InPrime: Indonesian Journal of Pure and Applied Mathematics
Vol. 6, No. 1 (2024), pp. 1 – 13, doi: 10.15408/inprime.v6i1.36049
p-ISSN 2686-5335, e-ISSN: 2716-2478

Generalized Space Time Autoregressive (GSTAR) Model for Air Temperature Forecasting in the South Sumatera, Riau, and Jambi Provinces

Ayu Aprianti*, Naflah Faulina and Mustofa Usman
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
University of Lampung, Bandar Lampung, Indonesia
Email: *ayuaprianti421@gmail.com

Abstract

Over the past few years, there has been a significant increase in air temperatures in regions such as South Sumatera, Riau, and Jambi, posing threats of drought, water resource crises, and erratic weather patterns. In response, developing air temperature forecasting techniques becomes imperative for effective climate change management. This study proposes implementing the Generalized Space Time Autoregressive (GSTAR) model as a practical approach for forecasting air temperatures in these regions using two weighting methods, i.e., inverse distance and normalized cross-correlation weighting. The GSTAR model, an extension of the Space Time Autoregressive (STAR) model, offers enhanced complexity by incorporating specific time and location factors, thereby increasing forecasting flexibility. The result reveals that GSTAR(1,1) with normalized cross-correlation weighting is the most optimal model, with a Root Mean Square Error (RMSE) value of 3.135, indicating high forecasting accuracy. The selection of this model is grounded in the geographical proximity and similarity of environmental characteristics of the three regions. This research contributes novel insights into the underlying mechanisms of air temperature dynamics in neighboring areas, providing a robust foundation for formulating effective policy and mitigation strategies in addressing climate change challenges.

Keywords: Air temperatures, Normalized cross-correlation weighting, GSTAR(1,1), Inverse distance weighting.

Abstrak

Dalam beberapa tahun terakhir, suhu udara mengalami peningkatan signifikan di wilayah-wilayah seperti Sumatera Selatan, Riau, dan Jambi, yang mengancam kekeringan, krisis sumber daya air, dan perubahan pola cuaca yang tidak terduga. Menghadapi situasi tersebut, pengembangan teknik peramalan suhu udara diperlukan untuk mengantisipasi dan mengelola dampak ekstrem dari perubahan iklim. Studi ini mengusulkan implementasi model Generalized Space Time Autoregressive (GSTAR) sebagai pendekatan praktis untuk meramalkan suhu udara di wilayah-wilayah tersebut menggunakan dua metode pembobotan yaitu pembobotan invers jarak dan normali korelasi silang. Model GSTAR, sebagai perluasan dari model Space Time Autoregressive (STAR), menawarkan kompleksitas yang lebih baik dengan menggabungkan faktor-faktor waktu dan lokasi tertentu, sehingga meningkatkan fleksibilitas dalam ramalan. Hasil analisis menunjukkan bahwa GSTAR(1,1) dengan pemberian bobot normalisasi korelasi silang merupakan model yang paling optimal, dengan nilai Root Mean Square Error (RMSE) sebesar 3.135, menandakan tingkat akurasi yang tinggi. Pemilihan model ini didasarkan pada kedekatan geografis dan kesamaan karakteristik lingkungan dari ketiga wilayah tersebut. Penelitian ini memberikan wawasan baru dalam mekanisme dinamika suhu udara di wilayah-wilayah yang berdekatan, serta memberikan dasar yang kuat bagi perumusan kebijakan dan strategi mitigasi yang efektif dalam menghadapi tantangan perubahan iklim.

Kata Kunci: *Bobot invers jarak, Bobot normalisasi korelasi silang, GSTAR(1,1), Suhu udara.*

2020MSC: 62P30

* Corresponding author
Submitted November 23th, 2023, Revised March 31st, 2024,
Accepted for publication April 17th, 2024, Published Online May 31st, 2024

1. INTRODUCTION

Indonesia is located near the equator, thus having a tropical climate renowned for its consistently high temperatures year-round. Its geographical position allows for high-intensity sunlight, leading to consistently high temperatures throughout the year. In Indonesia, air temperatures can significantly vary between day and night, especially during the dry season, ranging from an average of 23ºC to 38ºC [1]. There has been a consistent upward trend in air temperatures in recent years. Characteristic of the tropical climate, increasing temperatures coincide with heightened rainfall intensity due to high relative humidity (RH), signifying a higher presence of water vapor in the air and increased seawater evaporation [2]. The main consequences of rising temperatures encompass drought, water shortages, and alterations in weather patterns [3]. Elevated temperatures in regions such as South Sumatera, Riau, and Jambi provinces can increase the likelihood of forest and land fires. These effects are notably significant, particularly when high temperatures coincide with dry conditions and strong winds, making vegetation, such as plants and trees, highly susceptible to fire. Proper anticipation of temperature changes requires effective mitigation measures. The development of air temperature forecasting methods is a crucial issue that needs attention. Therefore, choosing a method that accurately and precisely forecasts air temperatures is essential in anticipating temperature changes and minimizing their adverse effects.

Forecasting is a statistical analysis used to predict future events based on simultaneous observation data within the same time interval, known as time series data [4]. Forecasting methods are applied to univariate and multivariate time series data (involving one variable) and multivariate data (involving multiple variables). One method employed for forecasting multivariate time series data is the Vector Autoregressive (VAR) model. In the VAR model, variables within the dataset rely on their preceding values (lags) and other variables, allowing the model to capture dynamic relationships among the variables [5][6].

With advancements, research suggests that multivariate time series data are not only correlated with preceding periods but may also exhibit interconnections across different geographical locations. Geographically, the proximity of South Sumatera, Riau, and Jambi provinces allows for interrelated air temperature data across both time and location. Modeling multivariate time series data while considering linkages between time and location is commonly called a Space-Time model. One such model is the Space-Time Autoregressive (STAR) model introduced by Pfeifer and Deutsch (1980). However, the homogenous assumptions regarding autoregressive and spatial parameters across all locations in this model were addressed by Ruchjana (2002) through the development of the Generalized Space-Time Autoregressive (GSTAR) model [7][8].

The GSTAR model represents a significant advancement from the STAR model, providing enhanced realism in handling multivariate data encompassing diverse spatial and temporal variations across different locations. It enables the incorporation of heterogeneous autoregressive parameters specific to varying locations, thereby allowing for a more accurate representation of the interconnections between time and locations exhibiting different characteristics. These diverse parameters for distinct locations are presented as weight matrices [4]. Using weight matrices within the GSTAR model also facilitates improved spatial connections, leading to a more comprehensive understanding and improved forecasting of temperature fluctuations in regions characterized by diverse attributes.

The modeling using GSTAR follows the Box-Jenkins approach, encompassing model identification, parameter estimation, model diagnostic, and validation [9]. The application of the GSTAR model in multivariate time series data modeling has garnered significant attention due to its capacity to enhance forecasting accuracy. The GSTAR model offers a more flexible and realistic approach to addressing complex data with spatial and temporal interconnections, aiding in identifying trends and patterns that may emerge in the future [9]. An instance of the GSTAR model's application is found in the research conducted by Muzdhalifah, Tarno, and Kartikasari in 2022, where they forecasted domestic flights in three airports on Java Island. The accuracy achieved through the GSTAR method resulted in a MAPE value of <10% [10]. Similarly, a study by Ilmi, Aswi, and Aidid aimed at rainfall prediction in Makassar City yielded the GSTARIMA(1,0,0) $(1,1,0)^{12}$ model with the smallest RMSE value of 132.9661 [11].

Based on the aforementioned research, our study aims to leverage the GSTAR methodology to derive an optimal GSTAR model utilizing air temperature datasets from the South Sumatera, Riau, and Jambi provinces. Innovatively, we integrate inverse distance weighting and normalized cross-correlation weighting methods to the GSTAR framework, aiming to enhance forecasting accuracy by minimizing the Root Mean Square Error (RMSE). This approach offers a comprehensive solution for addressing the challenges posed by climate change in the studied regions, equipping government agencies, research institutions, and other stakeholders with robust data to formulate effective plans for mitigating and adapting to ongoing temperature changes.

2. METHOD

2.1. GSTAR Model

The general form of the GSTAR(p, λ_s) model, where p represents the time order (AR) and λ_s represents the spatial order, is as follows [12]:

$$Z(t) = \sum_{k=1}^{p} \left(\Phi_{k0} + \sum_{l=1}^{\lambda_s} \Phi_{kl} W^{(l)} \right) Z(t-k) + e(t) \quad t = 0 \pm 1 \pm 2, \ldots, \tag{1}$$

where, $Z(t)$ denotes a matrix of time series variables at time t with a size of $(N \times 1)$. $Z(t-k)$ represents a matrix of size $(N \times 1)$ containing N variables at time $t - i$, i = 1,2,…,N. Φ_{k0} is the parameter autoregressive matrix $diag(\Phi_{k0}^1, \ldots, \Phi_{k0}^n)$. Φ_{kl} is the parameter autoregressive matrix $diag(\Phi_{kl}^1, \ldots, \Phi_{kl}^n)$. $W^{(l)}$ is the location weighting matrix for the lth location with a size of $(N \times N)$, and $e(t)$ represents the residual at time t with a size of $(N \times 1)$.

The following is the general form of the GSTAR(1,1) model equation for the three locations.

$$\boldsymbol{Z_1(t)} = \Phi_{10^{(1)}} Z_1(t-1) + \Phi_{11^{(1)}} w_{12^{(1)}} Z_2(t-1) + \Phi_{11^{(1)}} w_{13^{(1)}} Z_3(t-1) + e_1(t), \tag{2}$$
$$\boldsymbol{Z_2(t)} = \Phi_{10^{(2)}} Z_2(t-1) + \Phi_{11^{(2)}} w_{21^{(1)}} Z_1(t-1) + \Phi_{11^{(2)}} w_{23^{(1)}} Z_3(t-1) + e_2(t), \tag{3}$$
$$\boldsymbol{Z_3(t)} = \Phi_{10^{(3)}} Z_3(t-1) + \Phi_{11^{(3)}} w_{31^{(1)}} Z_1(t-1) + \Phi_{11^{(3)}} w_{32^{(1)}} Z_2(t-1) + e_3(t). \tag{4}$$

2.2. Parameter Estimation of GSTAR Model

The significant parameters are estimated using the Ordinary Least Squares (OLS) method [13] when developing the GSTAR model. The GSTAR model equation for all locations can be expressed as follows.

$$Z = Z^{*}\Phi + e. \tag{5}$$

Equation (2) yields the parameter estimation of the model for Φ as follows [14].

$$\hat{\Phi} = (Z^{*\prime}Z^{*})^{-1}(Z^{*\prime}Z), \tag{6}$$

where

$$Z = \begin{bmatrix} Z_1(1) \\ Z_1(2) \\ \vdots \\ Z_1(T) \\ \vdots \\ Z_1(T) \\ Z_2(T) \\ \vdots \\ Z_N(T) \end{bmatrix},\quad Z^{*} = \begin{bmatrix} Z_1(T-k) & \cdots & V_1(T-k) & \cdots & 0 & \cdots & 0 \\ Z_1(T-k) & \cdots & V_1(T-k) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_1(T-k) & \cdots & V_1(T-k) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & Z_N(T-k) & \cdots & V_N(T-k) \\ 0 & \cdots & 0 & \cdots & Z_N(T-k) & \cdots & V_N(T-k) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & Z_N(T-k) & \cdots & V_N(T-k) \end{bmatrix},\ \Phi = \begin{bmatrix} \Phi_{k0}^{1} \\ \Phi_{k0}^{2} \\ \vdots \\ \Phi_{k0}^{N} \\ \vdots \\ \Phi_{kl}^{1} \\ \Phi_{kl}^{2} \\ \vdots \\ \Phi_{kl}^{N} \end{bmatrix},\ e = \begin{bmatrix} e_1(1) \\ e_1(2) \\ \vdots \\ e_1(T) \\ \vdots \\ e_2(1) \\ e_2(2) \\ \vdots \\ e_N(T) \end{bmatrix}.$$

The equation $V_i(t)$ represents $\sum_{k-1}^{p} W_{ij}(k)Z_i(t)$, where $i \neq j$.

2.3. Weight Matrix for GSTAR Model

In the GSTAR modeling, the interlocation relationships are paramount, necessitating the consideration of appropriate weighting factors incorporated into the GSTAR model through the weight matrix W [15]. Suhartono and Subanar (2006) proposed several methods for determining location weighting in the GSTAR model, such as uniform weighting, binary weighting, inverse distance weighting, and normalized cross-correlation weighting [16]. Uniform weighting is deemed unsuitable for GSTAR modeling due to the heterogeneous characteristics among locations [17]. Similarly, binary weighting is unfit for the GSTAR model as it is subjective, involving weight values adjusted according to the distances between time series variables [18]. Consequently, inverse distance weighting is often employed as a weighting mechanism in the GSTAR model. Inverse distance weighting results from the calculation of actual distances, considering geographic coordinates based on latitude and longitude, which are subsequently normalized. The inverse distance weighting can be expressed as follows [19].

$$W_{ij} = \frac{w_{ij}^{*}}{\sum_{k=1}^{p} w_{ik}^{*}}. \tag{7}$$

with,

$$w_{ij}^{*} = \begin{cases} \frac{1}{d_{ij}}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{8}$$

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}. \tag{9}$$

where d_{ij} represents the distance from location *i* to *j*, (u_i, u_j) denotes the latitude coordinates, and (v_i, v_j) represents the longitude coordinates. Besides inverse distance weighting, another alternative to consider as a weighting mechanism in the GSTAR model is the normalized cross-correlation weighting. This is because this weighting factor considers the correlation values of data that exhibit both spatial and temporal effects. Generally, the formula for normalized cross-correlation weighting is expressed as follows [17].

$$W_{ij}(k) = \frac{r_{ij}(k)}{\sum_{j\neq i}|r_{ij}(k)|}, \tag{10}$$

where $i \neq j$ and $\sum_{j\neq i}|W_{ij}| = 1$, $k = 1,2,\dots,p$, and the estimation of cross-correlation on the sample data is formulated as follows [7].

$$r_{ij}(k) = \frac{\sum_{t=k+1}^{n}[Z_i(t)-\bar{Z}_i][Z_j(t-k)-\bar{Z}_j]}{\sqrt{(\sum_{t=1}^{n}[Z_i(t)-\bar{Z}_i]^2)(\sum_{t=1}^{n}[Z_j(t-k)-\bar{Z}_j]^2)}}, \tag{11}$$

where $Z_i(t)$ represents the time-series data at time *t* in region *i*, Z_j represents the time-series data at time *t* in region *j*, and *k* is the time lag.

2.4. Data Analysis

The GSTAR model is a method used to estimate air temperature based on multivariate time series data, considering the spatial proximity and temporal dependencies between various locations. This approach integrates spatial and temporal aspects of the data to model the relationship between air temperature locations at different locations. There are four steps in the GSTAR method as follows.

The first stage is model identification:

a. Test the stationarity using the Augmented Dickey-Fuller (ADF) test. If hypotesis null failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time dependent structure.
b. Test the correlation between locations using the Pearson Product Moment test. If hypotesis null failed to be rejected, then spatial modeling cannot continue.
c. Test the spatial heterogeneity using the Gini Index value. If $G_n = 0$, modeling is limited to the Space Time Autoregressive (STAR) model because the expected value is $G_n \geq 1$; namely, spatial heterogeneity is met [20].
d. Identify the order of the GSTAR model. The time order is determined based on the smallest AICC value [21].

$$AICC = n\log(2\pi) + n\log(\hat{\sigma}^2) + n + 2\frac{n(m+1)}{n-m-2}, \tag{12}$$

where $\hat{\sigma}^2$ is the estimated error variance, *n* is the number of observations, and *m* is the number of estimated parameters. Meanwhile, determining the spatial order of the GSTAR model is only limited to spatial order 1 (λ_s=1) because spatial orders of more than one are difficult to represent [22].
e. Determines the general form of the GSTAR model.

The second stage of model estimation:

a. Calculate the weighted matrix using the inverse distance weighting and the normalized cross-correlation weighting.

b. Estimate the model parameters using the OLS method, and their significance was tested based on the two location weights used.
c. Determine the final model for the GSTAR based on the two weight methods.

The third stage is model diagnostic:
Model diagnostic testing uses residual data to determine the feasibility of the GSTAR model formed. It is carried out by testing white noise residuals using the Ljung Box-Pierce test and normally distributed residuals using the Kolmogorov-Smirnov test. H_0 fails to be rejected if the p-value > α. It's mean that the model is appropriate to use.

The fourth stage is model validation:
Determining the best GSTAR model based on the smallest RMSE value [23].

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(Z_t - \hat{Z}_t\right)^2}\ . \tag{13}$$

3. RESULTS AND DISCUSSION

This research used secondary data regarding monthly air temperatures (℃) in South Sumatera, Riau, and Jambi provinces. The data were obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG), available at https://www.bmkg.go.id/, totaling 288 datasets from January 2015 to December 2022. Each variable in the study comprises 96 datasets divided into two parts: 95% training data used for GSTAR model formation and 5% testing data used for validating the most optimal GSTAR model. The dataset is depicted in Figure 1.

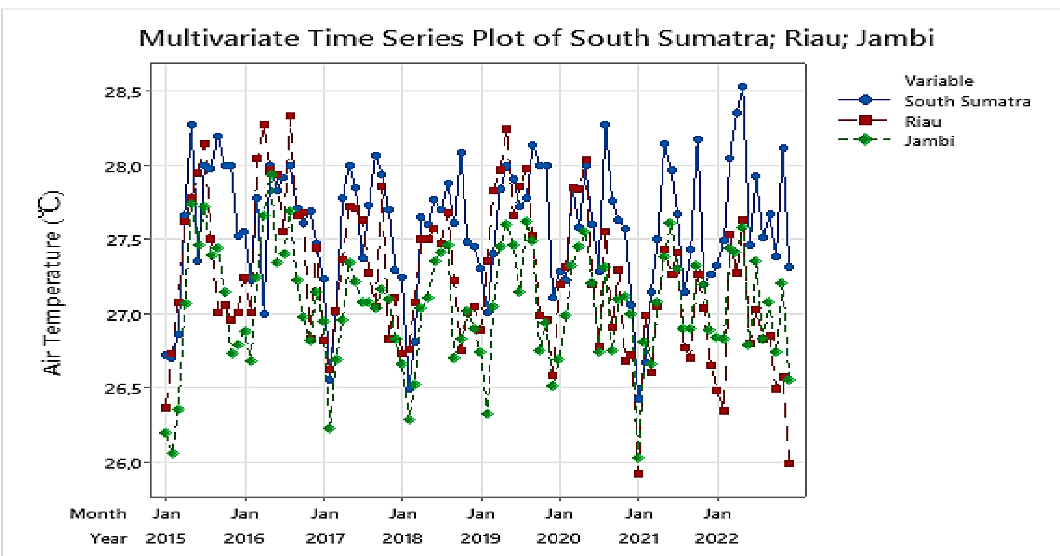

Figure 1. Multivariate time series plot of air temperature (℃) in the South Sumatera, Riau, and Jambi Provinces.

3.1. Descriptive Statistics

The initial step in forming the GSTAR model involves conducting descriptive statistics on the data. Descriptive statistical information is presented in Table 1. Based on Table 1, the highest average air temperature is recorded in South Sumatera Province, at 27.59℃, with a standard deviation of 0.447℃, indicating a relatively wide range of variability in air temperatures in that province. Conversely, the lowest average air temperature is recorded in Jambi Province, at 27.06℃, with a standard deviation of 0.399℃, suggesting that the air temperature data cluster at the mean value.

Table 1. Descriptive statistics of temperature in the South Sumatera, Riau, and Jambi Provinces.

Location	Average (℃)	StDev (℃)	Min (℃)	Max(℃)
South Sumatera	27.59	0.447	26.42	28.53
Riau	27.27	0.495	25.92	28.34
Jambi	27.06	0.399	26.03	27.94

3.2. Identifying the Model

The first step in this modelling is test the data stationarity. We use the ADF test to determine the stationarity in mean of the three locations. Based on the 5% significance level employed, the test indicates that the data of the three research locations have been stationary concerning the mean.

Table 2. Results of ADF test for stationary of air temperature data

Location	P-value	Decision about H_0
South Sumatera	0.01	Reject H_0
Riau	0.01	Reject H_0
Jambi	0.01	Reject H_0

The correlation test is conducted to determine the significant relationship among research locations. If the relationship among locations is not met, modeling using the space-time method cannot proceed. Based on the analysis of hypothesis testing using Pearson Product Moment with a significance level of $\alpha = 5\%$, it is indicated that the three research locations exhibit a significant correlation with a wide range of relationships ranging from moderate to strong, as follows.

Table 3. Results of Pearson Product Moment test

Location	South Sumatera	Riau	Jambi
South Sumatera	1	0.516	0.752
p-value	0.000	0.000	0.000
Riau	0.516	1	0.763
p-value	0.000	0.000	0.000
Jambi	0.752	0.763	1
p-value	0.000	0.000	0.000

The spatial heterogeneity test using the Gini Index yielded results in Table 4 indicating that all three locations have fulfilled the assumption, allowing the continuation of GSTAR modeling. After fulfilling the spatial conditions, the next step is to identify the order of the GSTAR model. Table 5 show the AICC values for each lag period. Based on Table 5, it is evident that the smallest AICC value is found at lag 1, which is -7.1942. Therefore, the selected time order is 1 $(p = 1)$. Additionally, the spatial order applied to the GSTAR model is 1 ($\lambda_s = 1$), because spatial orders of more than 1 are difficult to represent [22]. Consequently, the formed GSTAR model is GSTAR (1,1).

Table 4. Results of the Spatial Heterogeneity test

Location	Gini Indeks	Decision about H_0
South Sumatera	1.00076313	Reject H_0
Riau	1. 00076313	Reject H_0
Jambi	1. 00076313	Reject H_0

Table 5. AICC values for each lag period

Lag	1	2	3	4
AICC	-7.1942	-7.1727	-7.1057	-6.9929

3.3. Location Weight Matrix

Geographic coordinates in degrees of latitude and longitude of the research locations are presented in Table 6.

Table 6. Geographic coordinates for the three provinces

Location	Latitude	Longitude
South Sumatera	-2.92732	104.7720
Riau	0.45924	101.4474
Jambi	-1.60190	103.4844

Next, the inverse distance weighting matrix among three locations using equation (7) is

$$W = \begin{bmatrix} 0 & 0.280 & 0.720 \\ 0.379 & 0 & 0.621 \\ 0.611 & 0.389 & 0 \end{bmatrix}. \tag{14}$$

Based on W, if the distance is closer, the inverse distance weighting will be more significant, and vice versa.

The results of normalized cross-correlation weighting among three provinces using equation (10) at corresponding time lags is

$$W = \begin{bmatrix} 0 & 0.541 & 0.459 \\ 0.397 & 0 & 0.603 \\ 0.370 & 0.630 & 0 \end{bmatrix}. \tag{15}$$

3.4. Parameters Estimation of GSTAR (1,1) Model

3.4.1. GSTAR (1,1) Model using Inverse Distance Weighting

At a significance level of $\alpha = 5\%$, the estimation results of the GSTAR(1,1) model utilizing inverse distance weighting are presented in Table 7. Based on this table, the final form of the GSTAR(1,1) model using inverse distance weighting on equation (14) for each location is as follows:

a. GSTAR(1,1) model for South Sumatera:

$$\boldsymbol{Z_1(t)} = 0.43462\, Z_1(t-1) + 0.161101\, Z_2(t-1) + \ 0.414259\, Z_3(t-1) + e_1(t). \quad (16)$$

b. GSTAR(1,1) model for Riau:

$$\boldsymbol{Z_2(t)} = 0.790134\, Z_2(t-1) + e_2(t). \quad (17)$$

c. GSTAR(1,1) model for Jambi:

$$\boldsymbol{Z_3(t)} = 0.53526\, Z_3(t-1) + 0.279880\, Z_1(t-1) + \ 0.178189\, Z_2(t-1)\ \ e_3(t). \quad (18)$$

Table 7. Parameter estimation of GSTAR(1,1) model using Inverse Distance Weighting

Parameter	Parameter Estimation	t - Value	p - Value
$\Phi_{10^{(1)}}$	0.434620	3.26	0.0016
$\Phi_{11^{(1)}}$	0.575360	4.25	< 0.0001
$\Phi_{10^{(2)}}$	0.790134	5.78	< 0.0001
$\Phi_{11^{(2)}}$	0.210028	1.53	0.1284
$\Phi_{10^{(3)}}$	0.535260	2.87	0.0052
$\Phi_{11^{(3)}}$	0.458069	2.49	0.0146

3.4.2. GSTAR (1,1) Model using Normalized Cross-Correlation Weight

Table 8 presents the estimated parameters of the GSTAR(1,1) modelusing the normalized cross-correlation weights with a significance level of $\alpha = 5\%$.

Table 8. Parameter estimation of GSTAR(1,1) model using Normalized Cross-Correlation Weighting

Parameter	Parameter Estimation	t - value	p - value
$\Phi_{10^{(1)}}$	0.484676	4.12	< 0.0001
$\Phi_{11^{(1)}}$	0.523375	4.38	< 0.0001
$\Phi_{10^{(2)}}$	0.790444	5.82	< 0.0001
$\Phi_{11^{(2)}}$	0.209645	1.54	0.1264
$\Phi_{10^{(3)}}$	0.420535	2.41	0.0181
$\Phi_{11^{(3)}}$	0.572691	3.23	0.0013

Based on the significant parameter estimates in Table 8, the final form of the GSTAR(1,1) model obtained by applying normalized weights to the normalized cross-correlations on equation (15) of is as follows.

a. GSTAR(1,1) model for South Sumatera:

$$\boldsymbol{Z_1(t)} = 0.484676\, Z_1(t-1) + 0.283010\, Z_2(t-1) + 0.240365\, Z_3(t-1) + e_1(t). \tag{19}$$

b. GSTAR(1, 1) model for Riau:

$$\boldsymbol{Z_2(t)} = 0.790444\, Z_2(t-1) + e_2(t). \tag{20}$$

c. GSTAR(1, 1) model for Jambi:

$$\boldsymbol{Z_3(t)} = 0.420535\, Z_3(t-1) + 0.211970\, Z_1(t-1) + 0.240365\, Z_2(t-1) + e_3(t). \tag{21}$$

3.5. Model Diagnostic

Model diagnostic aims to determine whether the assumptions of white noise and multivariate normality of testing data residual data have been met. If the GSTAR(1,1) forecasting model, employing both location weights, satisfies the diagnostic assumptions, then the forecasting model is deemed suitable for air temperature forecasting.

The residual white noise test aims to evaluate the possibility of correlation among residuals, as the expected nature of a GSTAR model involves independent residuals. With a significance level of $\alpha = 5\%$, the residual white noise test results conducted using the Ljung-Box-Pierce test are presented in Table 9. Based on this table, the p-value of both methods is greater than the α. Therefore, we concluded that the residual of the GSTAR(1,1) model with these two weights has satisfied the assumption of white noise.

Table 9. Result of Ljung-Box-Pierce test

Weighting Method	P-Value
Inverse Distance	0.0681
Normalized Cross-Correlation	0.0677

Based on the hypothesis testing conducted using the Kolmogorov-Smirnov test with a significance level of $\alpha = 5\%$. The result as in Table 10. Based on this table, we concluded that the GSTAR(1,1) model employing both location weights has satisfied the assumption of multivariate normal distribution for residuals. Therefore, the GSTAR(1,1) model utilizing inverse distance weighting and the GSTAR(1,1) model employing normalized correlation weights have fulfilled the diagnostic assumptions of the model. Consequently, both models are deemed appropriate and ready to forecast the three research locations.

Table 10. Result of Kolmogorov-Smirnov test

Weighting Method	P-Value
Inverse Distance	0.1709
Normalized Cross-Correlation	0.1652

3.6. Model Validation

After the model's diagnostic assumptions are met, the next step is to obtain the best GSTAR(1,1) model based on the smallest value of the model's RMSE (Root Mean Square Error) calculated using equation (13). The best model will be used to forecast these three locations. The RMSE values calculated for the data are presented in Table 11. Based on ths table, it is evident that the GSTAR(1,1) model utilizing normalized cross-correlation weights yields an RMSE of 3.135 for forecasting air temperature data in the provinces of South Sumatera, Riau, and Jambi. Meanwhile, the GSTAR(1,1) model employing inverse distance weighting shows a higher RMSE value, precisely 3.141. Considering the comparison of these RMSE values, it can be concluded that the GSTAR(1,1) model using normalized cross-correlation weights is the most appropriate for forecasting air temperature data in the provinces of South Sumatera, Riau, and Jambi.

Table 11. RMSE for GSTAR(1,1) model

Weighting Method	RMSE
Inverse Distance	3.141
Normalized Cross-Correlation	3.135

3.7. Result Analysis

In previous research, [12] predicted air temperature using ARIMA and multiple regression in East Kalimantan. Then, [13] used Extreme Value Theory to predict air temperature in Central Java. However, no research explicitly focuses on forecasting air temperature for South Sumatera, Riau, and Jambi provinces using the GSTAR model.

This research showed that using the GSTAR(1,1) model effectively forecasted air temperature patterns in the provinces of South Sumatera, Riau, and Jambi. Using secondary data from BMKG, this research illustrates a robust approach to weather data analysis and modeling. Careful diagnostic steps, such as tests for stationarity, correlation, and spatial heterogeneity, ensure the suitability of the GSTAR model for the dataset used. These results are essential to understanding the region's air temperature data.

This study emphasizes the importance of considering spatial proximity and temporal relationships in modeling. Using inverse distance and normalized cross-correlation weighting show that taking these two aspects into account can improve forecasting accuracy.

Only significant parameters are used to form the GSTAR(1,1) model. Parameter estimates from the GSTAR(1,1) model indicate the existence of significant spatial and temporal dependencies in air temperature at the study location. The significant coefficients for the lagged variables in this model illustrate the importance of including these two dimensions in the analysis. These findings strengthen the model's reliability in forecasting future air temperature patterns.

Model diagnostic, including tests for residual white noise and multivariate normality, show that the GSTAR(1,1) model proposed in this study meets the assumptions necessary for accurate forecasting. Thus, the research results provide a deeper understanding of air temperature variability at this location, make a valuable contribution to the field of climatology, and can be used to support data-based decision-making in the face of future climate change.

4. CONCLUSIONS

This study proposes implementing the GSTAR model as a practical approach for modeling air temperature in South Sumatra, Riau, and Jambi Provinces, and the appropriate model is GSTAR(1,1). We use two weighted matrices, i.e., inverse distance and normalized cross-correlation weighting. Based on the smallest RMSE, the normalized cross-correlation weighting is more effective in forecasting the air temperature in these three provinces. The choice of model and location weighting matrix can significantly impact the accuracy of air temperature forecasting. This forecasting model can serve as a solid foundation for government, research institutions, and other stakeholders to make informed decisions regarding management and strategic planning for mitigating and adapting to temperature changes in the region.

REFERENCES

[1] G. I. Lippsmeier, *Tropenbau Building in the Tropics*. German: Callway, 1980.

[2] C. Chou, J. D. Neelin, C. A. Chen, and J. Y. Tu, "Evaluating the 'rich-get-richer' mechanism in tropical precipitation change under global warming," *J. Clim.*, vol. 22, no. 8, pp. 1982–2005, 2009, doi: 10.1175/2008JCLI2471.1.

[3] A. Machmudin and B. S. S. Ulama, "Peramalan Temperatur Udara di Kota Surabaya dengan Menggunakan ARIMA dan Artificial Neural Network," *J. Sains Dan Seni ITS*, vol. 1, no. 1, pp. 118–123, 2012.

[4] A. F. Hadi, I. Yudistira, D. Anggraeni, and M. Hasan, "The Geographical Clustering of the Rainfall Stations on Seasonal GSTAR Modeling for Rainfall Forecasting," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012238.

[5] G. Kirchgassner and J. Wolters, *Introduction to Modern Time Series Analysis*. New York: Springer, 2007.

[6] M. Usman *et al.*, "Analysis of Some Energy and Economics Variables by Using VECMX Model in Indonesia," *Int. J. Energy Econ. Policy*, vol. 12, no. 2, pp. 91–102, 2022, doi: 10.32479/ijeep.11897.

[7] P. E. Pfeifer and S. J. Deutsch, "A Three-Stage Iterative Procedure for Space-Time Modeling," *Technometrics*, vol. 22, no. 1, p. 35, 1980, doi: 10.2307/1268381.

[8] B. N. Ruchjana, "PEMODELAN KURVA PRODUKSI MINYAK BUM1 MENGGUNAKAN MODEL GENERALISASI S-TAR," *Forum Stat. dan komputasi*, pp. 1–6, 2002.

[9] D. R. S. Saputro, I. M. Putri, Sutanto, N. H. Noor, and P. Widyaningsih, "Generalized space time autoregressive (gstar)-artificial neural network (ann) model with multilayer feedforward networks architecture," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 243, no. 1, 2019, doi: 10.1088/1755-1315/243/1/012039.

[10] A. P. Muzdhalifah, T. Tarno, and P. Kartikasari, "Penerapan Model Generalized Space Time Autoregressive (Gstar) Untuk Meramalkan Penerbangan Domestik Pada Tiga Bandar Udara Di Pulau Jawa," *J. Gaussian*, vol. 11, no. 3, pp. 332–343, 2023, doi: 10.14710/j.gauss.11.3.332-343.

[11] N. Ilmi, A. Aswi, and M. K. Aidid, "Generalized Space Time Autoregressive Integrated Moving Average (GSTARIMA) dalam Peramalan Data Curah Hujan di Kota Makassar," *Inferensi*, vol. 6, no. 1, p. 25, 2023, doi: 10.12962/j27213862.v6i1.14347.

[12] S. Borovkova, H. P. Lopuhaä, and B. N. Ruchjana, "Consistency and asymptotic normality of least squares estimators in generalized STAR models," *Stat. Neerl.*, vol. 62, no. 4, pp. 482–508, 2008, doi: 10.1111/j.1467-9574.2008.00391.x.

[13] B. N. Ruchjana, S. A. Borovkova, and H. P. Lopuhaa, "Least squares estimation of Generalized Space Time AutoRegressive (GSTAR) model and its properties," *AIP Conf. Proc.*, vol. 1450, no. May, pp. 61–64, 2012, doi: 10.1063/1.4724118.
[14] D. N. Gujarati and D. C. Porter, *The McGraw-Hill Series Economics*. 2009.
[15] N. M. Huda and N. Imro'ah, "Determination of the best weight matrix for the Generalized Space Time Autoregressive (GSTAR) model in the Covid-19 case on Java Island, Indonesia," *Spat. Stat.*, vol. 54, p. 100734, 2023, doi: 10.1016/j.spasta.2023.100734.
[16] Suhartono and Subanar, "The Optimal Determination of Space Weight in GSTAR Model by Using Cross-correlation Inference," *J. Quant. Method, J. Devoted to Math. Stat. Apl. Var. F.*, vol. 2, no. 2, pp. 45–53, 2006.
[17] I. Adam, D. Kusnandar, and H. Perdana, "Penerapan Model GSTAR(1,1) Untuk Data Curah Hujan," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 6, no. 3, pp. 159–166, 2017.
[18] S. S. Handajani, H. Pratiwi, Y. Susanti, S. Subanti, Respatiwulan, and Hartatik, "Rainfall model on area of rice production in Sragen, Karanganyar and Klaten by using Generalized Space Time Autoregressive (GSTAR)," *J. Phys. Conf. Ser.*, vol. 855, no. 1, 2017, doi: 10.1088/1742-6596/855/1/012015.
[19] U. S. Pasaribu, U. Mukhaiyar, N. M. Huda, K. N. Sari, and S. W. Indratno, "Modelling COVID-19 growth cases of provinces in java Island by modified spatial weight matrix GSTAR through railroad passenger's mobility," *Heliyon*, vol. 7, no. 2, p. e06025, 2021, doi: 10.1016/j.heliyon.2021.e06025.
[20] F. N. Aryani, S. S. Handajani, and E. Zukhronah, "Penerapan Model Generalized Space Time Autoregressive Pada Data Nilai Tukar Petani Di Tiga Provinsi Pulau Sumatera," *J. Ilmu Dasar*, pp. 97–104, 2020.
[21] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989, doi: 10.1093/biomet/76.2.297.
[22] A. Fadlurohman, T. Wahyu Utami, R. Wasono, K. Kunci, J. Tengah, and S. Biaya Hidup, "Generalized Space Time Autoregressive Modeling With Variable Exogenous (Gstar-X) (Case Study: Inflation In Six Cities Of Central Java)," *Pros. Semin. Nas. Unimus*, vol. 3, pp. 26–36, 2020.
[23] W. W. S. Wei, *Multivariate Time Series Analysis and Application*. New York: John Wiley and Sons, Inc, 2019.