CENTROID ANALYSIS FOR EARTHQUAKES IN TURKEY

Boran Simsek[1], Dursun Zafer Seker[2] and Caner Guney[2]
[1]ITU, Informatics Institute Applied Informatics Department Geographic Information Technologies Master Program 34469 Maslak Istanbul, Turkey
Email: simsekbo@itu.edu.tr
[2] [1]ITU, Civil Engineering Faculty, Department of Geomatics Engineering 34469 Maslak Istanbul, Turkey
Email: (seker, guneycan)@itu.edu.tr

KEY WORDS: Centrography, Weighted mean center (spatial mean), Standard distance deviation, Web GIS

ABSTRACT: Geo-statistics is one of the most powerful mathematical method used in analysis of a collected spatial data, which depends on different variables, such as the distribution or the density of natural disasters, crimes and population. Centrography is a one of the efficient GIS based spatial analysis technique which allows to determine the average position, the distribution and/or the spatial change of a phenomenon along time. In this technique, the spatial mean is calculated to define average location and the standard deviation is calculated to determine the distribution of a selected dataset. In this study, centrography is applied to define the density of locations and the distributions of the earthquakes occurred in Turkey between 2000 and 2016. For this, weighted mean center and the standard distance deviation methods were used. Since a geographical information system contains storing, querying and displaying steps, the analyses on a big data need a computer based application, which consists of a database and presentation parts. Therefore, a web application works with a database, which stores the collected data, to display and query the information was created. The density locations and the distributions of the related earthquake dataset were obtained. For this reason to visualize the obtained results from the dataset an application on web was realized. Due to centrography is a reusable method and provides more accurate predictions on recursive and cumulative data, the obtained results presented that it is very powerful method to be applied on a dataset of the natural disasters on account of their tends to iterate themselves.

1. INTRODUCTION

Centrography is one of the GIS techniques, which allows one to assess and measure the average location, dispersion, movements, and directional change of phenomenon through time (LeBeau, 1987). In this technique, the position of a point, which is calculated, represents the average location according to the entire spatial data. Centrography is used to analyze the spatial-temporal characteristics and properties of the data set in related fields. The terms “centrography” and “centrogram”, suggested by the Mendeleev Centrographical Laboratory and the term “centrography” expresses compactly the group of geographical studies in this field of two-dimensional statistical analysis (Sviatlovsky and Eells, 1937).

At the beginning, these analyses were bounded by the studies about human populations. For instance, the studies in 1937, which were about the movement of the center of population in Europe and Northern America, showed that there was a contrary on the direction of movements. The movement in Europe had a tendency toward the east while the movement in Canada and in the United States had toward the west. In addition, the information was obtained that the movement in the northern hemisphere had been away from the Atlantic Ocean and toward the Pacific Ocean. In the light of this information, the studies suggested that the Pacific Ocean may become the concentration of the population. The result of these studies can be seen in Figure 1.

Subsequent studies, which this method has been applied, have been addressed different issues such as agricultural crops, natural disasters, criminology and a variety of regional distributions of other social and economic factors. In 1987, this method had been used by LeBeau, who is a professor at the Department of Criminology and Criminal Justice in Southern Illinois University, in order to analyze the 5 year dataset of lone-assailant rapes classified by type of offender (LeBeau, 1987). Mamuse et al, 2009 made a spatial centrographic analysis of mineral deposit clusters, using the komatiite-hosted Kambalda nickel sulphide deposit cluster. This study shows that spatio-geometric partitions are plausible locales for spatial analysis of nickel orebodies and endowment (Figure 2). The centrographic approach is potentially useful in mineral resource estimations and in mineral exploration targeting and can be considered as a sample study in recent years.

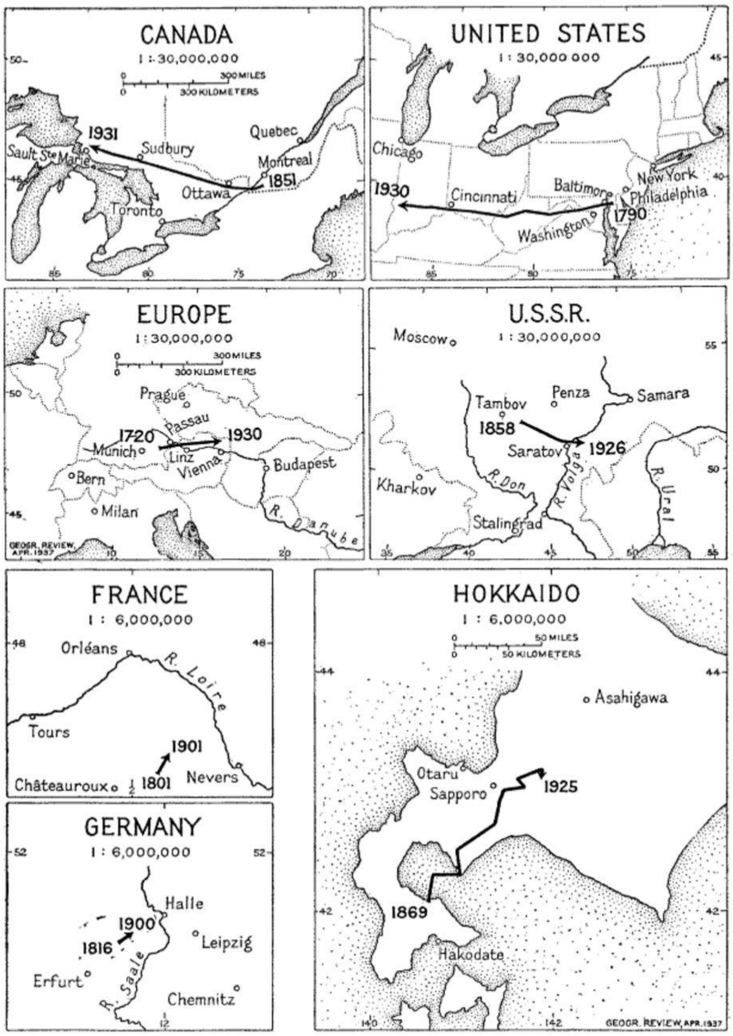

Figure 1. The trends of the center of the population (Sviatlovsky and Eells, 1937).

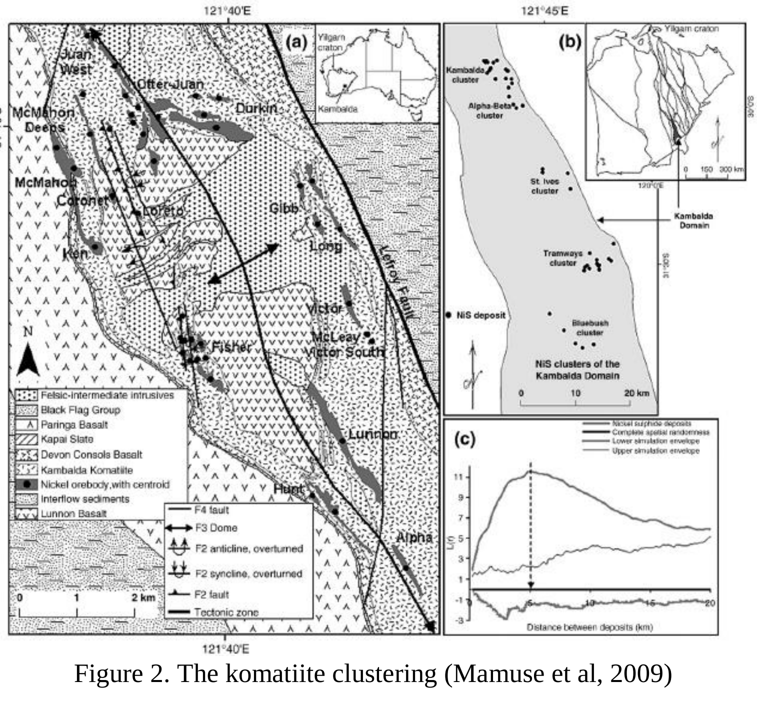

Figure 2. The komatiite clustering (Mamuse et al, 2009)

2. METHODOLOGY

There are several mathematical equation are available to calculate the center points according to the needs of the topic. Some of these methods are mean center, which is also known as spatial mean, weighted spatial mean and median center.

2.1 Mean Center (Spatial Mean)

The mean center is a central or average location of a set of points (Wong, 2005). The calculation of the mean center contains the summation of X and Y values of each coordinates individually and dividing each of these sums by the number of coordinates. This calculation generates the mean X and Y values, which produce the average coordinate according to the spatial dataset. The formula of this calculation is as follows:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Table 1. Sample Coordinates for Mean Center

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
X	17.112	21.002	18.442	18.856	20.518	19.794	17.483	19.639	20.168	18.964
Y	26.443	31.312	27.745	29.568	30.214	28.645	27.826	28.194	29.477	27.365

$$\sum_{i=1}^{10} Xi = 191.978 \qquad \bar{X} = 19.198$$

$$\sum_{i=1}^{10} Yi = 286.789 \qquad \bar{Y} = 28.679$$

The spatial mean point of the example dataset is located at (19.198, 28.679).

2.2 Weighted Mean Center

The weighted mean center is the average location, which is produced by weighting coordinates with another variable. In this case, the coordinates of the weighted mean center ($\bar{X}_W, \bar{Y}_w$), are given by:

$$\bar{X}_W = \frac{\sum_{i}^{N} X_i w_i}{\sum_{i}^{N} w_i} \qquad \bar{Y}_w = \frac{\sum_{i}^{N} Y_i w_i}{\sum_{i}^{N} w_i}$$

Where, w is the weight for each value.

Table 2. Sample Coordinates for Weighted Mean Center

X	4.56	5.21	3.89	6.05	6.75	4.99	7.01	5.5	4.75	7.44
Y	7.33	8.41	5.78	4.36	6.06	3.24	5.02	7.47	5.52	8.88
Magnitude	5.50	6.40	7.90	8.20	4.40	7.50	8.00	4.70	3.80	5.60

$$\sum_{i}^{N} X_i w_i = 347.53 \qquad \bar{X}_W = 5.60$$

$$\sum_{i}^{N} Y_i w_i = 372.49 \qquad \bar{Y}_W = 6.01$$

2.3 Median Center

The median center for a dataset is the point, which the half of the values in the dataset are smaller while the other half is larger than this value. When the dataset contains two-dimensional values such as locations then the median center is described as the point where 50% of the values fall east of north/south line and 50% fall west of that line, while an east/west line divides the values where 50% fall north of the line and 50% fall south of that line. The point, at which these lines intersect, is the median center of the locations (Winkle, n.d.).

The median center for a dataset is calculated by arranging in the values either ascending or descending order. The value at medium represents the median value. If the size of the dataset is even, the average of the two values at middle gives the median value. For a dataset contains two dimensional values such as locations, the median center is calculated by finding median values of each component of the points in the dataset individually and generating a new point with these median values (Figure 3).

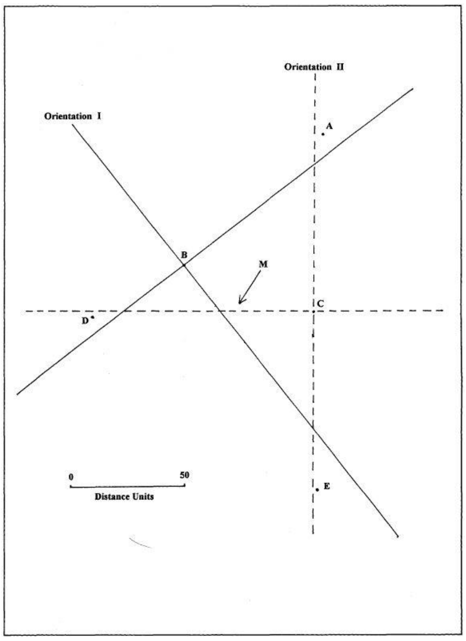

Figure 3. Median Center (Stoddart R., 1993)

2.4 Standard Distance Deviation

Standard distance deviation(also known as standard distance and alaogous to standard deviation inclassical statistics) (Furfey, 1927; Bachi, 1957), is a distance that is applied as the radius of a circle known as standard distance circle(SDC) centered at the spatial mean center of the spatial point pattern (Mamuse, 2009). A large standard distance deviation indicates that the data points can spread far from the mean center and a small standard distance deviation indicates that they are clustered closely around the mean center. The formula to calculate the standard distance deviation is given by:

$$\sigma_{\bar{X}\bar{Y}} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2 + \sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N}}$$

where and values represent the components of the center point, which is calculated with one of the “Mean Center”, “Median Center” and “Weighted Mean Center” methods.

Table 3. Sample Coordinates for Standard Distance Deviation

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
X	17.112	21.002	18.442	18.856	20.518	19.794	17.483	19.639	20.168	18.964
Y	26.443	31.312	27.745	29.568	30.214	28.645	27.826	28.194	29.477	27.365

$$\sum_{i=1}^{10} Xi = 191.978 \qquad \bar{X} = 19.198$$

$$\sum_{i=1}^{10} Yi = 286.789 \qquad \bar{Y} = 28.679$$

$$\sigma_{\overline{XY}} = 1.838$$

This result indicates that the points are mostly located in the circle, whose center is at (19.198, 28.679), with the radius of 1.838.

3. CASE STUDY: EARTHQUAKE ANALYSIS IN TURKEY

In this study, the spatial-temporal earthquake data with magnitude were gathered from the web services of United States Geological Survey (USGS). These web services provide the real time data and they are reachable from anywhere because of being open source.

In this study, a web application was created in order to gather the dataset from these web services and it is written by using Microsoft's C# language on ASP.NET web technologies on Visual Studio 2015 platform. Furthermore, Microsoft's SQL Server 2012(MS-SQL) was used to store earthquake information on database and provide ability to work offline. As a mapping server, Open Street Map (OSM)'s map was used and the javascript libraries of MapBox were implemented to manipulate and analyze the data on this map,

First of all, the dataset of earthquakes, which have magnitude higher than 3.0 and between (35, 25), (35, 46), (43, 25) and (43, 46) coordinates from 01.01.2000 to 01.04.2016, was selected to make centrography analysis. Then, the dataset, which was taken as a response in geojson format from the web services of USGS by using these parameters, were parsed into small objects in order to lessen the data volume by removing the unnecessary values for this analysis. These data were inserted into database to be used on the web application.

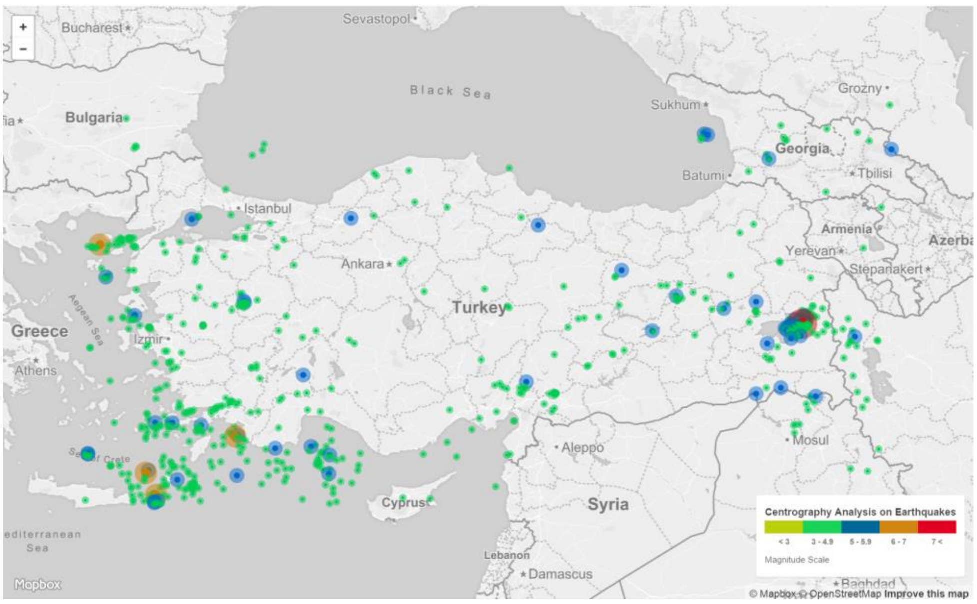

Figure 4. The earthquakes with magnitude higher than 4.0, from 2010 to 2016

Earthquakes were analyzed and centralized partially, which means that the whole field was divided into small latitudial and longitudial intervals. In order to calculate the center points of the earthquakes in these intervals, the weighted mean center method was used. The latitude and longitude values were representing X and Y values while the magnitude values were being used as the weight of each earthquake.

Subsequently, the standard distance deviation method was used to define the radius' of the circles, whose centers were calculated by weighted mean center method, in each interval. Firstly, the standard deviations of the latitudes and the longitudes were calculated in degrees. Since the distance between consecutive longitudes differs according to the latitude, the exact radius were calculated by applying the Haversine Formula to the latitude and longitude ranges, which were produced by adding and subtracting the standard deviations to the center point. The Haversine Formula is:

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right)$$

$$c = 2 \cdot \arcsin(\min(1, \sqrt{a}))$$

$$d = R \cdot c$$

Where φ is latitude, λ is longitude and R is the radius of the Earth, which is 6371 km. The unit of the latitude and longitude values were converted to radian unit. This calculation produces the radius of the distribution in km.

Finally, a confidence interval was applied to the circles in order to eliminate the ones, which doesn't cover enough earthquakes or covers wide area, to be considered as a correct centralization.

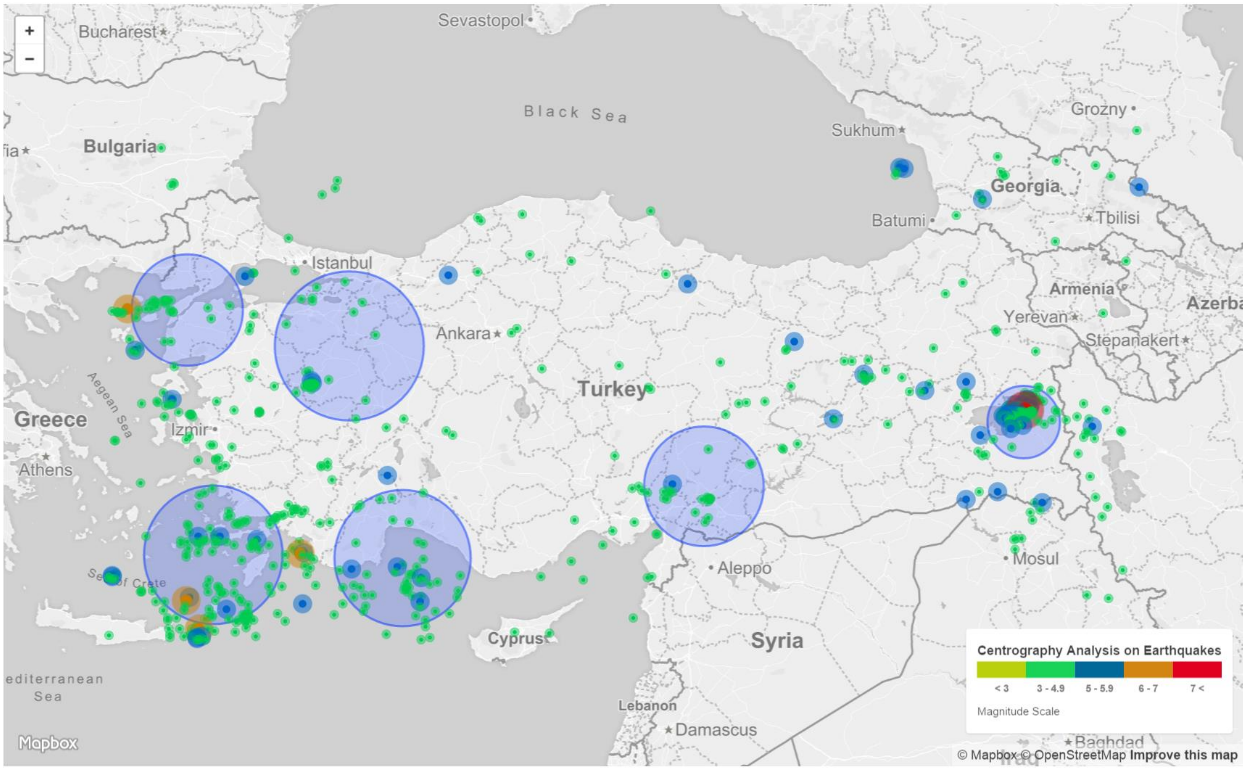

Figure 5. Centralization of the earthquakes with magnitude higher than 4.0, from 2010 to 2016

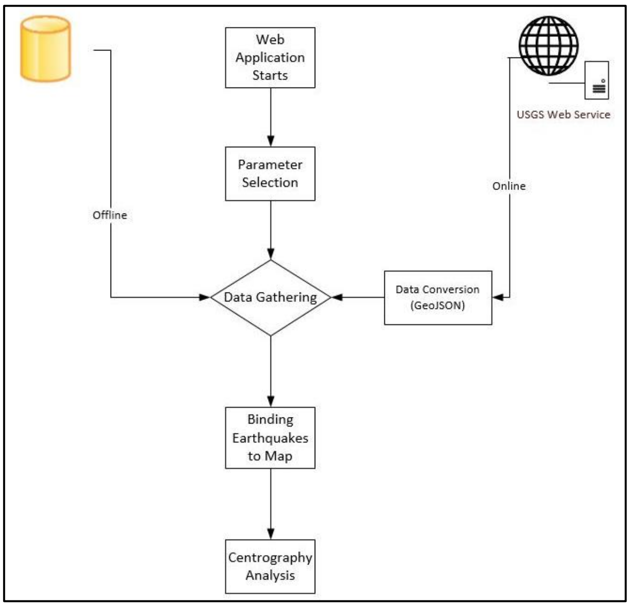

Figure 6. Flow diagram of the web application

4. DISCUSSIONS ABD CONCLUSIONS

We analyzed the earthquakes in between 2010 - 2016 in Turkey by applying a couple of centrography methods. In particular, we used (selected) weighted mean center and standard distance deviation technics: Firstly, the whole field which earthquakes happened was divided into small latitudinal and longitudinal intervals. The former technic is used for calculating the center of each interval and determining the earthquakes in which related intervals while the latter is applied to determine the radius of the distribution in each related intervals. Such an approach provides more precise analysis especially on wide areas. The results, which is related study can be summarized as follows:

1. This study shows that centrography can be a useful method with regards to centralize the natural disasters, analyze the spatial temporal distributions and interpret the results.

2. Since natural disasters tend to recur themselves, centrography method provides a powerful prediction about the future process of such recursive events, considering recent results.

REFERENCES

Bachi, R., 1957. Statistical Analysis of Geographical Series, Bulletin de l'Institut International de Statistique, Vol. 36, pp. 229-269.

Furfey, P. H., 1927. A Note on LEFEVER's 'Standard Deviation Ellipse, American Journal of Sociology, Vol. 33, pp. 94-102.

LeBeau, J. L., 1987. The methods and measures of centrography and the spatial dynamics of rape, Journal of Quantitative Crimonlogy, Vol. 3, Issue 2, pp. 125-141.

Mamuse, A., Porwal, A., Kreuzer, O., Beresford, S., 2009. A new method for spatial centrographic analysis of mineral deposit clusters, Ore Geology Reviews, Elsevier, pp. 293-305.
Sviatlovsky, E.E., Eels, W.C., 1937. The Centrographical Method and Regional Analysis, Geographical Review, Vol. 27, No. 2, pp.240-254.

Winkle, C., n.d., Center of Population Calculation for the City of Bloomington, Indiana, Department of Information and Technology Services of City of Bloomington, pp. 1-3, Retrieved: July 15, 2016, from: https://bloomington.in.gov/media/media/application/pdf/4013.pdf

Wong, D.W.S., Lee, J., 2005. Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS, John Wiley and Sons, pp. 52-55.