Regression Analysis: Predicting Depth of Earthquakes

Woosub Shin

July 22, 2014

Abstract

The importance of predicting Earthquakes is increasing due to recent catastrophes such as those in Haiti and Japan. In this project we tried to predict the depth of hypocenters of earthquakes based on ISE-GEM cagalog. Using Regression Tree we divided the Earth's surface into thirteen smaller regions, and classified them into five groups. On each group we applied General Additive Model(GAM) and Random Forests. In order to improve the result of the analysis we need to improve measuring methods as well as to filter to find more accurate data.

1 Introduction

As recent catastrophic earthquakes in Haiti and Japan show, understanding earthquakes is becoming an increasingly important task for geologists. However this phenomena are not fully understood - such as predicting the time and location of a major earthquake - by scientists yet. This is because earthquake occurrence is characterized by extreme randomness.[1] In our proejct we tried to predict the depth of earthquake hypocenters based on ISC-GEM Global Instrumental Earthquake Catalogue between 1962 and 2009.[2] This will contribute to the better understanding of global earthquake data if combined with further research.

2 Background

2.1 Earthquakes

Earthquake occurs when the energy in the Earth's crust is suddenly released. There are three basic types of faults. The first is the Reverse fault, which happens when two opposite forces approach to each other so that one side goes up. The Normal fault occurs when two forces diverge so that one side is pulled down below the other. Finally, Strike-slip fault occurs when two parallel forces make the crack of the surface. Because of the basic formation of faults, it is well known that most earthquakes occur around tectonic boundaries. Since tectonic

[1] Yan Y. Kagan, "Earthquakes: Models, Statistics, Testable Forecasts (Hoboken, NJ, Wiley & Sons, 2014), 4

[2] More information can be found at http://www.isc.ac.uk/iscgem/

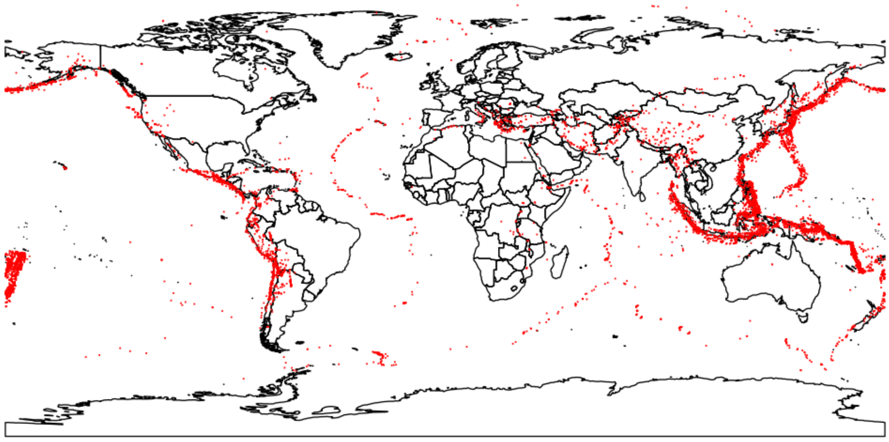

Figure 1: World earthquake map based on ISC-GEM

plates are constantly moving to certain direction the forces are stored until it reaches the threshold. Then finally earthquake happens.

2.2 Regression Analysis

Regression analysis is a statistical process for estimating the relationships among variables.[3] In prediction problem, our goal is to construct a model which predicts the response "y"(in our case, depth) using given predictors "x"(for example, moment tensor components, error ellipse parameters, latitude, etc.). R program is a strong tool which enables us to implement various techniques of regression analysis. Throughout our project we used R to analyze data.

3 Analysis and Results

3.1 Dividing Global Region

Figure 1 shows the plot of all earthquakesin the catalog between 1962 and 2009. We can easily recognize that this is closely related to plate tectonics since most of them are on the boundaries of tectonic plates. Figure 2 shows the nonparametric prediction based on longitude and latitude. Interestingly the contour lines have several centers, which means it would be appropriate to apply the tree model to divide the surface into several regions.
We applied the regression tree based on longitude and latitude, and the result is Figure 3. There are thirteen regions, which are labeled as 1 to 13 from left to right. Based on our knowledge of plate tectonics and geography, we could group them into five groups. They are shown in Figure 4 to Figure 8.

3.2 Comparing Groups

Our interest is in smaller regions, as they are sorted out first by the method of regression tree. We tried various techniques using GAM(Generalized Additive

[3] Wikipedia, "Regression Analysis", http://en.wikipedia.org/wiki/Regression_analysis

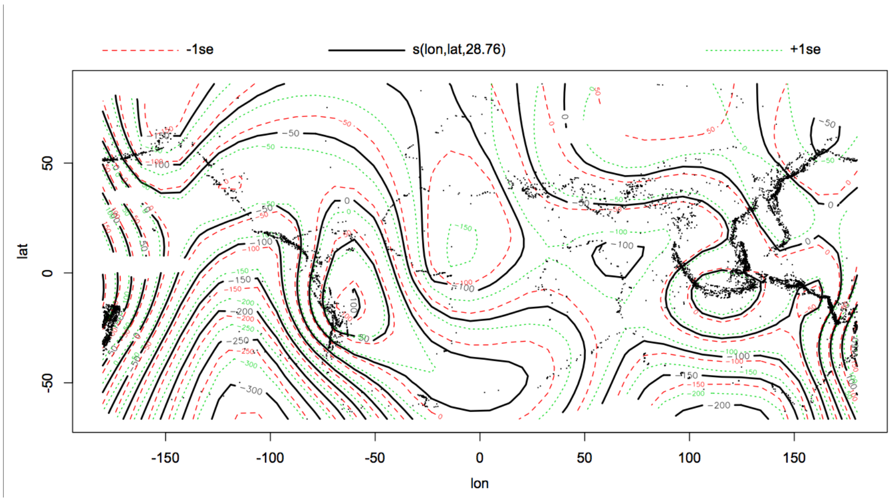

Figure 2: Nonparametric prediction of depth of earthquakes

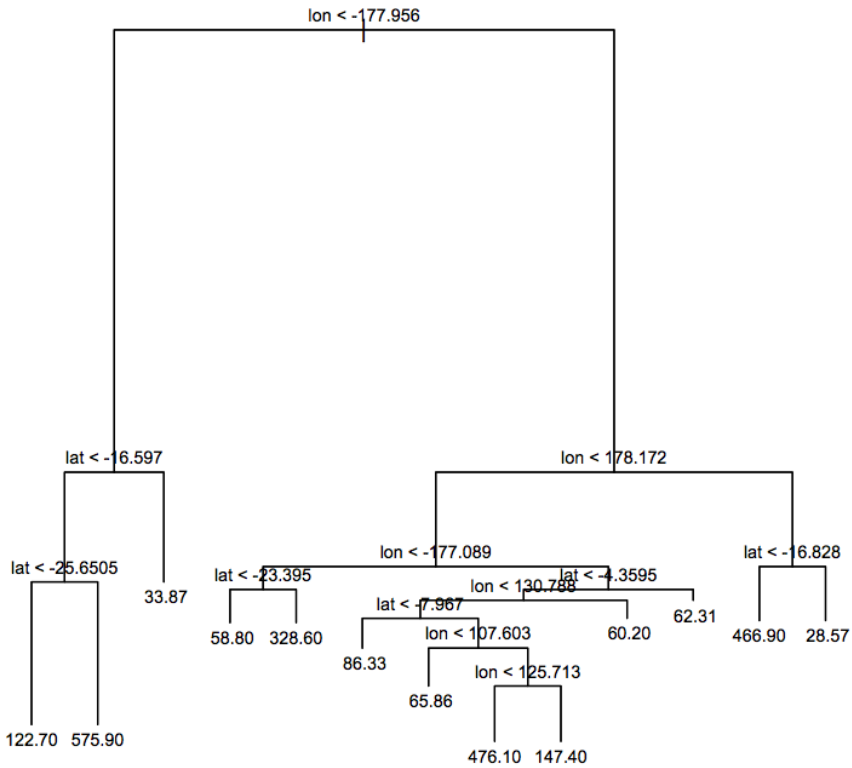

Figure 3: Regression Tree, based on lon(longitude) and lat(latitude)

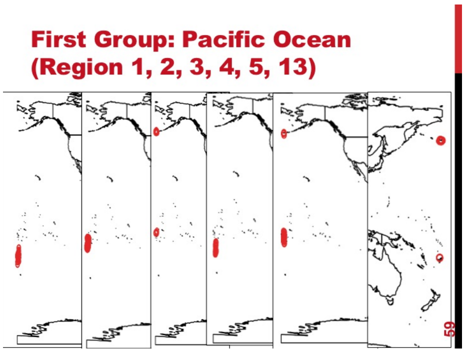

Figure 4: Group 1

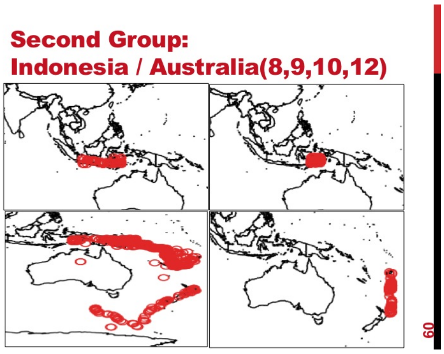

Figure 5: Group 2

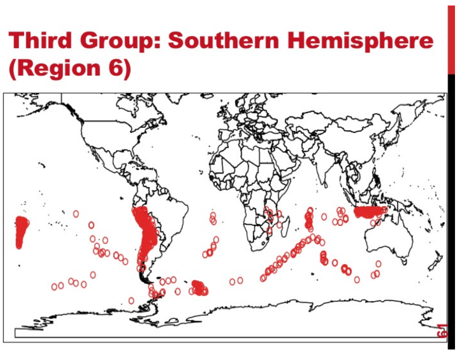

Figure 6: Group 3

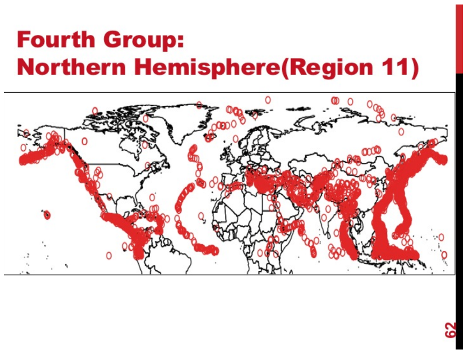

Figure 7: Group 4

Figure 8: Group 5

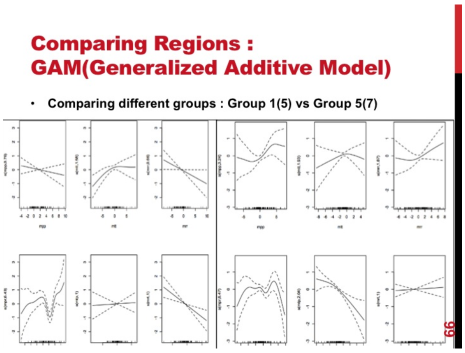

Figure 9:

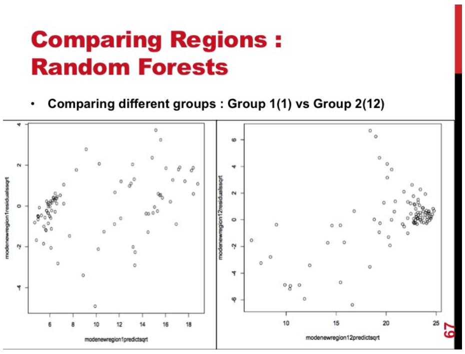

Figure 10:

Model) and Random Forests on those groups especially with six moment tensor components. Figure 9 and Figure 10 are such examples. As the figures show there are not coherent traits although each region shows certain characteristics.

4 Conclusion/Suggestion for Further Study

Given the previous results, we conclude that using tree model we could efficiently divide the Earth's surface into several regions/groups. However the figures also show that it is not sufficient to use GAM or Random Forests to predict the depth of earthquakes. More techniques can be applied to pedict depth in the future.

Based on our previous study we can suggest a few points to be improved. First, We can divide the surface using different variables such as magnitude, moment tensors, etc. Also for some regions such as californian fault, more detailed data are known to scientists. Combining individual in-depth data with global earthquake catalog will give us additional intuition to understand this phenomena more accurately. Finally we need to improve and expand our seismometers and filtering method as there are still some degree of uncertainty which makes things more difficult. If more data are stored and measuring tequniques are improved we could predict and understand the pattern of depth of earthquakes better.

5 Bibliography

Yan Y. Kagan, "Earthquakes: Models, Statistics, Testable Forecasts (Hoboken, NJ, Wiley & Sons, 2014)
Wikipedia, "Regression Analysis", http://en.wikipedia.org/wiki/Regression_analysis
http://www.isc.ac.uk/iscgem/