Global Scientific and Academic Research Journal of Multidisciplinary Studies
ISSN: 2583-4088 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journals-gsarjms-home/

FORECASTING SOLAR SHORTWAVE RADIATION FOR THREE LOCATIONS IN LAMPUNG PROVINCE USING HYBRID GSTARX-SVR

By

Intan Putri Hutami[1], Netti Herawati[1*], Tiryono Ruby[1], Khoirin Nisa[1]

[1]Department of Mathematics, University of Lampung

Article History

Received: 18/11/2024
Accepted: 23/11/2024
Published: 26/11/2024

Vol – 3 Issue –11

PP: - 81-85

Abstract

This study aims to forecast solar shortwave radiation at three locations in Lampung Province using hybrid GSTARX-SVR. The data used are solar shortwave radiation as an endogenous variable and the solar irradiation duration as an exogenous variable from 3 locations in Lampung Province. The results of the study indicate that the best hybrid GSTARX-SVR model is GSTARX(6,1)-SVR with inverse distance spatial weight and ratio of training data and testing data is 90:10. These results indicate that by using the hybrid GSTARX-SVR model, the forecast range of solar shortwave radiation is between 15 and 35 MJ/m^2 for the three locations.

***Keywords**- GSTARX, SVR, Hybrid Model, Solar Shortwave Radiation, Solar Irradiation Duration*

Introduction

Generalized Space Time Autoregressive (GSTAR) is generalization of Space Time Autoregressive (STAR) model that can be used to model time series data with spatial effects (Borovkova, et al., 2008). In GSTAR model there is an assumption that the research locations are heterogeneous, so that the differences between the locations are shown in the form of spatial weighting (Hestuningtias & Kurniawan, 2023). GSTAR model then developed into Generalized Space Time Autoregressive with Exogenous Variable (GSTARX) where modeling is done by considering external variables that affect the system in addition to the factors of time and location (Astuti, et al., 2017).

In more complex cases, such as nonlinear data, hybrid GSTARX-SVR model can be used to improve forecasting accuracy (Suhartono, et al., 2019). Support Vector Regression (SVR) is one model that can handle nonlinear data well. SVR works by finding a hyperplane that predicts the target value as close to the data as possible, while limiting the margin of error within certain limits (Moguerza & Munoz, 2006).

Solar shortwave radiation data is one of the nonlinear data example. Solar shortwave radiation is the portion of solar radiation that includes the ultraviolet (UV), visible light, and some near infrared spectra with wavelengths less than 4 µm (Sianturi & Simbolon, 2021).In addition to time and spatial, the solar shortwave radiation data influenced by solar irradiation duration (Reddy, et al., 2018). Thus, a hybrid GSTARX-SVR model is applied to forecast solar shortwave radiation data by involving solar irradiation duration data as the exogenous variable.

MATERIALS AND METHODS

Data

The data used in this study are solar shortwave radiation data and solar irradiation duration data from three locations in Lampung Province, Indonesia (Bandar Lampung, Pringsewu, and Tanggamus) obtained from the website https://open-meteo.com/ for the period January 1st, 2022 to June 30th, 2024.

Method

Data modeling was carried out using the GSTARX-SVR hybrid model. To see the performance of the model, the Mean Absolute Percentage Error (MAPE) was used. The best GSTARX-SVR hybrid model was determined based on the smallest MAPE value.

Hybrid GSTARX-SVR

Hybrid GSTARX-SVR model is a hybrid model that combine Generalized Space Time Autoregressive with Exogenous Variable (GSTARX) and Support Vector Regression (SVR). Let Z_t is linear component and Y_t is nonlinear component, then hybrid model for GSTARX-SVR can be written as follows (Zhang, 2003).

$$H_t = Z_t + Y_t \tag{1}$$

GSTARX is a statistical model used to analyze time series data with spatial elements. GSTARX combines elements of the GSTAR model and adds exogenous variables to capture the influence of external factors that affect the system. Let Zt is a multivariate time series of N locations, then GSTARX with time order p and spatial order written below (Suhartono, et al., 2016).

$$Z_t = \sum_{k=1}^{p} \left(\boldsymbol{\Phi}_{k0} \boldsymbol{W}^{(0)} + \sum_{l=1}^{\lambda_p} \boldsymbol{\Phi}_{kl} \boldsymbol{W}^{(l)} \right) \boldsymbol{Z}_{t-k} + \sum_{m=0}^{s} \boldsymbol{\beta}_m \boldsymbol{X}_{t-m} + \boldsymbol{\varepsilon}_t \quad (2)$$

where $\boldsymbol{\Phi}_{k0} = diag(\phi_{10}, \ldots, \phi_{N0}), \boldsymbol{\Phi}_{kl} = diag(\phi_{1s}, \ldots, \phi_{Ns})$, and $\boldsymbol{\varepsilon}_t$ is residual model that satisfies $\boldsymbol{\varepsilon}_t \sim iidN(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.
In modeling GSTARX, the weight matrix must be determined to capture spatial interactions between locations. There are three spatial weights that used in this research, including uniform, inverse distance, and normalized cross-correlation.
Uniform weights matrix is a matrix that reflects the assumption that all locations have equal influence on each other. The uniform weight defined as follows (Fadlurrohman, 2020).

$$w_{ij} = \frac{1}{n_i} \quad (3)$$

Inverse distance matrix is a matrix that capture spatial relationship between locations by assigning weights that are inversely proportional to the distance between the locations. Let d_{ij} is the distance between location i^{th} and location j^{th}, then the inverse distance weight written as follows.

$$w_{ij} = \frac{(d_{ij})^{-1}}{\sum_{i \neq j} (d_{ij})^{-1}} \quad (4)$$

Normalized cross-correlation matrix is weight calculated based on the cross correlation between two time series at different locations. Let r_{ij} is cross correlation value between location i^{th} and location j^{th}, then the matrix element for the normalized cross-correlation weight written as follows.

$$w_{ij} = \frac{r_{ij}(1)}{\sum_{i \neq j} r_{ij}(1)} \quad (5)$$

Next, SVR is a development of Support Vector Machine (SVM) model to solve regression problems (Cao & Tray, 2003). SVR works to find regression function that approximates the relationship between input and output variables by minimizing the prediction error below a certain limit called margin or epsilon. Regression function of SVR written as follows.

$$\boldsymbol{Y} = \boldsymbol{w}^T X + b \quad (6)$$

where $\boldsymbol{w}$ is weight vector and b is bias.

In modeling SVR, kernel function must be obtained to controls how much the observations affects the model. Kernel function that used in this research is Radial Basis Function (RBF). Let γ is the parameter that control the width of basis function, then RBF can be written as follows (Ramedani, et al., 2014).

$$K(x_t, x) = \exp\left\{-\gamma \left|\left|x_t - x\right|\right|^2\right\} \quad (7)$$

RESULT AND DISCUSSION

A descriptive analysis was first performed on the solar shortwave radiation and solar irradiation duration data to get a clear picture of the data. The results of descriptive analysis can be seen in Table 1.

Table 1: Statistical Descriptive

Variable	Location	Mean	Standar Deviation	Median
Solar shortwave radiation (MJ/m²)	Bandar Lampung	17.86	3.32	17.99
	Pringsewu	17.74	3.30	17.76
	Tanggamus	18.24	3.09	18.19
Solar irradiation duration (Hour)	Bandar Lampung	9.66	1.50	10.01
	Pringsewu	9.01	2.09	9.49
	Tanggamus	9.09	1.80	9.29

Table 1 shows that Pringsewu has the lowest average solar shortwave radiation (17.74 MJ/m²), which is influenced by the shorter average solar irradiation duration (9.01 hours) per day. Meanwhile, Tanggamus shows the highest average solar shortwave radiation (18.24 MJ/m²) with average solar irradiation duration of 9.09 hours. The standard deviation shows that solar shortwave radiation data of Bandar Lampung has the highest variability (3.32), followed by Pringsewu (3.30) and Tanggamus (3.09). For the solar irradiation duration data, Pringsewu shows the highest variability (2.09), followed by Tanggamus (1.80) and Bandar Lampung (1.50). Last, median value for solar shortwave radiation and solar irradiation duration data for all locations are close to the mean, which means a nearly symmetrical distribution.

Next, time series regression model are built for the three locations. Time series regression model for Bandar Lampung, written as follows.

$$L_t^1 = 0.02964T_t^1 + 0.14065S_{1,t}^1 + 1.27766S_{2,t}^1 + N_t^1$$

Time series regression model for Pringsewu written as follows.

$$L_t^1 = 0.02964T_t^1 + 0.14065S_{1,t}^1 + 1.27766S_{2,t}^1 + N_t^1$$

Time series regression model for Tanggamus written as follows.

$$L_t^1 = 0.02964T_t^1 + 0.14065S_{1,t}^1 + 1.27766S_{2,t}^1 + N_t^1$$

Now, check whether the residuals of time series regression models are autocorrelated and nonlinear. The p-value result for autocorrelation test and linearity test are less than 0.05 which means that the residuals of the time series regression models are autocorrelated and nonlinear. Other assumptions that must be satisfied are data stationarity and location heterogeneity. Using Augmented Dickey Fuller (ADF) test, obtained that the p-value is less than 0.05 which means the data is stationary. Meanwhile, the gini ratio result is 1.000122 or it can be interpreted that the locations are heterogenous. Therefore, all conditions have been satisfied and can be continued with hybrid GSTARX-SVR. The solar shortwave radiation data modeled by GSTARX and the order determined by AIC value.

Orde	AIC	Orde	AIC
1	4.006420	7	3.970980
2	3.959694	8	3.975458
3	3.962760	9	3.979540
4	3.962124	10	3.977255
5	3.959061	11	3.988501
6	3.957264	12	3.992772

Based on Table 2, it can be seen that the lowest AIC value is in the sixth order. Therefore, the model order for the data will use GSTARX(6,1). After obtaining the best model order, the next step is to determine the appropriate spatial weights to overcome heterogeneity between locations using three spatial matrix weights (uniform spatial weight, inverse distance spatial weight, cross-correlation normalized spatial weight). Uniform spatial weight is determined using equation (3), while inverse distance spatial weight is determined using equation (4), and cross-correlation normalized spatial weight is determined using equation (5). The results of spatial weighting can be seen in Table 3.

Table 3: Spatial Weight and MAPE

Location	Uniform			MAPE
	Bandar Lampung	*Pringsewu*	*Tanggamus*	
Bandar Lampung	0	0.5	0.5	49.5346
Pringsewu	0.5	0	0.5	48.6562
Tanggamus	0.5	0.5	0	48.9392
	Inverse Distance			
	Bandar Lampung	*Pringsewu*	*Tanggamus*	
Bandar Lampung	0	0.6551	0.3449	49.5297
Pringsewu	0.4875	0	0.5125	48.6547
Tanggamus	0.3337	0.6663	0	48.9376
	Normalized Cross-Correlation			
	Bandar Lampung	*Pringsewu*	*Tanggamus*	
Bandar Lampung	0	0.5434	0.4566	49.5339
Pringsewu	0.5252	0	0.4748	48.6565
Tanggamus	0.4978	0.5022	0	48.9390

From Table 3, it can be seen that the best weight is the inverse distance spatial weight compared to other spatial weight methods because it has the lowest MAPE value for Bandar Lampung, Pringsewu, and Tanggamus with MAPE values of 49.5297, 48.6547, and 48.9376, respectively. Furthermore, GSTARX(6,1) modeling is carried out with inverse distance spatial weight for the three locations. The following is GSTARX(6,1) with inverse distance spatial weight for Bandar Lampung:

$$Z_t^1 = 0.224Z_{t-1}^1 + 0.137Z_{t-1}^2 + 0.072Z_{t-1}^3 + 0.131Z_{t-2}^1 + 0.070Z_{t-2}^2 + 0.037Z_{t-2}^3 + 0.018Z_{t-3}^1 + 0.029Z_{t-3}^2 + 0.015Z_{t-3}^3 - 0.162Z_{t-4}^1 + 0.212Z_{t-4}^2 + 0.112Z_{t-4}^3 + 0.189Z_{t-5}^1 - 0.103Z_{t-5}^2 - 0.054Z_{t-5}^3 + 0.065Z_{t-6}^1 + 0.077Z_{t-6}^2 + 0.041Z_{t-6}^3 + 1.839X_{t-1}^1$$

The GSTARX(6,1) model with inverse distance spatial weight for Pringsewu is:

$$Z_t^2 = 0.169Z_{t-1}^1 - 0.024Z_{t-1}^2 + 0.178Z_{t-1}^3 + 0.046Z_{t-2}^1 + 0.075Z_{t-2}^2 + 0.048Z_{t-2}^3 - 0.010Z_{t-3}^1 + 0.051Z_{t-3}^2 - 0.011Z_{t-3}^3 + 0.009Z_{t-4}^1 + 0.073Z_{t-4}^2 + 0.009Z_{t-4}^3 + 0.088Z_{t-5}^1 - 0.141Z_{t-5}^2 + 0.093Z_{t-5}^3 + 0.035Z_{t-6}^1 + 0.035Z_{t-6}^2 + 0.037Z_{t-6}^3 + 1.982X_{t-1}^2$$

The GSTARX(6,1) model with inverse distance spatial weight for Tanggamus is:

$$Z_t^3 = 0.057Z_{t-1}^1 + 0.115Z_{t-1}^2 + 0.270Z_{t-1}^3 + 0.027Z_{t-2}^1 + 0.054Z_{t-2}^2 + 0.112Z_{t-2}^3 + 0.011Z_{t-3}^1 + 0.022Z_{t-3}^2 + 0.064Z_{t-3}^3 - 0.017Z_{t-4}^1 - 0.033Z_{t-4}^2 + 0.148Z_{t-4}^3 + 0.007Z_{t-5}^1 + 0.013Z_{t-5}^2 + 0.053Z_{t-5}^3 + 0.027Z_{t-6}^1 + 0.053Z_{t-6}^2 + 0.070Z_{t-6}^3 + 1.929X_{t-1}^3$$

After modeling with the GSTARX(6,1) model with inverse distance spatial weights, the residuals are divided into training data and testing data. By using various comparisons of training data and testing data. Then modeling is carried out with the SVR method for Bandar Lampung, Pringsewu, and Tanggamus using training data with the grid search method and the results are presented in Table 4.

Table 4: SVR Parameter and MAPE

Data Training Size	Data Testing Size	Locations	Cost	γ	ε	MAPE
70%	30%	Bandar Lampung	0.001	0.001	0.8	19.79
		Pringsewu	0.1	100	0.8	37.49
		Tanggamus	0.001	1000	0.8	26.91
80%	20%	Banda	0.1	10	0.7	21.99

		r Lampung				
		Pringsewu	0.01	100	0.8	109.95
		Tanggamus	0.01	100	0.8	60.62
90%	10%	Bandar Lampung	0.001	0.001	0.8	19.74
		Pringsewu	1	0.001	0.8	24.39
		Tanggamus	1000	0.001	0.8	26.31

Table 4 shows that the best SVR model is the model that has the smallest MAPE value obtained in the ratio of training data and testing data 90:10 with MAPE of 19.74, 24.39, 26.31 for Bandar Lampung, Pringsewu, and Tanggamus cities, respectively.

After obtaining the best SVR model, the GSTARX(6,1) model is combined with the SVR model to obtain a hybrid model GSTARX(6,1)-SVR. The hybrid model GSTARX(6,1)-SVR for Bandar Lampung is:

$$H_t^1 = 0.224Z_{t-1}^1 + 0.137Z_{t-1}^2 + 0.072Z_{t-1}^3 + 0.131Z_{t-2}^1 + 0.070Z_{t-2}^2 + 0.037Z_{t-2}^3 + 0.018Z_{t-3}^1 + 0.029Z_{t-3}^2 + 0.015Z_{t-3}^3 - 0.162Z_{t-4}^1 + 0.212Z_{t-4}^2 + 0.112Z_{t-4}^3 + 0.189Z_{t-5}^1 - 0.103Z_{t-5}^2 - 0.054Z_{t-5}^3 + 0.065Z_{t-6}^1 + 0.077Z_{t-6}^2 + 0.041Z_{t-6}^3 + 1.839X_{t-1}^1 - 3.63\exp\left(-10^{-3}||x_1 - x||^2\right) - 4.03\exp\left(-10^{-3}||x_2 - x||^2\right) + \cdots - 4.07\exp\left(-10^{-3}||x_{815} - x||^2\right) - 0.12$$

Hybrid GSTARX(6,1)-SVR model for Pringsewu is:

$$H_t^2 = 0.169Z_{t-1}^1 - 0.024Z_{t-1}^2 + 0.178Z_{t-1}^3 + 0.046Z_{t-2}^1 + 0.075Z_{t-2}^2 + 0.048Z_{t-2}^3 - 0.010Z_{t-3}^1 + 0.051Z_{t-3}^2 - 0.011Z_{t-3}^3 + 0.009Z_{t-4}^1 + 0.073Z_{t-4}^2 + 0.009Z_{t-4}^3 + 0.088Z_{t-5}^1 - 0.141Z_{t-5}^2 + 0.093Z_{t-5}^3 + 0.035Z_{t-6}^1 + 0.035Z_{t-6}^2 + 0.037Z_{t-6}^3 + 1.982X_{t-1}^2 - 4.09\exp\left(-10^{-3}||x_1 - x||^2\right) - 5.96\exp\left(-10^{-3}||x_2 - x||^2\right) + \cdots - 6.10\exp\left(-10^{-3}||x_{815} - x||^2\right) + 0.66$$

Hybrid GSTARX(6,1)-SVR model for Tanggamus is:

$$H_t^3 = 0.057Z_{t-1}^1 + 0.115Z_{t-1}^2 + 0.270Z_{t-1}^3 + 0.027Z_{t-2}^1 + 0.054Z_{t-2}^2 + 0.112Z_{t-2}^3 + 0.011Z_{t-3}^1 + 0.022Z_{t-3}^2 + 0.064Z_{t-3}^3 - 0.017Z_{t-4}^1 - 0.033Z_{t-4}^2 + 0.148Z_{t-4}^3 + 0.007Z_{t-5}^1 + 0.013Z_{t-5}^2 + 0.053Z_{t-5}^3 + 0.027Z_{t-6}^1 + 0.053Z_{t-6}^2 + 0.070Z_{t-6}^3 + 1.929X_{t-1}^3 - 7.15\exp\left(-10^{-3}||x_1 - x||^2\right) - 3.78\exp\left(-10^{-3}||x_2 - x||^2\right) + \cdots - 7.20\exp\left(-10^{-3}||x_{815} - x||^2\right) + 0.54$$

Based on the Hybrid GSTARX(6,1)-SVR model obtained for each of the above locations, the estimated solar shortwave radiation for the period from July 1st, 2024 to December 31st, 2024 at the three locations is visualized in Figure 1.

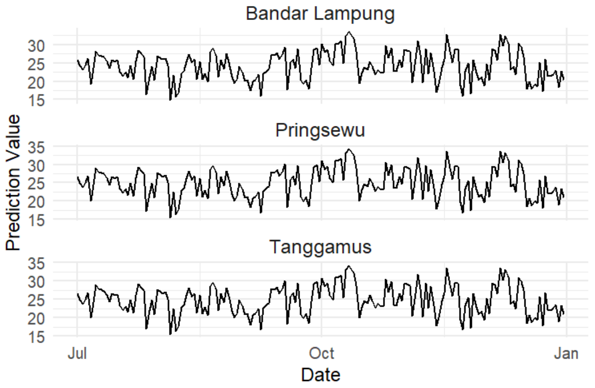

Figure 1: Forecasting Results of Solar Shortwave Radiation Data

Figure 1. shows the similarity of the fluctuation pattern of solar shortwave radiation in Bandar Lampung, Pringsewu, and Tanggamus. It can also be seen that the forecast data for the solar shortwave radiation ranges from 15 to 35 MJ/m^2 for the three locations.

CONCLUSION

From the research results obtained, it can be concluded that the best GSTARX-SVR hybrid model for the Bandar Lampung, Pringsewu, and Tanggamus areas is the GSTARX(6,1)-SVR hybrid model with inverse distance spatial weight with a ratio of 90:10. In addition, the forecast of solar shortwave radiation for the next 6 months for the three locations based on this best model shows that solar shortwave radiation ranges from 15 to 35 MJ/m^2.

REFERENCES

1. Borovkova, S., Lopuhaa, H.P., & Ruchjana, B.N. (2008). Consistency and Asymptotic Normality of Least Square Estimators in Generalized STAR Models. Statistica Neerlandica, 62(4), 428-508.
2. Hestuningtias & Kurniawan, M.H.S. (2023). The Implementation of The Generalized Space Time Autoregressive (GSTAR) Model for Inflation Prediction. Enthusiastic, 3(2), 176-188.
3. Astuti, D., Ruchjana, B.N., & Soemartini. (2017). Generalized Space Time Autoregressive with Exogenous Variable Model and Its Application. Journal of Physics, 893(1).
4. Suhartono, Maghfiroh, B., and Rahayu, S.P. (2019). Hybrid VARX-SVR and GSTARX-SVR for Forecasting Spatio Temporal Data. International Journal of Innovative Technology and Exploring Engineering, 8(4), 212-218.
5. Moguerza, J.M. & Munoz, A. (2006). Support Vector Machines with Applications. Statistical Science, 21(3), 322-336.
6. Sianturi, Y. & Simbolon, C.M. (2021). Pengukuran dan Analisa Data Radiasi Matahari di Stasiun Klimatologi Muaro Jambi. Megasains, 12(1): 40-47.

7. Reddy, M.V., Suvarana, R.P., Kumari, S.P., Balakrishnaiah, G., & Gopal, K.R. (2018). Estimation of Global Solar Radiation using Sunshine and Temperature Based Models Over Anantapur Region. International Journal of Creative Research Thoughts, 6(2), 607-618.
8. Zhang, G.P. (2003). Time Series Forecasting using A Hybrid ARIMA and Neural Network Model. Neurocomputing, 50, 159-175.
9. Suhartono, Wahyuningrum, S.R., Setiawan, & Akbar, M.S. (2016) GSTARX-GLS Model for Spatio Temporal Data Forecasting. Malaysian Journal of Mathematical Sciences, 10, 91-103.
10. Fadlurrohman, A. (2020). Integration of GSTAR-X and Uniform Location Weights for Forecasting Inflation Survey of Living Costs in Central Java. Journal of Intelligent Computing and Health Informatics, 1(1), 23-29.
11. Cao, L.J. & Tray, F.E.H. (2003). Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. IEEE Transactions on Neural Networks, 14(6), 1506-1518.
12. Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of Radial Basis Function Based Support Vector Regression for Global Solar Radiation Prediction. Renewable and Sustainable Energy Reviews, 39, 1005-1011.

*Corresponding Author: **Netti Herawati**