www.czasopisma.pan.pl PAN www.journals.pan.pl

ADVANCES IN GEODESY AND GEOINFORMATION
(formerly GEODESY AND CARTOGRAPHY)
Vol. 71, no. 2, article no. e24, 2022

https://doi.org/10.24425/agg.2022.141298

Original paper
Part of Topical collection:
"Advancements in Applied Geoinformatics"

Clustering analysis of the seismicity of Van Province and its surroundings via spatial autocorrelation techniques filters

Güzide Miray Perihanoglu[1]*, Ömer Bilginer[2], Elif Akyel[2]

[1]Van Yüzüncü Yıl University, Van, Turkey,
e-mail: gm.perihanoglu@yyu.edu.tr; ORCID: http://orcid.org/0000-0002-2730-590X

[2]Izmir Katip Çelebi University, Izmir, Turkey,
e-mail: omer.bilginer@ikcu.edu.tr; ORCID: http://orcid.org/0000-0001-5789-7929
e-mail: elif.akyel@ikcu.edu.tr; ORCID: http://orcid.org/0000-0002-9355-7478

*Corresponding author: Güzide Miray Perihanoglu, e-mail: gm.perihanoglu@yyu.edu.tr

Received: 2022-02-27 / Accepted: 2022-04-27

Abstract: Destructive aftershocks such as the Mw 7.2 Van earthquake on October 23, 2011, and the Hoy (Iran) earthquake with Mw 5.9 on February 23, 2020, occurred in the province of Van and its surroundings. In earthquake studies, the issue of examining the distribution and homogeneity of earthquake incidences with Geographic Information Systems (GIS) based via spatial autocorrelation techniques is frequently investigated. Van province and its surroundings are among the areas with high earthquake risk due to its location on the East Anatolian Compressive Tectonic Block. The aim of this study is to analyze the spatial patterns of earthquakes with magnitude Mw 4 and above that occurred in the province of Van and its surroundings during the instrumental period and to determine to cluster. Spatial cluster analyses play an important role in examining the distribution of seismicity. The data used in the study have been taken from the database system of the Earthquake Department of the Republic of Turkey Ministry of Interior Disaster and Emergency Management Presidency. Moran's I and Getis-Ord Gi methods from spatial autocorrelation techniques were preferred on the earthquake data set to be used in this research. It has aimed to determine the dangerous areas by testing the earthquake distributions in clustered regions via spatial autocorrelation techniques.

Keywords: earthquake, geographical information system (GIS), spatial autocorrelation, Van Province

1. Introduction

The city of Van is one of the most important urban centers in the Eastern Anatolian Region of Turkey. The Van earthquake region covers the area between 37.50°–40.00° N latitudes and 41.00°–45.00° E longitudes. Since 1900, 713 $M \geq 4.0$ earthquakes, the largest of which is 7.6, have occurred in the region. In addition, the earthquake records of 23 historical periods belonging to the mentioned region before 1900 are available. The Van-Erciş-centered earthquake that occurred on 23 October 2011 and the Van-Edremit-centered earthquake on 9 November 201 were felt very strongly in Van and its districts, causing devastating damage to the building stock of the region and causing many casualties. In the earthquakes of October 23 and November 9, 2011, 644 of our citizens lost their lives, 1,966 of our citizens were injured, and 252 of our citizens were rescued alive (AFAD, 2011). On 23.02.2020, earthquakes (Hoy/Iran) occurred on the Turkey-Iran border at 08.55 ($Mw = 5.9$) and 19.00 ($Mw = 6.0$). In the statement made by the Governorship of Van on the first earthquake that occurred on February 23, 2020 at 08:55, 10 citizens lost their lives and 53 people were rescued with injuries (Sezer, 2010; AFAD, 2020). Among the scientific literature that has been reviewed, there are many studies on the extent to which earthquake clusters can be detected and the spatial distribution of these clusters. Researching the clustering of earthquakes is an important feature that impacts fundamental and applied analysis of seismicity. Affan et al. (2016) performed spatial statistical analyses of earthquakes that occurred in Aceh region of Sumatra island between 1921–2014 using Nearest Neighbor, Moran's I, Getis-Ord Genel G, Anselin Local Moran's I and Getis-Ord Gi* methods. In this study; the clusters of earthquakes occurring in study area were examined and their local and general spatial patterns were determined. Zaliapin and Ben-Zion (2022) offer perspectives on the clustering of earthquakes and propose a quantitative clustering measure of space-time earthquake clusters. In the proposed approach in this study, it is stated that the ROC-based Gini coefficient is an effective and efficient tool to measure the degree of aggregated space-time clustering. Spatial patterns of earthquakes magnitudes that occurred around the Tripa Fault in Aceh province between 1990–2017 were investigated by Sofyan et al. (2019). As a result of this research, it was concluded that there is a relationship between the earthquake magnitudes globally in the Tripa fault region (Sofyan et al., 2019).

In this research, data are analyzed based on the seismic classification according to the location and magnitude of the earthquake, obtained from earthquakes recorded by the database of AFAD Earthquake Department using the geographic information system (GIS). Many different methods are applied in the analysis of the spatial patterns of the magnitudes of earthquakes. The aim of this research is to apply these methods to earthquakes with a magnitude of Mw 4 and above that occurred in and around the province of Van between 1900–2021. After the application, the global and local spatial patterns were examined and clustering was determined.

2. Spatial autocorrelation analysis

The data used in this study are the locations and earthquake magnitudes of the earthquakes that occurred in and surrounding the province of Van. These data were obtained from the database of AFAD Earthquake Department. Earthquake occurrence data of $Mw = 4$ and above, between 1900 and 2021, were obtained from the database. The total number of earthquakes with a magnitude of $M = 4$ and above in this time period is 713. The spatial distribution of earthquakes with a magnitude of $Mw = 4$ and above between 1900–2021 in the study area is shown in Figure 1. A spatial statistical analysis was carried out using a geographic information system (ESRI ArcGIS) by processing seismic and spatial data.

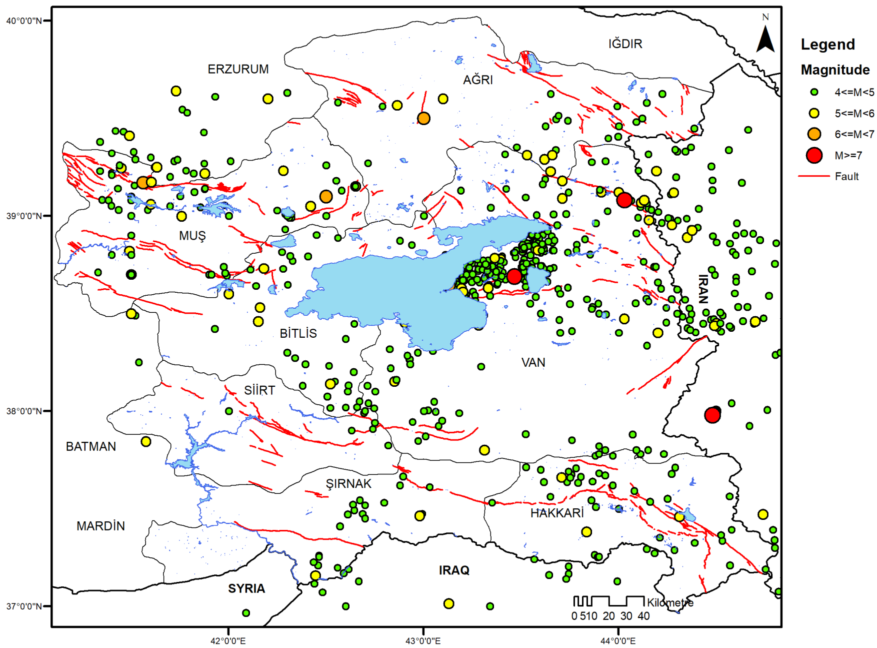

Fig. 1. Earthquake distribution between 1900–2021 in the study area

The foundation of spatial statistics was laid by Tobler. Tobler stated that everything is related to each other, but that what is near is more related than far (Tobler, 1970). Spatial autocorrelation indicates that the feature values in certain regions are related to these feature values in other adjacent or adjacent regions. If there is a systematic pattern in the distribution of a variable, it can be said that there is a spatial autocorrelation (Salima and Bellefon De, 2018; Hayati et al., 2019). These spatial autocorrelation methods are categorized as global and local. Global spatial autocorrelation methods give autocorrelation statistics over the whole study area. On the other hand, local spatial autocorrelation methods are used to determine the spatial differences in the relationships between the variables, especially the presence of clusters or hot spots (Al-Ahmadi et al., 2014; Affan et al., 2016).

Global and local statistics are used to gain a comprehensive understanding of spatial patterns. Of the four statistical methods discussed in this study, two are global and two are local. Globally; Global Moran's I and Getis Ord General G method, Anselin's local Moran's and Getis Ord Gi* method were applied locally. These methods are briefly described below. Global and local spatial patterns were determined by examining the clustering conditions of the earthquakes that occurred in the study area.

2.1. Moran's I Index

The global Moran's I statistic (GMI) gives the spatial autocorrelation level, taking into account data point locations and their simultaneous event-related value. GMI provides the spatial aggregation or dispersion level of datasets of a given point. The statistical significance of the Moran's I Test results and the distribution type of the pattern are determined by accepting or rejecting the following null hypothesis.

The hypotheses covered in the test are as follows:

Zero Hypothesis (H_0): The distribution of variables/attributes in the sample is random; that is, there is no spatial autocorrelation.

Alternative Hypothesis (H_A): The distribution of variables/attributes in the sample is not random; that is, there is spatial autocorrelation. In other words, clustering or dispersed is observed. Z value is calculated to indicate that any spatial pattern is not solely due to chance. Z value; It is obtained by subtracting the expected value from the observed value and dividing it by the standard deviation (Cerci, 2019). The Z value calculated in the GMI show whether the earthquake points are random or clustered. In autocorrelation research, values in the range of –1.96 to +1.96 give a random pattern and values above or below this range indicate whether there is clustering. When the z value is greater than 1.96 or less than –1.96, the null hypothesis is rejected, meaning that the model has a clustered spatial distribution. In this case, the index value is statistically significant, and if the p value is statistically significant and positive in parallel, the pattern is said to be not randomly distributed. In order for the p value to be defined as statistically significant, it must be less than 0.05 in all cases (Aslam and Naseer, 2020).

It is used to distinguish local clusters and outliers according to Anselin's Local Moran's I method. To calculate Moran's I, first the mean value is calculated, then the values of both the neighboring features and the target feature are compared with the mean value. Moran's I equation is shown below:

$$\mathbf{I} = \frac{N \sum\limits_{I} \sum\limits_{j} w_{ij} \left(x_I - \bar{x}\right) \left(x_j - \bar{x}\right)}{\left(\sum\limits_{I}^{n} \sum\limits_{j}^{n} w_{ij}\right) \sum\limits_{I} \left(x_I - \bar{x}\right)^2}. \tag{1}$$

In Moran's I equation, N is the sum of the number of data points, W_{ij} is the spatial weighting matrix, x is the variable and x_i, x_j is its values at ij positions, and $\bar{\boldsymbol{x}}$ is the mean value of x in all observations. The spatial weight matrix (W) shows the spatial relationship between the observation at position i and the observation at position j. The calculated **I** value takes a value in the range $[-1; 1]$. A positive autocorrelation when I value converges to 1, negative autocorrelation when it converges to -1 means that there is no spatial autocorrelation here if this value is equal to zero. Positive spatial autocorrelation in the dataset indicates that similar values tend to cluster spatially, but in negative spatial autocorrelation, higher values tend to accumulate near low values, in other words, these values are not spatially clustered (Mitchell, 2015; Aksha et al., 2018).

2.2. Getis Ord Genel G ve Getis Ord Gi

The Getis-Ord General G (GOGG) statistic was developed by Getis and Ord as a global statistic to analyze spatial patterns (Getis and Ord, 1992). GOGG identifies cold and hot spots based on high or low clustering of data values. High-clustered values are marked as hot spots, while low-clustered values are marked as cold spots in the region. A positive Z value indicates that the expected GOGG is lower than the calculated GOGG, meaning that the clustered data points have higher values in the area (Al-Ahmadi et al., 2014; Aslam and Naseer, 2020). The Getis Ord General G statistic is calculated as follows:

$$\mathrm{GOGG} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}. \tag{2}$$

In Eq. (2), x_i and x_j represent the feature values of i and j locations, and w_{ij} refers to the spatial weights between x_i and x_j features. As in other autocorrelation techniques, index, p and Z statistics values are produced separately for each feature. According to these values, hot spots and cold spots in the pattern are determined. Hot spots indicate clusters of high-value features, and cold spots indicate clusters of low-value features. Z value is obtained by subtracting the expected value from the observed value and dividing it by the standard deviation. The Getis Ord Gi* statistic is a local autocorrelation method used to detect hot and cold spots. If the calculated Gi* value is greater than the expected value, clustering of high values together indicates hot spots. Conversely, if the value of Gi* is less than the expected value, clustering together of low values indicates cold spots. Getis Ord Gi* index is calculated according to Eq. (3).

$$\mathrm{Gi}^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \bar{x}\sum_{j=1}^{n} w_{ij}}{s\sqrt{\frac{\left[\sum_{j=1}^{n} w_{ij}^2\left(\sum_{j=1}^{n} w_{ij}\right)^2\right]}{n-1}}}, \tag{3}$$

where x_j is the value of the data point, w_{ij} is the weight between the data point i and j for the n data point. For a positive Z value, it can be said that the larger the value of Z is, the more the high values cluster (hot spots), while for a negative Z value, the smaller the value is, the more the low values are clustered (cold spots) (Mitchell, 2005; Ren et al., 2020).

3. Results and discussion

Between 1900 and 2021, 713 earthquakes with a magnitude of $Mw = 4$ and above occurred in the province of Van. Descriptive statistics about earthquakes are summarized in Table 1. According to Table 1, the highest percentage of $M = 4$ and above total earthquakes is 88.22%, with earthquakes in the $4 \leq M < 5$ magnitude range. Secondly, earthquake magnitudes vary between $5 \leq M < 6$, which make up 10.66% of them. According to the Richter scale, there were 8 earthquakes of 6 and above. Earthquake dimensions were analyzed separately in four categories and their statistical values are shown in Table 1. The results of the analysis are given comprehensively in Table 2 and Table 3.

Table 1. Statistics of earthquakes between 1900–2021

Magnitude (M)	Earthquake total	% earthquake	Minimum magnitude	Maximum magnitude	Mean magnitude	Standard deviation
$4 \leq M < 5$	629	88.22	4	4.9	4.3	0.3
$5 \leq M < 6$	76	10.66	5	5.9	5.3	0.3
$6 \leq M < 7$	5	0.7	6	6.9	6.2	0.4
$M \geq 7$	3	0.42	7	7.6	7.2	0.3
$4 \leq M \leq 7.6$	713	100	4	7.6	4.4	0.5

Moran's I and Getis Ord General G statistical methods, which are global spatial autocorrelation methods, were studied on each of the four earthquake magnitude classes. The results of the analysis are given in Table 2 and Table 3. When the results in Table 2 are

Table 2. Analysis results of Global Moran Index

Magnitude (M)	Global Moran's I Index			
	Index	Z statistic	p value	Pattern type
$4 \leq M < 5$	0.11	7.95	0.00	clustered
$5 \leq M < 6$	0.01	0.32	0.74	random
$6 \leq M \leq 7.6$	0.36	1.63	0.10	random
$4 \leq M \leq 7.6$	0.15	11.12	0.00	clustered

examined according to Moran's I method, earthquakes with a magnitude of $4 \le M < 5$ have a clustering feature, while $5 \le M < 6$ and $6 \le M \le 7.6$ among the magnitude classes have a random distribution. Based on Table 2, it can be seen that earthquake magnitude points with $4 \le M < 5$ have spatial autocorrelation since the p value is less than 0.05 ($p \le 0.05$) and the Z value is greater than 1.96 ($Z > 1.96$). Since the p value of earthquake magnitude points of magnitude classes $5 \le M < 6$ and $6 \le M \le 7.6$ is greater than 0.05 ($p > 0.05$) and the Z value is between –1.96 and 1.96 ($-1.96 < Z < 1.96$), there is no spatial autocorrelation, that is, the magnitudes appear to have a random distribution. According to the Getis Ord General G method, when Table 3 is examined, earthquakes in the range of $4 \le M < 5$ and $5 \le M < 6$ show high clustering, while earthquakes in the $6 \le M \le 7.6$ class range show random distribution. Based on Table 2, $4 \le M < 5$ and $5 \le M < 6$ earthquake magnitude points show spatial autocorrelation since the p value is less than 0.05 ($p \le 0.05$) and the Z value is positive, in other words, high clustering is seen. It is seen that the earthquake magnitude points with $6 \le M \le 7.6$ have a p value greater than 0.05 ($p > 0.05$) and a negative Z value, that is, the magnitude has a random distribution. It is seen that different global indices can give different results when compared with each other. Since Moran's I and Getis Ord General G are global statistics, they provide only one measure that summarizes the model across the entire scope of the study area. Therefore, the local statistics Anselin local Moran's I and Getis Ord Gi* were applied to the data set as they were able to determine the presence and location of clusters or hotspots in the study region.

Table 3. Analysis result of Getis-Ord General

Magnitude (M)	Global Moran's I Index			
	Index	Z statistic	p value	Pattern type
$4 \le M < 5$	0.001	5.17	0.00	highly clustered
$5 \le M < 6$	0.013	2.15	0.03	highly clustered
$6 \le M \le 7.6$	0.138	–1.29	0.19	random
$4 \le M \le 7.6$	0.001	3.01	0.00	highly clustered

Figure 2 shows the locations of earthquakes from Anselin's local Moran's I statistics applied to identify significant clusters or spatial outliers using the spatial autocorrelation statistical method weighted by the magnitude of the earthquakes. Spatial clusters are indicated by dark red and dark blue dots. Dark red dots indicate high-magnitude earthquakes surrounded by high-magnitude earthquakes. In contrast, dark blue dots indicate low-magnitude earthquakes surrounded by other low-magnitude earthquakes. Light blue and light yellow dots indicate the presence of spatial outliers, i.e. high magnitude events surrounded by low events and vice versa. There are clearly spatial clusters of earthquakes occurring in certain regions of Van province and its surroundings. When the results are examined, the two local statistics show similar patterns. According to Anselin Moran's I method, it is seen that high values clustered on the Varto Fault zone around the Varto district of Muş province, while low values ($4 \le M < 5$) clustered around the Van and

Erciş fault line which is in the city center and northern parts of Van province. In other parts of the Van province, the occurrence and distribution of earthquakes show random patterns. It is seen that low ($4 \leq M < 5$) values are clustered around Hakkari province. On map in Figure 3, the red and blue dots indicate the major hot and cold spots, respec-

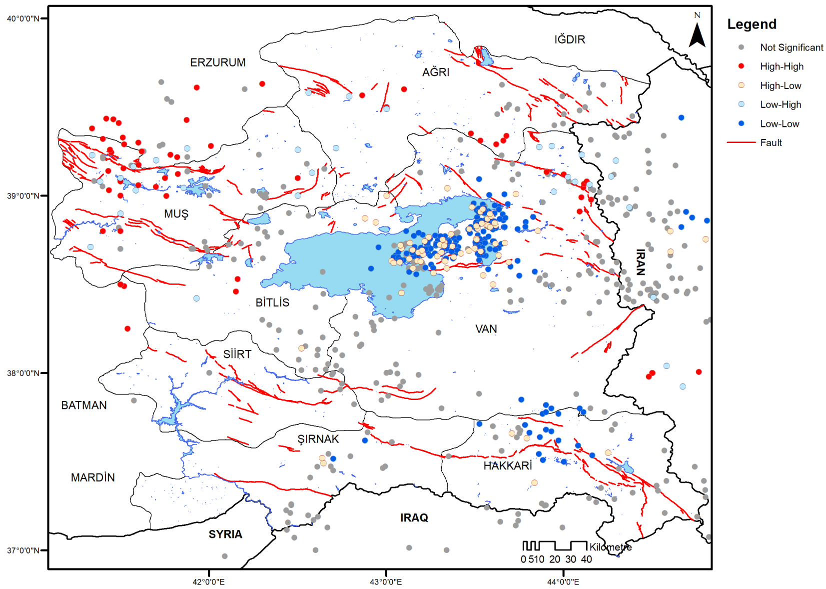

Fig. 2. Cluster analysis (Anselin Moran's I)

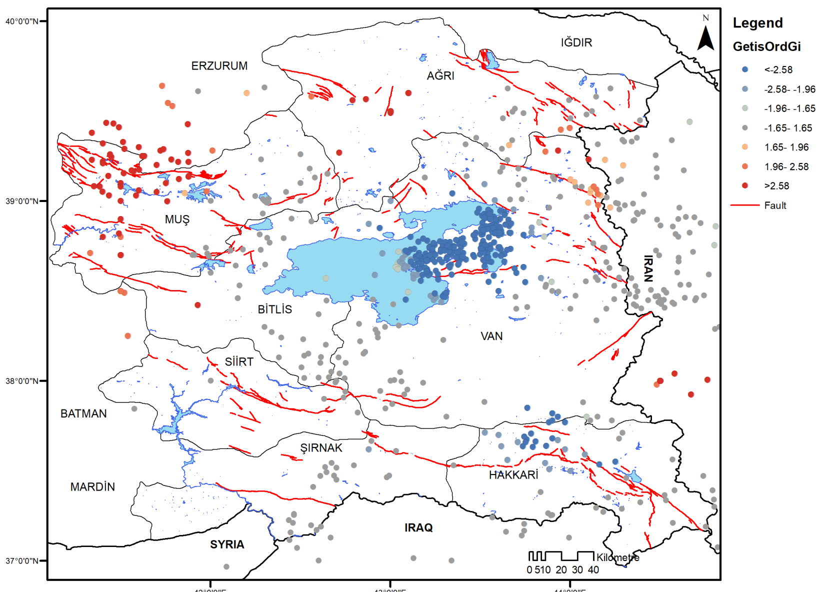

Fig. 3. Cluster analysis (Getis Ord Gi*)

tively. The results of the analysis of the Getis Or Gi* method show that there are similar clusters in the same regions. It is seen that high values, in other words, hot spots cluster around the Varto district of Muş province, while low values in other words cold spots cluster around the Van province and Erçiş district and around the provincial border of Van and Hakkari.

4. Conclusions

In this study, spatial statistical methods were applied to the earthquakes with a magnitude of $Mw = 4$ and above that occurred between 1900 and 2021 in the province of Van and its surroundings. It was examined whether the earthquakes that occurred in this study area showed clustering or random distribution. The spatial autocorrelation techniques of this research were able to detect clusters in earthquakes that occurred from 1900 to 2021, while the spatial pattern of earthquakes with higher magnitudes compared to the Richter scale was significantly concentrated in the Varto district of Muş and its northern parts, a rather high number of moderate-sized earthquakes occurred in the provincial center of Van and its northern. Spatial statistical analysis of the province of Van and its surroundings can be a valuable method for demonstrating earthquake clusters and complex spatial concepts between locations and seismic variables. Based on Anselin Moran's I and Getis Ord Gi* methods, it is seen that Muş Varto district and central and northern parts of Van have high seismic activity.

Author contributions

Conceptualization: G.M.P.; methodology development: G.M.P and O.B.; writing – original draft: G.M.P. and E.A.; writing – review and editing: G.M.P., O.B., and E.A.

Data availability statement

The raw data is available from the online database of the Republic of Turkey, Ministry of Interior, Department of Disaster and Emergency Management, Department of Earthquake. (https://deprem.afad.gov.tr/?lang=en).

Acknowledgements

The authors would like to thank the Republic of Turkey Ministry of Internal Affairs, Department of Disaster and Emergency Management, Earthquake Department for their support in obtaining the data.

References

AFAD. Republic of Turkey Ministry of Interior Disaster and Emergency Management Presidency. (2011). About the Van Earthquake. Retrived 2022 from https://www.afad.gov.tr/van-depremi-hakkinda.

AFAD. Republic of Turkey Ministry of Interior Disaster and Emergency Management Presidency. (2020). Evaluation Report on 23 February 2020 (IRAN) Mw 5.9.

Affan, M., Syukri, M., Wahyuna, L. et al. (2016). Spatial statistic analysis of earthquakes in aceh province year 1921–2014: Cluster seismicity. *Aceh Int. J. Sci. Technol.*, 5(2). DOI: 10.13170/aijst.5.2.4878.

Aksha, S.K., Luke, J., and Lynn M.R. (2018). Spatial and temporal analysis of natural hazard mortality in Nepal. *Environ. Hazards*, 17(2), 163–79. DOI: 10.1080/17477891.2017.1398630.

Al-Ahmadi, K., Al-Amri, A., and See, L. (2014). A spatial statistical analysis of the occurrence of earthquakes along the red sea floor spreading: Clusters of Seismicity. *Arab. J. Geosci.*, 7(7), 2893–2904. DOI: 10.1007/s12517-013-0974-6.

Aslam, B., and Naseer, F. (2020). A statistical analysis of the spatial existence of earthquakes in Balochistan: Clusters of seismicity. *Environ. Earth Sci.*, 79(1). DOI: 10.1007/s12665-019-8790-2.

Cerci, K.N. (2019). *Analysis of daily transportation movements in Istanbul with spatial autocorrelation methods*. Istanbul: Yildiz Technical University.

Getis, A., and Ord, J.K. (1992). The Analysis of spatial association by use of distance statistics. *Geogr. Anal.*, 24(3), 189–206. DOI: 10.1111/j.1538-4632.1992.tb00261.x.

Hayati, M.N., Purnamasari, I., and Wahyuningsih, S. (2019). Analysis of spatial autocorrelation of diarrhea events in East Kalimantan in 2015–2017 using a local indicator of spatial autocorrelation. *Journal of Physics: Conference Series*, vol. 1277. Institute of Physics Publishing.

Mitchell, A. (2005). *The ESRI Guide to GIS Analysis (1st ed)*. California, USA: ESRI.

Ren, H., Shang, Y., and Zhang, S. (2020). Measuring the spatiotemporal variations of vegetation net primary productivity in inner Mongolia using spatial autocorrelation. *Ecological Indicators*, 112, 106108.

Salima, A.B, and Bellefon De, M.P. (2018). *Spatial autocorrelation indices*. Handbook of Spatial Analysis, pp. 52-68. Montrouge, France: Insee, Eurostat.

Sezer, L.İ. (2010). Seismicity in the Van Seismotectonic Region. *Aegean Geogr. J.*, 19(1), 67–84.

Sofyan, H., Rahayu, L., and Lusiani, E. (2019). Spatial autocorrelation of earthquake magnitudes in tripa fault, Aceh Province, Indonesia. *IOP Conference Series: Earth and Environmental Science*, vol. 273. Institute of Physics Publishing.

Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.*, 46, 234–240.

Zaliapin, I., and Ben-Zion, Y. (2022). Perspectives on Clustering and Declustering of Earthquakes. *Seismol. Res. Lett.*, 93(1), 386–401. DOI: 10.1785/0220210127.