MAT5314 Graduate Project Team 2

Teng Li(7373086)
Shiya Gao(300381032)
Chuhan Yue(300376046)
Yang Lyu(8701121)

Introduction

Earthquakes are usually caused when the underground rock suddenly breaks and there is rapid motion along a fault. They can be categorized into natural earthquakes and artificial earthquakes. Natural earthquakes can be further classified into tectonic earthquakes and volcanic earthquakes. Tectonic earthquakes are caused by the rupture of rocks deep in the ground and the rapid release of energy accumulated over a long period of time. Volcanic earthquakes are caused by volcanic eruptions. In addition, artificial earthquakes are caused by human activities that alter the stresses and strains on Earth's crust. In our project we will only focus on the natural earthquake due to its probabilistic nature of occurrence.

Earthquakes are difficult to predict for three main reasons. Firstly, plate boundaries are prone to earthquakes, but the way and speed of plate movement is difficult to measure and predict accurately. Secondly, the low frequency of large earthquakes does not provide enough data for earthquake modeling. Thirdly, it is difficult to enter the inner crust of the earth to observe the data, so most of the detection of earthquakes is done by collecting some vibrations and signals on the surface to analyze and predict them, which leads to a decrease in the accuracy of earthquake prediction.

Frequent historical earthquakes give us a huge data set that can be used to study earthquake causes, correlations, space-time, and further earthquake prediction. Current researches have provided several statistical models to analyze the earthquake processes. Trigger model, a special case of the Neyman-Scott clustering model, can be used to estimate aftershocks after a major earthquake (Ogata 1988). Epidemic-Type Aftershocks Sequence (ETAS) model is also widely used to forecast earthquake occurrences (ROSS and KOLEV 2022). In addition, time series analysis can be done to explore cycles related to earthquake frequency and it is also effective in predicting large earthquakes (Amei, Fu, and Ho 2012). There are also several techniques from machine learning that provide us alternative ways other than conventional statistical models to analyze the earthquake data. For example, deep learning can be used to predict seismic events, including intensity and location (Nicolis, Plaza, and Salas 2021). Clustering model can identify regions with high-frequency earthquakes to upgrade building structures to reduce damage from impending earthquakes.

We focus on a data set of earthquake occurrence in Canada provided by the Government of Canada. The data set is located at https://open.canada.ca/data/en/dataset/4cedd37e-0023-41fe-8eff-bea45385e469. In this project we want to analyze the data to give a detailed review of this natural disaster in the country and make some inference about its characteristics. In particular, we want to check if there's any location-wise pattern in earthquake occurrence and how frequent it can happen. Then we want to analyze if the occurrence of earthquake is also correlated to some other variables such as time. Finally we want to apply some probabilistic models that have been extensively studies by other researchers to the data in order to assess the likelihood of future occurrence in high-risk area.

Method

First, we will demonstrate some aspects of the earthquake occurrence by visualizing the seismic data. Map charts and other related plots will be used to perform such initial analysis. Then we will use time series analysis to study whether there exist temporal trends, seasonal patterns, and certain frequency in seismic events. Third, we will use geographic characteristics such as longitude, latitude and magnitude to identify potential groups of clusters. We may use different techniques of cluster analysis to achieve this goal. This will allow us to compare the predicted clusters with the actual location of earthquakes. Lastly, we will try to estimate the probability of earthquakes in high-risk regions by using the variables from the data and the Epidemic-Type Aftershocks Sequence (ETAS) model.

References

Amei, Amei, Wandong Fu, and Chih-Hsiang Ho. 2012. "Time Series Analysis for Predicting the Occurrences of Large Scale Earthquakes." *International Journal of Applied Science and Technology* 2 (7).

Nicolis, Orietta, Francisco Plaza, and Rodrigo Salas. 2021. "Prediction of Intensity and Location of Seismic Events Using Deep Learning." *Spatial Statistics* 42.

Ogata, Yosihiko. 1988. "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes." *Journal of the American Statistical Association* 83 (401): 9–27.

ROSS, GORDON J., and ALEKSANDAR A. KOLEV. 2022. "SEMIPARAMETRIC BAYESIAN FORECASTING OF SPATIOTEMPORAL EARTHQUAKE OCCURRENCES." *The Annals of Applied Statistics* 16 (4): 2083–2100.