Mukhaiyar *et al.*
Environmental Sciences Europe (2024) 36:180
https://doi.org/10.1186/s12302-024-00979-6

Environmental Sciences Europe

RESEARCH

Open Access

The generalized STAR modelling with three-dimensional of spatial weight matrix in predicting the Indonesia peatland's water level

Utriweni Mukhaiyar[1*], Adilan Widyawan Mahdiyasa[1*], Tarasinta Prastoro[2], Udjianna Sekteria Pasaribu[1], Kurnia Novita Sari[1], Sapto Wahyu Indratno[1], Indratmo Soekarno[3], Devi Nandita Choesin[4], Isro Ismail[5], Dian Rosleine[4] and Danang Teguh Qoyyimi[6]

Abstract

The release rate of CO_2 gas can be influenced by peatlands' physical properties, such as water level and soil moisture, and rainfall. To anticipate the unstable condition which is when the peatland emit more carbon, we developed the Generalized Space Time Autoregressive (GSTAR) model in predicting these physical properties for the following weeks. As the innovation in modelling, the spatial weight matrix was based on three-dimensional coordinates with a modification on the height factor. The data we used are real-time data of water level on the peatlands in Pulang Pisau Regency, Central Kalimantan Province from 20 February 2021 to 18 March 2023. We then used Ordinary Kriging interpolation on the prediction results to create contour maps on different dates. There were empty data on several dates, especially from 24 March until 3 August 2022. To fill the empty data, we used linear interpolation and then we added white noise to the interpolation results. From the data, the water level has a downward trend pattern from around November to September and an upward trend pattern from October to November. Furthermore, we found that the best model for water level was GSTAR (2;0.1) with a modified matrix $a = 0.1$ and $b = 1.1$. Based on the predicted water level, there is a risk of changes in the properties of the peatlands in several areas in Pulang Pisau Regency.

Keywords Peatland, Prediction, Water level, GSTAR model, 3D-spatial weight matrix, Ordinary kriging

*Correspondence:
Utriweni Mukhaiyar
utriweni.mukhaiyar@itb.ac.id
Adilan Widyawan Mahdiyasa
adilan.widyawan@itb.ac.id
[1] Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
[2] Master Program in Actuarial Science, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
[3] Water Resources Engineering Research Group, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
[4] Ecology Research Division, School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
[5] Indonesian Peat Association, Bogor, Indonesia
[6] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara, Yogyakarta, Indonesia

Introduction

One of the most important roles of peatlands on the environment is their function as the largest carbon storage. On a global scale, peatlands can store twice as much carbon as other forests in total. Peatlands accumulate vast amounts of carbon in the long term over thousands of years, because the absorbed carbon through plant photosynthesis is not all released back into the atmosphere as a consequence of incomplete decomposition. However, the condition changes dramatically for peatlands in unstable conditions, because they could release huge amounts of carbon that has been stored. The unstable condition can be caused by drainage and fires, which commonly occur on degraded peatlands [6]. This condition could affect the global carbon cycle and contribute positively to global warming.

The rate of carbon input and release from the peatland is significantly affected by the water level position [5, 13, 15, 16, 35]. A shallow water level position leads to the shorter residence time of plant litter or organic matter in an unsaturated zone that supports carbon accumulation. Contrastingly, a deeper water level increases the thickness of the unsaturated zone, which results in a more significant carbon release from the peatland [17]. Furthermore, a deep water level position could produce peatland fire [33] and leads to the unstable condition of the peatland. Therefore, modelling and predicting the water level are crucial to anticipate the unstable condition of the peatland.

The water level can be modelled by deterministic or stochastic approaches. The deterministic models [5, 17, 35] do not capture the uncertainties of the system that could influence the model outputs. Contrastingly, the stochastic models involve uncertainties that might improve the model capability and accuracy, particularly for complex systems including peatland water level. Tamea et al. [32] developed a stochastic model of the water level for the wetlands that can describe a long-term probability distribution of water level depth in temporarily inundated sites. However, this model ignores the influence of spatial interactions on the water level position, which becomes a significant limitation. The field observation from Lewis et al. [14] showed that the water level is located in a deeper position at the margin compared to the centre on the peatland due to the variations in hydraulic conductivity and bulk density. This condition is also supported by the simulation from Mahdiyasa et al. [15, 16], who found the spatial variability of the peatland water level through the fully coupled mechanical-ecohydrological model.

To overcome these problems, we propose a novel stochastic approach through the Generalized Space Time Autoregressive (GSTAR) combined with Kriging to provide the future value and spatial interpolation of water level position. The GSTAR model can be employed to predict the future outcome of random variables that are affected not only by past observations but also by observations in neighbouring regions [4, 8, 22]. The GSTAR is the extension of the Space Time Autoregressive (STAR) model [29], because it allows the autoregressive parameters to vary across locations. The specific characteristic of GSTAR is the inclusion of a weight matrix that captures the spatial relationship between different locations [27].

The GSTAR model has rapidly developed both in theory and application. On the theoretical side, the development of weight matrices has been carried out using various approaches, including the kernel approach [37] and minimum spanning tree [24]. In application, the GSTAR model has been explored for economic [26], oil palm production [21], commodity prices [10], tea production [36], criminal cases [18], coffee berry borer attack [30], Dengue Fever cases [22], layer peat soil [12], rainfall [34], copper and gold grades [28], toll gates density [25], COVID-19 cases [8], and climate data [31]. Those applications considered the spatial dependency among time series observations in many locations and used the two-dimensional weight matrices. Here, by considering that the current water level can be seen as the impact of the previous water level in same locations as well its neighbour locations, then the GSTAR modelling be applied. Furthermore, since the peatlands have the thickness variable, it is deemed necessary to use a three-dimensional weight matrix.

The location for this study is the Pulang Pisau Regency, Central Kalimantan Province, Indonesia. The region features a mix of lowlands and wetlands, with a significant portion of its land covered by peatlands. Similar to the other area of Central Kalimantan, Pulang Pisau experiences a tropical rainforest climate, with monsoon season. One of the most significant issues in the

Central Kalimantan area is the peatland fires. For example, around 20324 peatland fires were detected in the Central Kalimantan in 2015, which becomes the highest density of peatland fires in South East Asia [7, 19]. The distribution of peatland fires indicate the maximum number of fire appear in August until October, which is related to the long dry season and strong El Nino phenomenon [9].

This paper, therefore, sets out to (1) develop a three-dimensional spatial weight matrix for Generalized Space Time Autoregressive (GSTAR) that involves the height factor, (2) estimate the future value of water level position through GSTAR model, and (3) produce water level map of the peatland in Pulang Pisau Regency, Central Kalimantan Province, Indonesia as the case study.

The spatial weight matrix

Suppose $\{Y_{i,t}\}$ denotes a process with zero mean at the i-th location. Suppose $\{\boldsymbol{Y}_t\}$ denotes a process vector with zero mean and $\boldsymbol{\Phi}_{k\ell} = diag\left(\phi_{k\ell}^{(1)}, \ldots, \phi_{k\ell}^{(N)}\right)$ where $\phi_{k\ell}^{(i)}$ denotes the autoregressive parameter at time lag k and spatial lag ℓ for location i. If $\{\boldsymbol{Y}_t\}$ follows the GSTAR(p; $\lambda_1, \ldots, \lambda_p$) model, we get

$$\boldsymbol{Y}_t = \sum_{k=1}^{p} \sum_{\ell=0}^{\lambda_k} \boldsymbol{\Phi}_{k\ell} \boldsymbol{W}^{(\ell)} \boldsymbol{Y}_{t-k} + \boldsymbol{\varepsilon}_t,$$

where the vector $\boldsymbol{Y}_t$ represents the observation vector at N locations with $\boldsymbol{Y}_t = \left(Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t}\right)'$; matrix $\boldsymbol{W}^{(\ell)}$ represents the spatial weight matrix at spatial lag ℓ with $\boldsymbol{W}^{(\ell)} = \left(w_{ij}^{(\ell)}\right)$; the notation $\boldsymbol{\varepsilon}_t$ states the error vector at the time t which is a white noise with $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$ [37].

The spatial weight matrix $\left(\boldsymbol{W}^{(\ell)} = \left\{w_{ij}^{(\ell)}\right\}\right)$ serves to describe the spatial relationship between locations. The elements of the spatial weight matrix $w_{ij}^{(\ell)}$ represent the weight of the spatial relationship between location j and location i. Determining the value of $w_{ij}^{(\ell)}$ is strongly influenced by the spatial lag order definition.

In this paper, we defined a spatial lag order which resembles a layered rhombus. This definition is a development of the square grid system [3] and the geographic location definition [26]. Suppose s_i represents the ith reference location's square in the grid with its two-dimensional coordinates, namely (p_i, q_i), for $i = 1,2,\ldots,N$. Assume there is only one location in s_i. Neighbours ordered $\ell = 1$ from the location located in s_i are locations with squares directly adjacent to s_i. This makes neighbours of order $\ell = 1$ located in squares with their two-dimensional coordinates, namely $(p_i - 1, q_i)$, $(p_i + 1, q_i)$, $(p_i, q_i - 1)$, or $(p_i, q_i + 1)$. For order $\ell = 2, 3, \ldots$, neighbours of order ℓ are all neighbours of order $(\ell - 1)$ from neighbours of order $(\ell - 1)$ of the location located in s_i other than the reference location itself. As an alternative definition, a location located in s_j is an ℓth order neighbour of a location located in s_i if

$$C(\ell - 1) < |p_j - p_i| + |q_j - q_i| \leq C\ell, \tag{1}$$

where C is a positive constant. An illustrative example of this definition is shown in Fig. 1.

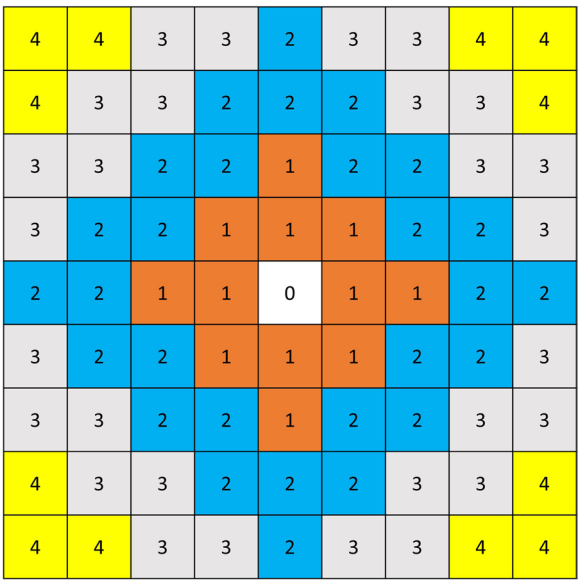

Fig. 1 Illustration of the spatial lag order definition resulting from the development for $C = 2$

After defining the spatial lag order, we can formulate the spatial weight matrix for each respective spatial lag.

To accommodate peatlands that have a thickness factor, a three-dimensional weight matrix was developed. In determining the weight value in three dimensions, we proposed the following three ways.

1) Three-dimensional distance inverse matrix with modifications to the height element

Let $\boldsymbol{W^{Mod^{(\ell)}}} = \left\{ w_{ij}^{Mod^{(\ell)}} \right\}$ represents a three-dimensional inverse distance spatial weight matrix with modifications to the height element having dimensions $N \times N$ with ℓth spatial lag order. The matrix $\boldsymbol{W^{Mod^{(\ell)}}}$ is the normalized form of $\boldsymbol{W^{SH^{(\ell)}}} = \left\{ w^{SH^{(\ell)}}_{ij} \right\}$ which is the Hadamard multiplication matrix between matrices $\mathcal{S}^{(\ell)}$ and $\mathcal{H}$ which has the same dimensions as $\boldsymbol{W^{Mod^{(\ell)}}}$.

The matrix $\mathcal{S}^{(\ell)}$ represents the spatial relationship between locations in two dimensions with ℓ spatial lag order with $\boldsymbol{\mathcal{S}^{(\ell)}} = \left\{ s_{ij}^{(\ell)} \right\}$ where

$$s_{ij}^{(\ell)} = \frac{1}{1+d_{ij}},$$

With d_{ij} denotes the distance between the centre of the reference location i and the neighbour j according to its position in the plane.

The matrix $\mathcal{H}$ represents the spatial relationship between locations according to the difference in height with $\boldsymbol{\mathcal{H}} = \{h_{ij}\}$. Thus, the h_{ij} element represents the weight by height of the neighbour j towards the reference location i, so it can be defined

$$h_{ij} = \begin{cases} 1, & \text{if } l_i = l_j, \forall i \neq j \\ 0, & \text{if } l_i = l_j, \forall i = j \\ \dfrac{1}{b\left|l_i - l_j\right|^a}, & \text{if } l_i \neq l_j, d_{ij} \leq d \\ 1, & \text{if } \ l_i \neq l_j, d_{ij} > d, \end{cases}$$

where $a, b \geq 0$ is a fixed parameter, l_i represents the height of the centre of the reference location i, l_j represents the height of the centre of the neighbouring location j, d_{ij} represents the euclidian distance between the centres of locations i and j on the plane, and ⌈ is a constant. In this paper, we define ⌈ as the average distance from each pair of reference locations and neighbouring locations.

By combining the matrices $\mathcal{S}^{(\ell)}$ and $\mathcal{H}$ through element-by-element multiplication, we get the matrix

$$\boldsymbol{W^{SH^{(\ell)}}} = \boldsymbol{\mathcal{S}^{(\ell)}} \odot \boldsymbol{\mathcal{H}} = \begin{pmatrix} 0 & s_{12}^{(\ell)} h_{12} & \cdots & s_{1N}^{(\ell)} h_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1}^{(\ell)} h_{N1} & s_{N2}^{(\ell)} h_{N2} & \cdots & 0 \end{pmatrix},$$

With operator $\odot$ represents the Hadamard matrix multiplication [20]. After obtaining $\boldsymbol{W^{SH^{(\ell)}}}$, $\boldsymbol{W^{Mod^{(\ell)}}}$ can be obtained through this equation for the elements

$$w_{ij}^{Mod^{(\ell)}} = \frac{w_{ij}^{SH^{(\ell)}}}{\sum_{j=1}^{N} w_{ij}^{SH^{(\ell)}}}. \tag{2}$$

To determine the value of h_{ij}, there are two alternatives in determining the value of a and b, and both are

a. Using $a = 0.1$ and $b = 1.1$ [20].
b. Using the assumption of the smallest distance value.

Suppose $z = \left|l_i - l_j\right|$.
Suppose we define a function

$$f(a, b, z) = \frac{1}{bz^a}, a \geq 0, b > 0, z > 0.$$

Assume the smallest z is $z = 10^{-4}$ and the difference between 1 and $\frac{1}{bz^a}$ with $z = 10^{-4}$ is 10^{-4}, we then determine the values of a and b, so that

$$\frac{1}{b\left(10^{-4}\right)^a} + 10^{-4} = 1.$$

With $a \geq 0$ and $b > 1$. Therefore, we obtain

$$\frac{1}{b\left(10^{-4}\right)^a} = 0.9999.$$

With $a \geq 0$ and $b > 1$. From this equation, the value of b can be obtained by the following equation:

$$b = \frac{1}{0.9999(0.0001)^a}.$$

Therefore, it can be defined

$$f = \frac{1}{\frac{1}{0.9999(0.0001)^a} z^a} = 0.9999\left(\frac{0.0001}{z}\right)^a.$$

Since $\frac{1}{b|l_i - l_j|^a} = f$, we obtain

$$\hbar_{ij} = \begin{cases} 1, & \text{if } l_i = l_j, \forall i \neq j \\ 0, & \text{if } l_i = l_j, \forall i = j \\ 0.9999\left(\dfrac{0.0001}{|l_i - l_j|}\right)^a, & \text{if } l_i \neq l_j, d_{ij} \leq d \\ 1, & \text{if } l_i \neq l_j, d_{ij} > d \end{cases}$$

with a a fixed parameter.

From the equation above, for $l_i \neq l_j$ and $d_{ij} \leq \lceil$, the value of $\langle_{ij}$ will decrease following the increase in the value of a. This states that the greater the value of a, the smaller the weight of location j towards the reference location i if locations i and j have different heights. Furthermore, the value $a = 0$ causes $h_{ij} = 0.9999$ for $l_i \neq l_j$ and $d_{ij} \leq d$ so $w_{ij}^{SH(\ell)} \approx \int_{ij}^{(\ell)}$. Hence, for $a = 0$, the difference in height has no effect in determining the weight of the neighbouring location towards the reference location.

2) Three-dimensional distance inverse matrix without modification

Let $\boldsymbol{W}^{\boldsymbol{Inv}(\ell)} = \left\{ w^{Inv(\ell)}_{ij} \right\}$ represents an unmodified three-dimensional inverse distance spatial weight matrix with $N \times N$ dimensions with ℓth spatial lag order. Then, it can be defined that

$$w_{ij}^{Inv(\ell)} = \frac{\frac{1}{d_{ij}^*}}{\sum_{j=1}^{N} \frac{1}{d_{ij}^*}} \tag{3}$$

if location j is the ℓ th neighbour of location i or $w^{Inv(\ell)}_{ij} = 0$ if otherwise where d_{ij}^* represents the euclidian distance between the centres of location i and j in three-dimensional space. The d_{ij}^* value can be obtained by the following equation:

$$d_{ij}^* = \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2 + (r_i - r_j)^2},$$

where (p_i, q_i) and (p_j, q_j), respectively, represent the coordinates of locations i and j on the plane, and r_i and r_j, respectively, represent the heights of locations i and j.

Uniform weight matrix

Let $\boldsymbol{W}^{\boldsymbol{Unif}(\ell)} = \left\{ w^{unif}{}^{(\ell)}_{ij} \right\}$ denotes a uniform weight matrix with ℓ th spatial lag order. Then, it can be defined that

$$w^{unif_{ij}^{(\ell)}} = \begin{cases} \frac{1}{n_i^{(\ell)}}, & \text{if location } j \text{ is the } \ell\text{th neighbour of location } i \\ 0, & \text{if location } j \text{ is not the } \ell\text{th neighbour of location } i \end{cases} \tag{4}$$

with $n_i^{(\ell)}$ denotes the number of locations which are the ℓ th neighbours of location i [26].

The GSTAR-Kriging modelling

In this paper, we analysed the water level using the GSTAR model. The GSTAR modelling consisted of model identification, parameter estimation, and diagnostic tests. Then, we used Ordinary Kriging interpolation to create contour maps of the water level on the observed area. The flowchart of GSTAR-Kriging modelling be presented in Fig. 2 and the detail steps are explained in Algorithm 1.

Algorithm 1 GSTAR-Kriging modelling

Algorithm 1. *GSTAR-Kriging Modelling.*

Step 1: Perform data preparation and construction of the weight matrix through the following stages.

a. Enter the data for a variable at each location along with the location coordinates
b. If the data is empty on some dates,
   1) Estimate AR(1) model parameters on the data per location.
   2) Calculate the error variance for each location from the estimated parameters using
   $$Var(Y_{i,t})(1-\hat{\theta}_i^2)$$
   with $\hat{\theta}_i$ representing the estimated autoregressive parameters in the AR(1) model for location i.
   3) Generate simulated data of the same size as the number of dates with empty data per location in the form of white noise.
   4) Perform linear interpolation on the empty data for each location.
   5) Add up the results of linear interpolation with the white noise that has been generated
c. Convert the location coordinates into the UTM (Universal Transverse Mercator) Zone 50S coordinate system.
d. Construct a grid which includes the observed locations
e. Construct modified spatial weight matrices (Equation (2)) with different a and b values, inverse distance matrix (Equation (3)), and uniform (Equation (4)) based on the spatial order definition (Equation (1)).
f. Perform data centralization
g. Divide the data into training data (centred data from the 1st to 93rd order) and testing data (Centred data from 94th to 109th order)
h. If the training data is not stationary, continue doing the following until the data is stationary:
   1) If the average of the training data is constant over time, perform a differentiation on the data.
   2) If the average of the training data is not constant over time, use another model and the stage is complete.

Step 2: Identify the GSTAR model through the following stages.

a. Calculate the Space-Time Autocorrelation Function (STACF) from training data that has not been differentiated and has been differentiated until stationary for each weight matrix.
b. Calculate the Space-Time Partial Autocorrelation Function (STPACF) from undifferentiated and differentiated training data until stationary for each weight matrix.
c. Identify model candidates based on the movement of STACF and STPACF values with increasing time lag.

Step 3: Estimate parameters for model candidates for each weight matrix using the Least Squares method.

Step 4: Perform diagnostic tests on model candidates through the following stages.

a. Perform stationarity test for model candidates with Inverse of Autocovariance Matrix (IAcM) (Mukhaiyar & Pasaribu, 2012).
b. Perform independence test of the residuals from the candidate models using the Ljung-Box Test.
c. Perform a normality test with zero mean and constant variance on the residuals of the candidate models with the Kolmogorov-Smirnov Test.

Step 5: Select the best model through the following stages.

a. Calculate Mean Squared Error (MSE) for training data.
b. Calculate MSE for testing data.
c. Choose the best model with the smallest MSE value for testing data.

Step 6: Perform parameter significance test on the best model.

Step 7: Predict up to five weeks in the future using the best model.

Step 8: Create a contour map using ordinary kriging through the following stages.

a. Create a polygon which resembles the area to be interpolated using ordinary kriging.
b. Determine the semivariogram model which fits the experimental semivariogram of the data.
c. Perform ordinary kriging interpolation on the polygon which have been created to form a contour map.

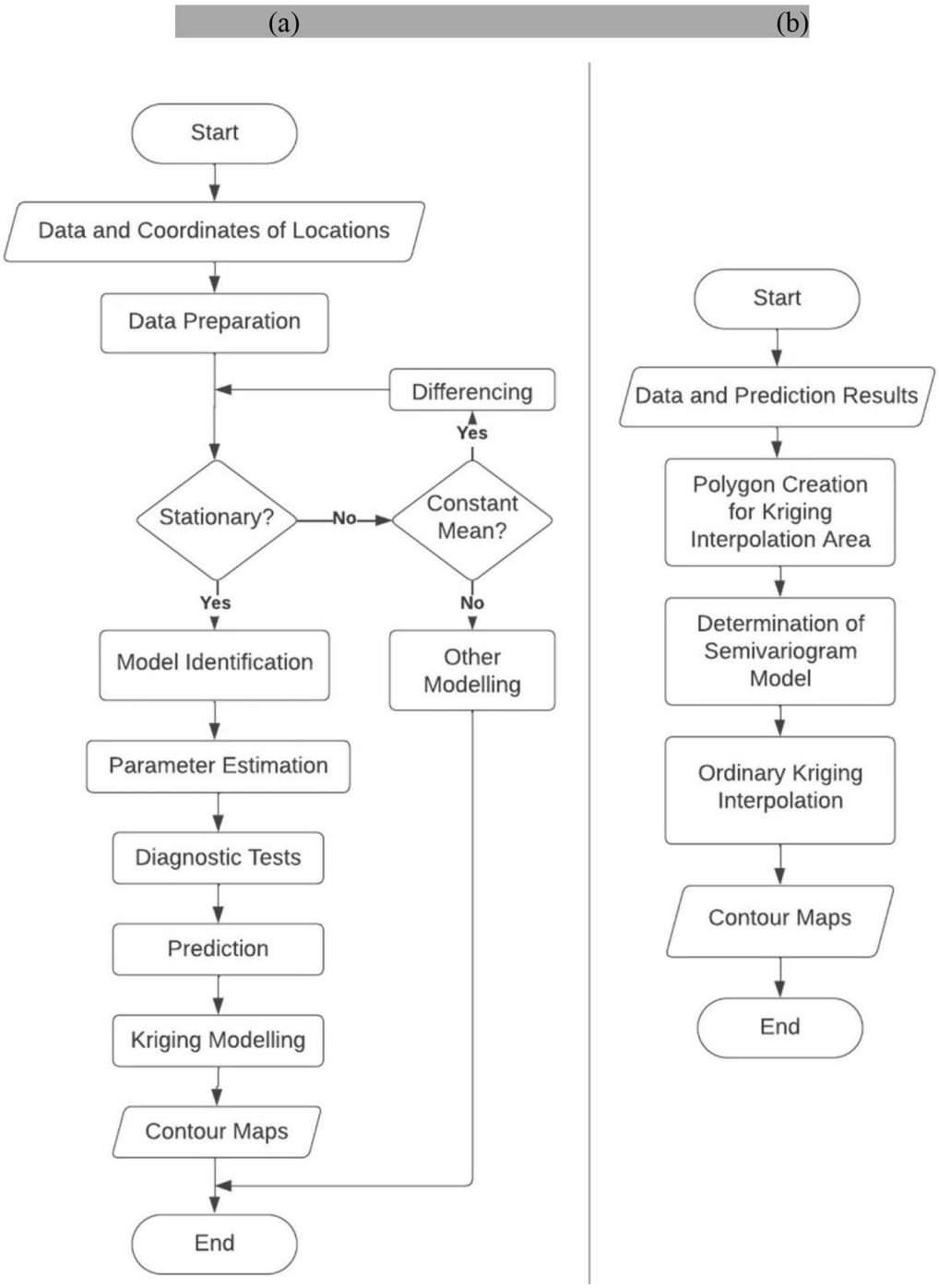

Fig. 2 Flowchart of GSTAR-Kriging modelling (**a**) and flowchart of kriging modelling in GSTAR-Kriging modelling (**b**)

Data

The data used in this paper are the average daily water level at peatland location points in Pulang Pisau Regency, Central Kalimantan Province from 20 February 2021 to 18 March 2023 with 7 days apart. The data were obtained from the Peatland and Mangrove Restoration Agency of Republic of Indonesia (BRGM Indonesia) via the prims.brg.go.id website. In addition, data on the distribution of peat thickness were also taken from the Indonesian Centre for Agricultural Land Resources Research and Development (BBSDLP) in 2019 from the same site. Information about the

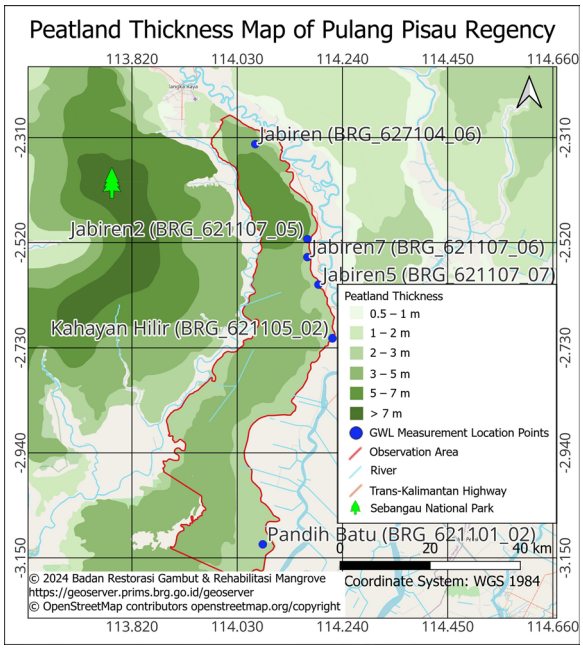

Fig. 3 Map of the observed groundwater level (GWL) measurement location points along with the distribution of peatland thickness according to BBSDLP in 2019

location points used is shown in Fig. 3 with the location names decided by the author. For water level, let locations 1, 2, 3, 4, 5, and 6, respectively, represent Jabiren, Jabiren2, Jabiren5, Jabiren7, Kahayan Hilir, and Pandih Batu. The position of the location points used along with their thickness level can also be seen in Fig. 3.

From the data obtained, it turns out that there are several dates with empty observation values. To overcome this problem, a linear interpolation was carried out to fill in the gaps as seen in Fig. 4. After that, the interpolation results were summed with white noise or the simulation results of a normally distributed data with zero mean and variance according to the data.

The combined numerical summary of the original data and interpolated results for water level is shown in Table 1. In Fig. 4, the water level tends to fall from November 2021 to August 2022 and then rises again as we enter October 2022 at all locations except Jabiren2. At Jabiren2, water level tends to fall from February 2022 to August 2022.

The longitude and latitude coordinates of each location were then converted according to the Universal Transverse Mercator (UTM) coordinate system, specifically in the 50S Zone. In addition, the elevation position of each location was obtained by subtracting the midpoint of the thickness range by 7 m (the largest number in the peat thickness range), so that a height position of 7 m was used as a reference point. Hence, with this definition, the coordinates of each location on the plane and the altitude position used were obtained, as shown in Table 2.

Results and discussion

The GSTAR modelling begins by defining the spatial lag order for the lag spatial weight matrix. For that purpose, we defined a grid as be shown in Fig. 5. To guarantee that each location has at least one neighbour location in its second spatial lag order, we use $C = 4$ for Eq. (1). It also can reduce the possibility of error in the parameter estimation process. In Fig. 5, each grid has a size of 7500m $\times$ 7500m.

The candidate models were identified using the STACF and STPACF [29]. The obtained models were GSTAR(2,0,1) and GSTAR(2;1,1) with modification and conventional weight matrices. After obtaining the candidate models from the model identification stage, parameter estimation and then diagnostic testing can be performed. For the diagnostic test, the Inverse of Autocovariance Matrix (IAcM) approach was used to test stationarity [23], the Ljung–Box test to examine the independent nature of the residuals, and the Kolmogorov–Smirnov test to check for normality with a zero mean of the residuals. For the Ljung–Box test and the Kolmogorov–Smirnov test, a significance level of 0.05 was used. The candidate models which passed the stationarity test and the independence test, be compared according to the Mean Squared Error (MSE) using the training data (in sample) and the testing data (out sample). In this paper, MSE in sample is defined as the average of squared difference between the training data and model estimation of the training data, while MSE out sample is defined as average of squared difference between the testing data and prediction from the training data. The MSE comparison between the GSTAR model candidates is shown in Table 3. The smaller the MSE, the better the model and the more valid the model is.

From Table 3, GSTAR(2;1,1) with a modified matrix $a = 0.1$ and $b = 1.1$ has the lowest MSE in sample value, while GSTAR(2;0,1) with a modified matrix $a = 0.1$ and $b = 1.1$ has the lowest MSE out sample value, indicating that GSTAR(2;1,1) with a modified matrix $a = 0.1$ and $b = 1.1$ is the best model in estimating the training data, while GSTAR(2;0,1) with a modified matrix $a = 0.1$ and $b = 1.1$ is the best model in predicting the testing data. Because the main objective is to estimate the future value of water level, GSTAR(2;0,1) with a modified matrix $a = 0.1$ and $b = 1.1$ was chosen as the best model for water level. This indicates that the water level is affected by the values of that location and the values of the neighbouring locations in the past. However, the difference in the altitude position between the locations only has

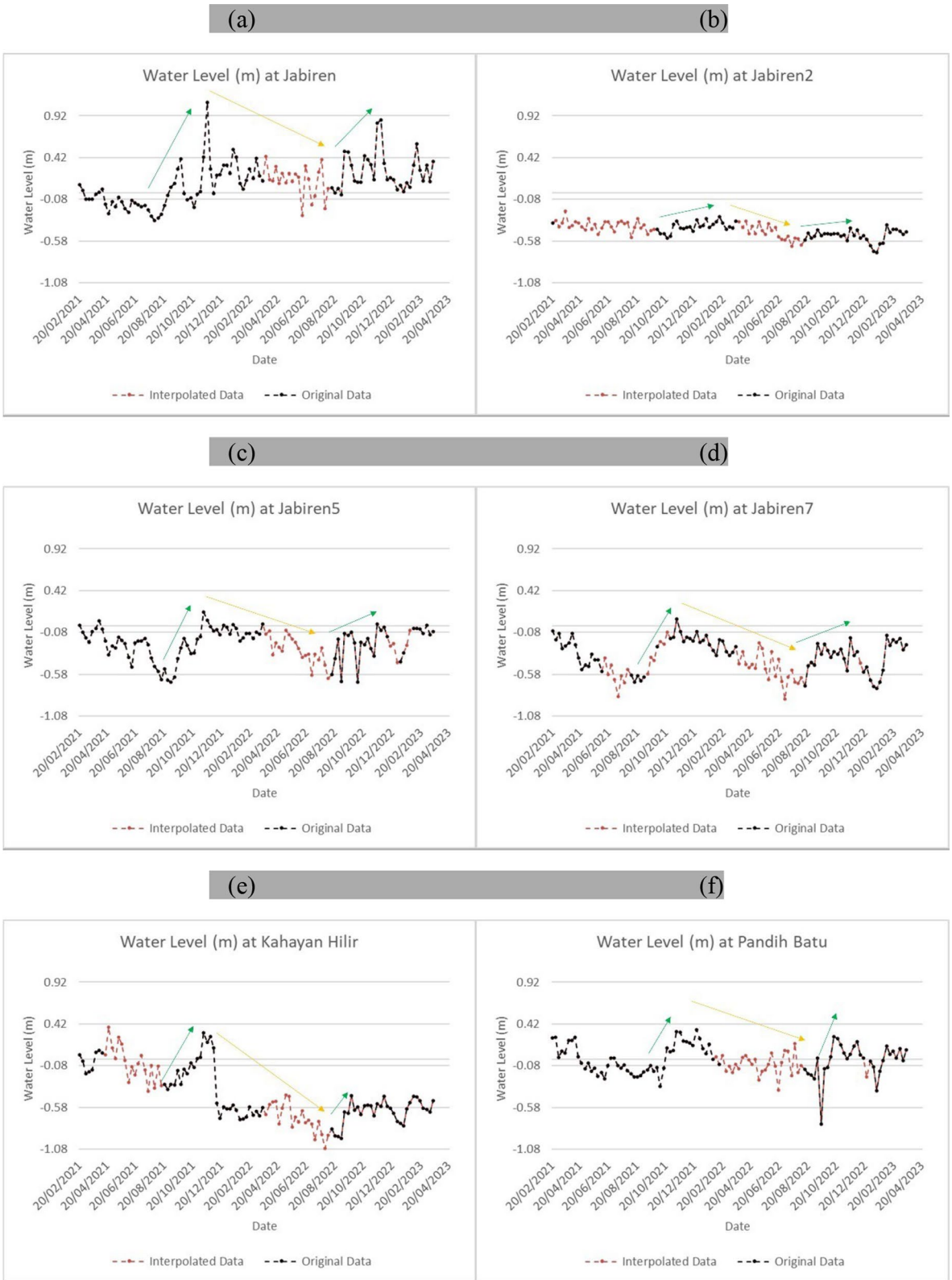

Fig. 4 Water level graphs in meters of the original data (black) and the interpolated results (red) for Jabiren (**a**), Jabiren2 (**b**), Jabiren5 (**c**), Jabiren7 (**d**), Kahayan Hilir (**e**), and Pandih Batu (**f**). The green arrows show an increasing trend, while the orange arrows show a decreasing trend

Table 1 Numerical summary of water level (m) at the observed location points

	Jabiren	Jabiren2	Jabiren5	Jabiren7	Kahayan Hilir	Pandih Batu
Sample size	109.000	109.000	109.000	109.000	109.000	109.000
Mean	0.123	−0.450	−0.219	−0.382	−0.424	−0.015
Standard deviation	0.249	0.089	0.188	0.201	0.335	0.174
Median	0.119	−0.442	−0.182	−0.356	−0.537	−0.019
Minimum	−0.335	−0.715	−0.672	−0.878	−1.073	−0.776
Maximum	1.079	−0.220	0.161	0.083	0.379	0.351
Skewness	0.814	−0.417	−0.736	−0.212	0.483	−0.549
Kurtosis	1.472	0.193	−0.138	−0.796	−0.743	2.053

Table 2 Coordinate of each location in meters (m)

Location name	Longitude (m)	Latitude (m)	Altitude (m)
Jabiren	173642.612	9742898.370	−3.0
Jabiren2	185264.310	9721895.187	−4.5
Jabiren5	187735.042	9711829.913	−4.5
Jabiren7	185273.007	9717911.155	−4.5
Kahayan Hilir	190877.919	9699995.884	−4.5
Pandih Batu	175527.045	9654360.647	−4.5

a significant effect on the weights for water level on the spatial weight matrix. The obtained models for water level in each location, which is marked with the dark blue for Jabiren, green for Jabiren2, brown for Jabiren5, grey for Jabiren7, purple for Kahayan Hilir, and blue for Pandih Batu, can be expressed as follows:

$$\hat{Y}_{1,t} = \mathbf{0.640}Y_{1,t-1} - 0.013Y_{1,t-2} - \mathbf{0.506}Y_{2,t-2}$$
$$\hat{Y}_{2,t} = 0.348Y_{2,t-1} + 0.253Y_{2,t-2} + 0.006Y_{1,t-2} + 0.017Y_{3,t-2} + 0.043Y_{4,t-2} + 0.008Y_{5,t-2}$$
$$\hat{Y}_{3,t} = \mathbf{0.521}Y_{3,t-1} + 0.118Y_{3,t-2} + 0.035Y_{2,t-2} + 0.056Y_{4,t-2} + 0.030Y_{5,t-2}$$
$$\hat{Y}_{4,t} = \mathbf{0.512}Y_{4,t-1} + \mathbf{0.384}Y_{4,t-2} - 0.129Y_{2,t-2} - 0.078Y_{3,t-2} - 0.027Y_{5,t-2}$$
$$\hat{Y}_{5,t} = \mathbf{0.714}Y_{5,t-1} + \mathbf{0.199}Y_{5,t-2} - 0.012Y_{2,t-2} - 0.021Y_{3,t-2} - 0.014Y_{4,t-2}$$
$$\hat{Y}_{6,t} = \mathbf{0.361}Y_{6,t-1} + \mathbf{0.182}Y_{6,t-2} + \mathbf{0.056}Y_{1,t-2} + \mathbf{0.056}Y_{2,t-2} + \mathbf{0.056}Y_{3,t-2} + \mathbf{0.056}Y_{4,t-2} + \mathbf{0.056}Y_{5,t-2}$$

In the above equation, black marks positive coefficients, red marks negative coefficients, and bold marks significant coefficients. The above equation indicates that water level at Jabiren 7 days prior ($Y_{1,t}$) contributes positively to the present water level at Jabiren ($\widehat{Y}_{1,t}$) with coeficient 0.640, while the 14 days prior water level at Jabiren ($Y_{1,t-2}$) and Jabiren2 ($Y_{2,t-2}$) contributes negatively to the present water level at Jabiren, with respective coefficients are −0.013 and −0.506. Overall, for each location $i = 1,2,\ldots,6$, the 7 days prior water level of respective location ($Y_{i,t-1}$) has the positive and biggest influence to the current water level ($Y_{i,t}$). As for the neighbour locations have less significant influences to its reference locations, except for Jabiren2 which has negative and positive contribute to Jabiren and Pandih Batu consecutively. As a note, all locations have smaller positive contribution to Pandih Batu. From those models, we suspect that Jabiren2 has important roles to predict the water level in Pulang Pisau Region.

After testing the significance of the parameters, predictions were made for the next 5 weeks using the best model. Predictions for water level are shown in Fig. 6. From the prediction results in Fig. 6, the water level tends to decrease at Jabiren, Jabiren7, and Pandih Batu, tends to increase at Jabiren2 and Kahayan Hilir, and tends to remain constant at Jabiren5. The results indicate that there is heterogeneity in the movement of the water level at various points of peatland locations in Pulang Pisau Regency. Besides that, the model is adequate in estimating the water level at Jabiren2 and Jabiren7. However, the model overestimates the water level at Jabiren5 and Kahayan Hilir and underestimates the water level at Jabiren and Pandih Batu. From Fig. 3, Jabiren2, Jabiren5, Jabiren7, and Kahayan Hilir are in proximity, with the shortest distance between locations

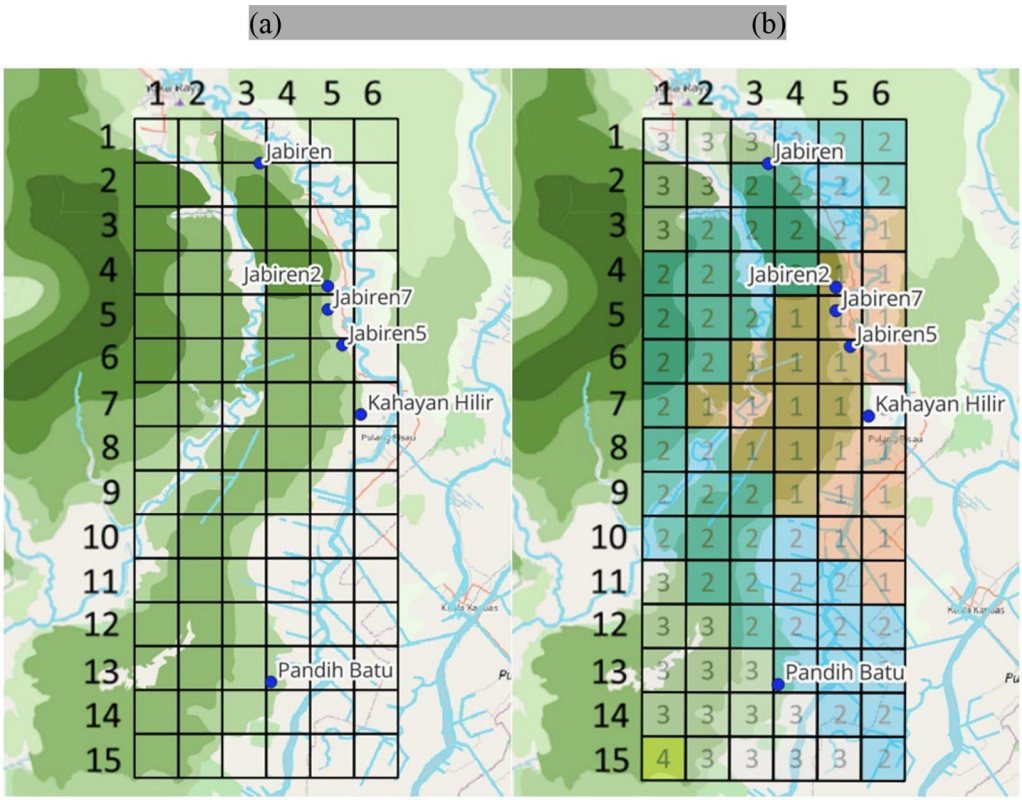

Fig. 5 Defining the 15×6 grid at the observed location points (**a**) and spatial lag order resulting from the development for $C = 4$ of reference location Kahayan Hilir (**b**)

Table 3 Selection of the best model for water level (no differencing)

GSTAR order	Spatial weight matrix		MSE in sample ($\times 10^{-4}$)	MSE out sample ($\times 10^{-4}$)
(2;0,1)	Modification	$a = 0.1, b = 1.1$	206.158	282.716
		$a = 0, b = \frac{(0.0001)^{-a}}{0.9999}$	206.164	282.783
	Conventional	Inverse distance	206.164	282.783
		Uniform	206.849	285.527
(2;1,1)	Modification	$a = 0.1, b = 1.1$	203.249	282.842
		$a = 0, b = \frac{(0.0001)^{-a}}{0.9999}$	203.254	282.979
	Conventional	Inverse distance	203.254	282.979
		Uniform	203.783	286.604

is the distance between Jabiren2 and Jabiren7, while Jabiren and Pandih Batu are far from other locations. There is a possibility that the goodness of the model at each location is related to the closeness of the location with other locations. From Fig. 6, the model estimation is adequate for the locations with the shortest distance to another location, overestimate for the locations with a distance to another location which is slightly further than the shortest distance, and underestimate for the locations with the longest distance to another location.

By obtaining the prediction results for each variable, we will see the interpolation of the observed values of location points that were not observed through ordinary kriging on the observation and prediction dates,

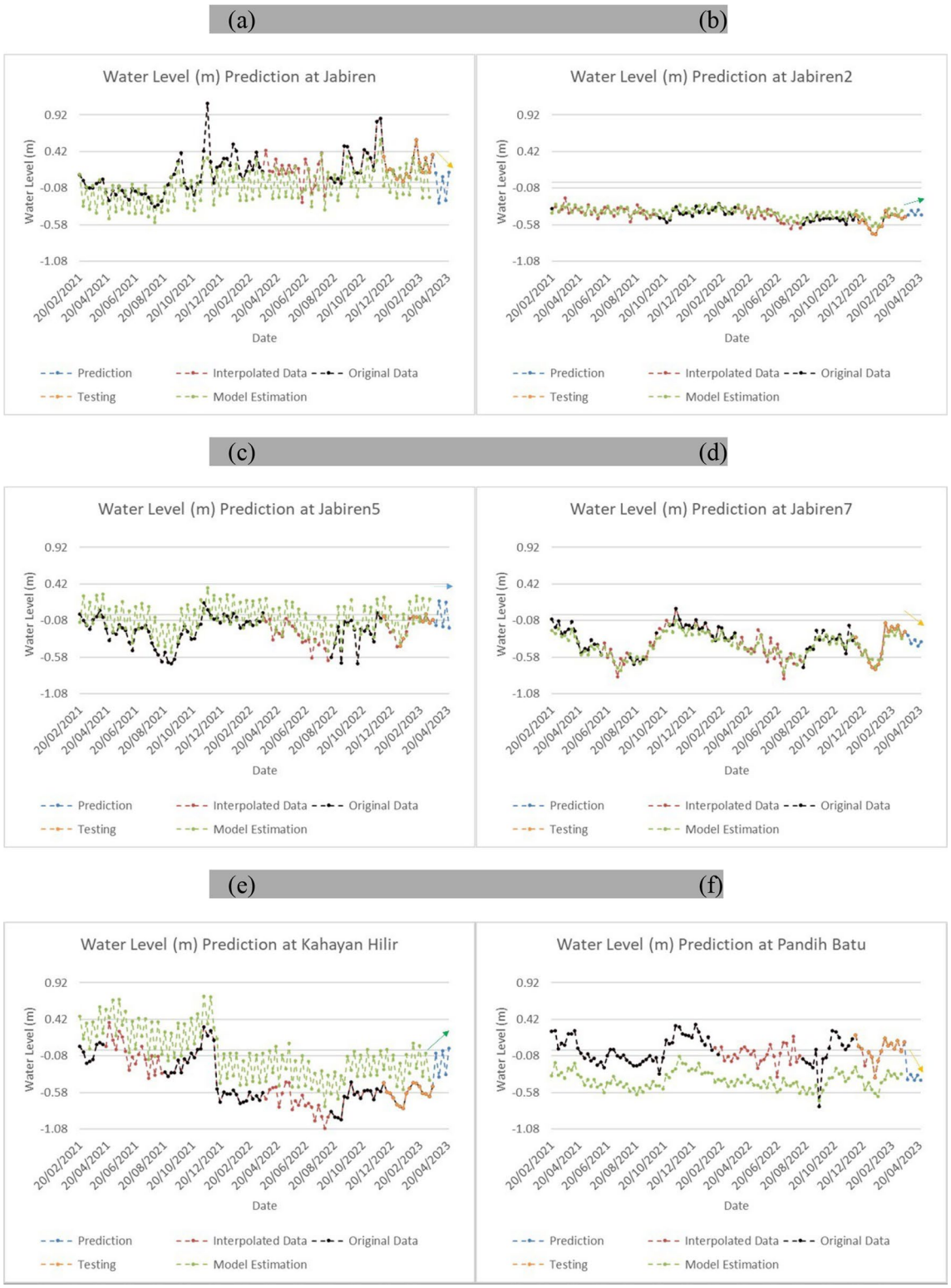

Fig. 6 Water level graphs in meters of the original data (training, black and testing, orange), the interpolation results (red), GSTAR(2;0,1) model with modified matrices $a = 0.1$and $b = 1.1$estimation results (green), and the prediction results (blue) for Jabiren (**a**), Jabiren2 (**b**), Jabiren5 (**c**), Jabiren7 (**d**), Kahayan Hilir (**e**), and Pandih Batu (**f**). The arrow at the end shows the trend of prediction. The green arrow shows an increasing trend, the orange arrow shows a decreasing trend, while the blue arrow shows no trend

Table 4 Ordinary Kriging parameters for water level with three neighbours

Date	Model	Maximum distance	Lag (h)	Nugget	Sill	Range	RMSE
04/03/2023	Linear	53134.665	5000	0.001	0.157	2.9×10^4	0.052
18/03/2023	Linear	53134.665	5000	0.000	0.151	3.0×10^4	0.044
01/04/2023	Exponential	88557.775	10000	0.002	0.054	0.326	0.029
15/04/2023	Linear to Sill	53134.665	10000	0.000	0.044	0.174	0.011

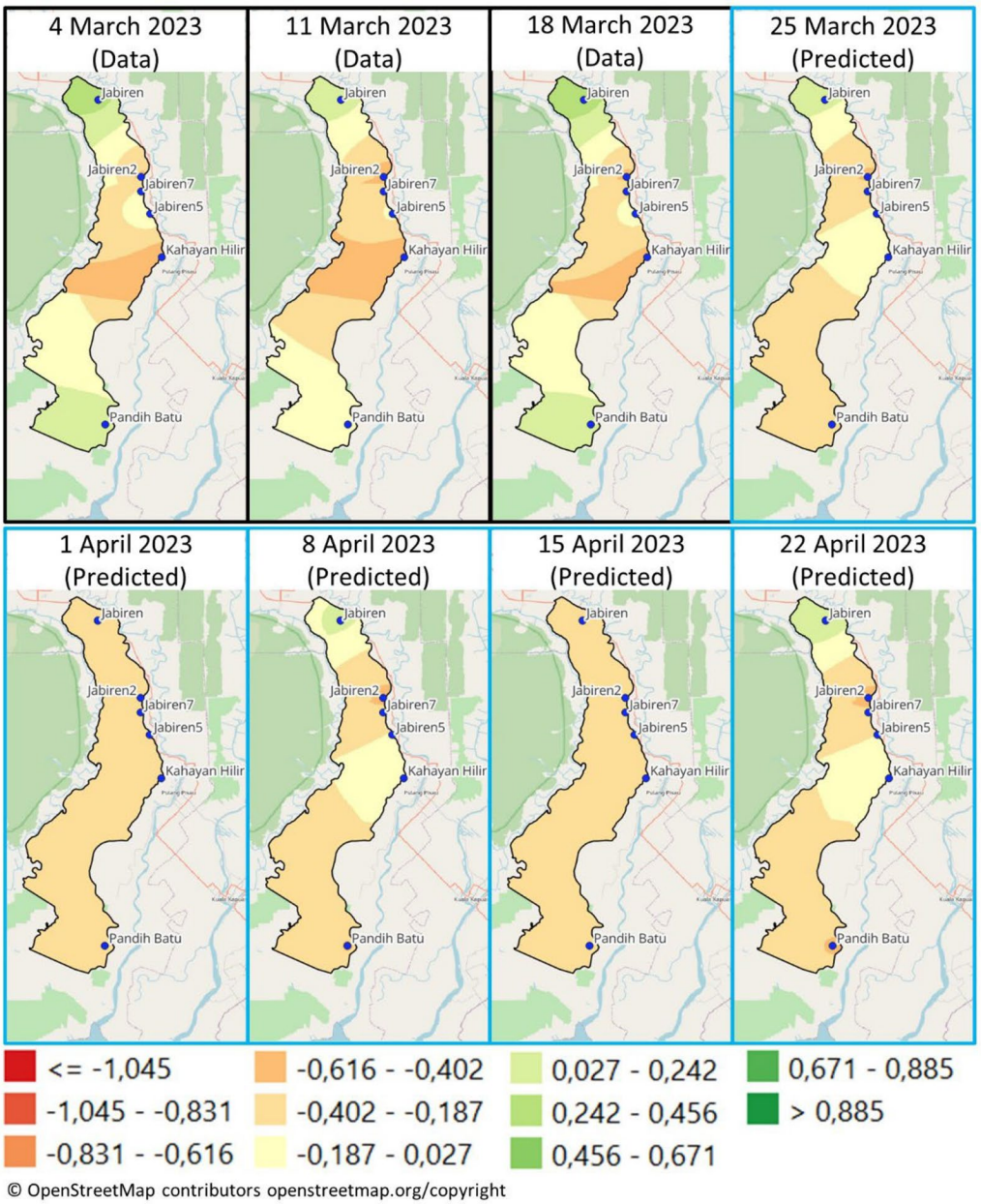

Fig. 7 Water level contour maps from data and predicted results on 4 March 2023, 11 March 2023, 18 March 2023, 25 March 2023, 1 April 2023, 8 April 2023, April 15, 2023, and 22 April 2023

namely from 4 March 2023 to 22 April 2023. Some set of parameters used in ordinary kriging for water level is shown in Table 4. Through ordinary kriging, the obtained contour maps are shown in Fig. 7.

From the contour maps made using the QGIS application, patterns of the water level can be obtained in March and April 2023. In Fig. 7, from 4 March 2023 to 25 March 2023, the water level in all areas tends to fall, while from 1 April 2023 to 22 April 2023, the water level in the areas around Jabiren and Kahayan Hilir fluctuates, while in other areas, it tends to be stable. Thus, the area in the middle is always the area with the lowest value indicating that the area is drier than the other areas. There is a possibility that this was caused by growing settlement development due to the existence of the Trans Kalimantan Road which connects Banjarmasin City and Palangka Raya City [1, 11]. This development may have caused the soil to harden which can cause drying and backfilling in several areas.

If the water level falls below 0.4 m below the ground surface, there can also be a change in the characteristics of the peatland [2]. In Fig. 7, the peatland area around Kahayan Hilir had a change in the characteristics from 4 March 2023 to 18 March 2023. The peatland around Jabiren2 had a change in the characteristics from 11 March 2023 to 22 April 2023. The peatland around Pandih Batu had a change in the characteristics at 22 April 2023.

Conclusion

In this paper, GSTAR modelling has been used to predict water level from peatlands in Pulang Pisau Regency, Central Kalimantan Province. Some conclusions can be written as follow:

1. The water level at almost every location has a downward trend pattern from around November to September and an upward trend pattern from October to November.
2. The best model for water level is GSTAR(2;0.1) with a modified matrix $a = 0.1$ and $b = 1.1$. It has been predicted that in the next 5 weeks, the water level will tend to decrease. In addition, in Pulang Pisau Regency, the area in the middle is the driest area.
3. Based on the results of the prediction of the water level, there is a risk of changes in the properties of the peatlands in several areas in Pulang Pisau Regency.

For further research, physical peat variables such as humidity and rainfall factors can be involved to measure the water level. Furthermore, these variables can be used to predict the risk of drought that will occur in the future.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12302-024-00979-6.

Supplementary material 1: Text S1. data collection and analysis methods.

Acknowledgements
This research is a collaboration between the Ministry of Education, Culture, Research and Technology and the Education Fund Management Institute (LPDP) through the RISPRO UKICIS Funding Program. The authors would like to thank Prof. David Large, Dr. Matteo Icardi, and Dr. Bagus Muljadi from University of Nottingham, UK, as the partners in UKICIS program. The authors would also like to thank to all reviewers who have read and commented on our manuscript to make it better.

Author contributions
U.M., A.W.M., and T.P. constructed the models, data analysed, interpreted the results, and wrote the main manuscript text; U.S.P., K.N.S., and S.W.I. evaluated the space–time, spatial kriging model; I.S., D.N.C., I.I., D. R., and D.T.Q. gave insights in the environmental and risk aspect; D.N.C. and D. R. did survey to the field for data confirmation; All authors reviewed the manuscript.

Funding
This work was funded by LPDP RISPRO-UKICIS (United Kingdom-Indonesia Consortium for Interdisciplinary Sciences) green economy under Grant No. 4345/E4/AL.04. It was also partially funded by FMIPA ITB through PPMI FMIPA 2023 and LPPM ITB through Riset Unggulan 2024 (contract number: 959/IT1.B07.1/TA.00/2024).

Data availability
Sequence data that support the findings of this study were obtained from the Peatland and Mangrove Restoration Agency of Republic of Indonesia (BRGM Indonesia) via the prims.brg.go.id website (open accessed). The water level data are openly available at https://bit.ly/PulangPisauWaterLevel.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
All authors declared that they have no competing interests.

Received: 17 April 2024 Accepted: 16 August 2024
Published online: 14 October 2024

References

1. Alamgir M, Campbell MJ, Sloan S, Suhardiman A, Supriatna J, Laurance WF (2019) High-risk infrastructure projects pose imminent threats to forests in Indonesian Borneo. Sci Rep 9(1):140. https://doi.org/10.1038/s41598-018-36594-8
2. Barani AM, Dariah A, Suryotomo AP, Mulyani A, Apriyanto A, Hidayat A, Sumawinata B, Kartiwa B, Taniwiryono D, Sadono D, Fahamsyah E, Widiastuti H, Hermantoro, Pulunggono HB, Ismail I (2022) Gambut, Sawit, dan Lingkungan. Bogor: IPB Press. ISBN: 978-623-256-855-6, eISBN: 978-623-467-377-7
3. Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J Roy Stat Soc Ser B (Methodol) 36(2):192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x
4. Borovkova S, Lopuhaä HP, Ruchjana BN (2008) Consistency and asymptotic normality of least squares estimators in generalized STAR models. Stat Neerl 62(4):482–508. https://doi.org/10.1111/j.1467-9574.2008.00391.x

5. Cobb AR, Harvey CF (2019) Scalar simulation and parameterization of water table dynamics in tropical peatlands. Water Resour Res 55(11):9351–9377. https://doi.org/10.1029/2019WR025411
6. Hooijer A, Silvius MJ, Woesten HAB, Page SE (2006) PEAT-CO2. Assessment of CO2 emissions from drained peatlands in SE Asia. https://api.semanticscholar.org/CorpusID:201308974
7. Horton AJ, Virkki V, Lounela A, Miettinen J, Alibakhshi S, Kummu M (2021) Identifying key drivers of peatland fires across Kalimantan's ex-mega rice project using machine learning. Earth Space Sci 8(12):e2021EA001873. https://doi.org/10.1029/2021EA001873
8. Huda NM, Imro'ah N (2023) Determination of the best weight matrix for the Generalized Space Time Autoregressive (GSTAR) model in the Covid-19 case on Java Island, Indonesia. Sp Stat 54:100734. https://doi.org/10.1016/j.spasta.2023.100734
9. Huijnen V, Wooster MJ, Kaiser JW, Gaveau DLA, Flemming J, Parrington M, Inness A, Murdiyarso D, Main B, van Weele M (2016) Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. Sci Rep 6(1):26886. https://doi.org/10.1038/srep26886
10. Ilmi NFF, Mukhaiyar U, Fahmi F (2015) The generalized STAR(1,1) modeling with time correlated errors to red-chili weekly prices of some traditional markets in Bandung, West Java. AIP Conf Proc 1692(1):020014. https://doi.org/10.1063/1.4936442
11. Januar R, Sari ENN, Putra S (2021) Dynamics of local governance: the case of peatland restoration in Central Kalimantan, Indonesia. Land Use Policy 102:105270. https://doi.org/10.1016/j.landusepol.2020.105270
12. Jonathan R, Yundari N, Nada OYE (2021) Application of GSTAR(1,1) model for layer peat soil predicted based on resistivity log data. J Phys: Conf Ser 2106(1):012031. https://doi.org/10.1088/1742-6596/2106/1/012031
13. Kurnianto S, Warren M, Talbot J, Kauffman B, Murdiyarso D, Frolking S (2015) Carbon accumulation of tropical peatlands over millennia: a modeling approach. Glob Change Biol 21(1):431–444. https://doi.org/10.1111/gcb.12672
14. Lewis C, Albertson J, Xu X, Kiely G (2012) Spatial variability of hydraulic conductivity and bulk density along a blanket peatland hillslope. Hydrol Process 26(10):1527–1537. https://doi.org/10.1002/hyp.8252
15. Mahdiyasa AW, Large DJ, Icardi M, Muljadi BP (2023) MPeat2D – a fully coupled mechanical-ecohydrological model of peatland development in two dimensions. EGUsphere 2023:1–31. https://doi.org/10.5194/egusphere-2023-2535
16. Mahdiyasa AW, Large DJ, Muljadi BP, Icardi M (2023) Modelling the influence of mechanical-ecohydrological feedback on the nonlinear dynamics of peatlands. Ecol Model 478:110299. https://doi.org/10.1016/j.ecolmodel.2023.110299
17. Mahdiyasa AW, Large DJ, Muljadi BP, Icardi M, Triantafyllou S (2022) MPeat—A fully coupled mechanical-ecohydrological model of peatland development. Ecohydrology 15(1):e2361. https://doi.org/10.1002/eco.2361
18. Masteriana D, Riani MI, Mukhaiyar U (2019) Generalized STAR (1;1) Model with Outlier—Case Study of Begal in Medan, North Sumatera. J Phys: Conf Ser 1245(1):012046. https://doi.org/10.1088/1742-6596/1245/1/012046
19. Miettinen J, Shi C, Liew SC (2017) Fire distribution in peninsular malaysia, sumatra and borneo in 2015 with special emphasis on peatland fires. Environ Manage 60(4):747–757. https://doi.org/10.1007/s00267-017-0911-7
20. Mimis A (2016) 3D weight matrices in modeling real estate prices. Int Arch Photogramm Remote Sens Sp Inform Sci XLII-2/W2:123–125. https://doi.org/10.5194/isprs-archives-XLII-2-W2-123-2016
21. Mukhaiyar U (2015) The goodness of generalized STAR in spatial dependency observations modeling. AIP Conf Proc 1692(1):020008. https://doi.org/10.1063/1.4936436
22. Mukhaiyar U, Huda NM, Sari KN, Pasaribu US (2020) Analysis of generalized space time autoregressive with exogenous variable (GSTARX) model with outlier factor. J Phys Conf Ser 1496(1):012004. https://doi.org/10.1088/1742-6596/1496/1/012004
23. Mukhaiyar U, Pasaribu US (2012) A new procedure for generalized STAR modeling using IAcM approach. ITB J Sci 44(2):179–192. https://doi.org/10.5614/itbj.sci.2012.44.2.7
24. Mukhaiyar U, Bilad BI, Pasaribu US (2021) The Generalized STAR Modelling with Minimum Spanning Tree Approach of Weight Matrix for COVID-19 Case in Java Island. J Phys: Conf Ser 2084(1):012003. https://doi.org/10.1088/1742-6596/2084/1/012003
25. Mukhaiyar U, Nabilah FT, Pasaribu US, Huda NM (2022) The Space-Time Autoregressive Modelling with Time Correlated Errors for The Number of Vehicles in Purbaleunyi Toll Gates. J Phys: Conf Ser 2243(1):012068. https://doi.org/10.1088/1742-6596/2243/1/012068
26. Nurhayati N, Pasaribu US, Neswan O (2012) Application of generalized space-time autoregressive model on GDP data in West European Countries. J Probab Stat 2012(1):867056. https://doi.org/10.1155/2012/867056
27. Pasaribu US, Mukhaiyar U, Huda NM, Sari KN, Indratno SW (2021) Modelling COVID-19 growth cases of provinces in java Island by modified spatial weight matrix GSTAR through railroad passenger's mobility. Heliyon 7(2):e06025. https://doi.org/10.1016/j.heliyon.2021.e06025
28. Pasaribu US, Mukhaiyar U, Heriawan MN, Yundari (2022) Generalized Space-Time Autoregressive Modeling of the Vertical Distribution of Copper and Gold Grades with a Porphyry-Deposit Case Study. Int J Adv Sci Eng Inf Technol 12(5). https://doi.org/10.18517/ijaseit.12.5.14835
29. Pfeifer PE, Deutsch SJ (1980) A three-stage iterative procedure for space-time modeling. Technometrics 22(1):35–47. https://doi.org/10.2307/1268381
30. Pramoedyo H, Ashari A, Fadliana A (2020) Application of GSTAR kriging model in forecasting and mapping coffee berry borer attack in Probolinggo district. J Phys: Conf Ser 1563(1):012005. https://doi.org/10.1088/1742-6596/1563/1/012005
31. Salsabila AB, Ruchjana BN, Abdullah AS (2024) Development of the GSTARIMA(1,1,1) model order for climate data forecasting. Int J Data Netw Sci 8:773–788. https://doi.org/10.5267/j.ijdns.2024.1.001
32. Tamea S, Muneepeerakul R, Laio F, Ridolfi L, Rodriguez-Iturbe I (2010) Stochastic description of water table fluctuations in wetlands. Geophys Res Lett. https://doi.org/10.1029/2009GL041633
33. Taufik M, Veldhuizen AA, Wösten JHM, van Lanen HAJ (2019) Exploration of the importance of physical properties of Indonesian peatlands to assess critical groundwater table depths, associated drought and fire hazard. Geoderma 347:160–169. https://doi.org/10.1016/j.geoderma.2019.04.001
34. Utami R, Nurhayati N, Maryani S (2021) Forecasting the amount of rainfall in West Kalimantan using Generalized Space-time Autoregressive model. IOP Conf Ser: Earth Environ Sci 746(1):012035. https://doi.org/10.1088/1755-1315/746/1/012035
35. Young DM, Baird AJ, Morris PJ, Dargie GC, MampouyaWenina YE, Mbemba M, Boom A, Cook P, Betts R, Burke E, Bocko YE, Chadburn S, Crabtree DE, Crezee B, Ewango CEN, Garcin Y, Georgiou S, Girkin NT, Gulliver P, Lewis SL (2023) Simulating carbon accumulation and loss in the central Congo peatlands. Glob Change Biol 29(23):6812–6827. https://doi.org/10.1111/gcb.16966
36. Yundari Y, Pasaribu US, Mukhaiyar U (2017) Error Assumptions on Generalized STAR Model. J Math Fundam Sci 49(2):136–155
37. Yundari, Pasaribu US, Mukhaiyar U, Heriawan MN (2018) Spatial weight determination of GSTAR(1;1) model by using Kernel function. J Phys Conf Ser 1028:012223. https://doi.org/10.1088/1742-6596/1028/1/012223

Publisher's Note