1;1) model, location dependence is represented by a spatial weight matrix. Research on the GSTAR model points out that the first-order spatial weight matrix  $\mathbf{W}$  is assumed to have been fixed before modeling begins.

In this modeling, the kernel function approach method was used in determining the value of the spatial weight matrix. This kernel function is commonly used as a weight function to estimate the density function and the regression function. Kernel functions work by adding up some kernel functions for each point with each point around it (see Figure 2).

In general, the kernel function ( $k(\cdot)$ ), for a point with the closest point  $x$  and  $y$  is  $k\left(\frac{x-y}{h}\right)$ .

The notation of  $h$  is a *bandwidth* that controls the smoothness of the kernel function.



**Figure 2.** Kernel Density Estimation Plot.

The kernel location weights are obtained by adopting the Nadaraya-Watson kernel estimator (Wand & Jones, 1995), which is mathematically illustrated as:

$$m_h(x) = \frac{n^{-1} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right) Y_i}{n^{-1} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right)} \quad (2)$$

It also uses the average value of the observations of each location  $\bar{Y}_i$ , with  $n$  defines the quantity of data (time). As a result, the weight of location  $j$  to location  $i$  can be written as follows:

$$\tilde{W}_{ij} = \frac{k\left(\frac{\bar{Y}_i - \bar{Y}_j}{h}\right)}{\sum_{l=1, l \neq i}^n k\left(\frac{\bar{Y}_i - \bar{Y}_l}{h}\right)}. \quad (3)$$

The average value selection of the observations of each location is intended to obtain the characteristics of the overall data (data centering) by ignoring outliers from observation data. In the form of a weight matrix, it can be written as the following Equation:

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

The weight matrix results with the kernel function approach revealed that the result still fulfills the properties of the weight matrix, including random, because the weights come from random variable data, namely observation data, and meet the characteristics of  $\sum_{j=1}^N \tilde{W}_{ij}, N > 1$ . On the other hand, the kernel function approach estimates the

GSTAR(1;1) process. The method used in this research was a semiparametric method. It is a combination of parametric and non-parametric methods. Specifically, it used partial linear for the autoregression model (Gao & Yee, 2000). The first-order partial linear model, AR(1), is defined as follows:

$$Y_t = \Phi Y_{t-1} + f(Y_{t-2}) + \varepsilon_t, t = 3, 4, \dots, T$$

where  $\Phi$  as a parameter that must be estimated,  $f$ -function is an unknown real function, and  $\varepsilon_t$  is an error vector *iid* with zero mean and finite variance of  $\sigma^2$ .

The semiparametric GSTAR(1;1) model is introduced as follows:

$$Y_i(t) = \Phi_{0i} Y_i(t-1) + \Phi_{1i} \sum_{j=1}^N w_{ij} Y_j(t-1) + f(Y_i(t-2)) + \varepsilon_i(t) \quad (4)$$

where  $f$ -function is a smoothing function (kernel) and  $t = 3, 4, \dots, T$ .

There are four stages of semiparametric method modeling. The first stage is the conventional or parametric step using the smallest squares parameter estimation. The next stage is a non-parametric step which is applied to the  $f$ -function in Equation (4), namely

$$f(Y_i(t-2)) = f_1(Y_i(t-2)) - \Phi_{0i} f_2(Y_i(t-2)) - \Phi_{1i} \sum_{j=1}^N \tilde{W}_{ij} f_2(Y_i(t-2))$$

The estimation of the  $f_i$  function uses a kernel function approach, called the Nadaraya-Watson estimator. Further, the third stage is fitting and validation. The validation stage uses a residual test, including the Autocorrelation function (ACF) plot, histogram, and qq-normal. Finally, the last stage is to make predictions from the GSTAR Semiparametric model results that have been validated.

#### 4. RESULTS AND DISCUSSION

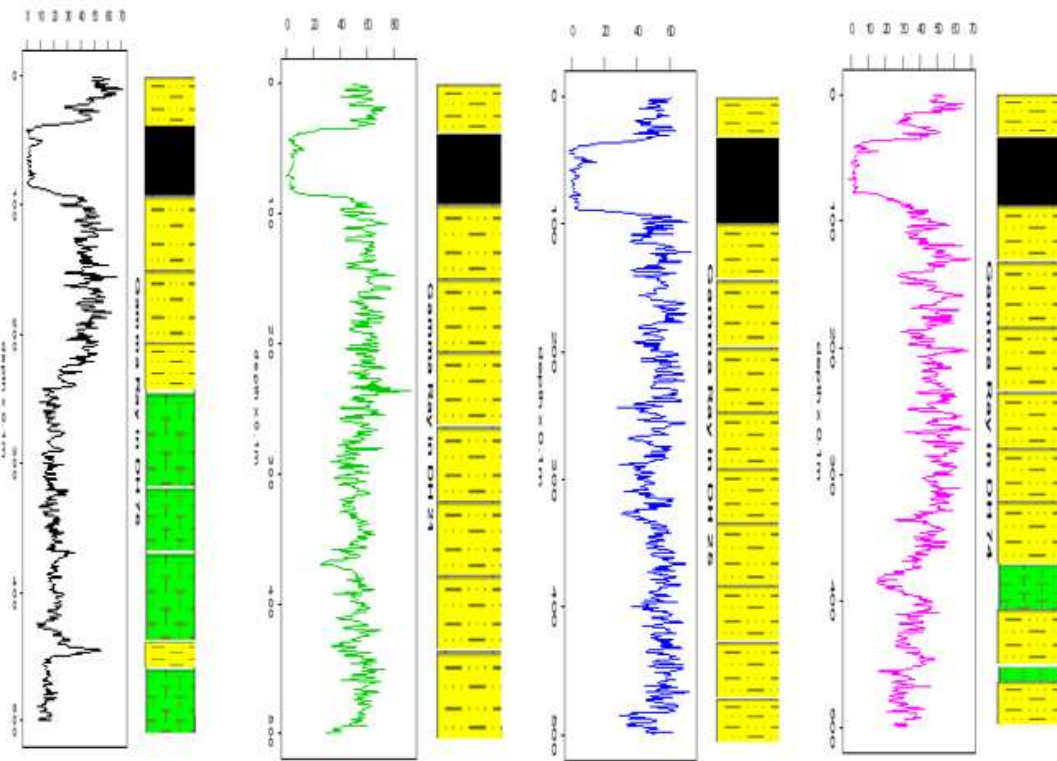
This Semiparametric GSTAR space-time modeling was applied to secondary log GR data in four coal boreholes. The data is presented in Figure 3, where the black layer is a coal seam. This Semiparametric GSTAR modeling is a stationary model. As illustrated in Figure 3, the data was still not stationary (weak). Thus, it was essential to transform the data to make it stationary and able to be modeled. After the data became stationary, the next step was forming a weight matrix with the kernel function. In this study, the uniform kernel function was employed as follows:

$$k(x) = \frac{1}{2} I(|x| < 1)$$

Spatial weight matrix with uniform kernel for the parametric method was obtained as the following:

$$\tilde{W} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0.50 & 0.50 \\ 0.33 & 0.33 & 0 & 0.33 \\ 0 & 0.50 & 0.50 & 0 \end{bmatrix}$$





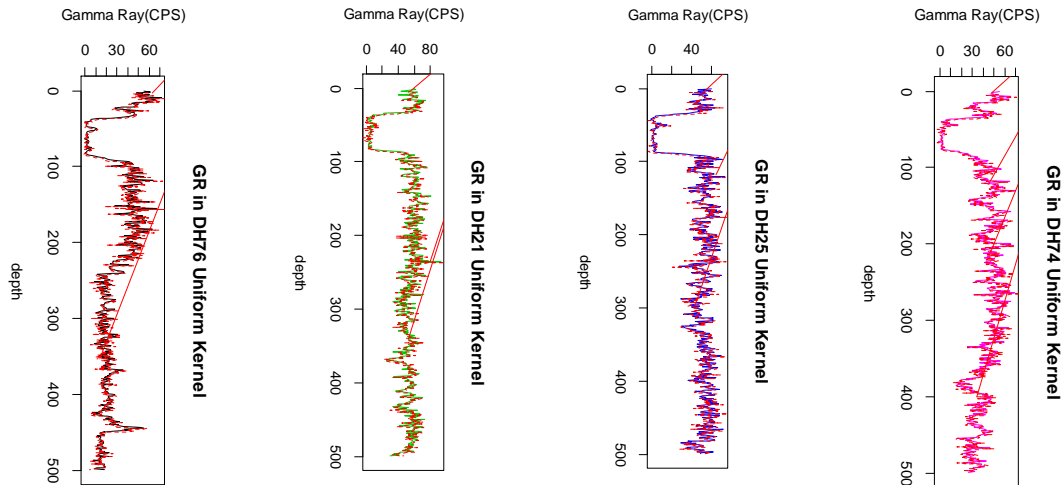
**Figure 3.** Plot Data Used for Modeling GSTAR(1;1) on 4 Coal Drill Hole/DH.

In the next step, the GSTAR modeling stage was conducted after obtaining the spatial weight matrix. The initial modeling stage was parametric modeling that was carried out for the order of spatial lag-1 and time lag-1. The results of the estimated smallest squares can be seen in Table 2.

After the parametric estimation stage was carried out, it was crucial to perform the model validation stage by validating the result parameters. The validation stage utilized the Eigenvalue method from the GSTAR(1;1) model parameter matrix between values -1 and 1. The next estimation stage was non-parametric modelling which was based on the results of the parametric method. Using the estimation of the uniform kernel function in Equation (4), the data estimation results were obtained, as shown in Figure 4.

**Table 2.** Parameter Estimation Results Using Least Squares in Parametric Stages In 4 Wells (Drill-Hole/DH)

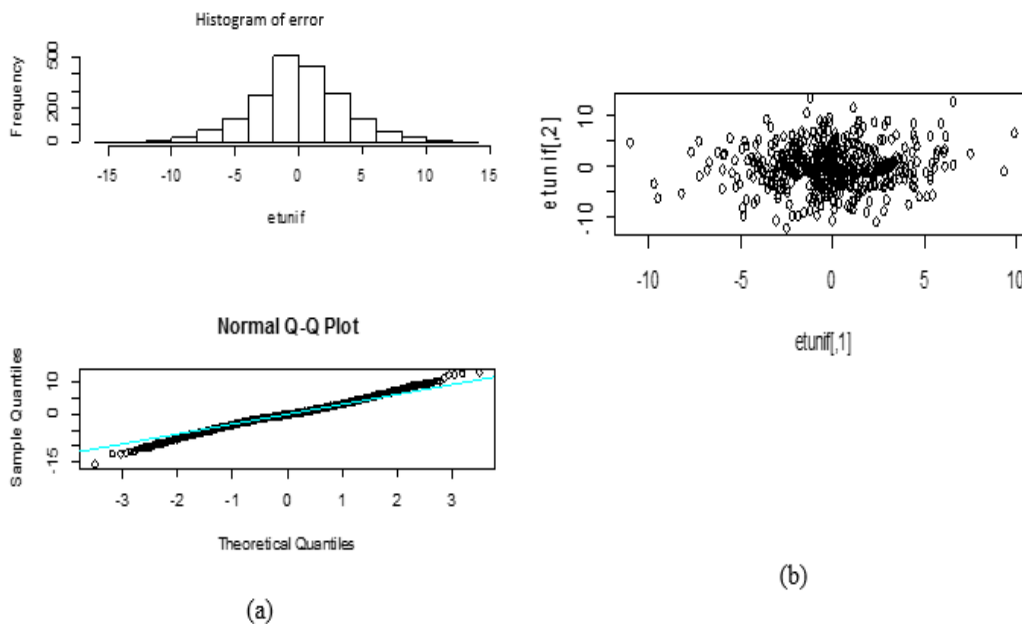
Location	$\phi_{0i}$	$\phi_{1i}$	Parameter Validation
$i$ =DH76	-0.402	0.009	Valid
$i$ =DH21	-0.372	0.020	Valid
$i$ =DH25	-0.424	-0.010	Valid
$i$ =DH74	-0.387	0.005	Valid



**Figure 4.** Plot Results of Original and Estimated Data (Red) for 4 Coal Wells

As provided in Figure 4, the estimation results followed the original data pattern. The Mean Average Percentage Error (MAPE) obtained for location DH76, DH21, DH25, and DH74 were 15.39%, 13.51%, 16.40%, and 11%. The obtained model is presented in the following Equation:

$$\hat{Y}_i(t) = \hat{\Phi}_{0i}Y_i(t-1) + \hat{\Phi}_{1i} \sum_{j=1}^4 \hat{W}_{ij}Y_j(t-1) + \hat{f}(Y_i(t-2)) \quad (5)$$



**Figure 5.** (a) Histogram and QQ-Normal Plots of Errors (b) Scatterplots of Errors

Based on the MAPE results, each location can be categorized as a good model because the error is below 20%.



**Table 3.** The Prediction Results of Unobserved GR Log Values (API) for 1 Meter and Below

Depth Location	DH76	DH21	DH25	DH74
50.1 m	12.22	30.22	19.90	34.59
50.2 m	12.40	30.74	20.64	35.31
50.3 m	12.21	30.51	20.22	35.02
50.4 m	12.14	30.52	20.27	35.07
50.5 m	12.03	30.45	20.13	35.01
50.6 m	11.93	30.42	20.07	34.98
50.7 m	11.83	30.37	19.97	34.94
50.8 m	11.73	30.32	19.89	34.90
50.9 m	11.63	30.28	19.80	34.87
51.0 m	11.53	30.23	19.72	34.83

A residual test is a step that is carried out after the model has been obtained. This test is performed visually, as shown in Figure 5. The figure shows that the residuals have met the assumptions of randomness and normality. The estimation results are categorized as 'very good' so that the next step can be done, namely prediction. The predictions using GSTAR Semiparametric modeling. The results are shown in Table 3. The table reveals that DH76 has a log GR value of less than 15 API. In other words, the seam at a depth of 50.1-51 meters is a coal seam. This is in accordance with the lithological criteria mentioned in Table 1.

## 5. CONCLUSION

The Semiparametric GSTAR method with a uniform kernel produces a reasonable estimate of the MAPE result of less than 20%. This method can predict coal seam reserves at the prediction stage according to the out-sample data on DH76. However, this model can only predict data that is not very long/deep and is only effective for 12 steps forward (short time). However, the model can be updated every time it makes predictions.

## REFERENCES

- Abdullah, A. S., Matoha, S., Lubis, D. A., Falah, A. N., Jaya, I. G. N. M., Hermawan, E., & Ruchjana, B. N. (2018). Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging Model for Predicting Rainfall Data at Unobserved Locations in West Java. *Applied Mathematics and Information Sciences*, 12(3), 607–615. <https://doi.org/10.18576/amis/120316>
- Box, G., Jenkins, G., & Reinsel, G. (1994). *Time Series Analysis Forecasting and Control* (3rd ed.). Prentice-hall International.
- Gao, J., & Yee, T. (2000). Adaptive Estimation in Partially Linear Autoregressive Models. *Canadian Journal of Statistics*, 28(3), 571–586. <https://doi.org/10.2307/3315966>
- Harsono, A. (1993). *Pengantar Evaluasi Log. Schlumberger Data Services*. Mulia Center.
- Kuzairi, Miswanto, & Mardianto, M. F. F. (2021). Semiparametric Regression Based on Fourier Series for Longitudinal Data with Weighted Lest Square (WLS) optimization. *Journal of Physics: Conference Series*, 1836(1). <https://doi.org/10.1088/1742-6596/1836/1/012038>
- Kyriakidis, P. C., & Journel, G. (1999). *Geostatistical Space – Time Models : A Review I*. 31(6), 651–684.

- Matuszak, D. R. (1972). Stratigraphic Correlation of Subsurface Geologic Data by Computer. *Journal of the International Association for Mathematical Geology*, 4(4), 331–343. <https://doi.org/10.1007/BF02114094>
- Muliadi, M., Zulfian, Z., & Muhandi, M. (2019). Identifikasi Ketebalan Tanah Gambut Berdasarkan Nilai Resistivitas 3D: Studi Kasus Daerah Tempat Pembuangan Akhir Batu Layang Kota Pontianak. *Positron*, 9(2), 86. <https://doi.org/10.26418/positron.v9i2.34821>
- Nichols, G. (1999). *Sedimentology and Stratigraphy*. Willey-Blackwell.
- Nurcahyani, H., Budiantara, I. N., & Zain, I. (2021). The Semiparametric Regression Curve Estimation by Using Mixed Truncated Spline and Fourier Series Model. *AIP Conference Proceedings International Conference on Mathematics, Computational Sciences and Statistics 2020*. <https://doi.org/10.1063/5.0042870>
- Pasaribu, U. S., Mukhaiyar, U., Huda, N. M., Sari, K. N., & Indratno, S. W. (2021). Modelling COVID-19 Growth Cases of Provinces in Java Island by Modified Spatial Weight Matrix GSTAR Through Railroad Passenger's Mobility. *Heliyon*, 7(2), e06025. <https://doi.org/10.1016/j.heliyon.2021.e06025>
- Pramoedyo, H., Ashari, A., & Fadliana, A. (2020). Interpolasi Kriging dalam Pemodelan GSTAR-SUR dan GSTARX-SUR Pada Serangan Hama Penggerek Buah Kopi. *Media Statistika*, 13(1), 25–35. <https://doi.org/10.14710/medstat.13.1.25-35>
- Sahu, B. K. (1982). Stochastic Modelling of Mineral Deposits. *Mineralium Deposita*, 17(1), 99–105. <https://doi.org/10.1007/BF00206378>
- Thomas, L. (2013). Coal geology. In *Choice Reviews Online* (Vol. 50, Issue 12). <https://doi.org/10.5860/choice.50-6779>
- Wand, M., & Jones, M. (1995). Kernel Smoothing. In *Angewandte Chemie International Edition*, 6(11), 951–952. Springer-Sciences+Business Media BV.
- Wu, Y., Wang, X., Chen, L., & Jiang, K. (2021). The Consistency for the Estimators of Semiparametric Regression Model with Dependent Samples. *Acta Mathematicae Applicatae Sinica, English Series*, 37, 299–318.
- Yundari, Pasaribu, U. S., & Mukhaiyar, U. (2017). Error Assumptions on Generalized STAR Model. *Journal of Mathematical and Fundamental Sciences*, 49(2), 136–155. <https://doi.org/10.5614/j.math.fund.sci.2017.49.2.4>
- Yundari, Pasaribu, U. S., Mukhaiyar, U., & Heriawan, M. N. (2018). Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function. *Journal of Physics: Conference Series*, 1028(1). <https://doi.org/10.1088/1742-6596/1028/1/012223>