Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00139

Research Article

Spatial analysis of earthquakes in and around the northern Anatolian Fault Zone

Cenk İÇÖZ[1,*], Kadir Özgür PEKER[1]

[1]*Department of Statistics, Faculty of Science, Eskişehir Technical University, Eskişehir, 26555, Türkiye*

ARTICLE INFO

Article history
Received: 02 June 2023
Revised: 18 August 2023
Accepted: 06 January 2024

Keywords:
Intensity; Kernel Density Estimation; North Anatolian Fault Zone; Spatial Data; Spatial Point Processes

ABSTRACT

Earthquakes are seen to be complex phenomena when their occurrence reasons are investigated. Factors such as the presence of a fault line and between earthquake relationships like foreshocks, mainshocks, and aftershocks may be considered as some of the sources for this complexity. Estimating the earthquake risk and modeling the earthquake intensities over a region is vital in minimizing future tangible and intangible losses. The principle aim of this study is to examine the clusters and seismicity of moderate to major earthquake occurrences in the North Anatolian Fault Zone (NAFZ). For this purpose, a rectangular region including the NAFZ is selected as a study region to analyze the earthquake patterns. Attributes of moderate to major earthquakes with a magnitude higher than 5, which are listed regarding time interval and space domain in the earthquake catalog, are visualized owing to exploratory data analysis techniques. Spatial patterns for the earthquakes are revealed, and simulations of the earthquakes are realized using spatial processes within the specified time and space domain. Intensity changes and related earthquake risks are disclosed. The western part of the region is classified as having a higher risk for future earthquakes because of having higher previous earthquake and magnitude intensities.

Cite this article as: İçöz C, Peker KÖ. Spatial analysis of earthquakes in and around the northern Anatolian Fault Zone. Sigma J Eng Nat Sci 2024;42(6):1813–1825.

INTRODUCTION

The reach for spatial data is getting more accessible with the advancements in technology and GIS (Geographic Information Systems). Cressie [1] proposed a still valid classification for spatial data according to study region (or domain) properties. According to this classification, spatial data can be grouped into geostatistical data, spatial point patterns, and lattice data. This classification is essential for the right choice of analysis according to data type.

Earthquakes according to the domain properties fit into the class of spatial point patterns as spatial data. In spatial point pattern terminology, these incidences are noted as events to discriminate the other arbitrary points in the study region. There is no restricted region for earthquake occurrences, and earthquake epicenters can be discriminated from random points in the area. The main difference in spatial point patterns compared to other types of spatial data is the randomness of the domain.

***Corresponding author.**
*E-mail address: address:cicoz@eskisehir.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Al- Ahmadi et al. [2] used global and local spatial point pattern techniques to detect clusters of seismicity of the earthquake occurrences in the Red Sea for the period 1900-2009 according to the specific magnitude ranges. They also produced spatial versions of the central tendency and dispersion measures according to ranges of magnitudes for the selected time and space domains.

Öztürk [3] analyzed characteristics of seismic activity along the North Anatolian Fault Zone between 1970 and 2010 using several numbers of statistical parameters such as namely size-scaling parameters (such as the slope of recurrence curve *b*-value), seismic quiescence *Z*-value, the temporal and spatial distribution of earthquakes with the characteristic of fractal correlation dimension, *Dc*, as well as the histograms of temporal, spatial and magnitude distributions.

Akol and Bekler [4] applied a probabilistic hazard analysis to earthquakes with magnitude M.C ≥ 3.0 occurred in the region of NW Turkey covering Gelibolu and Biga Peninsulas period between 1912 and 2007. The region is divided into four subduction zones regarding seismotectonic properties. Sayil [5] estimated the recurrence intervals for large earthquakes at 7 different zones in the Marmara region limited with coordinates of 39°-42°N, 25°- 32°E. Coban and Sayil [6] modeled earthquake hazard by using 3 different probability model along the North and East Anatolian fault zones. Sayil [7], used statistical approaches using a data consist of a historical and instrumental period in six sub-regions of Marmara region. Earthquake recurrences were obtained by using Poisson statistical distribution models meanwhile Gutenberg-Richter relationship were employed to determine the seismic hazard parameters.

Turkey has long been suffered from devastating earthquakes and has active fault zones like North Anatolian Fault Zone (NAFZ). The NAFZ is the most important one amongst them, extending approximately 1200 km from the east through the country's west coast. It was described in the late 1940s, and it became one of the best-studied strike-slip fault zones on Earth [8-9] In addition, it has got one of the most seismologically active faults in the world with a broad-arc shape [10-11].

The most recent major earthquakes that occurred over the region were İzmit (M_s=7.8) and Düzce (M_s =7.2) earthquakes of 17 August and 12 November 1999 respectively that they caused enormous economic loss and death of many residents around the region. Nowadays, a major earthquake is expected in the western region of the NAFZ through the Marmara-Sea, which may affect one of the most populated metropolitans of the world's İstanbul.

Bohnhoff et al. [10] compiled and examined the historical earthquake catalogs by dividing the NAFZ into three sub-regions. They found out that the expected earthquake in the Marmara- Sea region would not exceed 7.5, most probably according to their work on historical earthquake catalogs (maximum observed event magnitudes) and other structural analysis and parameters.

Study region including the NAFZ is selected from the database with longitudes 26°- 40°E and latitudes 39.5°-41.5°N. Study region and North Anatolian Fault Zone with faults and deformations are given in Figure 1.

Recent spatial analysis of seismicity is required for NAFZ, especially from major earthquakes since there is an expectancy of a major earthquake near Istanbul. This study examines the characteristics of moderate to major earthquakes in the NAFZ. Clusters of moderate to high magnitude earthquakes are obtained. The higher intensity and lower intensity regions of the NAFZ is determined using kernel density estimation methods. The earthquake simulations according to past earthquake intensities are performed to predict possible patterns of earthquake occurrences. The spatial point pattern type of the earthquake point pattern is revealed. This study region can be divided into sub-regions for the detailed application of earthquake simulations. This might change the spatial point pattern type, which can be different from our region choice.

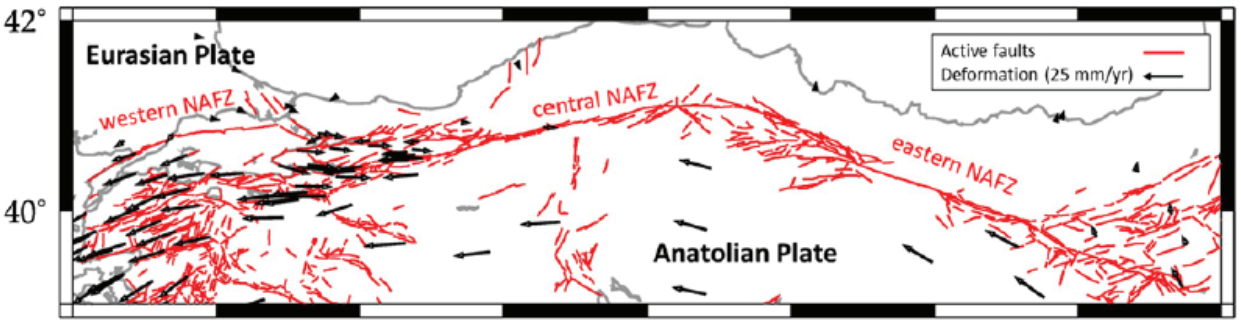

Figure 1. North Anatolian Fault zone along with the rectangular study region (modified from Bohnhoff et al. [9], with permission from Elsevier).

SPATIAL POINT PATTERNS AND PROCESSES

Stochastic mechanisms underlying a spatial point pattern are described as a spatial point process. A spatial point pattern realizes a process in two-dimensional space $Z(\mathbf{s})$: $\mathbf{s} \in D \subset \mathbb{R}^2$, while D is a random domain and $Z(\mathbf{s})$ is the attribute value-dependent to location **s**. Spatial point patterns are classified into three groups: Complete Spatial Random (CSR) patterns, Regular patterns, and Clustered patterns. Shabenberger and Gotway [12] describe CSR with two important conditions satisfied:

i) Intensity $\lambda(s)$, the average occurrence number of events in a unit area must be homogenous over domain D.
ii) Let A_1 and A_2 be two non-overlapping sub-regions. The event numbers in these two sub-regions must be independent and fit into a Poisson distribution.

$\lambda(\boldsymbol{s})$, the (first-order) intensity of a spatial point process is described as a limit and as a function of **s** as follows:

$$\lambda(\boldsymbol{s}) = \lim_{v(d\boldsymbol{S})\to 0} \frac{E\big(N(d\boldsymbol{s})\big)}{v(d\boldsymbol{s})} \tag{1}$$

In equation (1), $v(d\boldsymbol{s})$ is the area of an infinitesimal disc in $\mathbb{R}^2$ centered at **s.** The CSR's stochastic mechanism is the spatial Poisson process (homogenous Poisson process).

Many point patterns are somehow generated from CSR. Events can be independent in the non-overlapping region in a study domain; however, they do not satisfy the condition of having homogenous event numbers in the unit area. In some areas of the field, the intensity might be higher, whereas in some might be lower. Events might also have an interaction with each other. The presence of an event might prevent another event occurrence or might cause the occurrence of another event. Deviations from CSR patterns are clustered and regular patterns [12].

If above condition (i) is not satisfied when condition (ii) is satisfied, an inhomogeneous Poisson process is realized. The two following properties characterize the inhomogeneous Poisson process.

If the conditions of CSR are not satisfied, a clustered pattern is considered when there might be some sub-regions of a domain with a higher average event count than other sub-regions. There is no homogenous intensity over the domain for such patterns. Also, some events can attract other events in nearby regions, resulting in a clustering pattern.

In the same manner, some events can prevent another event from occurring. A regular pattern follows a standard scheme. In this pattern, the average distances between an event and its nearest neighbor are lower than CSR distances.

TESTS FOR CSR

There are different types of tests based on point distances or interactions of points or basic tests like quadrat counting based on goodness of fit to check whether a pattern fits into CSR or not. In these tests, a test statistic or generally function values under the null hypothesis when the theoretical pattern is a realization of a homogenous Poisson process is compared with the observed pattern's test statistic or functional importance. According to the specific function values or test statistic value, the null hypothesis will be rejected or fail to be rejected. When an observed pattern is found to fit into CSR, it is assumed to realize the homogenous Poisson process.

The alternative hypothesis of CSR testing depends on the background knowledge about the point pattern, which can usually be regularity or clustering [13]. For a regular-shaped spatial domain, distributional properties of quadrat counts are easy to establish. However, some approximations like the closed form of a test statistic and its asymptotical properties may not perform well for irregular-shaped spatial domains because of the rare events in quadrats and edge effects. In such circumstances, relying on simulation techniques like Monte Carlo tests and simulation envelopes may be the options to test CSR patterns [12].

A Monte Carlo test is a simulation test based on the simulations from the null hypothesis; for example, a theoretical distribution or point process in our case. Monte Carlo tests are frequently used in nonparametric statistics, a particular case of randomization tests [14]. Applications of Monte Carlo tests to spatial statistics can be found in Ripley's studies. [15-16]

For instance, one generates m independent simulations from CSR inside the study domain D and estimates the functions for each of the realizations. After evaluating these functions, one can obtain the pointwise lower and upper envelopes of these simulated functions, which are curves of distance-based functions in this case. The null hypothesis of CSR is rejected when the point pattern's estimated function values remain outside these envelope values [14].

Quadrat Counting Test for CSR

Quadrat testing is one of the simplest and primitive tests for evaluating CSR patterns. This test is considered for only regular shaped and bounded study domains. Quadrats are standard rectangular or squared shapes of equal area. Study domain D is split up into non-overlapping quadrats row-wise and column-wise with similar areas, and then the event numbers for each quadrat are counted. The union of these non-overlapping sub-regions must form the study domain D.

The bounded study domain D is partitioned into r rows and c columns. Let n_{ij} is the number of events in the quadrat ij and $\bar{n} = n/(rc)$ the expected number of events in each quadrat under CSR, while n is the total number of events in the study domain D. The chi-square test is given in equation (1) [12].

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\big(n_{ij} - n\big)^2}{\bar{n}} \tag{2}$$

This test statistic is a Chi-square test for the goodness of fit test of a null hypothesis. The n point is distributed uniformly and independently in *D*. In other words, quadrat counts are independent Poisson variates with a typical mean CSR pattern. The disadvantage of the test is the heavy influence of the subjective choice of the quadrat size.

G Function (nearest neighbor distances)

Nearest neighbor distance analysis examines the most relative event distances for each event in a study domain. In a closest neighbor-based test of CSR pattern, the theoretical most relative neighbor distances of CSR pattern are compared with the observed patterns nearest neighbor distances [17].

The empirical distribution function of *G* is given below for n events when r_i denotes a distance from an event to its nearest neighbor event in *D*.

$$\hat{G}(r) = \frac{1}{n}\sum_i I(i: r_i \leq r) \tag{3}$$

*I*t is the indicator function; the summation counts the number of nearest neighbor distances for all events when the condition $r_i \leq r$ holds and r is an arbitrary distance in equation (3). The empirical nearest neighbor function $\hat{G}(r)$ is interpreted as the probability of an observed nearest neighbor of a point appearing at any given distance r. Therefore, it is used as a function that determines the level of clustering or regularity [17].

One important thing to consider is the edge effects when applying tests based on distances. There could be the nearest neighbor outside our study domain *D,* causing biased nearest neighbor distances near the domain's boundary. Therefore applying edge effects are important [17-18].

The distribution function under CSR is given below when edge effects are ignored:

$$P(N(D) = n) = \frac{e^{-\lambda v(D)}\lambda^n v(D)^n}{n!} \tag{4}$$

Here λ is intensity while $N(D)$ is several events in D and $v(D)$ is the area of *D.* Under CSR, the probability of having no events for an event is $e^{-(\lambda\pi r^2)}$ if πr^2 the area of a disc is centered on the target event for the searching nearest neighbor in the defined disc.

Hence, the nearest neighbor distribution function for CSR is given by:

$$G(r) = 1 - e^{-(\lambda\pi r^2)}, r \geq 0 \tag{5}$$

Higher values for small r from an empirical distribution function for a clustered pattern when compared with function values under CSR and lower values for a regular pattern.

F function (empty space function)

This test is also known as the point to nearest event distance test. This test measures the distance from the arbitrary point to its nearest event in a study domain.

The empirical distribution function of *F* is given below from sample points (arbitrarily chosen in the region) when r_i denotes a distance from an arbitrary point to its nearest neighbor event in *D*.

$$\hat{F}(r) = \frac{1}{m}\sum_i I(i: r_i \leq r) \tag{6}$$

In the same manner, ignoring the edge effects under CSR, the distribution function for space distances are given below:

$$F(r) = 1 - e^{-(\lambda\pi r^2)}, r \geq 0 \tag{7}$$

It is expected that for a regular pattern, $\hat{F}(r)$ observed empty space distances could be greater than its theoretical counterpart $F(r)$. On the other hand, values $\hat{F}(r) < F(r)$ advocate that the point pattern's empty space distances are shorter than the empty space distances for a Poisson process, resulting in a clustered pattern.

J Function

J function is a mixture of both *F* and *G* functions that are described below [19]:

$$J(r) = \frac{1-G(r)}{1-F(r)} \tag{8}$$

This equation holds only when $r \geq 0$ and $F(r) \neq 1$. In the case of CSR, $F(r) = G(r)$ is expected so that the value of $J(r) = 1$ if $G(r)$ is greater than (r), which means $J(r) > 1$ then point pattern is a clustered pattern, whereas if $F(r)$ is greater than $G(r)$ so that $J(r) > 1$ then point pattern is a clustered pattern.

K Function

The weakness of the F and G functions described before is their dependence on nearest neighbor distances for each event. Therefore, it can be a major problem, considering where nearest neighbor distances are very short relative to other distances in the pattern for clustered patterns. K functions rely on all distances in a sub-region of a study window [16].

Ripley [16] described the *K*- function as $\lambda K(r)$ is the expected number of other points of the process within a distance r of a specific point of the process.

Observed pairwise distances described as $d_{ij} = ||x_i - x_j||$ form a biased sample of pairwise distances in the point process, with a bias on the side of smaller distances because the maximum distance can be observed the diameter of the window for a pairwise distance [14].

The expected number of points for a disc containing a typical point of the process is $\lambda\pi r^2$ when r is the radius of the disc and λ is the intensity of the homogenous Poisson process. Therefore, it is $K(r) = \pi r^2$ independent from the intensity.

Numerous estimators of K have been proposed. The general form of the weighted and renormalized empirical distribution functions of pairwise distances is given below, where (EU,v,r) is the edge correction weight:

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2 v(D)} \sum_i \sum_{J \neq i} I\left\{\left||x_i - x_j|\right| \leq r\right\} e(x_i, x_j, r) \quad (9)$$

Likewise, for other functional values, we compare $\hat{K}(r)$ with $K(r)$, the theoretical counterpart under the Poisson process. If $\hat{K}(r) > \pi r^2$, then the observed pattern is a clustered one, whereas if $\hat{K}(r) < \pi r^2$ is a clustered one.

Intensity Estimation with Kernel Density Functions

Nonparametric smoothing methods are popular because of assumption-free properties and the use of original data set to estimate the probability distribution function. One of the most familiar methods of this data for univariate density is histograms. Kernel functions are probability density functions themselves. They must be even functions. Moving kernel functions can be used to estimate the number of events for an area in the study region instead of using fixed grids to determine the number of events in quadrat counting.

This method is mostly used to obtain smooth estimates of univariate (or multivariate) probability densities from an observed sample of observations [20]. Gatrell [21] explained that the estimation of intensity for a spatial point pattern resembles estimating bivariate probability density.

The Gaussian kernel function is given in equation (10) and is selected as a kernel function to estimate the intensity for this study for the study window:

$$K(\mathbf{x}) = \frac{1}{2\pi} \exp\left\{-\frac{(\mathbf{x} - \mathbf{x}_i)^2}{2}\right\} \quad (10)$$

A two-dimensional kernel density estimator with a fixed bandwidth is given below in equation (11) [20]:

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left\{\frac{1}{h}(\mathbf{x} - \mathbf{x}_i)\right\} \quad (11)$$

In equation (11), $\mathbf{x}$ is the location of any location in the study area, the bandwidth of the kernel function K, n is the number of earthquakes, and $\mathbf{x}_i$ is the location of each earthquake. The drawback of this method is to find the optimum bandwidth (h) to estimate the intensity. There are also adapted smoothing methods in which the bandwidth in the study region changes according to the low and high density of the moving windows for the whole study region.

APPLICATION

Earthquakes higher than magnitude 5.0 occurred between the years 1950 and 2017 were obtained from the database of KOERI. (Bogaziçi University, Kandilli Observatory and Earthquake Research Institute).

Data is downloaded as an earthquake catalog with several variables, including occurrence date and time, magnitude and depth of earthquakes, and spatial information of latitudes and longitudes. Declustering algorithms was not performed, and the article is presented as a numerical example. The biggest earthquake scales were selected for the study because all earthquakes in downloaded data was measured in a different scale. As we are examining a narrow range of magnitudes which are moderate to major earthquakes homogeneity of magnitudes would not be a great issue. Spatial analyses, basic statistics about earthquakes, and earthquake visualizations were performed in core R, *ssbase*, and *spatstat* and *aspace* packages [22-26]. In the first part of the application section, exploratory data analysis will be mentioned. Spatial point pattern analysis for the earthquake pattern will be given in the second part.

Exploratory Data Analysis

To discriminate the magnitude distribution over the region, distinct point characters are employed in Figure 2. It is clearly seen that from the map in Figure 2, earthquakes greater than magnitude 7.0 only occurred in the western part of the study region. In addition, it can be inferred that the middle part and eastern part of the region is unproductive about treating moderate to major earthquakes.

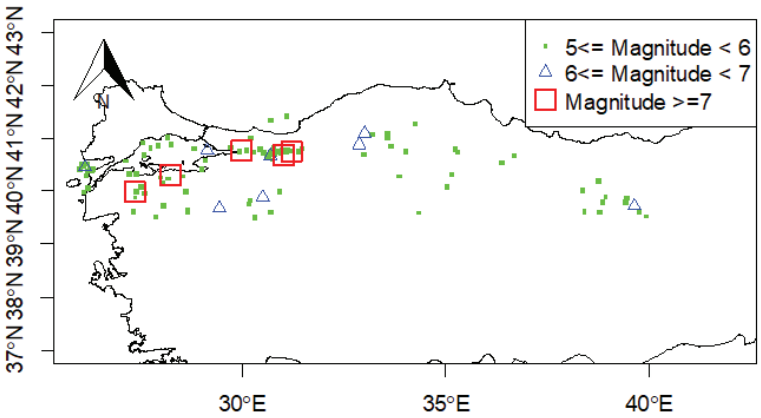

Figure 2. Distribution of earthquake epicenters according to three different magnitude ranges over the selected region.

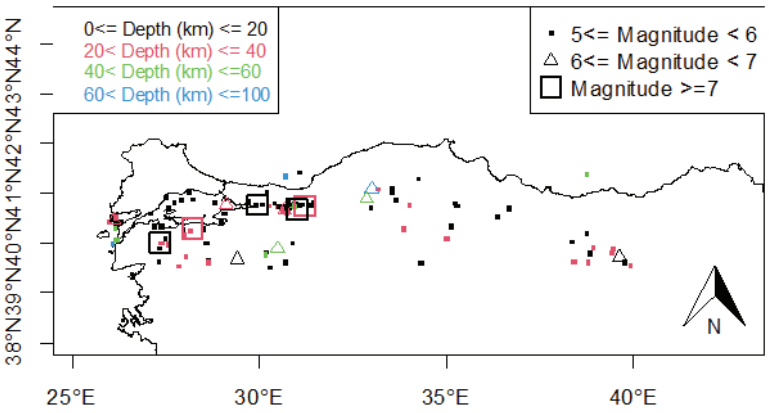

Figure 3. Distribution of earthquake epicenters according to three different magnitude ranges and four different focal depth ranges over the selected region.

In Figure 3, both magnitude and depth distributions are given in a map with different color coding and point characters. In this way, it is easy to recognize both features for a given earthquake. Most of the earthquakes can be classified as shallow earthquakes because of the low depth earthquake occurrences under 20 km.

In Figure 4, central tendency measures like mean center and variability measures like a standard ellipse for earthquake point patterns are given. The change in magnitudes over the region is low because of both the choice of the data and rare frequencies of magnitude 7 higher earthquakes. Therefore, the mean center and weighted mean center according to magnitudes overlap approximately the center of the study region. However, the standard deviational ellipse reflects the variability far better than the standard distance by taking into consideration both axes separately.

In Figure 5, the number of earthquake incidences are given yearly. The devastating İzmit and Gölcük earthquakes triggered another seismicity period of aftershocks in 1999. Therefore, there are high event occurrences because of the high magnitude events' aftershock effect

In Figure 6, the depths of earthquakes are given. It can be inferred that most of the earthquakes occurred between 0 and 40 km depths. Hence, they can be classified as shallow events from this figure too.

One of the important parameters reflecting this relationship that is employed by seismologists frequently in

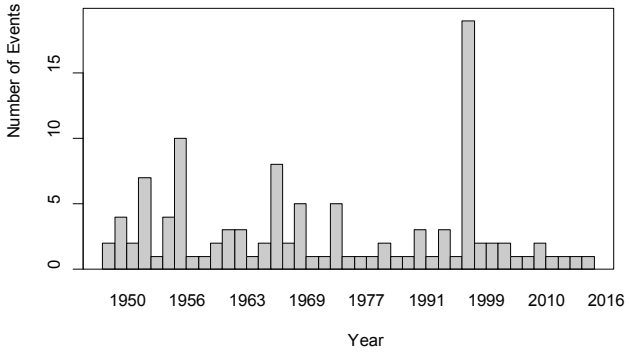

Figure 5. Yearly earthquake incidences.

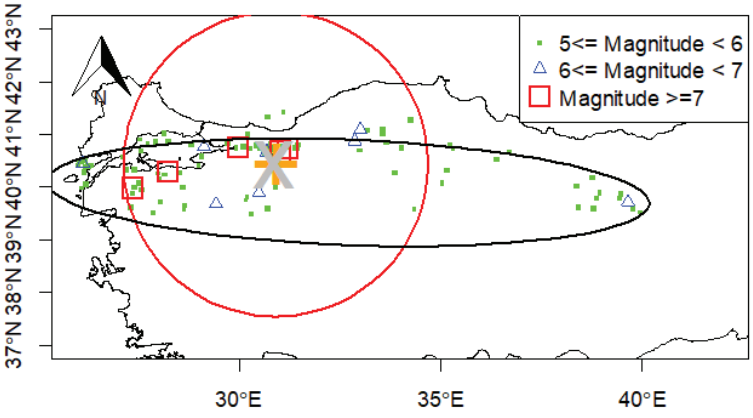

Figure 4. Central tendency and variability measures for earthquake point pattern (ellipse: standard deviational ellipse, circle: standard circle, plus: mean center; cross: weighted mean center of earthquake point process).

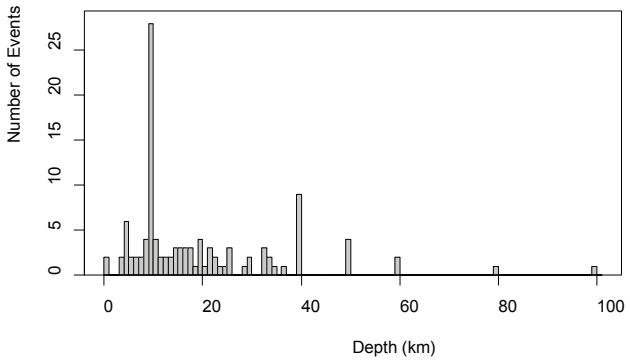

Figure 6. Depth distribution of earthquake occurrences.

literature is Gutenberg-Richter law, and it is commonly expressed as below:

$$logN(m_i > m) = -bm + a \tag{12}$$

In equation (12): N denotes the number of events with magnitude m_i greater than some threshold m, and b is the log-linear constant which is a determinant of proportionality called the b-value. The constant a is described as a function of the total number of counts in the sample at m=0 is the cumulative structure [25].

A large value of b can be interpreted as more alike events in terms of magnitudes are likely, while a small value can be interpreted as a larger range of magnitudes. To calculate b value, Aki [27] proposed a maximum likelihood approach:

$$\hat{b} = \frac{\log(e)}{\overline{m} - (m_{min} - \Delta m/2)}, \tag{13}$$

where $\overline{m}$ is the magnitude average, m_{min} is the magnitude threshold $\Delta m/2$ is a correction for finite binning.

In addition, Aki [27] described the standard error of this estimation as:

$$\sigma_{sterr} = \hat{b}/\sqrt{n} \tag{14}$$

In Figure 7, histograms of magnitude are given. As it is seen clearly, both log frequencies and frequencies decrease when the magnitudes are high. A benefit of this representation is that it gives insight into the right-skewed exponential relationship about frequency-magnitude distribution.

The change of b -values within their standard error bars according to the different magnitude completeness values is given in Figure 8. The b-value is relatively stable approximating to 1 below a cutoff m_c = 6.5; the b-value estimate at m_c = 5 is 0.92 with a standard error of 0:08.

In Figure 9, a scatter plot of magnitude and depth is given with the measured linear correlation coefficient. A low positive correlation coefficient (r=0.09) is the indicator that there is no relationship between depths and magnitudes of moderate to major earthquakes.

In Figure 10, deviations from the mean earthquake frequency graph are given. In this graph, earthquake frequencies in a training period are taken into consideration, and this period's earthquake frequency is compared with

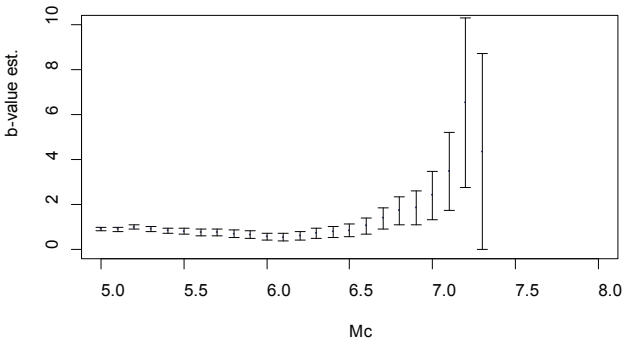

Figure 8. Magnitude cut-off values with associated b- value estimations.

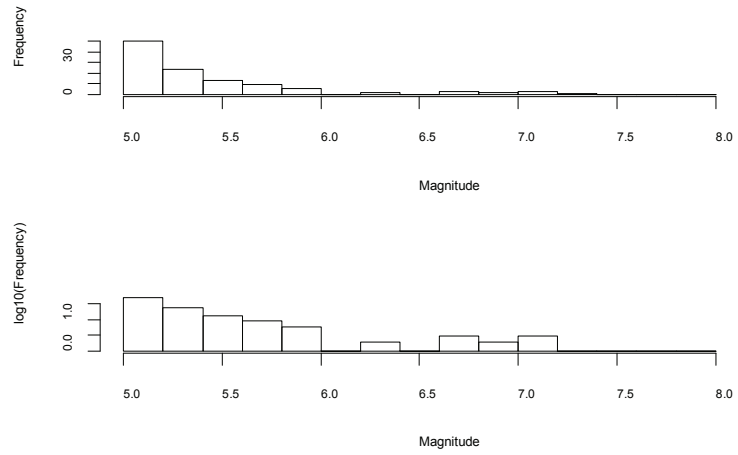

Figure 7. Magnitude distribution and log magnitude distribution of earthquake occurrences.

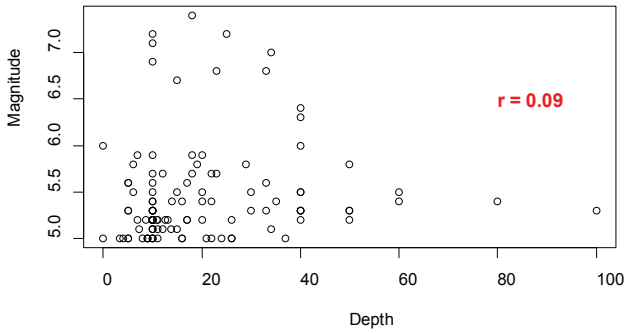

Figure 9. Scatterplot of magnitude and Depth variables.

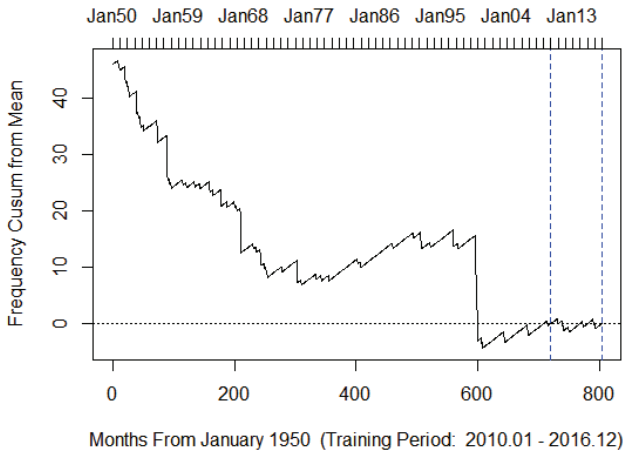

Figure 10. Reverse cumulative sum graph of deviations from the mean frequency.

other periods' earthquake frequencies. It is an example of time series graphs called reverse cumulative sum graphs. It must be examined left to right direction-wise and increases in mean value is a sign of the increase in frequency when compared with the training period. On the other hand, the decline of the mean value is a sign of a decrease in earthquake frequency when compared to the training period. In the 600-800 days' period, the time mean earthquake occurrences is less than the occurrences of the training period. Meanwhile, mean earthquake occurrences are increasing after the earthquakes that occurred in 1999 that started with 600th day today 0.

Spatial Point Pattern Analysis

It is an interesting subject whether CSR pattern holds or not for earthquakes as a spatial point pattern. The subject of having CSR for an earthquake spatial point pattern depends on several criteria. Study window, time domain of the study, and magnitude range of the earthquakes are some examples that can be given for these criteria. Also, earthquake occurrence types contribute to this matter. Devastating earthquakes may create aftershocks of lower magnitudes may provide a clustering scheme in the given study window.

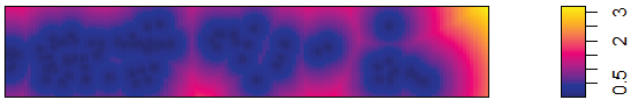

Figure 11. Empty space distances.

Visually examining some of the distance measures defined in the previous sections may give researchers a lead about the point pattern type before moving on to the statistical tests. In Figure 11, empty space distances of the earthquake pattern are given. A high distance from an arbitrary point to an event is shown in hot colors, whereas low distances are shown in cold colors and are a sign of a neighborhood event close to an arbitrary point.

The nearest neighbor distances are given in Stienen Diagram in Figure 12. In this diagram, each event's distance to a neighbor event is shown as a circle whose diameter changes proportionately on each event. Therefore, small nearest neighbor distances are illustrated as a point.

A realization of the homogenous Poisson process in the selected study window and the earthquake pattern are given together in Figure 13. It is highly important to visually examine and compare patterns visually before performing the tests about determining the pattern type of a spatial point pattern.

Determination of intensity anomalies in a point pattern is one of the most prominent problems in point pattern

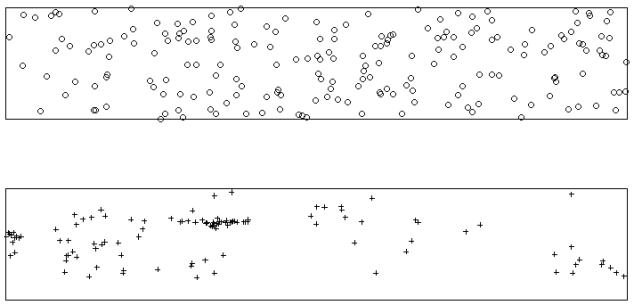

Figure 12. Stienen Diagram.

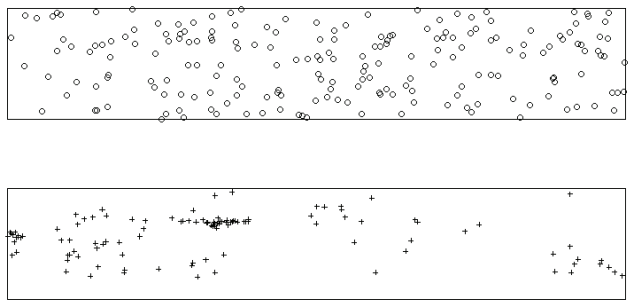

Figure 13. Realization of homogenous Poisson process (CSR pattern above) and the earthquake pattern (below).

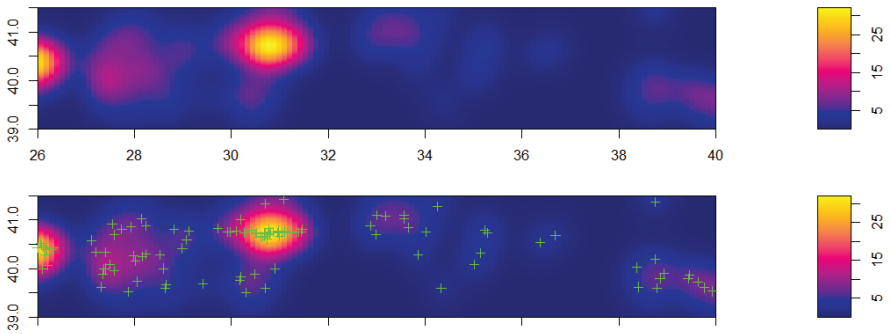

Figure 14. Intensity image alone (above) and superimposed earthquake pattern together with the intensity image (below).

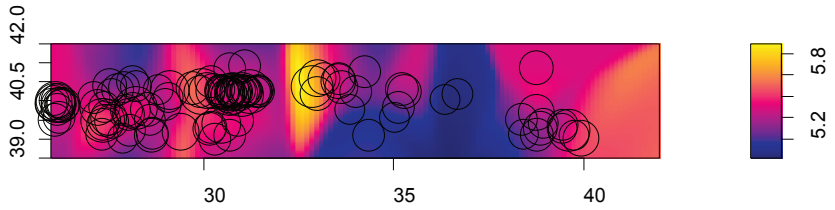

Figure 15. Magnitude intensities over the region.

analysis. Hot spots and cold spots are defined as the regions which have higher and lower intensities compared to the mean intensity of the study window, respectively. Hot spots are also considered as a sign of the possible formation of clustering and the regions of possible clustering. The easiest way to achieve obtaining hot and cold spots in the estimation of intensities through the study region is via kernel density estimation using kernel functions. [14]. Intensity estimation with kernel density estimation and the superimposed earthquake pattern to intensity image is given in Figure 14.

In Figure 14, it can be inferred from superimposed pattern to intensity image how well the pattern coincides with the estimated intensity along the study region. From cold colors to hot colors, the intensity increase is observed within the several high-intensity clusters. If the NAFZ is classified into three groups as western, middle, and eastern parts, the middle part has got the lowest intensity when compared to other parts.

In Figure 15, the magnitude intensity image together with the earthquakes represented as circles according to their magnitude size over the region is given. In the middle region of the NAFZ, cold colors are dominant. Therefore, it can be inferred that earthquakes of magnitude around 5 occurred in neighborhood locations. In the western and eastern parts of the NAFZ, magnitude intensities increase approximately 6 with a few areas of cold color, which means magnitude intensities approximately 5.

Quadrat counts, along with their expected event count and residual values, are given in Figure 16. The whole study region is divided into 24 quadrants of equal area. In the CSR pattern, a similar event count is expected in each quadrat. However, important differences in the event count through quadrats are detected. As a result of the chi-square test, the probability is found to be approximately zero. Hence, the pattern in the analysis does not belong to a homogenous poison process which is the stochastic mechanism under the CSR pattern.

In Figure 17 and Figure 18, distance-based tests for CSR and simulation envelopes for these tests are given. From both figures, according to the results, the earthquake pattern is revealed as a clustered pattern. The H_0 hypothesis of the observed pattern is CSR is rejected because of separately higher G values, lower F values, lower J, and higher K values in the whole 4 tests.

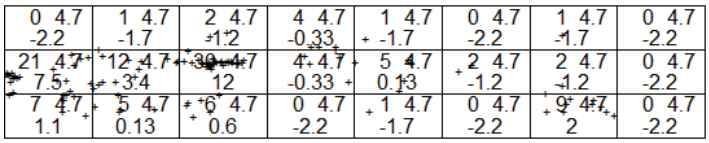

Figure 16. Quadrat counts.

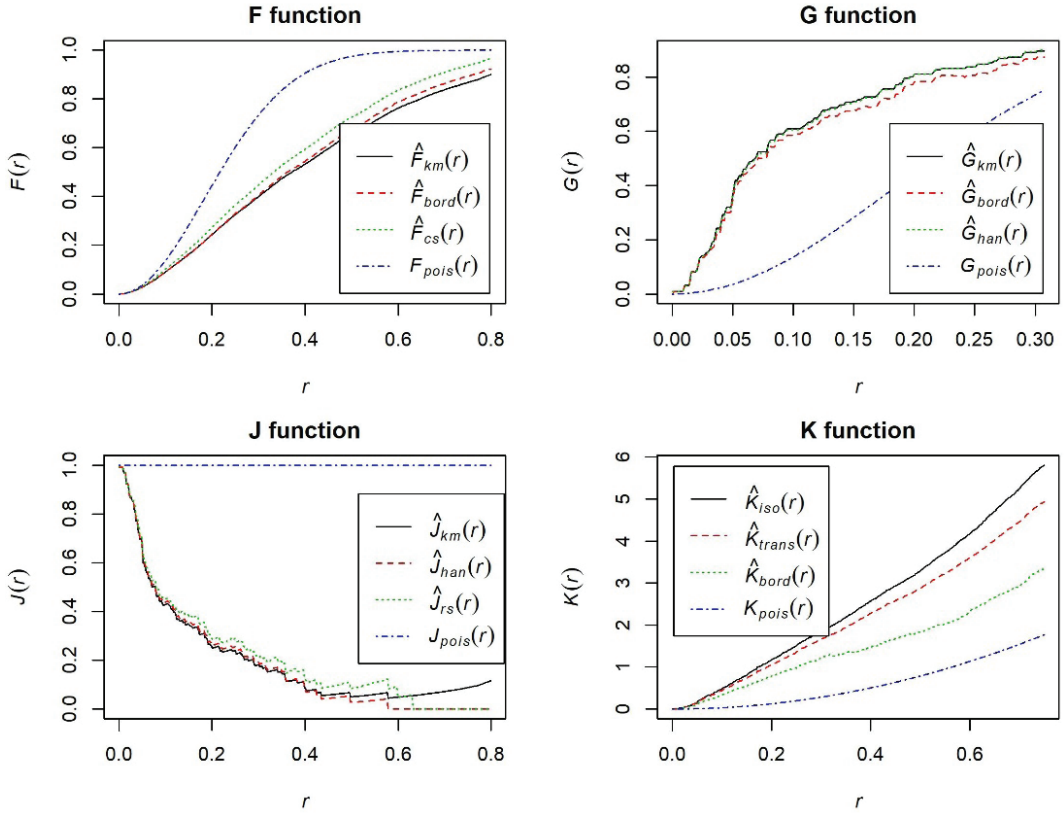

Figure 17. Distance-based tests for CSR.

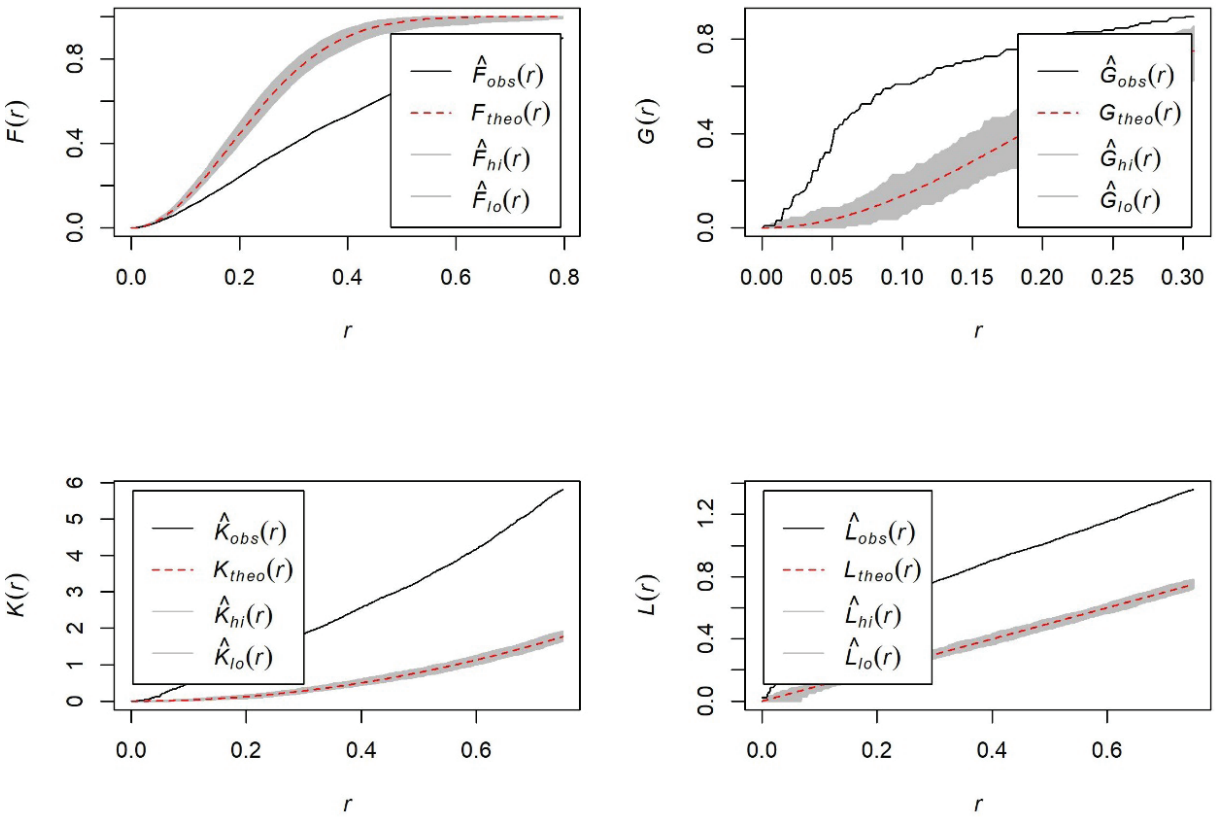

Figure 18. Simulation envelopes for CSR.

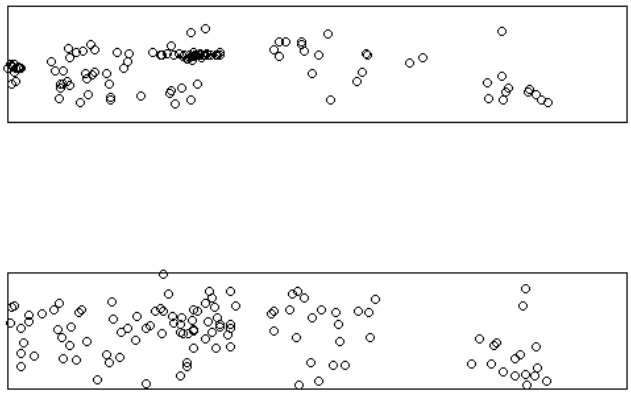

Figure 19. Earthquake pattern (above) and simulated inhomogeneous Poisson process (below).

In simulation envelopes, none of the observed functional values are between the lower and upper boundaries of theoretical counterparts. Therefore, the initial results are supported by using simulation envelopes.

In Figure 19, earthquake patterns and simulations of earthquakes with inhomogeneous Poisson processes are given. The similarity between the two patterns is remarkable. Changing intensities over the region are taken as changing intensities for simulation of inhomogeneous Poisson process. This simulation may give future predictions of the possible earthquakes by taking past earthquake occurrences into account.

RESULTS AND DISCUSSION

In this study, clusters of seismicity for moderate to major earthquakes in and around the NAFZ between the years 1950 and 2017 were aimed to obtain. Moreover, point pattern analysis was applied to determine the pattern of the selected earthquakes. With the use of distance-based tests and quadrat analysis, a clustered pattern was revealed for the pattern in the selected study region within the selected time domain. Both kernel density estimation and the quadrat analysis indicate that the higher seismicity intensity and seismicity clusters are formed in the middle -western through the Marmara Sea and in the south-eastern parts of the NAFZ. The middle part of the NAFZ is the least productive part of the NAFZ, both according to the intensity and total earthquake count measures. Earthquakes with higher magnitudes were concentrated through the western parts. Exploratory data analysis was performed to reveal the relationships between magnitude and depth distributions, frequency magnitude, frequency depth, and recent earthquake frequency and time. Useful distance measures were plotted to examine the empty space and nearest neighbor distances over the region.

The consistency of the magnitude cut-offs and *b*-values were examined. Through the magnitudes between 5 – 6 and six *b*-values are consistent. Simulations for the study region were performed both to determine the pattern type and to predict future earthquakes. Inhomogeneous Poisson process was employed for the future earthquake predictions with changing intensity according to the earthquake pattern's intensity. Therefore, spatially similar earthquake clusters were formed with higher intensities.

Upon the study region, magnitude range, and time-domain changes, the findings of our study would change. For instance, eastern middle, and western parts of the NAFZ can be examined individually for cluster formations, exploratory data analysis, earthquake parameters like b-values for both moderate to major and whole earthquakes above a certain magnitude rather than our choice of moderate to major earthquakes.

Akol and Bekler [4] found that the triangle of Gelibolu–Tekirdağ western part of Marmara Sea is the dangerous area concerning earthquake occurrence of the largest magnitude with 7.3 within a 100-year period. This occurrence was determined to be 46 % according to the Poisson distribution. According to Sayil [5] the total earthquake activity constant a and *b*-value was higher in the golf of Saros sub region which located in the most western part of the fault zone. Sayıl [7] stated that next earthquakes can occur in seismogenic source 6 (Bandırma) with the highest probabilities. Öztürk [3] analyzed the NAFZ by using the seismotectonic b-value, fractal dimension *Dc*-value, precursory seismic quiescence Z-value. In addition, correlations of these parameters were compared with each other. Regions that has Highest Z-value with the lowest b-value and the highest *Dc*-value are the vicinity of the Düzce fault and the Black Sea coast. By using orthogonal regression, a strong correlation negative correlation coefficient was found between *Dc* and *b*-values for the NAFZ and its vicinity (r=-0.98). Öztürk [28] examines the seismic activity characteristics along the NAFZ between 1970 and 2010. *b*-values were typically close to 1 according to the study. Eight significant abnormal zones throughout the North Anatolian Fault Zone were detected after a statistical analysis to spot the seismic quiescence after 2010. Coban and Sayil [6] calculated the conditional probability for earthquake occurrence (Mw ≥ 5.5) employing according to three different probabilistic models for the three regions along the NAFZ and EAFZ. By using the interevent time of earthquakes, model parameters were calculated under the condition of MLE (Mw ≥ 5.5) between 1900 and 2016 in the regions. As a result, they found out that the short-time earthquake probabilities are high in regions. The limitation of the study is the assessment of earthquakes specific parameters (only whole region-specific *b*- value is used without interpreting temporal spatial forms along the region) with exploratory spatial data analysis. However, spatial point pattern analysis and exploratory spatial analysis might give insights to researchers studying seismology. Therefore, these type of studies would be a plot for statical analysis of earthquakes, probabilistic hazard analysis and interpretation of seismic parameters temporally and spatially. Our exploratory data

results are consistent with literature in terms of high earthquake activity in western parts of the fault and large earthquakes were expected with a higher quality in this region with a higher probability according to other studies.

CONCLUSION

In this study, large earthquakes on the NAFZ were analyzed with the help of point processes and descriptive spatial data analysis. Graphical analyses were used and tools such as spatial distribution of these earthquakes, intensities, magnitude intensities, earthquake-free areas, inter-earthquake gap graphs were used. In addition to seismotectonic characteristics, these tools can also be used to sub-regionalize study areas, especially along fault lines. The characteristics of high magnitude earthquakes in the study area along and around the fault line were analyzed. The *b*-value was calculated as 0.92 for the lowest magnitude cut off value. Large earthquakes can be fit to a spatially inhomogeneous Poisson process because they are affected by the shape of the fault line and have highly diverging spatial intensities. With the help of these analyses, it can be concluded that large magnitude earthquakes occur mostly in the western part and the central part has less seismicity.

ACKNOWLEDGEMENTS

This work is supported by Anadolu University Project Unit under project no 1601F002. The article is derived from Cenk İçöz's Ph.D. dissertation. R codes for the reproducibility of the figures can be found in the below github link.

https://github.com/cenkicoz26/doktora-tez-deprem-sekiller

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Cressie NAC. Statistics for Spatial Data Revised Edition. Hoboken, New Jersey: Wiley; 1993. [CrossRef]

[2] Al-Ahmadi K, Al-Amri A, See L. A spatial statistical analysis of the occurrence of earthquakes along the Red Seafloor spreading: clusters of seismicity. Arab J Geosci 2014;7:2893–2904. [CrossRef]

[3] Öztürk S. Kuzey Anadolu Fay Zonu ve civarındaki güncel deprem aktivitesinin bölgesel ve zamana bağlı analizleri. Yerbilimleri, Hacettepe Üniversitesi Yerbilimleri Uygulama ve Araştırma Merkezi Bülteni 2017;38:193–228.

[4] Akol B, Bekler T. Assessment of the statistical earthquake hazard parameters for NW Turkey. Nat Hazards 2013;68:837–853. [CrossRef]

[5] Sayil N. Long-term earthquake prediction in the Marmara region based on the regional time- and magnitude-predictable model. Acta Geophys 2013;61:338–356. [CrossRef]

[6] Coban KH, Sayil N. Different probabilistic models for earthquake occurrences along the North and East Anatolian fault zones. Arab J Geosci 2020;13:1–16. [CrossRef]

[7] Sayil N. Evaluation of the seismicity for the Marmara region with statistical approaches. Acta Geod Geophys 2014;49:265–281. [CrossRef]

[8] Ketin I. Über die tektonisch-mechanischen Folgerungen aus den grossen anatolischen Erdbeben des letzten Dezenniums. Geologische Rundschau 1948;36:77–83. [Deutsch] [CrossRef]

[9] Bohnhoff M, Martínez-Garzón P, Bulut F, Stierle E and Ben-Zion Y (2016). Maximum earthquake magnitudes along different sections of the North Anatolian fault zone. Tectonophysics 2016;674:147–165. [CrossRef]

[10] Bayrak Y, Öztürk S, Çınar H, Kalafat D, Tsapanos TM, Koravos GC, et al. Estimating earthquake hazard parameters from instrumental data for different regions in and around Turkey. Eng Geol 2009;105:200–210. [CrossRef]

[11] Bozkurt E. Neotectonics of Turkey-a synthesis. Geodinamica Acta 2001;14:3–30. [CrossRef]

[12] Schabenberger O, Gotway CA. Statistical methods for spatial data analysis. Boca Raton, FL, USA: Chapman and Hall/CRC; 2017. [CrossRef]

[13] Illian J, Penttinen A, Stoyan H, Stoyan D. Statistical Analysis and Modelling of Spatial Point Patterns (Vol. 70). Hoboken, New Jersey: John Wiley & Sons; 2008. [CrossRef]

[14] Baddeley A. Analyzing spatial point patterns in R. Workshop Notes 2008;12:1–199.

[15] Ripley BD. Spatial Processes. New York: Wiley; 1981.

[16] Ripley BD. Modeling spatial patterns. J R Stat Soc Series B Stat Methodol 1977;39:172–192. [CrossRef]

[17] Aryal NR. Point Pattern Analysis. Dissertation submitted for Master of Science in Mathematics with Modern Applications, Department of Mathematics The University of York UK, 2011. http://citeseerx.ist.psu.edu/viewdoc/download?-doi=10.1.1.459.8842&rep=rep1&type=pdf

[18] Diggle PJ. Spatial analysis of Spatial Point Patterns (second Edition), London: Edward Arnold; 2009.

[19] Van Lieshout MNM, Baddeley AJ. A nonparametric measure of spatial interaction in point patterns. Stat Neerland 1996;50:344–361. [CrossRef]

[20] Silverman BW. Density estimation for statistics and data analysis. Oxfordshire: Routledge;1986

[21] Gatrell AC. Density Estimation and the visualization of point patterns, in: H.J. Hearnshaw, D.J. Unwin, (Eds.), Visualization in geographical information systems. Chicheste: John Wiley; 1994. p. 65–75.

[22] R. Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. Available at: https://www.R-project.org/ Last Accessed Date: 11.11.2024.

[23] Randy B, Buliung RN. and Remmel TK. Aspace: A collection of functions for estimating centrographic statistics and computational geometries for spatial point patterns. R package version 3.2, 2012. https://CRAN.R-project.org/package=aspace Last Accessed Date: 11.11.2024.

[24] Harte D, Brownrigg R. ssBase: Base Functions for SSL. R package version 2.3-6. Statistics Research Associates, Wellington, 2017. ftp://ftp.gns.cri.nz/pub/davidh/sslib/r-repo Last Accessed Date: 11.11.2024.

[25] Naylor M, Orfanogiannaki K, Harte D (2010). Exploratory data analysis: magnitude, space, and time. Community Online Resource for Statistical Seismicity Analysis. doi:10.5078/corssa-92330203

[26] Baddeley A, Rubak E, Turner R. Spatial Point Patterns: Methodology and Applications with R. Boca Raton, FL: Chapman and Hall/CRC; 2015. [CrossRef]

[27] Aki K. Maximum likelihood estimate of b in the formula log N= a-bM and its confidence limits. Bull Earthq Res Inst Tokyo Univ 1965;43:237–239.

[28] Öztürk S. Characteristics of seismic activity in the Western, Central and Eastern parts of the North Anatolian Fault Zone, Turkey: Temporal and spatial analysis. Acta Geophysica 2011. [CrossRef]