Exploratory Analysis of the Last 20 Years of Earthquakes in Turkey

Nilay Çiçekli
Computer Engineering Department
FMV Işık University
İstanbul, Turkey
nilay.cicekli@isik.edu.tr

Faik Boray Tek
Computer Engineering Department
FMV Işık University
İstanbul, Turkey
boray.tek@isikun.edu.tr

***Abstract*— Earthquakes are natural disasters that are unpredictable and have sociological and economical costs. There are studies to predict earthquakes using several methods such as statistical and deep learning methods. Yet, there is no truly successful method for such a prediction. Earlier studies show that better results for prediction require better data analysis on bigger historical data and require more parameters (e.g. weather data) are involved in the research process rather than using merely earthquake data. This paper represents a comprehensive analysis of earthquakes in Turkey in the past 20 years and, also associates the earthquake occurrence with climate using data mining techniques.**

Key words: *Earthquake, data science, data analysis, polynomial regression*

I. INTRODUCTION

Earthquakes are the results of the Earth's crust movement. At the surface, they appear by shaking. They might lead to loss of life and destruction of property. So, they have sociological and economical costs.

The World's costliest earthquake is recorded as the 2011 Tōhoku Earthquake and Tsunami in Japan. According to the Japanese National Police Agency the earthquake caused over 15,000 deaths, and the property damage valued at about $235 billion making it not only the costliest earthquake ever but also the most costly natural disaster ever [1].

In Turkey, in 1999 Izmit Earthquake over 17,000 people died, even more than the Tōhoku earthquake. And it resulted in more than 250,000 people becoming homeless [2].

There are some studies which approach earthquakes from a data analysis perspective like our study. For example, Yılmaz et al. [3] studied the prediction of the next earthquake in the North Anatolian Fault Zone in Turkey with statistical methods. Goswami et al. [4] studied the literature of data mining techniques on natural disasters where they reviewed the data mining models that were applied to various types of disaster data. Wang et al. [5] studied earthquake prediction with a deep learning technique to learn spatial-temporal relationship among earthquakes in different locations to confirm their claim that earthquakes does not have only local effects and they resulted that two-dimensional input made better earthquake predictions. Jin A. et al. tried to find a relationship between date, location, magnitude, and depth of earthquake using global earthquake data since 1973 [7]. Polat and colleagues [6] analyzed the earthquake hazard of the Aegean Region of Turkey. Their analysis covered earthquakes from 1900 to 2002 for the West Anatolian Region.

We found no comprehensive study of earthquakes or its relation to climate data from a data science perspective. In this study, we hope to fulfill this gap and investigate some interesting research questions.:

1. How does the number of earthquake occurrence change over the years in Turkey? Is there a trend?
2. Which region of Turkey has the highest number of earthquakes?
3. Is there an "earthquake season"?
4. How is Marmara Region's and Istanbul's situation in particular?
5. Drawing Turkey's fault-line map by using the 20-years of data.
6. Is there a pattern between the weather and the earthquake occurrence?
7. Can we predict the average daily number of earthquakes monthly for the upcoming months in Istanbul?

II. DATA

In this study two different data sources are used, which are explained below. There are 159584 rows for the earthquake and 26000 rows for the climate data.

A. Kandilli Observation Center Data

A dataset of earthquakes with a magnitude of 2 to 9 in Turkey over the years 2000-2019. The dataset is generated by Kandilli Observation Center's data engine (see http://www.koeri.boun.edu.tr). The data have 19 columns.

- *ID:* Unique row number of the event.
- *Code*: Unique ID for the event that is assigned by Kandilli Observation Center.
- *Date*: Date of the event.
- *Time:* Origin time of the event (UTC).
- *Latitude, Longitude*: Coordination of event (in decimal degrees)
- *Depth*: Depth of the event in kilometers.
- *Type*: Ke (Earthquake) or Sm (Suspected Explosion).
- *xM*: Biggest magnitude value in specified magnitude values.

- *MD, ML, Mw, Ms, Mb*: Magnitude types (MD: Duration, ML: Local, Mw: Moment, Ms: Surface wave, Mb: Body-wave).

While conducting the study we added 2 more columns for region and season.

B. *NASA's Weather Data of Turkey*

A dataset generated using Turkey's coordinates by NASA's climate database engine. The dataset represents the climate data of Turkey in between 2000-2019. It is coordinate-based and shows 0.5 x 0.5 Degree Interannual Averages/Sums. The data have 16 columns that represent the values for each month, annual average, latitude, longitude, and parameter. Parameters are as follows:

- Minimum Temperature at 2 Meters (C)
- Maximum Temperature at 2 Meters (C)
- The Temperature at 2 Meters (C)
- Temperature Range at 2 Meters (C)
- Wet Bulb Temperature at 2 Meters (C)
- Relative Humidity at 2 Meters (%)
- Precipitation (mm)
- Maximum Wind Speed at 10 Meters (m/s)
- Minimum Wind Speed at 10 Meters (m/s)
- Maximum Wind Speed at 2 Meters (m/s)
- Minimum Wind Speed at 2 Meters (m/s)
- Dew/Frost Point at 2 Meters (C)
- Surface Pressure (kPa)

III. METHODS

All data analysis and experiments were done in Python using numpy [8] , pandas [9], matplotlib [10] , sklearn [11] libs. The analysis was prepared as a Jupyter Notebook and shared on Nilay Çiçekli's GitHub page. [12] The weather-earthquake correlation hypothesis was tested both by comparing the patterns for weather parameters and earthquake graphs and using numpy's correlation method. Linear and polynomial regressions [13][14] are used for earthquake prediction using 20 years of historical data.

IV. EXPERIMENTS & ANALYSIS

The data is grouped by day and the daily number of earthquake occurrences is found and shown in Fig. 1. The x-axis shows the order of days (20 years * 365 days = 7000 days) and the y-axis shows the number of occurrences. The greatest number of occurrences in one day is 505 on the 21st of July 2017. On the same day, there are several earthquakes with a magnitude of over 4. They are close by each other, around Muğla and Gökova Körfezi (Mediterranean). This is called foreshock; small earthquakes trigger larger ones.

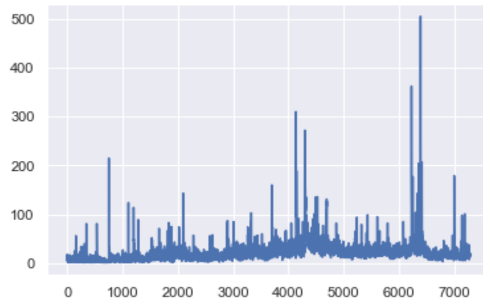

Fig. 1. Daily Number of Quake Occurrences

What is the distribution of earthquake magnitudes and how do they change in time? In Fig. 2. distribution of small and big earthquakes is seen over the years. Small earthquakes reach up to 16000 in a year while big ones are about 25 at most, and sometimes even less than 5. That means Turkey does not have many strong earthquakes. Turkey is a country that has smaller but frequently occurring earthquakes.

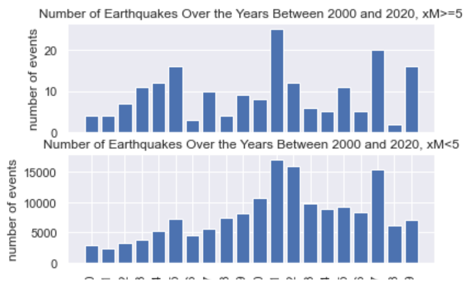

Fig. 3. Yearly Number of Small (xM<5) and Big (xM>5) Earthquakes in Turkey in One Bar Chart

Next graph combines the previous results in one chart to show the relation between small and big earthquakes. In Fig. 3. yearly number of small and big earthquakes are shown next to each other. As seen on this bar chart, sometimes small size earthquakes get the upper hand, sometimes great ones. Therefore, this study is not enough on its own to come to the conclusion that great magnitude earthquakes trigger small-sized earthquakes or vice versa.

The next analysis explores the regional distribution of earthquakes of Turkey. The experiment made by 159584 data points and the results showed that over 20 years, the earthquakes occurred in the Aegean region the most, with a number 54658 in total, that is over 34% of all the data. The following regions are Eastern Anatolia, Mediterranean, Marmara, Central Anatolia, Black Sea, and Southeastern Anatolia in decreasing order. A pie chart representing this result is shown in Fig. 4. with percentages. There is also a grey slice in the pie chart that shows the borders of Turkey with 1.22%.

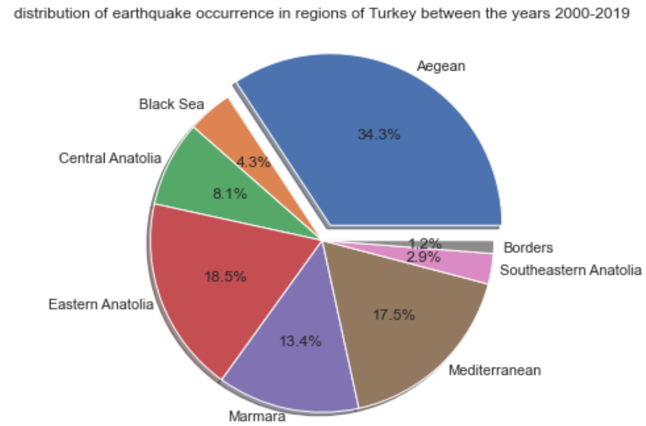

Fig. 4. Distribution of Earthquake Occurrence in Regions of Turkey between 2000-2019

The distribution of the earthquakes according to the regions of Turkey is also made for each year between 2000-2019. Over the years, it is observed that the Aegean region generally dominates other regions, it reaches a peak in 2002 and 2005 with almost 50%. However, there are times that the Marmara region increases unexpectedly. It even dominates the Aegean region by 43% in 2001 while it is 10% in 2005 and 2007; and 7% in 2011 and 2012. This situation proves that the Marmara region does not have a specific interval and it is unstable. This instability of the Marmara and regional distribution of earthquakes over 20 years is shown in figures Fig. 5.a – Fig. 5.c. You can see the rest of the plots of regional distribution of earthquakes in the Jupyter Notebook on GitHub. [12]
Due to observing the changes better, line charts are drawn for each region separately. Some of those charts are represented in Fig. 6.a - Fig. 6.b. You can see the charts on GitHub. [12]
It is observed that the Aegean region is moving around 35% and the Marmara region is around 15% mostly, but the Marmara region has a dramatic difference between its highest and lowest points.

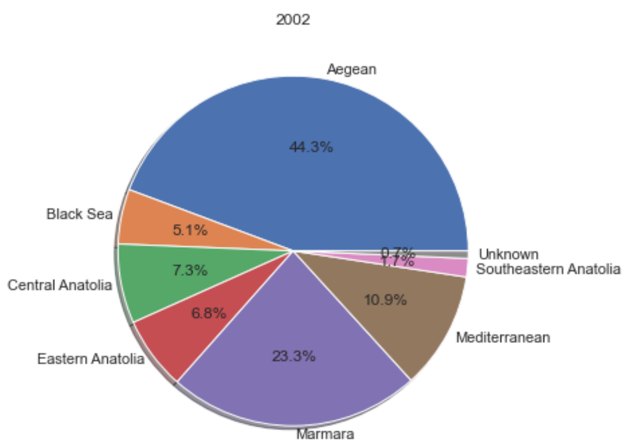

Fig. 5.a. Regional Distribution of Earthquakes of Turkey in 2002

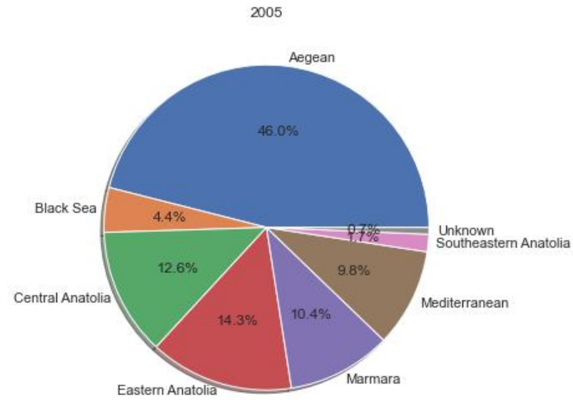

Fig. 5.b. Regional Distribution of Earthquakes of Turkey in 2005

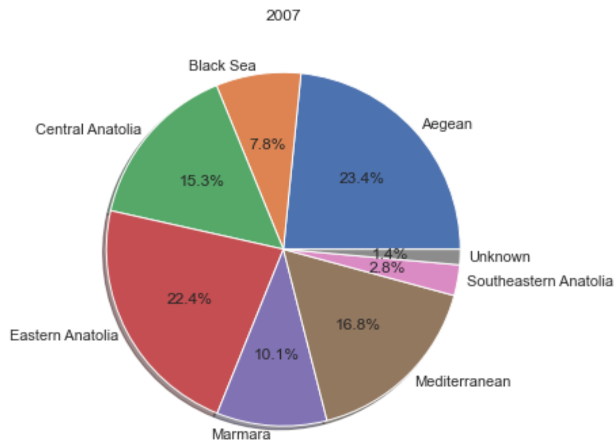

Fig. 5.c. Regional Distribution of Earthquakes of Turkey in 2007

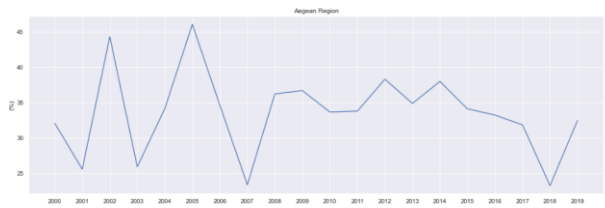

Fig. 6.a. Percentages of Earthquakes Over 20 Years for Aegean Region

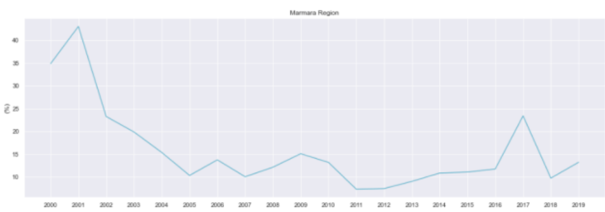

Fig. 6.b. Percentages of Earthquakes Over 20 Years for Marmara Region

Another research is to find if there is any specific "earthquake season", meaning "is there a season where earthquakes occur more often than other seasons?". Due to find an answer, we classified the earthquakes into months. December, January, February are winter; March, April, May are spring; June, July, August are summer and September, October, November are autumn in Turkey. We did not observe a specific season where earthquakes occur more than other seasons (see Fig. 7).

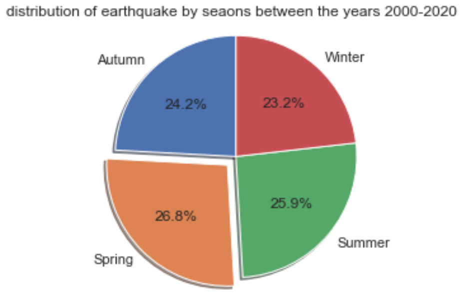

Fig. 7. Distribution of Earthquakes by Seasons in Turkey

After analyzing the Istanbul data, a graph is drawn that shows the number of earthquakes that occurred in Istanbul between 2000 and 2019. (Fig. 8). The result shows that Istanbul was stable between the years 2004-2013, but around 2016 and 2017, earthquakes increased dramatically.

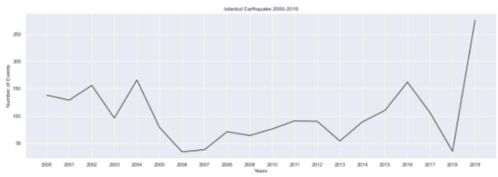

Fig. 8 Earthquake Occurrence in Istanbul between 2000-2019

Drawing the graph again daily, not only with the number of occurrences but with their magnitude results as in Fig. 9. This figure tells that magnitudes of earthquakes in Istanbul are smaller after 2010 compared to the period 2000-2010.

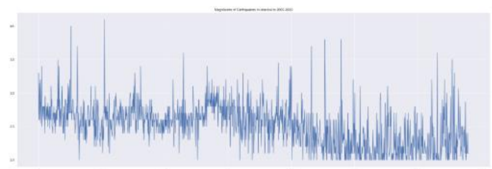

Fig. 9. Magnitudes of Earthquakes in Istanbul in 2000-2020, Daily

Another study we made was about the depth of the earthquakes. According to Kandilli Observatory, there are three depth zones. Shallow, intermediate, and deep earthquakes. Deep earthquakes are 300-700 km deep in the ground and they are not seen in Turkey. Intermediate is 70-300 km deep, and shallow ones are 0-70 km deep. The analysis shows that the earthquakes in Turkey are mostly shallow quakes. And, sadly they cause the most damage. As seen in Fig. 10 and Fig. 11, intermediate earthquakes are less frequent and range is 80-180 km, shallows are dense in the interval 0-30 km.

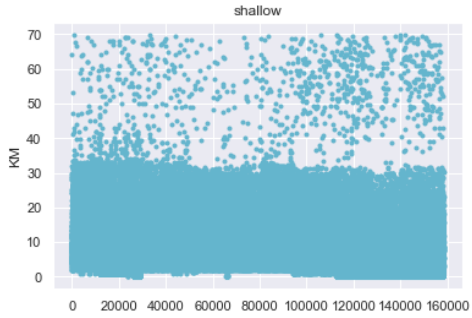

Fig. 10. Shallow Earthquakes of Turkey

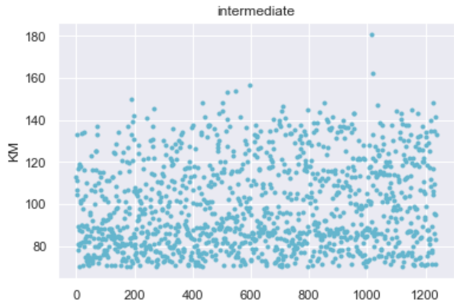

Fig. 11. Intermediate Earthquakes of Turkey

Then we studied the distribution of shallow and intermediate earthquakes according to regions. Since shallow ones dominate and the Aegean region is already found as the most earthquake occurring region before in another experiment in this paper (Fig.4), the distribution of shallow earthquakes is not so different from the previous experiment's result. In Fig. 12.a this result is shown and while this figure is compared with Fig. 4., similarity can be observed.

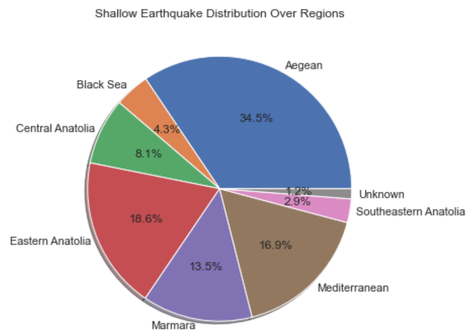

Fig. 12.a. Shallow Earthquake Distribution for Regions of Turkey

However, the result for intermediate earthquakes is more interesting. The result shows that most intermediate ones, 89.5%, happened in the Mediterranean region in the past 20 years. This result proves that Turkey is a shallow earthquake area except for the Mediterranean. (Fig. 12.b)

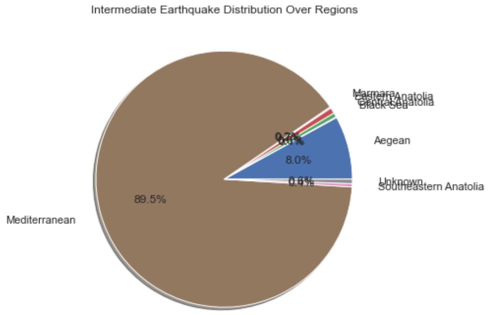

Fig. 12.b. Intermediate Earthquake Distribution for Regions of Turkey

Finally, we plotted the depth distributions of the Marmara region in Fig. 13. The plot suggests that 2017 was a year where the number of earthquakes increased drastically in the Marmara. The same relationship was found in previous experiments for Istanbul in this paper. So, it can be concluded that Istanbul (and its close area) dominates the earthquakes of the Marmara region.

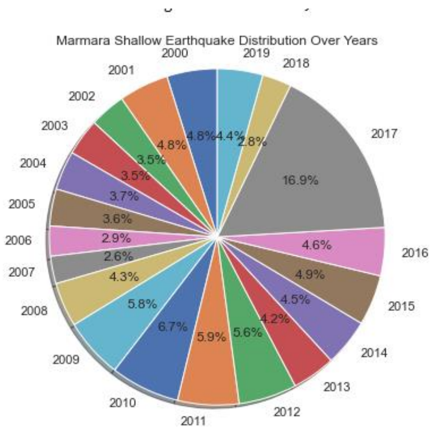

Fig. 13. Marmara Region's Shallow Earthquakes

Using all the data it can be drawn Turkey's fault line map by marking the coordinates for each earthquake. The result is convenient and represents a roughly drawn map. The map is given in Fig. 14.

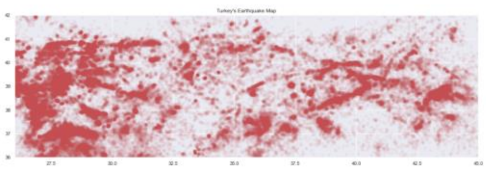

Fig. 14. Turkey Fault Line Map

In the next experiment, we study earthquake and temperature data together to see if there is a correlation on yearly basis. In Fig. 15 max, min, average temperatures, and the number of earthquakes are seen in the same graph. The number of earthquakes is downscaled to fit into the graph, the main idea was to see if there is a matching pattern.
The correlation coefficient is also investigated for min temp-number of earthquakes and max temp-number of earthquakes. Coefficients were -0.045 and -0.015 which are very low to call the relation a non-coincidental relation.

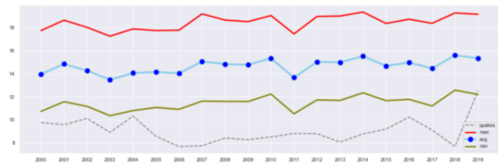

Fig. 15. Temperature and Earthquake

The same experiment is made with precipitation and humidity data combined with earthquake data separately. Resulting graphs are shown in Fig. 16 and Fig. 17. Either Fig. 15, Fig. 16 or Fig. 17 do not suggest any pattern between earthquake and weather data. The correlation scores for humidity and precipitation were 0.059 and -0.18 respectfully. Even though the correlation for precipitation level appears to be higher than the others, the score is not high enough to worth investigating. However, when the experiment is repeated for magnitude and depth of the earthquakes with weather parameters, some interesting results are found. A correlation score of 0.309 over 1 for humidity-magnitude relation, a score of -0.330 for max temperature-magnitude, and a score of -0.362 for min temperature-magnitude are found. These scores suggest that there might be a relation between weather and earthquakes somehow. Due to a lack of data, this paper has not taken further analysis on the topic, so it does not claim to have exhaustive conclusions but it does urge future research on the topic.

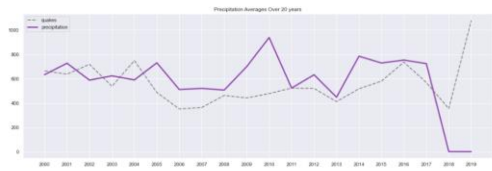

Fig. 16. Precipitation Levels and Earthquake

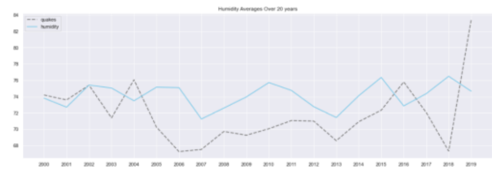

Fig. 17. Humidity Level and Earthquake

The very final experiment was on making a prediction of future earthquakes for Istanbul. Linear and polynomial regression techniques are used. Linear regression did not provide a good fit, but polynomial did. But towards the ends, an increase in the year 2017 affected the line wrongly. Linear and polynomial results are given on the same graph in Fig. 18. In conclusion, the prediction should be improved with other methods.
The linear prediction provided us a trend. There was a positive slope. Which means number of earthquakes increases over the years. The polynomial prediction suggested that from January to June in 2020, the daily number of earthquakes in Istanbul would be around 25. Normally, real values are in a range from 0 to 25. So, predicted numbers look viable but there are still question marks since 25 is the highest value normally.

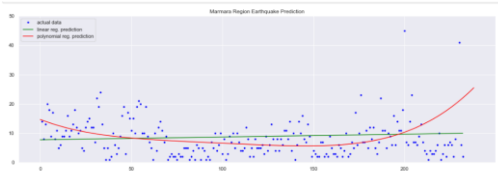

Fig. 18. Prediction Using Linear and Polynomial Regression

V. DISCUSSION

In this paper, we conducted an extensive earthquake analysis of Turkey using data mining techniques. Earthquakes over the past 20 years are in Turkey are investigated and represented in graphs. The regional analysis of earthquakes was also made. We observed that the Aegean region is the region that has the highest number of earthquakes in Turkey. But the Marmara is also a region at risk, due to the unstable nature of the region. We have shown that Istanbul and its nearby area dominate the earthquakes in the Marmara region. Also, we have seen that Turkey is a shallow earthquake area, and shallow earthquakes are the most damaging ones among all types.

Weather data of Turkey was studied in combination with earthquake data to find a correlation and there found possible relations between humidity-magnitude and temperature-magnitude. The paper does not have a certain conclusion, but the results are worth working on more.

Because the climate data did not seem correlated, it did not prove itself as a good candidate to be a second dimension for earthquake prediction. Hence, a prediction for earthquakes was made by using only earthquake data. Linear regression is used to observe the trend. The plot suggested that we will have more earthquakes. Polynomial regression is used to make the prediction. Because of lack of the second-dimensional data, prediction was not successful. But it was a good step to start with. This paper can encourage further studies to work relational analysis on earthquakes.

REFERENCES

[1] Sawe, B. (2018, July 12). The World's Costliest Earthquakes. Retrieved June 24, 2020, from https://www.worldatlas.com/articles/the-world-s-costliest-earthquakes.html

[2] Gurenko, Eugene; Lester, Rodney; Mahul, Olivier; Gonulal, Serap Oguz (2006). Earthquake Insurance in Turkey: History of the Turkish Catastrophe Insurance Pool. World Bank Publications. p. 1. ISBN 9780821365847.

[3] YILMAZ, Veysel; ERİŞOĞLU, Murat; ÇELİK, H. Eray. Probabilistic Prediction of the Next Earthquake in the NAFZ (North Anatolian Fault Zone), Turkey. Doğuş Üniversitesi Dergisi, [S.l.], v. 5, n. 2, p. 243-250, mart 2011. ISSN 1308-6979.

[4] Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Chakraborty, A review on application of data mining techniques to combat natural disasters, Ain Shams Engineering Journal, Volume 9, Issue 3, 2018, Pages 365-378, ISSN 2090-4479.

[5] Q. Wang, Y. Guo, L. Yu and P. Li, "Earthquake Prediction Based on Spatio-Temporal Data Mining: An LSTM Network Approach," in IEEE Transactions on Emerging Topics in Computing, vol. 8, no. 1, pp. 148-158, 1 Jan.-March 2020, doi: 10.1109/TETC.2017.2699169.

[6] Polat, Orhan, Elcin Gok, and Doğuser Yilmaz. "Earthquake hazard of the Aegean extension region (West Turkey)." Turkish Journal of Earth Sciences 17.3 (2008): 593-614.

[7] Lee, Jin A., JongGyu Han, and Kwang Hoon Chi. "Mining quantitative association rule of earthquake data." Proceedings of the 2009 International Conference on Hybrid Information Technology. 2009.

[8] Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. **The NumPy Array: A Structure for Efficient Numerical Computation**, Computing in Science & Engineering, **13**, 22-30 (2011), DOI:10.1109/MCSE.2011.37 (publisher link)

[9] Wes McKinney. **Data Structures for Statistical Computing in Python**, Proceedings of the 9th Python in Science Conference, 51-56 (2010) (publisher link)

[10] John D. Hunter. **Matplotlib: A 2D Graphics Environment**, Computing in Science & Engineering, **9**, 90-95 (2007), DOI:10.1109/MCSE.2007.55 (publisher link)

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. **Scikit-learn: Machine Learning in Python**, Journal of Machine Learning Research, **12**, 2825-2830 (2011) (publisher link)

[12] Çiçekli Nilay, Earthquake, (2020), GitHub repository, https://github.com/nilaycicekli/earthquake

[13] Duda, R. O., Hart, P. E.,, Stork, D. G. (2000). *Pattern Classification (2nd Edition)*. Wiley-Interscience. ISBN: 0471056693

[14] VanderPlas, J. (2017). *Python data science handbook: Essential tools for working with data*. Sebastopol, CA: O'Reilly.