Web-based geographic information system for flood disaster using spatial autoregressive method

M A Nurrijal, A Helen, and A S Abdullah

Department of Computer Science, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia

*atje.setiawan@unpad.ac.id

Abstract. The territory of Indonesia is classified as one of the countries prone to disasters, both natural disasters and disasters caused by human activities. Flood disaster is one example of a disaster that often occurs, especially in the DKI Jakarta area. The common cause of the flooding is usually caused by high rainfall exceeding the average in general so that water sources such as rivers, reservoirs, and lakes cannot accommodate the amount of rainwater. This is also supported because the water absorption system is minimal so that it is unable to absorb excess water, causing flooding. The method used to define the relationship between various factors and the relationship between regions is to use the Spatial Autoregressive (SAR) method. In order to get a better relationship between regions, the existing spatial data will be visualized with the Geographical Information System. This research focuses on creating a web-based Geographic Information System application. Applications created using the programming language PHP and Javascript with the Codeigniter framework. The results obtained by the SAR equation to predict the height of flooding in the DKI Jakarta Province and the test results using the Mean Absolute Deviation (MAD) were 1.354656% while using the Mean Absolute Percentage Error (MAPE) was 2.157288%. Testing applications using the black box testing method was obtained by 93.35%. This shows that the SAR Model is suitable for predicting flood heights in the DKI Jakarta Province. This flood prediction can be used one year from the prediction model made.

1. Introduction
The territory of Indonesia is classified as one of the countries prone to disasters, both natural disasters and disasters caused by human activities. Flood disaster is an example of a disaster that can be caused by natural or human factors. Flood disasters are events that can occur at any time and often result in losses in the form of damage to buildings, loss of valuables, to losses that result in being unable to go to work and attend school [1]. The factors that cause flood disasters include excessive exploitation of groundwater which results in high ground levels, the number of pollutants in channels and water reservoirs such as rivers or reservoirs which result in unable to accommodate large volumes of water from extreme rainfall [2].

One of the methods used in defining the relationship of various factors is by means of regression analysis. Spatial linkages between regions that need attention because they are linked to location / region. The spatial relationship between these regions tends to occur frequently because of their close proximity to each other. To determine the effect of the relationship between existing locations, spatial analysis is required. In order to better see the relationship between regions, the existing spatial data
will be visualized with a geographic information system. The location that will be the research area is DKI Jakarta province where there are 43 sub-districts [3].

2. Method
2.1 Geographic information system
A geographic information system is a computer-based system or technology that was built with the aim of collecting, storing, processing, and analyzing, as well as presenting data and information from an object or phenomenon related to its location on earth [4].

2.2. Spatial weighted matrix
The spatial influence between locations in the model is formed in a weighted matrix $W$ with size data $n \times n$ [5].

$$
W = \begin{pmatrix}
W_{11} & \cdots & W_{1n} \\
\vdots & \ddots & \vdots \\
W_{n1} & \cdots & W_{nn}
\end{pmatrix}.
$$

(1)

The types of matrices used for spatial weighting matrices are sided contact (rook contiguity), corner contact (bishop contiguity), and side-corner intersection (queen contiguity). The spatial weighting matrix is used to determine the weight between observed locations based on the neighborhood relationship between locations. The diagonal element in the weighting matrix $W$ is given a weight of 0,

$$
w_{ij} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ are neighbours} \\
0 & \text{for non-neighbour.}
\end{cases}
$$

(2)

2.3. Ordinary Least Squares Method (OLS)
Regression analysis is an analysis to obtain the relationship between the dependent variable ($Y$) and the independent variable ($X$). The relationship between one dependent variable and one or more can be expressed by a linear regression model [6].

$$
\beta = (X^T X)^{-1} X^T (I - \rho W) Y
$$

(3)

$\beta$ is the parameter estimation value, $X$ is the independent variable matrix with size $n \times (k + 1)$, $I$ is the identity matrix, $\rho$ is the spatial lag coefficient parameter of the dependent variable, $W$ is the form of the spatial weighting matrix, and $Y$ is the dependent variable matrix with size $n \times 1$.

2.4. Spatial Autoregressive (SAR) model
The Spatial Autoregressive (SAR) model is a model that combines a simple regression model with spatial lag on the dependent variable using a cross-section [7].

$$
Y = \rho Wy + X\beta + \epsilon
$$

(4)

$$
\epsilon \sim N (0, \sigma^2 I)
$$

$Y$ is the dependent variable vector with size $n \times 1$, $\rho$ is the parameter of the dependent variable lag spatial coefficient, $W$ is the weighting matrix with size $n \times n$, $X$ is the independent variable matrix with size $n \times (k + 1)$, $\beta$ as the coefficient vector of the regression parameter with size $(k + 1) \times 1$, and $\epsilon$ is an error vector with size $n \times 1$. 
2.5. Mean absolute deviation (MAD)
A Mean Absolute Deviation (MAD) is a method for evaluating the method of forecasting using the sum of the difference between the actual and predicted values. The smaller the MAD value, the better the model will be. The MAD method measures the accuracy of the forecast by the average error [8].

\[
MAD = \frac{1}{n} \sum_{t=1}^{n} \left| A_t - F_t \right|
\]  

where \( A_t \) for actual value, \( F_t \) is predicted value, and \( n \) is sample size.

2.6. Mean absolute percentage error (MAPE)
Mean Absolute Percentage Error (MAPE) is a statistical measure to calculate how accurate a forecasting method is. In the MAPE method, the measure of accuracy is measured as a percentage, and can be calculated as the mean absolute percent error for each time period minus the actual value and divided by the actual value. The model is said to have very good forecasting ability if it has a MAPE value of less than 10% [9].

\[
MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|
\]

where \( y_t \) for actual value, \( \hat{y}_t \) is predicted value, and \( n \) is sample size.

2.7. Black box testing
Testing the system using the black box method is a high-level design evaluation stage with specifications according to user needs. Black box testing is carried out to ensure the system can do what must be done based on the scenario that has been created [10].

2.8. The step of processing data
In this study, the data used are geographic data and flood events that occurred in the DKI Jakarta Province in January 2020. The steps taken in model building and implementation are as follows:
1. The digital map design uses the administrative map of DKI Jakarta Province
2. The formation of a spatial weighted matrix using queen contiguity
3. The calculation of the estimated value of the SAR model parameters uses the least squares method
4. Establishment of the SAR model
5. Implementation of digital maps and SAR models on the website
6. Prediction accuracy testing uses MAD and MAPE
7. Testing applications on users using black box testing

3. Result and Discussion
3.1. Map design
In designing the map in this geographic information system, ARCGIS software is used for making the map. Referring to the administrative map for the DKI Jakarta province in 2018, the thematic map for the DKI Jakarta province after being created using ARCGIS can be seen in Figure 1.
3.2. The SAR model for flood data in DKI Jakarta

In this study, to determine the spatial weighted matrix (W), information is needed about the proximity between regions and distances. The relationship of neighbours reflects one location to another, while distance is information obtained from latitude and longitude. The following is the spatial weighting matrix of the initial 10 neighbouring regions:

\[
\begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

To find out the estimated parameter values from the data in Table 1, the least squares method is needed to get the results. The results of calculations using equation (3) are as follows:

\[
\hat{\beta}_0 = -32.59133 \\
\hat{\beta}_1 = 0.125487 \\
\hat{\beta}_2 = -0.46718 \\
\hat{\beta}_3 = 1.19343 \\
\hat{\beta}_4 = -0.774542.
\]

Using the estimator of the ordinary least squares method, we have the SAR estimation model as follow:

\[
\hat{Y} = -32.59133 - 0.00432 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} y_j + 0.125487 x_1 - 0.46718 x_2 + 1.19343 x_3 - 0.774542 x_4.
\]

Based on the previously calculated model it can be interpreted as follows:

1. The constant value is -32.59133, meaning that if the value of the other factor variable is 0, then the flood height is negative, which is -32.59133 cm.
2. If the rainfall value (X_1) in a district increases by one unit, and other factors are constant or constant, it can increase the flood height by 0.12487 cm.
3. If the area height value \((X_2)\) in a district increases by one unit and other factors are constant or constant, it can reduce the flood height by 0.046718 cm.

4. If the value of the level of pollution \((X_3)\) in a sub-district increases by one unit and other factors remain or constant, it can increase the flood height by 1.19343 cm.

5. If the value of the water absorption rate \((X_4)\) in a district increases by one unit and other factors remain or constant, it can reduce the flood height by 0.774542 cm.

3.3. Implementation on website

After designing the digital map and the model, both are implemented on the website where the map will be created using Java script and the data will be processed using PHP. The colours for determining the flood height range from green indicating low flood height to dark red for high flood height.

![Figure 2. Implementation of the predicted results](image)

3.4. Prediction accuracy of testing uses MAD and MAPE

In testing the accuracy of the prediction results, this application will be tested for its predictive ability using testing data. The testing data used to test the accuracy of the prediction results is the geographical data of the DKI Jakarta province in April 2020.

In this test, it can be seen how much the similarity between the predicted \(\hat{Y}\) value and the actual \(Y\) value. To find out this, the Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) methods are used to test the accuracy of the prediction results. The calculation for MAD uses equation 6, while for MAPE uses equation 7. The actual data uses January data, while the prediction data uses April data. In Table 1, you can see the results of testing using MAD and MAPE using testing data.

|          | MAD         | MAPE        |
|----------|-------------|-------------|
| Table 1  | 1.354656    | 2.157288    |

Based on the results of the accuracy test in Table 1, it shows that the Spatial Autoregressive model is able to predict the flood height with an error percentage level of 1.354656% based on the MAD test method and 2.157288% based on the MAPE test method. It can be seen that this model has the ability...
to predict the value of $\hat{Y}$ with a very minimal level of error. It is also in accordance with the theory if the MAD value is smaller and if the MAPE value is less than 10% then the model can forecast well. In other words, the model has high accuracy.

3.5. Testing applications on users using black box testing

In testing applications on a web-based Geographical Information System for this flood disaster using the black box testing method. During the testing phase, this application will have two conditions, namely "success (S)" if the user can run the scenario, and "fail (F)" if the user cannot run the scenario.

| No | Testing Scenarios | Expected Results | Test Results |
|----|-------------------|------------------|--------------|
| 1  | Run the application program | User can open the start page of the application | 100% |
| 2  | View geographic data pages | Users can see geographic data that has been provided on the application | 100% |
| 3  | Perform data input and submit for predictive results | The user successfully inputs data and makes predictions against the data that has been entered | 86.7% 13.3% |
| 4  | See the prediction results | The user sees the prediction results | 86.7% 13.3% |

4. Conclusion

Based on the results of analysis and testing on the accuracy and feasibility of the application, this model can be used to predict the flood height in the DKI Jakarta province. Because the test results show the accuracy of the prediction results using MAD is 98.6453%, while using the MAPE method is 97.8427%. This means that the Spatial Autoregressive model is feasible to be used to predict flood heights. While the test results based on black box testing showed an average percentage of success reached 93.35%. This means that this application is in accordance with user needs. For this flood prediction can be used one year from the prediction model made. This is because the geographic data of an area can change from year to year.

Acknowledgments

The authors thank Rector Universitas Padjadjaran who provided financial assistance for the dissemination of research Academic Leadership Grant 2020 contract number: 1427/UN6.3.1/LT/2020. Also thanks to the reviewers for the valuable review of this paper.

References

[1] Findayani A 2015 Community Preparedness in Flood Management in the City of Semarang Journal of Geography 12 5561
[2] Adiyoso W 2020 Big Flood in Jakarta (Jakarta:The Conversation)
[3] Ahmar, Saleh A, Adiatama, Aidid and Kasim M 2013 Crime Modeling using Spatial Regression Approach Journal of Physics Conference Series 954 012013
[4] Ekadinata A, Dewi S, Hadi D P, Nugroho D K and Johana F 2008 Geographical Information
System for Management of Landscapes Based on Natural Resources (Bogor: World Agroforestry Centre)

[5] Anselin L 1998 *Spatial Econometrics: Methods and Models* (Dordrecht: Kluwer)

[6] Mawar S 2016 Spatial Regression Model on the Number of Unemployment in South Sulawesi Province with Spatial Dependency Examination through Moran's I Test *Unpublished Script* (Makassar: Alauddin State Islamic University)

[7] Gaitan C, and Guyon X 2010 *Spatial Statistics and Modeling* (New York: Springer)

[8] Lamusa F 2017 Forecasting the Number of Passengers at PT. Angkasa Pura (PERSERO) Sultan Hasanuddin International Airport Branch Office Makassar Using the Holt-Winters Exponential Smoothing Method *Unpublished Script* (Makassar: Alauddin State Islamic University)

[9] Margi K, and Pendawa S 2015 Analysis and Application of the Single Exponential Smoothing Method for Prediction of Sales in Certain Periods *Prosiding Snatif* 2 349

[10] Larrea M 2017 Black-Box Testing Technique for Information Visualization. Sequencing Constraints with Low-Level Interactions *Journal of Computer Science and Technology* 17 1