Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project
Project Plan

Forecasting induced seismicity in Oklahoma

by

Zhiyong Liu

zl1220@ic.ac.uk
GitHub login: acse-liuzyon

Supervisors:
Dr. Stephen P. Hicks

June 2021

Abstract

1 Introduction

It is well known that humans can cause earthquakes through fluid injection and extraction. Fluid injection includes wastewater disposal and hydraulic fracturing, while fluid extraction includes oil, gas and groundwater. Such cases of induced seismicity have been recorded and proven in Oklahoma, where seismicity has been increased dramatically since 2010 (Ellsworth 2013). In cases like Oklahoma, because the rate of earthquakes was very low before high-rate wastewater injection, it is relatively easy to determine the fundamental causes of induced seismicity. Therefore, due to the low background stress in intraplate regions, the human triggering signatures are clearly identifiable. However, in tectonically-active areas, it is more challenging to distinfuish between natural and triggered seismicity. Big-data and statistical approaches will be crucial in separating natural causes from potential human triggering factors in these areas. Besides, large-scale big-data and statistical approaches are effecient method to predict the locations where human triggered seismicity may be expected.

2 Literature Review

Ellsworth stated the current understanding of the causes and mechanics of human-induced earthquakes. It includes wastewater injection, emerging oil and gas recovery technologies, and other indirect induced activities, including deep fluid injection (Ellsworth 2013). Hincks et al. developed a Bayesian network to for quantitative evaulation of correlations between well operational parameters, geological formation, and seismicity. The injection depth relative to crystalline basement was found to be the most significant parameter for seismic moment release (Hincks et al. 2018). Wozniakowska and Eaton performed a machine learning estimation of the seismogenic activation potential for each well using logistic regression and found that the injection depth influence greatly the seismogenic activation potential (Wozniakowska & Eaton 2020). Norbeck and Rubinstein genreated a model based on fluid flow and seismic physics, which associate the past injection trends with the seismicity patterns (Norbeck & Rubinstein 2018).

The correlations of induced earthquakes with fluid extraction and injection are well established. Some physics models are also developed in previous research for earthquake prediction. However, these researches do not explain why seismicity does not always occured close to an injection well. Also, these studies did not develop a statistical model or a machine-learning model to predict earthquakes based on well documented data (such recorded data in Oklahoma)

3 Description of Problem and Objectives

Although a causal relationship has been established between fluid injection and earthquakes, it is still not known why some areas are more susceptible to induced earthquakes and why seismicity does not always occur close to an injection well. To solve these problems, statistical methods and machine learning methods are required. In this project, a unique and rich dataset of industrial activities from regions in Oklahoma will be used to statistically evaluate and retrospectively forecast possible signatures of human-induced seismicity from existing high-quality earthquake in these regions. The database used contains well locations, geological metadata, monthly injection and extration
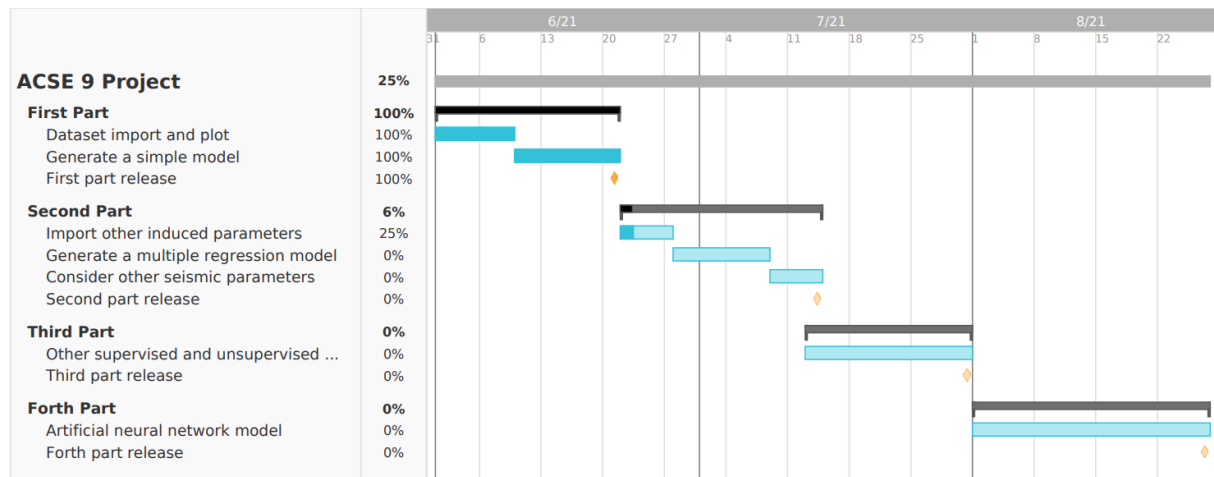
Figure 1: Gantt chart.

volumes. To test the existence of correlations of humen induced parameters with seismicity and predict the earthquake occurrences, some machine learning techniques will be developed to generate a model with predictive capabilities. Linear Regression and logistic Regression will be performed as one starting point to produce a simple model of earthquake occurence. Other supervised and unsupervised technologies will also be implemented to obtain the higher performance model. The main program will be coded in python, including using specific Python libraries such as Pandas, Scikit-Learn, statsmodels, matplotlib and some GIS modules.

4   Progress to Date and Future Plan

In this project, there are many potential parameters to investigate correlations with earthquake activity, which are injection volume, well depth and geological formation being injected into. The parameters of earthquake activity considered are earthquake occurence, maximum earthquake magnitude and earthquake moment rate. As Figure 1 shown, This project is divided into four parts. In the first part, dataset is successfully imported and plotted with geopandas. Then, as a starting point, injection volume will be concentrated on to find the correlation with the earthquake occurence and produce a simple model. Here, some regression techniques such as linear or logistic regression will be used for the model generateing. In the second part, other potential parameters like well depth and geological formation (sedimentary basin thickness, and geological faults) will be introduced to consider together and multiple regression models will be generated. Besides, other earthquake activity parameters (maximum earthquake magnitude and earthquake moment rate) will also be investigated correlations with induced potential parameters. In the third part, Other supervised and unsupervised technologies (K-means, PCA, etc.) will be used on the dataset to generate better models compared with linear regression and logistic regression. In the final part, Deep learning technologies such as Neural network models will be trained with all potential parameters and generated with predictive capabilities.

The development model used in this project is iterative model (Alshamrani & Bahattab 2015). In each part, requirements analysis, design, coding, testing will be carried out. Each iteration will result in a model which can be used to predict. As is demonstrated in the Figure 1, Each part will generate available models. For each generated model, a lot of evaluation criterions will be used to test predictive capability, such as MSE, RMSE, MAE, accuracy, confussion matrix, precision, recall and so on.

The datasets used for this project are extremely huge and varied. They are split into training dataset and testing dataset for generating models and testing. The preliminary main part of them are wells
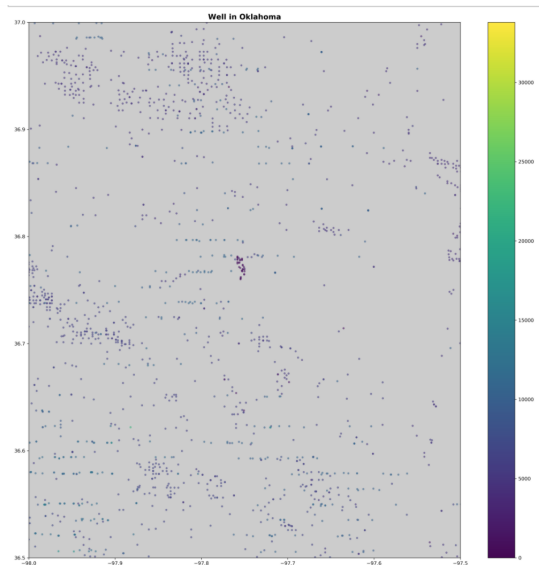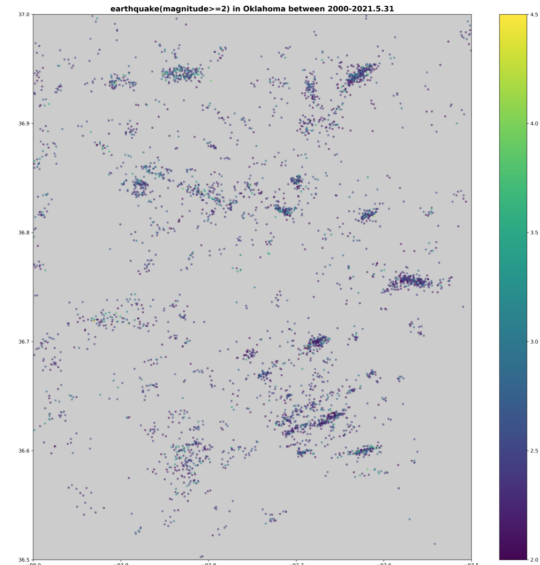
2

Figure 2: Well Depth in Oklahoma

Figure 3: Earthquake Magnitude in Oklahoma

and earthquakes, which are mapped and shown in Figure 2 and Figure 3. Other data such as injection volume and geological formation are also provided in csv and xlsx format.

References

Alshamrani, A. & Bahattab, A. (2015), 'A comparison between three sdlc models waterfall model, spiral model, and incremental/iterative model', *International Journal of Computer Science Issues (IJCSI)* **12**(1), 106.

Ellsworth, W. L. (2013), 'Injection-induced earthquakes', *Science* **341**(6142).

Hincks, T., Aspinall, W., Cooke, R. & Gernon, T. (2018), 'Oklahoma's induced seismicity strongly linked to wastewater injection depth', *Science* **359**(6381), 1251–1255.

Norbeck, J. & Rubinstein, J. L. (2018), 'Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity in oklahoma and kansas', *Geophysical Research Letters* **45**(7), 2963–2975.

Wozniakowska, P. & Eaton, D. W. (2020), 'Machine learning-based analysis of geological susceptibility to induced seismicity in the montney formation, canada', *Geophysical Research Letters* **47**(22), e2020GL089651.

Date	Attendees	Type	Subject
2021/06/01	Stephen P. Hicks, Zhiyong Liu	Email	Obtain Dataset
2021/06/05	Stephen P. Hicks, Zhiyong Liu	Teams	Dataset correction and interpretation
2021/06/09	Stephen P. Hicks, Zhiyong Liu	Teams	Project progress report (dataset import and visualization)
2021/06/21	Stephen P. Hicks, Zhiyong Liu	Teams	Project update and results presentation, next missions to do

Figure 4: Contact Log.