frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 29 August 2024
DOI 10.3389/fams.2024.1417037

OPEN ACCESS

EDITED BY
Ronald Wesonga,
Sultan Qaboos University, Oman

REVIEWED BY
Zakariya Yahya Algamal,
University of Mosul, Iraq
Mohamed R. Abonazel,
Cairo University, Egypt

*CORRESPONDENCE
Utriweni Mukhaiyar
✉ utriweni.mukhaiyar@itb.ac.id

RECEIVED 13 April 2024
ACCEPTED 05 August 2024
PUBLISHED 29 August 2024

CITATION
Mukhaiyar U, Mahdiyasa AW, Sari KN and Noviana NT (2024) The generalized STAR modeling with minimum spanning tree approach of spatial weight matrix. *Front. Appl. Math. Stat.* 10:1417037. doi: 10.3389/fams.2024.1417037

The generalized STAR modeling with minimum spanning tree approach of spatial weight matrix

Utriweni Mukhaiyar[1]*, Adilan Widyawan Mahdiyasa[1], Kurnia Novita Sari[1] and Nur Tashya Noviana[2]

[1]Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia, [2]Undergraduate Program in Mathematics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia

The weight matrix is one of the most important things in Generalized Space–Time Autoregressive (GSTAR) modeling. Commonly, the weight matrix is built based on the assumption or subjectivity of the researchers. This study proposes a new approach to composing the weight matrix using the minimum spanning tree (MST) approach. This approach reduces the level of subjectivity in constructing the weight matrix since it is based on the observations. The spatial dependency among locations is evaluated through the centrality measures of MST. It is obtained that this approach could give a similar weight matrix to the commonly used, even better in some ways, especially in modeling the data with higher variability. For the study case in traffic problems, the number of vehicles entering the Purbaleunyi toll was modeled by GSTAR with several weight matrix perspectives. According to Space–Time ACF-PACF plots, GSTAR(1;1), GSTAR(1,2), and GSTAR(2;1,1) models are the candidates for appropriate models. Based on the root mean square errors and mean absolute percentage errors, it is concluded that the GSTAR(2,1,1) with MST approach is the best model to forecast the number of vehicles entering the Purbaleunyi toll. This best model is followed by GSTAR(1,1) with an MST approach of spatial weight matrix.

KEYWORDS

autoregressive, correlation, distance, minimum spanning tree, weight matrix

1 Introduction

The Generalized Space–Time Autoregressive (GSTAR) is one of the methods to analyze the space–time series. This model adapted the vector autoregressive model, which, instead of involving many variables, considered time series in many locations simultaneously. The development of the GSTAR model in Indonesia is very fast, both in theory and application. The theory includes process stationarity properties using inverse autocovariance matrix in (1) and the kernel approach in (2); the GSTAR with heteroscedastic effect in (3, 4); and the weight matrix construction of the GSTAR model using the kernel approach in (5) and graph in (6). The development of the GSTAR model with correlated errors is in (7, 8). The GSTAR model with exogenous variables and outliers is in (9), the Poisson GSTAR in (10), the optimal spatial aggregation of GSTARMA model in (11), the higher order model in (12), and the GSTAR for discrete data in (13). The application of the GSTAR model has been carried out on economic data in (14), tea plantation production in (15), oil palm production in (16), commodity prices of red chilies in (17), rainfall data of West Java in (18), number of dengue fever cases in (19), "Begal" criminal cases in Medan, North Sumatera, in (20), variation of Northern Ethiopia's

temperature in (21), and the number of COVID-19 cases in Java island in (6, 22).

This model's basic assumption is the existence of spatial dependence among observed locations. In addition to the spatial autoregressive parameters, the dependence among locations is described by the spatial weight matrix. Most of the time, this matrix is simply composed based on the researcher's subjectivity. This matrix can be classified as uniform, binary, and non-uniform weight matrix (23). The development of a weight matrix using the kernel approach is to less subjectivity (5). The weight matrix is composed based on the observations. Thus, the obtained weight matrix can be considered more objective than the conventional weight matrix. This study proposes another way to construct the weight matrix using the observations. This method constructs the minimum spanning tree (MST) graph to obtain the spatial dependence among locations through its adjacency matrix.

To obtain the description of this new approach application, the number of vehicles entering the Purbaleunyi toll gates of Bandung—West Java data is used. As the capital city of West Java, Indonesia, Bandung is one of the busiest cities on Java Island. Bandung is well-known as the center of art and music, education, business, culinary, and tourism. However, its citizens experience traffic jams every day, and the peak congestion occurs on the weekends because many people come from outside Bandung. Due to its geographic location, land transportation is more favorable to access this city. Thus, the Purbaleunyi toll gates are crowded by numerous vehicles passing every day. The space–time analysis is applied to capture this phenomenon since the occurrences are related to time and locations. There are eight toll gates involved, and the locations of those eight gates can be seen in Figure 1. It can be seen that all toll gates are connected to each other with various distances and eccentricities. Thus, the traffic route modeling of those toll gates satisfies the basic graph properties necessary to make use of MST. Some research studies regarding traffic flow forecasting with different methods are the neural network bagging ensemble hybrid modeling (24), threshold autoregression (TAR) (25), stacked autoencoder (SAE) of deep learning model (26), GSTAR model with time-correlated errors (27), segment-based data imputation (28), queueing networks (29), long short-term memory network (LSTM) (30), and multi-view travel time prediction (MVPPT) (31).

This study aims to determine the best space–time model that can be used to predict the number of vehicles entering the Purbaleunyi toll. There are eight toll gates involved: Padalarang (PDL), Pasteur (PST), Baros (BAR), Pasir Koja (PKJ), Kopo (KPO), M. Toha (MTH), Buah Batu (BBT), and Cileunyi (CLY). First, the new procedure is applied to obtain the spatial weight matrix using the adjacency matrix and MST graph approach. The novelty of this research is that both theory and application are new in statistical space–time modeling. This procedure is explained in Section 2. Then, in Section 3, the model selection improvement is investigated by checking the stationarity of all possible models using the inverse of autocovariance matrix (IAcM). Finally, the case study is discussed in Section 4.

2 Materials and methods

2.1 The spatial weight matrix

2.1.1 The spatial lag and conventional weight matrix

The spatial lag is constructed for all observation locations and can be obtained in many ways. A well-known method is applying the radius system, as illustrated in Figure 2. The locations closer to the reference location (s_0s$_0$) will have a smaller spatial lag, distinguished by a fixed distance, d_0d$_0$. The distinct configuration of each spatial lag order may give different weight matrices.

The weight matrix is a square matrix of size N, with the following properties: (i) the diagonal entries of weight matrix W are zeros, (ii) the sum of weight values in one row must equal to zero or $\sum_{j=1}^{N} w_{ij} = 1$. For all locations S_i, $i = 1, 2, \ldots N$, and (iii) every weight value is non-negative, or $w_{ij} \geq 0$.

In general, there are three weight matrix types: binary, uniform, and non-uniform. The binary weight matrix consists of 0 and 1 values in off-diagonal entries, of which 1 represents the most influential

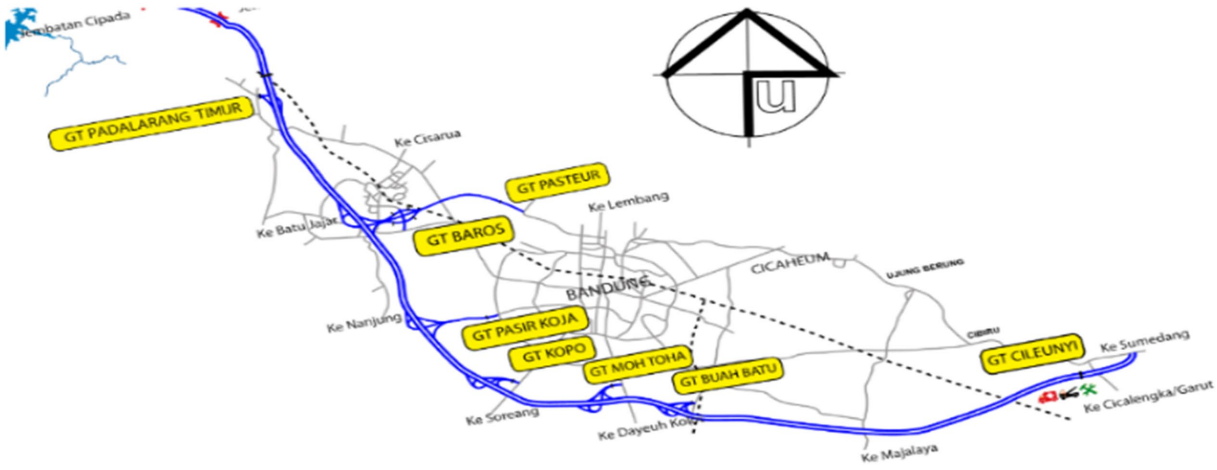


FIGURE 1
Map of eight gates in Purbaleunyi Toll (source: www.jasamarga.com/public/id/infolayanan/TOLl/ruas.aspx, Accessed 29 April 2018).

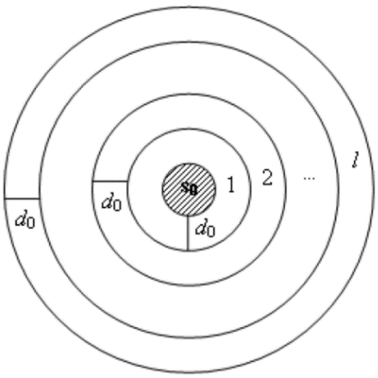


FIGURE 2
The radius system in determining the spatial lag. The neighbors of a reference location are classified in the same spatial lag if they lie in the same radius (23).

neighbor. The element in the uniform weight matrix is defined as follows:

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

where $n_i^{(\ell)}$ represents the number of neighbors for location i in ℓ th spatial lag.

There are some ways to determine the non-uniform weight matrix. The simplest way is by involving the distance between locations, such as inverse distance weight, that is

$$w_{ij}^{(\ell)} = \begin{cases} \dfrac{\dfrac{1}{1+d_{ij}^{(\ell)}}}{\sum_{j=1}^{N} \dfrac{1}{1+d_{ij}^{(\ell)}}}, & j \text{ is the neighbor of } i \text{ in } \ell\text{th order} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

for location i and j in ℓth spatial lag which Euclidean distance is $d_{ij}^{(\ell)}$. Since the distance is fixed, then the obtained weight matrix is also fixed. This construction could not accommodate some changes that happened in the observations. The new approach is constructed to accommodate the pattern of the observations by using an MST graph. Some previous research studies regarding the MST approach for spatial weight matrix have been investigated by (6, 32).

2.1.2 Minimum spanning tree

A graph (G) is not an empty set that makes a node at the endpoint vertices. The G is a simple graph if there are no circles and two lines that merge into a pair of vertices. If an edge can combine every vertice in a simple graph, it is called a completed graph. A tree (T) is a connecting graph without any repetition/cycle. Thus, every two vertices will be connected by one unique path. The minimum spanning tree (MST) is a connected and undirected graph. MST is a collection of lines and dots with a line consisting of an undirected weight that connects all vertices and does not contain a cycle that produces the smallest value (33). Some methods to construct the MST include the Prim and Kruskal algorithms. The common idea of both algorithms is selecting the graph with the smallest weight and connected vertices that do not form a circle. Kruskal's algorithm is formed by adding the smallest weight of lines into the tree one by one, while the Prim algorithm is built by minimizing the weight of the connected lines. Both algorithms are simple and popular methods for constructing MST. However, they produce only an MST from all possible MSTs; thus, the uniqueness of MST could not be obtained. Therefore, an algorithm was proposed to construct a forest graph consisting of all possible MSTs using fuzzy relations (34). This algorithm is called sub-dominant ultrametric (SDU).

One of the important steps in MST construction is defining the weight of lines. Here, the weight is determined by using a distance matrix. This distance matrix can be extracted from the correlation matrix. For example, supposedly two random variables are X and Y, then, the correlation between both variables is defined as follows:

$$\rho_{X,Y} = corr(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - (E[X])^2}\sqrt{E[Y^2] - (E[Y])^2}}$$

The correlation represents the linear dependence between variables, with values in $0 \leq |\rho_{X,Y}| \leq 1$. If both variables have strong linear dependence, the correlation values tend to be one and otherwise zero. In space–time data, the random variables are represented by the time series of each location. Consider N locations with time series realizations $\{z_i(t)\}, i = 1,2,\ldots,N\}$, then the correlation between series in locations i and j is defined as follows:

$$c(i,j) = \frac{z_i z_j - z_i z_j}{\sqrt{z_i^2 - z_i^2}\sqrt{z_j^2 - z_j^2}}$$

with z_i is the average of $z_i(t)$. The $c(i,j)$ can be a sample correlation between series in locations i and j. The distance matrix $D = \{d(i,j)\}$ is symmetric whose entries are defined as follows:

$$d(i,j) = \sqrt{2(1 - c(i,j))}$$

for all $i, j = 1,2,\ldots,N$. This distance matrix is symmetric and anti-reflexive fuzzy.

2.1.3 Spatial lag classification using MST

The spatial lag membership for a reference location using MST is based on the number of lines connected from the reference location to its neighbors. If a location is directly connected to a reference location, that location is in the first spatial lag. Thus, if ℓ lines connect the neighbor to the reference location, then this neighbor is classified as the member of spatial lag ℓth of the reference location.

Figure 3 shows the procedure of this spatial lag classification using MST. If the Pasir Koja toll gate is considered as the reference location, then the Pasteur toll gate is a member of its first spatial lag, and Padalarang, Kopo, M. Toha, and Baros are the members of the second spatial lag. Finally, the Cileunyi toll gate will be in the third spatial lag, and Buah Batu is the only member of the fourth spatial lag. After determining the number of memberships in the spatial lag ℓth for

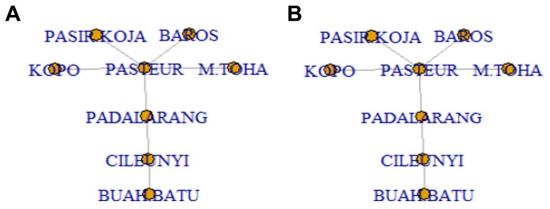


FIGURE 3
MST in determining: **(A)** First and **(B)** Second spatial lag for Pasir Koja. Since Pasteur has a direct line connected to Pasir Koja, it becomes a neighbor in the first lag spatial of Pasir Koja.

TABLE 1 STACF and STPACF theoretical patterns in identifying space–time model.

Model	STACF	STPACF
G-STAR(p; λ_1, λ_2,…,λ_p)	*(tail-off) exponentially decreases or* sinus wave.	(*cutoff*) cut after lag time pth, and lag spatial λ_pth
GSTMA(q; λ_1, λ_2,…,λ_q)	(*cutoff*) cut after lag time qth, and lag spatial λ_qth	*(tail-off) exponentially decreases or* sinus wave.
GSTARMA(p; λ_1, λ_2,…,λ_p; q; λ_1, λ_2, …,λ_q)	*(tail-off) exponentially decreases or* sinus wave.	*(tail-off) exponentially decreases or* sinus wave

location i ($n_i^{(\ell)}$), the spatial weight can be composed using Equations 1, 2, consecutively to obtain MST uniform and MST inverse distance weight matrix.

2.2 The generalized STAR modeling

The GSTAR model is obtained through Box-Jenkins' three-stage-iterative model identification, parameter estimation, and diagnostic checking (23). First, all appropriate models can be identified through Space–Time ACF and PACF or named STACF and STPACF. This identification stage has been derived briefly in Pfeifer and Deutsch (35). Then, all possible models may be identified by observing the STACF and STPACF plot pattern, as stated in Table 1.

The observation at location i and time t, $Z_i(t)$ follows the GSTAR model if there is a linear combination of past observations for both time and spatial indices. Suppose, a random vector process $\{\boldsymbol{Z}(t)\}$, $\boldsymbol{Z}(t)=\left(Z_1(t),Z_2(t),\ldots,Z_N(t)\right)'$ with time $t=\{1,2,\ldots,T\}$ follows GSTAR$\left(p;\lambda_1,\lambda_2,\ldots,\lambda_p\right)$ model, then $\boldsymbol{Z}(t)$ can be expressed as follows:

$$\boldsymbol{Z}(t)=\sum_{k=1}^{p}\left[\boldsymbol{\Phi}_{k0}\boldsymbol{Z}(t-k)+\sum_{\ell=1}^{\lambda_k}\boldsymbol{\Phi}_{k\ell}\mathbf{W}^{(\ell)}\boldsymbol{Z}(t-k)\right]+\boldsymbol{\varepsilon}(t) \quad (3)$$

The $\mathbf{W}^{(\ell)}$ is ℓth order weight matrix whose main diagonal is zero and the sum of each row is one, $\boldsymbol{\Phi}_{k\ell}=diag\left(\phi_{k\ell}^{(1)},\phi_{k\ell}^{(2)},\ldots,\phi_{k\ell}^{(N)}\right)$ is (N×N) diagonal matrix which presents autoregressive parameter of kth time order and ℓth spatial order for each location $i=1,2,\ldots,N$, and $\boldsymbol{\varepsilon}(t)=\left(\varepsilon_1(t),\varepsilon_2(t),\ldots,\varepsilon_N(t)\right)$ is $(N\times1)$-dimensional vector of errors that is assumed which has i.i.d. normal distribution with null mean and constant variance. Thus, for each location $i=1,2,\ldots,N$, the model can be written as follows:

$$Z_i(t)=\sum_{k=1}^{p}\left[\phi_{k0}^{(i)}Z_i(t-k)+\sum_{\ell=1}^{\lambda_p}\phi_{k\ell}^{(i)}\left(\sum_{j=1}^{N}w_{ij}^{(\ell)}Z_j(t-k)\right)\right]+\varepsilon_i(t) \quad (4)$$

Another important representation of the GSTAR model is the linear model. This form is indispensable when using the least-square (LS) method for parameter estimation. For each location i Equation 4 can be represented by a linear model $\boldsymbol{Y}_i=\mathbf{X}_i\phi_i+\boldsymbol{\varepsilon}_i$, with $\boldsymbol{Y}_i=\left(Z_i(p),Z_i(p+1),\ldots,Z_i(T)\right)'$, $\mathbf{X}_i=\left(\mathbf{U}_{i1}\ \mathbf{U}_{i2}\ \cdots\ \mathbf{U}_{im}\right)$, $\phi_i=\left(\phi_{10}^{(i)},\ldots,\phi_{1\lambda_1}^{(i)},\phi_{20}^{(i)},\ldots,\phi_{2\lambda_2}^{(i)},\ldots,\phi_{p0}^{(i)},\ldots,\phi_{p\lambda_p}^{(i)}\right)'$, and $\boldsymbol{\varepsilon}_i=\left(\varepsilon_i(p),\varepsilon_i(p+1),\ldots,\varepsilon_i(T)\right)$. The matrix $\mathbf{X}_i$ has size $\left(T-(p+1)\right)\times\left(p+\sum_{m=1}^{p}\lambda_m\right)$, and for $m=1,2,\ldots,p$, defines

$$\mathbf{U}_{im}=\begin{pmatrix} Z_i(p-m) & \sum_{j=1}^{N}w_{ij}^{(1)}Z_j(p-m) & \cdots & \sum_{j=1}^{N}w_{ij}^{(\lambda_m)}Z_j(p-m) \\ Z_i(p-m+1) & \sum_{j=1}^{N}w_{ij}^{(1)}Z_j(p-m+1) & \cdots & \sum_{j=1}^{N}w_{ij}^{(\lambda_m)}Z_j(p-m+1) \\ \vdots & \vdots & \ddots & \vdots \\ Z_i(T-m) & \sum_{j=1}^{N}w_{ij}^{(1)}Z_j(T-m) & \cdots & \sum_{j=1}^{N}w_{ij}^{(\lambda_m)}Z_j(T-m) \end{pmatrix}$$

The linear model for all locations is defined as $\boldsymbol{Y}=\mathbf{X}\phi+\boldsymbol{\varepsilon}$, with $\boldsymbol{Y}=\left(\boldsymbol{Y}_1',\boldsymbol{Y}_2',\ldots,\boldsymbol{Y}_N'\right)$, $\mathbf{X}=diag\left(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_N\right)$, $\phi=\left(\phi_1',\phi_2',\ldots,\phi_N'\right)$, and $\boldsymbol{\varepsilon}=\left(\boldsymbol{\varepsilon}_1',\boldsymbol{\varepsilon}_2',\ldots,\boldsymbol{\varepsilon}_N'\right)$. Thus, least-square estimators are obtained by $\phi=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{Y}$.

In the diagnostic checking stage, the error assumption is examined through the normality and uncorrelated test of residuals. The best appropriate model is also obtained by evaluating the model whose residual is minimum. The minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) may be assessed to choose the best model. Here, the stationarity of the process is reevaluated using parameter estimations to obtain the most appropriate model. Mukhaiyar and Pasaribu (1) introduced the inverse of the autocovariance matrix (IAcM) for evaluating the GSTAR model stationarity through the following propositions.

Proposition 1: Suppose that $(N\times N)$-dimensional matrix $\mathbf{A}=\boldsymbol{\Phi}_{10}+\boldsymbol{\Phi}_{11}\mathbf{W}$ and IAcM, $\mathbf{M}_1=\mathbf{I}_N-\mathbf{A}'\mathbf{A}$, has elements which consist of the parameters of GSTAR(1;1) process, the GSTAR(1;1) model is stationary if the determinants of all the leading principal submatrices of IAcM are positive.

Proposition 2: For GSTAR(1;λ_1) models with $\lambda_1\geq0$, define $\mathbf{A}=\sum_{\ell=0}^{\lambda_1}\boldsymbol{\Phi}_{1\ell}\mathbf{W}^{(\ell)}$ and IAcM, $\mathbf{M}_1=\mathbf{I}_N-\mathbf{A}'\mathbf{A}$, the model is stationary if the determinants of all the leading principal submatrices of $\mathbf{M}_1$ are positive.

Proposition 3: For GSTAR(2; λ_1, λ_2) model, define $\mathbf{A}_k = \sum_{\ell=0}^{\lambda_k} \Phi_{k\ell} \mathbf{W}^{(\ell)}$, $k = 1, 2$ and IAcM, $\mathbf{M}_2 = \begin{pmatrix} \mathbf{I}_N - \mathbf{A}_2' \mathbf{A}_2 & -\mathbf{A}_1' - \mathbf{A}_2' \mathbf{A}_1 \\ -\mathbf{A}_1 - \mathbf{A}_1' \mathbf{A}_2 & \mathbf{I}_N - \mathbf{A}_2' \mathbf{A}_2 \end{pmatrix}$. The models are stationary if the determinants of all the leading principal submatrices of its IAcM are positive.

The proofs of those propositions have been evaluated by Mukhaiyar and Pasaribu (1).

3 Results and discussion

The GSTAR model is obtained through Box-Jenkins' three-stage-iterative model identification, parameter estimation, and diagnostic checking. First, a weight matrix should be developed; thus, all appropriate models can be investigated through Space–Time ACF and PACF or named STACF and STPACF. Then, since the error assumption is uncorrelated and has constant mean and variance, the ordinary least-square method is applied for parameter estimation.

To get an overview of the GSTAR modeling using the MST weight matrix and its comparison to other approaches, the number of vehicles entering Bandung city through the eight toll gates was investigated. The data retrieved are from January to December 2018, and the plot series and boxplot of each location can be seen in Figure 4. Based on Figure 4, the busiest toll gate is Cileunyi, followed by Pasteur and Padalarang gates. However, some trends are found in some periods (Figure 4A), which means the process is not stationary. Meanwhile, in Figure 4B, some outliers are found in every toll gate, but they will be ignored and still be involved in data analysis. Figure 5 shows the correlation matrix among toll gates observations and its visualization through scatter plots.

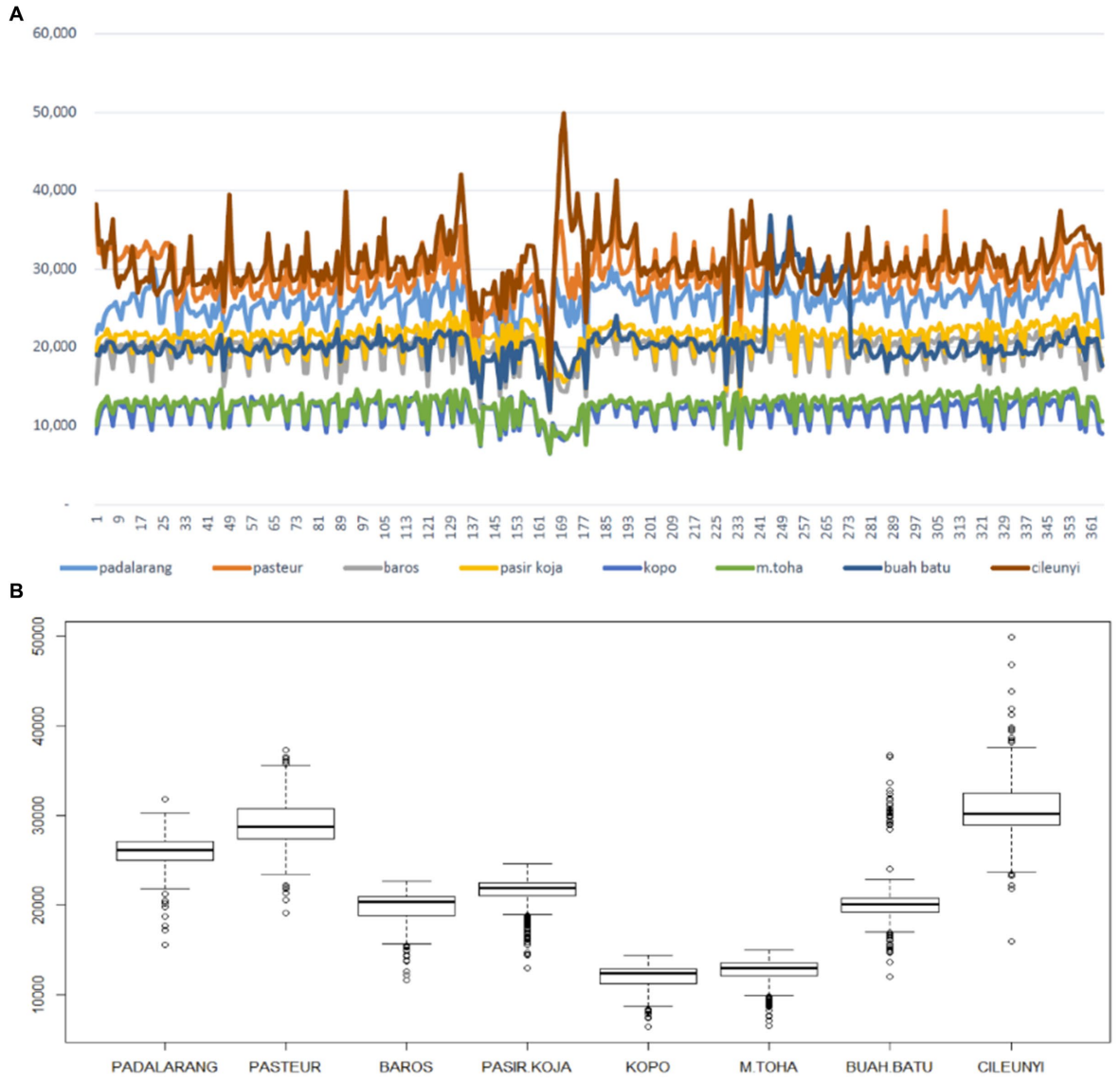


FIGURE 4
(A) The daily number of vehicles entering Bandung city through eight toll gates. Some trends found in all locations in some periods indicate a non-stationary process. **(B)** The boxplot for each gate shows the different mean and variability of observations. Some outliers are found in every gate.

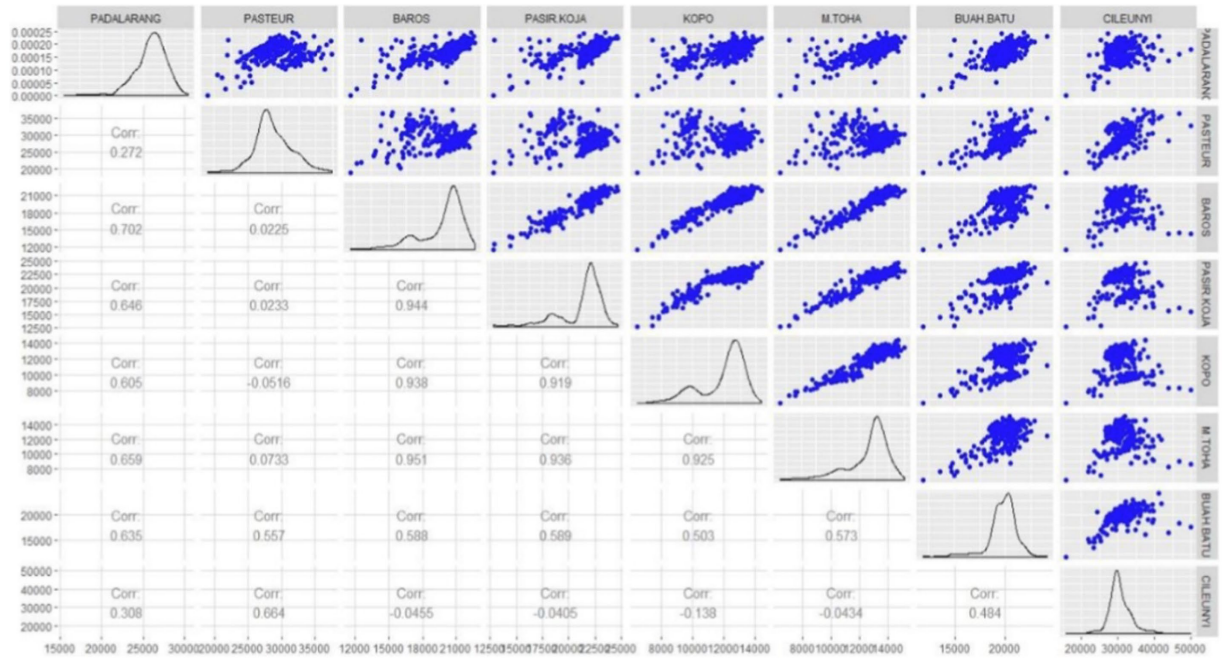


FIGURE 5
Scatter plot and correlation matrix among toll gate observations.

The highest correlations were obtained among the Baros, Pasir Koja, and Kopo toll gates. These three gates have closer distances in an exact line of the middle toll's route (see Figure 1). These highest correlations indicate the highest spatial dependencies among other toll gates. It can be seen in the weight matrix (see Table 2). Meanwhile, although Baros' and Pasteur's gates are closer, the correlation is the smallest. It is possible since both gates are not in the same line.

The first stage in modeling is identification, in which a plot of STACF and STPACF is performed. To construct both plots, data should be stationary. Then, to make data weak and stationary, differentiation is applied. However, it is not enough, so log transformation is used. Thus, the modeling is carried out to this log-differencing data. After that, a spatial weight matrix should be composed.

There are two types of weight matrices to be applied. Both are uniform and non-uniform (inverse distance), and each type has two approaches, namely, radius system and MST. To build the matrix, the number of neighbors for each location in every spatial lag (there are four spatial lags to be used) must be investigated for each approach as follows:

a The radius system way needs the distance (mileage) among toll gates and uses a radius (d_0) equal to 10 km (see Figure 2) to obtain the configuration of neighbors.
b MST way needs a correlation matrix to define the distance matrix based on Equation 3. The distance is the weight in the graph that connects all toll gates. Based on the graph, the MST is obtained every month, and the overall period of observations is consecutively illustrated in Figures 3, 6.

The majority of the months have a similar configuration to MST. In contrast, Pasteur toll gate is most frequently the center from February until April and July until December. In addition to the Pasteur toll gate, Padalarang and Cileunyi are also the centers, consecutively in January and June. As a result, those three toll gates are the top three busiest among the observed locations.

The most interesting configuration happened in May, since there are almost no connected graphs among locations, except for Cileunyi, Kopo, and Pasteur, which are connected with a line, and Kopo is the center. It is indicated that the observations' dependence on toll gates is not significant. There was a long holiday (Eid Al-Fitr celebration) this month, and people simultaneously went in and out of Bandung for homecoming. These observations can also be seen in Figure 4A, during 161–177 days. From both approaches, the configuration of neighbors for every location and spatial lag is obtained as shown in Table 3.

However, the radius system and the MST approach have different neighbor configurations. For example, as shown in Table 3, the Padalarang gate, using the radius system and MST approach consecutively, has one (3:Baros) and two (2:Pasteur and 8:Cileunyi) neighbors in the first spatial lag and four and five neighbors in the second spatial lag. Meanwhile, in the third and fourth spatial lags, for each, there is only one neighbor obtained using the radius system but no neighbor for the MST approach. Overall, there is a significant difference in the number of neighbors for every spatial lag between the radius system and the MST approach. Therefore, it will have an impact on weight matrix construction.

Based on neighbors' configurations, the uniform and inverse distance weights are composed using Equations 1, 2. The obtained weight matrices for each spatial lag are concluded in Table 2. It shows a significant difference between the weight matrix's radius system and the MST approach in line with the configuration.

From the weight matrix, the STACF and STPACF can be calculated. The significance of the autocorrelation is illustrated in Table 4 with symbols "X" for significance and "0" for no significance.

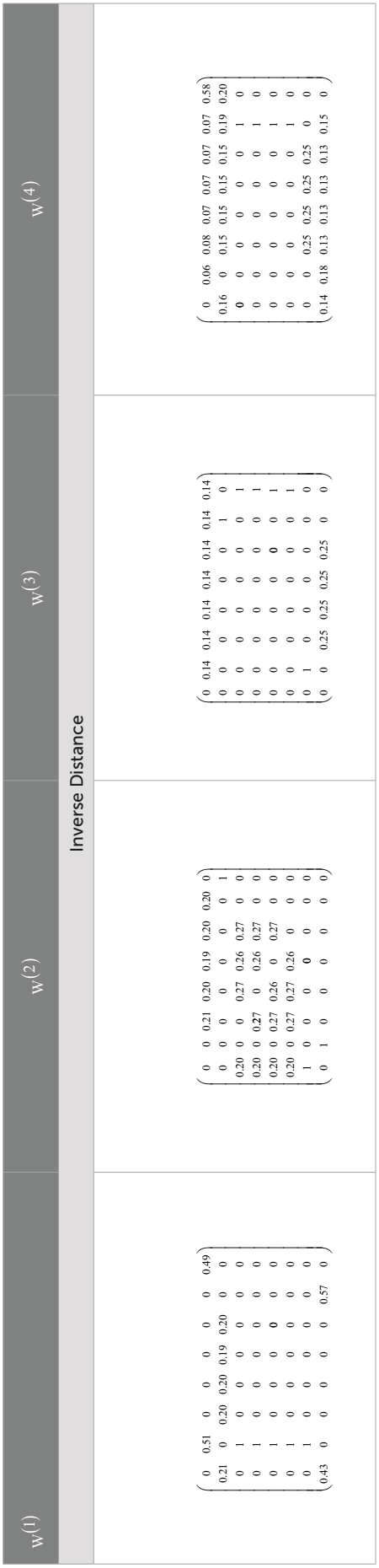
TABLE 2 Uniform and inverse distance weight matrix for each approach, radius system, and MST.

$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$
RADIUS SYSTEMS			
Uniform			
$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.33 & 0.33 & 0 & 0.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0 & 0.33 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0.33 & 0 & \mathbf{0}.33 & 0.33 & 0 \\ 0 & 0 & 0 & 0.33 & 0.33 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & 0.50 & 0 & 0 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.25 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0.25 & 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0.33 & 0.33 & 0 \\ 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0.33 & 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0.25 \\ 0 & 0 & \mathbf{0}33 & 0.33 & 0 & 0 & 0 & 0.33 \\ 0 & 0 & 0 & 0 & 0 & 0.50 & 0.50 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 & 0.14 \\ 0.46 & 0.54 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.28 & 0.33 & 0.39 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.14 & 0.14 & 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0 & \mathbf{0}.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & \mathbf{0}.14 & 0.14 & 0 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 \\ 0.47 & 0.53 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 \end{pmatrix}$
Inverse Distance			
$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.24 & 0.49 & 0 & 0.27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.29 & 0 & 0.44 & 0.27 & 0 & 0 \\ 0 & 0 & 0 & 0.37 & 0 & \mathbf{0}.38 & 0.25 & 0 \\ 0 & 0 & 0 & 0.23 & 0.38 & 0 & 0.39 & 0 \\ 0 & 0 & 0 & 0 & 0.40 & 0.60 & 0 & 0 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.25 & 0 & 0.27 & 0.27 & 0.22 & 0 & 0 \\ 0.21 & 0 & 0 & 0.31 & 0.26 & 0.22 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.42 & 0.32 & 0.26 & 0 \\ 0.27 & 0.38 & 0 & 0 & 0 & 0 & 0.35 & 0 \\ 0.26 & 0.32 & 0.42 & 0 & 0 & 0 & 0 & 0 \\ 0.20 & 0.26 & 0.29 & 0 & 0 & 0 & 0 & 0.24 \\ 0 & 0 & \mathbf{0}29 & 0.39 & 0 & 0 & 0 & 0.32 \\ 0 & 0 & 0 & 0 & 0 & 0.46 & 0.54 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 & 0.14 \\ 0.50 & 0.50 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0.33 & 0.33 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.14 & 0.14 & 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0 & \mathbf{0}.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & \mathbf{0}.14 & 0.14 & 0 & 0.14 & 0.14 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 & 0.14 \\ 0.50 & 0.50 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 \end{pmatrix}$
MST			
Uniform			
$\begin{pmatrix} 0 & 0.50 & 0 & 0 & 0 & 0 & 0 & 0.50 \\ 0.20 & 0 & 0.20 & 0.20 & 0.20 & 0.20 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.50 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0.20 & 0.20 & 0.20 & 0.20 & 0.20 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.25 & 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0.25 & 0 & \mathbf{0}.25 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0.25 & 0 & 0.25 & 0.25 & 0 & 0.25 & 0 & 0 \\ 0.25 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ 0.14 & 0 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 \\ \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0.14 & 0 \end{pmatrix}$

(Continued)

10.3389/fams.2024.1417037

TABLE 2 (Continued)



There is almost no difference between uniform and inverse distance for each radius system and the MST approach.

Here, the significance value used is 5%. It can be seen from Table 4 that the STACF gives a tail-off pattern because the values are significant for many first lags. Meanwhile, the STACF is cut off in some lags. Therefore, it is indicated that the autoregressive GSTAR model should be considered. Since the pattern of cutoff is quite random, then it is suggested to use the GSTAR model with various simple orders. It is recommended to consider GSTAR(1;1), GSTAR(1,2), and GSTAR(2;1,1) to be involved in the next steps of modeling. After identifying the model, the next step is estimating the model parameters. These parameters were estimated using the least-square method. The estimators were obtained for all parameters of the GSTAR (1; 1), GSTAR (1; 2), and GSTAR (2;1,1) models.

Table 5 shows the parameter estimators using the MST weight matrix. The uniform and inverse distance weight matrix gives similar estimators in some locations, and it depends on the order of the model. It is suspected that larger orders will show less dissimilarity among both types of matrices, especially in a location with low to intermediate variability.

After obtaining the estimated parameters, the stationarity of the process can be detected by the inverse of autocovariance matrix (IAcM). This step is part of diagnostic checking. Based on Propositions 1–3, the obtained GSTAR models are stationary since every leading principal submatrix of IAcM has positive determinants. The checklist of this property is summarized in Table 6. The next step is obtaining the least-square estimators.

Consider z_i is the ith real observation of several vehicle numbers and is estimated by z_i using the obtained models, for $i = 1, 2, \ldots n$ with n as the number of observations, regardless of the time and spatial. Then, the residual is the difference of z_i and z_i. The residuals are uncorrelated and follow a normal distribution with zero mean and constant variance for a significance value of less than 10%. The best prediction model is obtained by evaluating each model's root mean square errors (RMSEs) and mean absolute percentage errors (MAPE).

First, the RMSE is defined as $RMSE = \sqrt{\sum_{i=1}^{n} \left(x_i - \hat{x}_i\right)^2 / n}$. After that, the MAPE is defined as $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Z_i(T+1) - \hat{Z}_i(T+1)}{Z_i(T+1)}\right| \times 100\%$.

The values of RMSE are reported in Table 7.

It can be seen from Table 7 that the MST inverse distance weight matrix gives the smallest RMSE much more than other types of weight matrices. For example, in the type of model perspective, GSTAR(1;1) has the smallest RMSE values among the obtained models.

The comparison models are also executed for prediction. The observation of $Z_i(T+1)$ is saved by comparing the one-step prediction result of all models in each location and its MAPE, as shown in Table 8. A slightly different conclusion was obtained from this one-step ahead forecasting. The GSTAR(2;1,1) gives the closest prediction in four (Pasteur, Pasir Koja, Buah Batu, and Cileunyi) of eight toll gates. In Table 8, the MST approach gives the best predictions, especially with the inverse distance approach. Those four toll gates are included in the top five of the busiest toll gates. The GSTAR(1;1) model gives the best prediction in Kopo and M. Toha toll gates using an inverse distance weight matrix.

Overall, the least average of MAPE is presented by GSTAR(2;1,1) with MST, consecutively by uniform (MAPE is 2.39%) and inverse

TABLE 3 Neighbor configuration in every spatial lag using radius system and MST for every location with four different spatial lags.

i	Gate *i*th	Method		1st lag	2nd lag	3rd lag	4th lag
1	PDL	Rad	(a)	1	4	1	1
			(b)	{3}	{2,4, 5,6}	{7}	{8}
		MST	(a)	2	5	0	0
			(b)	{2,8}	{3,4,5,6,7}	0	0
2	PST	Rad	(a)	1	4	1	1
			(b)	{3}	{1,4,5,6}	{7}	{8}
		MST	(a)	5	1	1	0
			(b)	{1,3,4,5,6}	{8}	{7}	0
3	BAR	Rad	(a)	3	3	1	0
			(b)	{1,2,4}	{5,6,7}	{8}	0
		MST	(a)	1	4	1	1
			(b)	{2}	{1,4,5,6}	{8}	{7}
4	PKJ	Rad	(a)	3	3	1	0
			(b)	{3,5,6}	{1,2,7}	{8}	0
		MST	(a)	1	4	1	1
			(b)	{2}	{1,3,5,6}	{8}	{7}
5	KPO	Rad	(a)	3	3	1	0
			(b)	{4,6,7}	{1,2,3}	{8}	0
		MST	(a)	1	4	1	1
			(b)	{2}	{1,3,4,5}	{8}	{7}
6	MTH	Rad	(a)	3	4	0	0
			(b)	{4,5,7}	{1,2,3,8}	0	0
		MST	(a)	1	2	1	1
			(b)	{2}	{4,8}	{8}	{7}
7	BBT	Rad	(a)	2	3	2	0
			(b)	{5,6}	{3,4,8}	{1,2}	0
		MST	(a)	1	1	1	4
			(b)	{8}	{1}	{2}	{3,4,5,6}
8	CLY	Rad	(a)	0	2	3	2
			(b)	0	{6,7}	{3,4,5}	{1,2}
		MST	(a)	1	1	4	0
			(b)	{7}	{2}	{3,4,5,6}	0

It consists of (a) the number of neighbors and (b) the related neighbors (Gate *i*th).

distance approach (MAPE is 2.46%). The next best model is GSTAR(1;1) with MST spatial weight. Both uniform and inverse distance weight matrices give the same MAPE, that is, 3.57. This result aligns with the principle of parsimony in modeling that the model with the least parameters will be better than complex, on condition that the performances are not significantly different for the modeling. Thus, the GSTAR(1;1) with MST inverse distance weight matrix is recommended.

The predicted model using GSTAR(1;1) is formulated as follows:

$$\hat{Z}_{\mathrm{PDL}}(t) = 0.273 Z_{\mathrm{PDL}}(t-1) - 0.207 \left(0.51\, Z_{\mathrm{PST}}(t-1) + \mathbf{0.49}\, \boldsymbol{Z}_{\mathbf{CLY}}(t-1)\right)$$

$$\hat{Z}_{\mathrm{PST}}(t) = \mathbf{0.228}\, \boldsymbol{Z}_{\mathbf{PST}}(t-1) - 0.031(0.21 Z_{\mathrm{PDL}}(t-1) + 0.20 Z_{\mathrm{BAR}}(t-1) + 0.20\, Z_{\mathrm{PKJ}}(t-1) + 0.19\, Z_{\mathrm{KPO}}(t-1) + 0.20\, Z_{\mathrm{MTH}}(t-1))$$

$$\hat{Z}_{\mathrm{BAR}}(t) = -\mathbf{0.159}\, \boldsymbol{Z}_{\mathbf{BAR}}(t-1) + 0.133\, Z_{\mathrm{PST}}(t-1)$$

$$\hat{Z}_{\mathrm{PKJ}}(t) = -1.025\, Z_{\mathrm{PKJ}}(t-1) + \mathbf{1.289}\, Z_{\mathrm{PST}}(t-1)$$

$$\hat{Z}_{\mathrm{KPO}}(t) = 0.042\, Z_{\mathrm{KPO}}(t-1) - \mathbf{0.136}\, Z_{\mathrm{PST}}(t-1)$$

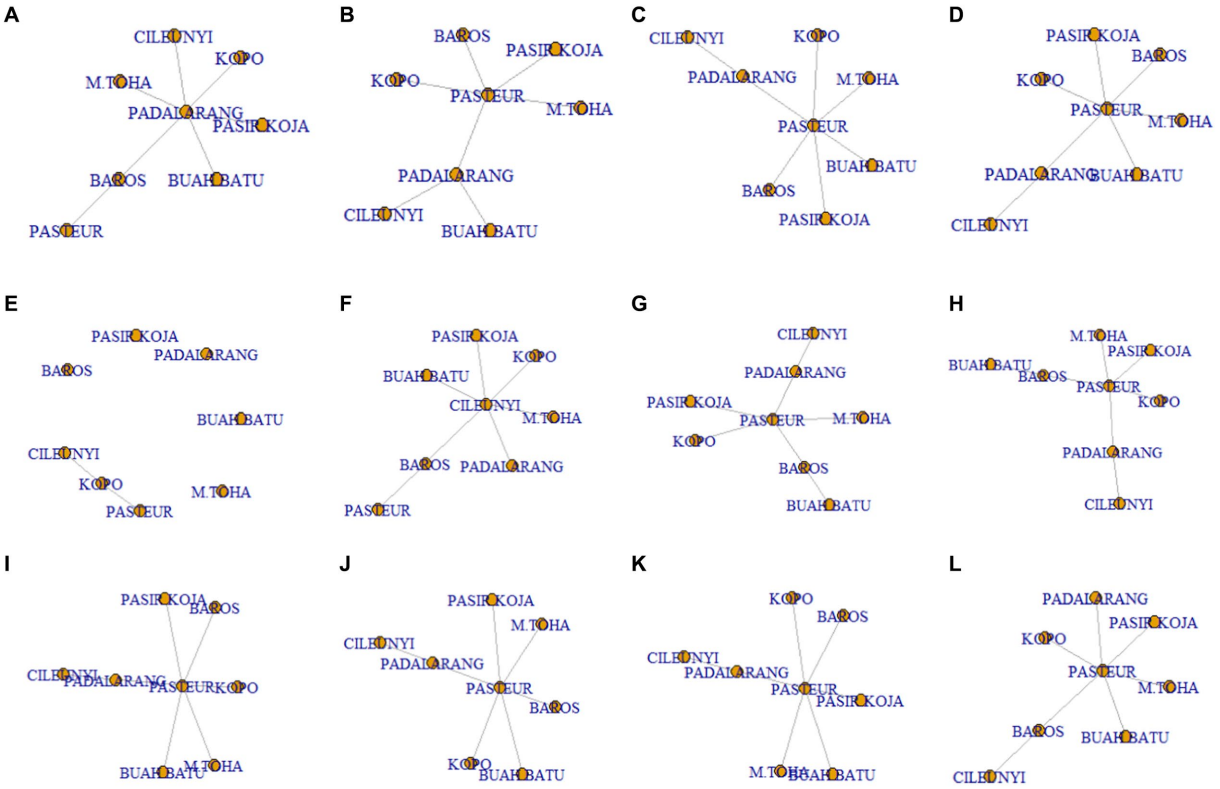


FIGURE 6
MST was obtained for every month of observations. The Pasteur toll gate is the center of MST most of the time. **(A)** January. **(B)** February. **(C)** March. **(D)** April. **(E)** May. **(F)** June. **(G)** July. **(H)** August. **(I)** September. **(J)** October. **(K)** November. **(L)** December.

TABLE 4 STACF and STPACF of weight matrices.

		STACF												STPACF										
		Time lag												Time lag										
(a) Uniform																								
			0	1	2	3	4	5	6	7	8	9	10		0	1	2	3	4	5	6	7	8	9	10
RS	Spatial lag	0	X	X	X	X	X	X	X	X	0	0	0	0	X	X	0	0	0	0	0	X	X	0	0
MST			X	X	X	X	0	X	X	X	0	0	0		X	X	0	0	0	0	0	X	X	0	0
RS		1	X	X	X	X	X	0	0	X	0	0	0	1	X	X	X	X	0	0	0	0	X	0	0
MST			X	X	X	X	X	X	0	0	0	0	0		X	0	X	X	0	0	0	0	X	0	0
RS		2	X	X	X	X	X	X	0	0	0	0	0	2	X	0	0	X	0	X	0	X	X	0	0
MST			X	X	X	X	0	0	0	X	0	0	0		X	X	0	0	0	0	0	0	0	0	0
RS		3	X	X	X	X	X	X	0	0	0	0	0	3	X	0	0	0	0	X	X	0	0	0	0
MST			X	0	0	0	0	X	0	0	0	0	0		X	0	0	0	0	0	0	0	0	0	0
RS		4	X	X	X	X	X	X	X	X	0	0	0	4	X	0	0	0	0	X	0	0	0	0	0
MST			X	X	X	X	X	X	X	X	0	0	0		X	0	X	0	0	0	0	X	X	0	0
(b) Inverse distance																								
			0	1	2	3	4	5	6	7	8	9	10		0	1	2	3	4	5	6	7	8	9	10

(Continued)

TABLE 4 (Continued)

		STACF											STPACF										
		Time lag											Time lag										
RS	Spatial lag	0	X	X	X	X	0	X	X	X	0	0	0	X	X	0	0	0	0	0	X	X	0	0
MST			X	X	X	X	X	X	X	X	X	0	0	X	X	0	0	0	0	0	X	X	0	0
RS		1	X	X	X	X	X	X	0	0	0	0	0	1	X	0	X	X	0	0	0	0	X	0	0
MST			X	X	X	X	X	X	X	0	0	0	0		X	0	X	X	0	0	0	0	X	0	0
RS		2	X	X	X	X	X	X	X	X	0	0	0	2	X	0	0	X	0	0	0	X	X	0	0
MST			X	X	X	X	0	0	0	X	0	0	0		X	X	X	0	0	0	0	0	0	0	0
RS		3	X	X	X	X	X	X	0	0	0	0	0	3	X	0	0	0	0	0	0	0	0	0	0
MST			X	0	0	0	0	X	0	0	0	0	0		X	X	0	0	0	0	0	0	0	0	0
RS		4	X	X	X	X	X	X	X	X	0	0	0	4	X	0	0	0	0	0	0	0	0	0	0
MST			X	X	X	X	X	X	0	0	0	0	0		X	X	X	0	0	0	0	X	X	0	0

The symbol "X" means significant and "0" otherwise.

TABLE 5 Least-square estimators of GSTAR parameters using MST−u (uniform) and MST−id (inverse distance) weight matrix.

Model			PDL	PST	BAR	PKJ	KPO	MTH	BBT	CLY
GSTAR (1;1)	u	$\hat{\phi}10$	0.050	0.226	−0.154	−0.082	−0.074	−0.020	0.248	0.696
		$\hat{\phi}11$	0.135	0.029	0.070	0.138	0.076	0.090	0.006	−0.304
	id	$\hat{\phi}10$	0.273	0.228	−0.159	−1.025	0.042	0.767	0.334	0.698
		$\hat{\phi}11$	−0.207	−0.031	0.133	1.289	−0.136	−0.756	−0.114	−0.383
GSTAR (1;2)	u	$\hat{\phi}10$	0.099	0.167	−0.435	−1.320	0.173	0.104	0.404	0.684
		$\hat{\phi}11$	0.145	−0.407	0.016	0.061	0.223	0.201	0.049	0.176
		$\hat{\phi}12$	−0.104	0.705	0.573	0.148	0.327	−0.197	−0.189	−0.655
	id	$\hat{\phi}10$	0.099	0.179	−0.252	−0.132	0.174	0.104	0.297	0.679
		$\hat{\phi}11$	0.147	−0.542	0.093	0.061	0.223	0.201	0.762	0.238
		$\hat{\phi}12$	−0.106	0.881	0.282	0.148	−0.327	−0.197	−0.111	−0.717
GSTAR (2;1,1)	u	$\hat{\phi}10$	0.307	0.183	−0.187	−1.252	0.186	0.977	0.258	0.683
		$\hat{\phi}11$	−0.229	−0.070	0.195	1.759	−0.283	−0.896	−0.070	−0.394
		$\hat{\phi}20$	−0.145	0.221	0.146	1.327	−0.426	−0.835	0.174	0.021
		$\hat{\phi}21$	0.127	0.152	−0.156	−1.453	0.512	0.938	0.021	0.169

TABLE 6 Determinant values of leading principal IAcM submatrices show the stationarity of processes.

GSTAR order	Uniform		Distance	
	Radius	MST	Radius	MST
(1;1)	>0	>0	>0	>0
(1;2)	>0	>0	>0	>0
(2;1,1)	>0	>0	>0	>0

10.3389/fams.2024.1417037

TABLE 7 RMSE from eight Purbaleunyi toll gates.

Tollgate	Order	Uniform		Distance	
		Radius	MST	Radius	MST
PDL	(1;1)	2616.661	2586.390	2616.661	**2585.834**
	(1;2)	**2546.412**	2559.655	2548.948	2569.522
	(2;1,1)	**2623.792**	2686.045	**2623.792**	2686.088
PST	(1;1)	3443.521	3435.682	3443.521	**3435.634**
	(1;2)	3320.072	3394.913	**3294.799**	3394.910
	(2;1,1)	3604.767	3518.186	3604.767	**2518.698**
BAR	(1;1)	2630.346	2626.811	2619.479	**2599.394**
	(1;2)	2626.287	2626.593	2623.845	**2605.288**
	(2;1,1)	2657.802	**1640.275**	2664.289	2662.164
PKJ	(1;1)	2599.527	2679.495	**2613.121**	2679.495
	(1;2)	**2620.340**	2659.426	2625.978	2648.594
	(2;1,1)	3099.357	**2747.165**	3090.523	**2747.165**
KPO	(1;1)	1912.118	**1892.634**	1909.651	**1892.634**
	(1;2)	1881.117	1867.131	1877.790	**1866.727**
	(2;1,1)	1919.934	1939.994	**1918.663**	1939.994
MTH	(1;1)	**1750.203**	1769.667	1776.659	1769.667
	(1;2)	1730.875	1709.350	1751.897	**1700.146**
	(2;1,1)	1856.581	**1791.338**	1846.763	**1791.338**
BBT	(1;1)	1644.857	1650.299	1644.915	**1637.668**
	(1;2)	1655.997	1658.806	1655.825	**1648.768**
	(2;1,1)	1720.936	1983.612	**1720.269**	1720.390
CLY	(1;1)	**3863.379**	3871.781	3863.679	3871.949
	(1;2)	**3951.813**	3977.738	3954.516	3975.918
	(2;1,1)	4457.320	**4358.423**	4457.320	4378.970

The majority of the smallest RMSE is given by inverse distance weight matrix with the MST approach. Meanwhile, the GSTAR(1;1) has the smallest RMSE.
The bold values indicate the smallest value for each possible model of every location.

TABLE 8 Predicted value of the GSTAR model with various weight matrices and its MAPEs.

	Location			PDL	PST	BAR	PKJ	KPO	MTH	BBT	CLY	MAPE
	Real data (Z)			28,050	31,395	21,290	23,956	13,164	14,339	21,705	32,207	
Prediction (Z)	GSTAR (1;1)	u	radius	26,287	29,494	19,805	21,595	**12,117**	12,822	20,421	30,803	4.36%
			MST	26,089	29,426	19,897	21,425	11,881	12,492	20,492	31,057	**3.57%**
		id	radius	26,287	29,494	19,805	21,595	**12,117**	**12,823**	20,421	30,803	4.36%
			MST	26,089	29,426	19,913	21,425	11,881	12,492	20,492	31,056	**3.57%**
	GSTAR (1;2)	u	radius	26,261	29,342	19,807	21,132	11,726	12,009	20,653	30,774	4.45%
			MST	26,291	29,413	19,855	21,477	11,713	12,340	20,483	30,749	4.53%
		id	radius	26,257	29,311	19,812	21,135	11,754	12,108	20,652	30,774	4.45%
			MST	**26,293**	29,413	**19,977**	21,414	11,705	12,299	20,468	30,718	4.62%
	GSTAR (2;1,1)	u	radius	25,968	29,579	19,686	21,859	12,021	12,742	20,286	30,839	4.25%
			MST	26,111	29,668	19,834	21,724	11,810	12,783	20,780	**31,437**	**2.39%**
		id	radius	25,968	29,579	19,733	**21,866**	11,980	12,775	20,295	30,839	4.25%
			MST	26,113	**29,679**	19,819	21,724	11,810	12,783	**20,853**	31,416	**2.46%**

The MST gives a closer prediction than the radius approach in five of eight locations (bold printed). Those are the Padalarang, Pasteur, Baros, Buah Batu, and Cileunyi toll gates.

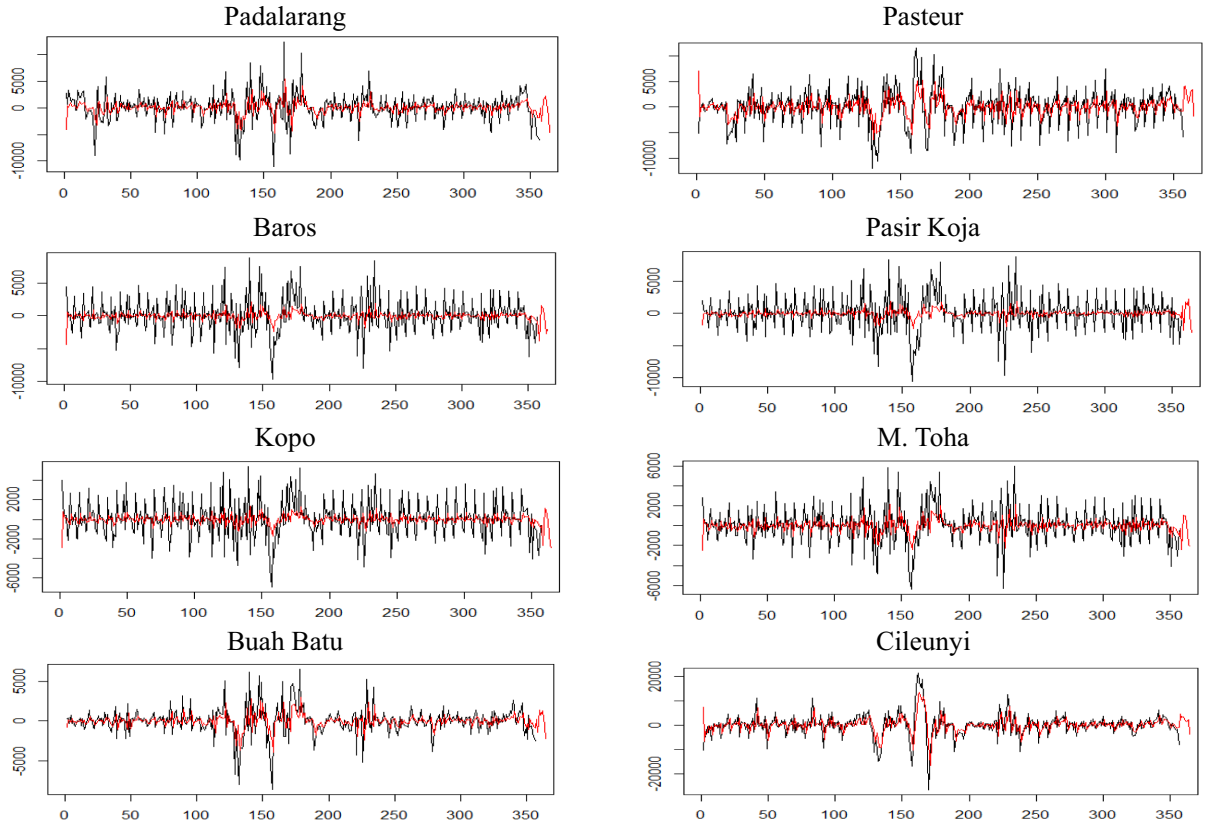


FIGURE 7
A plot of re-estimation and prediction of observations using the GSTAR(1;1) model with MST inverse distance weight matrix for every toll gate. The real observations (black line) have higher variability than the estimated/predicted observations (red line).

$$\hat{Z}_{\mathrm{MTH}}(t) = \mathbf{0.767}\, Z_{\mathrm{MTH}}(t-1) - 0.746\, Z_{\mathrm{PST}}(t-1)$$

$$\hat{Z}_{\mathrm{BBT}}(t) = \mathbf{0.334}\, Z_{\mathrm{BBT}}(t-1) - 0.114\, Z_{\mathrm{PST}}(t-1)$$

$$\hat{Z}_{\mathrm{CLY}}(t) = \mathbf{0.698}\, Z_{\mathrm{CLY}}(t-1) - 0.383\left(0.43\, Z_{\mathrm{PDL}}(t-1) + 0.57\, Z_{\mathrm{CLY}}(t-1)\right)$$

Based on those equations, every location in a certain time (t) is most influenced by the one previous observation in the same locations, except for Padalarang and Pasir Koja. For both toll gates, consecutively, the previous observation of Cileunyi and Pasteur plays a big role. Different from others, the Baros toll gate has a negative coefficient (−0,159) for its previous observation, which means that the current number of vehicles will decrease linearly to the previous one. With Pasir Koja, Kopo, M. Toha, and Buah Batu, Baros has Pasteur as the only neighbor whose previous observations influenced the respective toll gates. Padalarang, Pasteur, and Cileunyi, the top three busiest toll gates, have their own gates in previous times as the biggest influence on the current observation, except for Padalarang, which has Cileunyi for this case. Pasteur toll gate is influenced by all toll gates, except Buah Batu and Cileunyi toll gates.

The prediction using GSTAR(1;1)–MST inverse distance of weight matrix is plotted in Figure 7. Here, the pattern of re-estimation follows the real observations, although the variability of the estimated observations tends to be smaller than the real observations.

4 Conclusion and remarks

The MST, as a new approach to building the weight matrix of the GSTAR model, can capture the spatial configuration based on space–time observations. This approach is competitive with the conventional approach's good performance, which is purely based on the inverse distance, in predicting. The MST inverse distance weight matrix is suspected to be appropriate for data with larger variability. The most important thing in building this weight matrix is that the correlation among spatial observations must exist.

Furthermore, since the real observations are the realization of stochastic processes, then the MST approach may give different configurations in different time frames. In this study, the MST approach was built based on all history observations to obtain a fixed weight matrix that can be used to do short-time forecasting. It can be considered as both an advantage and a limitation of this approach. In the future, the weight matrix can be evaluated as a random matrix. It can be the future work of this research.

On the other hand, for the number of vehicles entering Bandung through some toll gates data, the GSTAR(1;1) is the most appropriate to be used. This model is chosen based on the RMSE and AIC-BIC values. Although the prediction result is tight competing with the

GSTAR(2;1,1) model, the GSTAR(1;1) is selected due to the parsimony principle in modeling. However, the results of re-estimation and prediction have not been maximal (still not close to the real observations); therefore, the method should be improved. One of the ways to improve is using the SDU approach to obtain a unique MST.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

UM: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. AM: Supervision, Writing – review & editing. KS: Formal analysis, Supervision, Writing – review & editing. NN: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors would like to extend their gratitude to the PDUPT research grant of RISTEK-BRIN 2020–2021 (Contract No. 2/E1/KP.PTNBH/2020 and No. 2/E1/KP.PTNBH/2021) for supporting this research and PT. Jasamarga, Purbaleunyi branch for the data. This research is also supported by LPPM ITB through Riset Unggulan ITB 2022 (Contract No. 293/IT1.B07.1/TA.00/2022) and 2024 (Contract No. 959/IT1.B07.1/TA.00/2024).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Mukhaiyar U., Pasaribu U. S. The use of IAcM to identify stationarity of the generalized STAR models. (2012) In: *2012 IEEE conference on control, Systems & industrial informatics* (pp. 255–260). IEEE.

2. Yundari Y., Huda N. A. M., Pasaribu U. S., Mukhaiyar U., Sari K. N. Stationary process in GSTAR (1; 1) through kernel function approach. (2020). In: *AIP conference proceedings* (Vol. 2268, No. 1). AIP Publishing.

3. Nainggolan N., Titaley J. Development of generalized space time autoregressive (GSTAR) model. (2017). In: *AIP conference proceedings* (Vol. 1827). AIP Publishing.

4. Mukhaiyar U, Ramadhani S. The generalized STAR modeling with heteroscedastic effects. *CAUCHY J Matematika Murni dan Aplikasi*. (2022) 7:158–72. doi: 10.18860/ca.v7i2.13097

5. Yundari PUS, Mukhaiyar U, Heriawan MN. Spatial weight determination of GSTAR (1; 1) model by using kernel function. *J Phys*. (2018) 1028:012223. doi: 10.1088/1742-6596/1028/1/012223

6. Mukhaiyar U, Bilad BI, Pasaribu US. The generalized STAR modelling with minimum spanning tree approach of weight matrix for COVID-19 case in Java Island. *J Phys*. (2021) 2084:012003. doi: 10.1088/1742-6596/2084/1/012003

7. Masteriana D., Mukhaiyar U. Monte Carlo simulation of error assumptions in generalized STAR (1; 1) model. (2019). In: *Proceedings of the Jangjeon Mathematical Society* (Vol. 22, pp. 43–50).

8. Yundari PUS, Mukhaiyar U. Error assumptions on generalized STAR model. *J Mathemat Fundamental Sci*. (2017) 49:136. doi: 10.5614/j.math.fund.sci.2017.49.2.4

9. Mukhaiyar U, Huda NM, Sari KN, Pasaribu US. Analysis of generalized space time autoregressive with exogenous variable (GSTARX) model with outlier factor. *J Phys*. (2020) 1496:012004. doi: 10.1088/1742-6596/1496/1/012004

10. Wardhani LP, Kuswanto H. Poisson GSTAR model: spatial temporal modeling count data follow generalized linear model and count time series models. *J Phys*. (2020) 1490:012010. doi: 10.1088/1742-6596/1490/1/012010

11. Gehman A, Wei WW. Optimal spatial aggregation of space–time models and applications. *Comput Stat Data Analys*. (2020) 145:106913. doi: 10.1016/j.csda.2020.106913

12. Mukhaiyar U., Pasaribu U. S., Budhi W. S., Syuhada K. The stationarity of generalized STAR (2; λ1, λ2) process through the invers of autocovariance matrix. (2014). In: *AIP conference proceedings* (Vol. 1589, No. 1, pp. 484–487). American Institute of Physics.

13. Huda NM, Mukhaiyar U, Pasaribu US. The approximation of GSTAR model for discrete cases through INAR model. *J Phys*. (2021) 1722:012100. doi: 10.1088/1742-6596/1722/1/012100

14. Nurhayati N, Pasaribu US, Neswan O. Application of generalized space-time autoregressive model on GDP data in west European countries. *J Probabil Stat*. (2012) 2012:1–16. doi: 10.1155/2012/867056

15. Mukhaiyar U. The goodness of generalized STAR in spatial dependency observations modeling. (2015). In: *AIP conference proceedings* (Vol. 1692). AIP Publishing.

16. Nugraha R. F., Setiyowati S., Mukhaiyar U., Yuliawati A. Prediction of oil palm production using the weighted average of fuzzy sets concept approach. (2015). In: *AIP conference proceedings* (Vol. 1692). AIP Publishing.

17. Mukhaiyar U., Fahmi F. The generalized STAR (1, 1) modeling with time correlated errors to red-chili weekly prices of some traditional markets in Bandung, West Java. (2015). In: *AIP conference proceedings* (Vol. 1692). AIP Publishing.

18. Abdullah AS, Matoha S, Lubis DA, Falah AN, Jaya IGNM, Hermawan E, et al. Implementation of generalized space time autoregressive (GSTAR)-kriging model for predicting rainfall data at unobserved locations in West Java. *Appl Maths Informat Sci*. (2018) 12:607–15. doi: 10.18576/amis/120316

19. Mukhaiyar U, Sari RKN, Pasaribu US. Modeling dengue fever cases by using GSTAR (1; 1) model with outlier factor. *J Phys*. (2019) 1366:012122. doi: 10.1088/1742-6596/1366/1/012122

20. Masteriana D, Riani MI, Mukhaiyar U. Generalized STAR (1; 1) model with outlier-case study of begal in Medan, north Sumatera. *J Phys*. (2019) 1245:012046. doi: 10.1088/1742-6596/1245/1/012046

21. Zewdie MA, Wubit GG, Ayele AW. G-STAR model for forecasting space-time variation of temperature in northern Ethiopia. *Turk J Forecast*. (2018) 2:9–19. doi: 10.34110/forecasting.437599

22. Pasaribu US, Mukhaiyar U, Huda NM, Sari KN, Indratno SW. Modelling COVID-19 growth cases of provinces in java island by modified spatial weight matrix GSTAR through railroad passenger's mobility. *Heliyon*. (2021) 7:e06025. doi: 10.1016/j.heliyon.2021.e06025

23. Mukhaiyar U, Pasaribu US. A new procedure for generalized STAR modeling using IAcM approach. *ITB J Sci*. (2012) 44:179–92. doi: 10.5614/itbj.sci.2012.44.2.7

24. Moretti F, Pizzuti S, Panzieri S, Annunziato M. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing*. (2015) 167:3–7. doi: 10.1016/j.neucom.2014.08.100

25. Giacomazzo M, Kamarianakis Y. Bayesian estimation of subset threshold autoregressions: short-term forecasting of traffic occupancy. *J Appl Stat*. (2020) 47:2658–89. doi: 10.1080/02664763.2020.1801606

26. Wang P, Xu W, Jin Y, Wang J, Li L, Lu Q, et al. Forecasting traffic volume at a designated cross-section location on a freeway from large-regional toll collection data. *IEEE Access*. (2019) 7:9057–70. doi: 10.1109/ACCESS.2018.2890725

27. Mukhaiyar U, Nabilah FT, Pasaribu US, Huda NM. The space-time autoregressive modeling with time correlated errors for the number of vehicles in Purbaleunyi toll gates. *J Phys*. (2022) 2243:012068. doi: 10.1088/1742-6596/2243/1/012068

28. Jedwanna K, Athan C, Boonsiripant S. Estimating toll road travel times using segment-based data imputation. *Sustain For*. (2023) 15:13042. doi: 10.3390/su151713042

29. Shi T, Wang P, Qi X, Yang J, He R, Yang J, et al. CPT-DF: congestion prediction on toll-gates using deep learning and fuzzy evaluation for freeway network in China. *J Adv Transp*. (2023) 2023:1–16. doi: 10.1155/2023/2941035

30. Niu J, He J, Li Y, Zhang S. Highway temporal-spatial traffic flow performance estimation by using gantry toll collection samples: a deep learning method. *Math Probl Eng*. (2022) 2022:1–10. doi: 10.1155/2022/8711567

31. Shevchenko P, Peters GW, Matsui T, Septier F. Multi-view travel time prediction based on electronic toll collection data. *Entropy*. (2022) 24:1050. doi: 10.3390/e24081050

32. Huda NM, Imro'ah N. 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*. (2023) 54:100734. doi: 10.1016/j.spasta.2023.100734

33. Bondy JA, Murty USR. Graph theory with applications (Vol. 290). London: Macmillan (1976).

34. Djauhari MA, Gan SL. Optimality problem of network topology in stocks market analysis. *Phys A Stat Mechanics Applic*. (2015) 419:108–14. doi: 10.1016/j.physa.2014.09.060

35. Pfeifer PE, Deutsch SJ. A three-stage iterative approach for space-time modelling. *Technometrics*. (1980) 22:397–408. doi: 10.1080/00401706.1980.10486172