

Injection-Induced Seismicity Forecast using Analytical and Machine-Learning-Based Approaches in Northeast British Columbia, Canada

Alireza Babaie Mahani¹, Fatemeh Esfahani¹, and Honn Kao²

¹Mahan Geophysical Consulting Inc.

²Geological Survey of Canada, Natural Resources Canada

Summary

What is the likelihood of causing an earthquake with a specific magnitude following injection? This question is of great importance to oil and gas operators, regulators, and the public. Concerns about injection induced seismicity (IIS) have long been a focal point within the Western Canada Sedimentary Basin (WCSB). While the probability of inducing a significant earthquake following a single injection may be small, the cumulative effect of thousands of such injections could potentially increase the overall likelihood of significant IIS. Over the past decade, the WCSB has experienced multiple noteworthy, induced earthquakes, some reaching magnitudes as high as 5+. In this study, our focus is on forecasting IIS following hydraulic fracturing within the Montney unconventional resource play in northeast British Columbia (NEBC). We focus on two approaches. One is an analytical approach aiming at finding a solution for time-dependent forecast of IIS and the other is a machine learning approach to analyze the controlling factors of IIS for prediction of seismicity parameters, including the number of earthquakes and earthquake magnitude. Using seismicity and operational data combined with well-known empirical seismological formulas, our analytical approach provides a time-dependent exceedance rate for the occurrence of earthquakes with magnitudes equal to or above a given value following injection. Our machine learning model is trained on extensive datasets corresponding to several spatial and temporal grids to predict the number of earthquakes and earthquake magnitude based on regression and classification schemes and given operational and stratigraphic features controlling the occurrence of IIS.

Analytical Approach

Our analytical approach for time-dependent exceedance rate of earthquakes with magnitudes equal to or above a given value relies on three components. The first component is related to the ability of injection process to generate earthquakes. We call this component the Seismic Productivity of Injection (SPI). There has been extensive research into the relationship between fluid injection and seismicity by considering various operational parameters such as injection rate, pressure, and fluid volume. Despite the uncertainties associated with empirical relationships, an overall correlation can be obtained. For our analytical approach, we take the ratio of the number of earthquakes and the number of hydraulic fracturing stages to estimate SPI. The second component of our analytical approach is a model to estimate the seismicity rate of different magnitudes. A well-known approach is the Gutenberg-Richter relationship (Gutenberg and Richter, 1944). The Frequency-Magnitude Distribution (FMD) of earthquakes can be modeled by the Gutenberg-Richter relationship as:

$$\log_{10}(N|m_c \leq M \leq m_{max}) = a - bM \quad (1)$$

where N is the number of earthquakes and a and b are the seismicity parameters. In a cumulative FMD, Equation (1) provides a linear relationship between the number of earthquakes and their magnitudes that occur between m_c and m_{max} . m_c is the completeness magnitude and m_{max} is the largest magnitude after which data deviates from the linear trend of the FMD.

The third component of our analytical approach provides information on the temporal characteristics of seismicity. Another well-known relationship widely used in seismology is the modified Omori law, which predicts the occurrence rate of earthquakes (r) following a main-shock event as (Utsu, 1961)

$$r(t, M \geq m_c) = \frac{k}{(t+c)^p} \quad (2)$$

where t is time in days after the occurrence of the main-shock event. In the Omori law, k is the overall productivity of the sequence, c is a time delay until the sequence starts to decay, and p is the rate at which aftershocks decay with time (Sieh et al., 1993).

We test our analytical approach on IIS from hydraulic fracturing operations within the Montney unconventional resource play in NEBC. Figure 1 shows the estimated SPI. This is done by calculating the ratio of the number of earthquakes to the number of hydraulic fracturing stages within each day. SPI is given as a log-10 based value marked as SPI_{mean} (mean values) and SPI_{sd} (standard deviation).

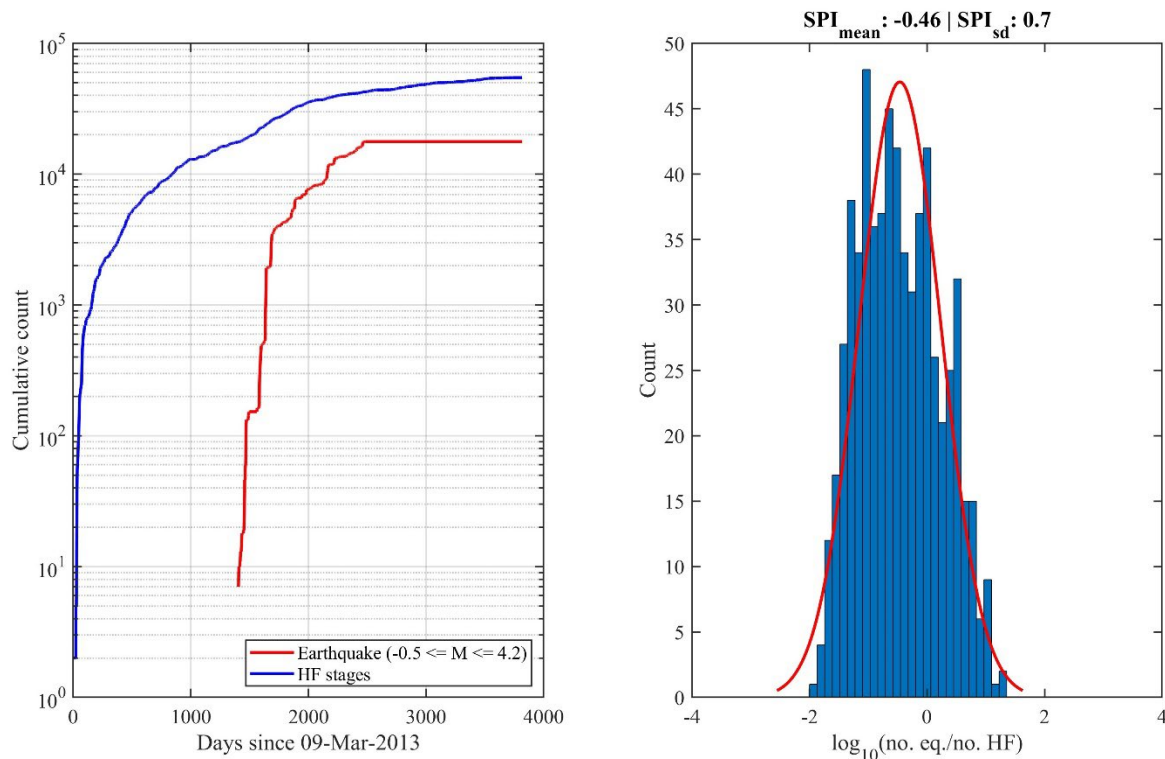


Figure 1. Mean and standard deviation (sd) of Seismic Productivity of Injection (SPI). HF = hydraulic fracturing.

Figure 2 shows the exceedance rate for different magnitude levels versus days after hydraulic fracturing injection. The results shown in figure 2 are calculated for the following input parameters: $m_c = 1$, $m_{max} = 5$, $b = 1.4$, $c = 0$, $p = 1$, $SPI_{mean} = -0.46$, $SPI_{sd} = 0.7$, and number of hydraulic fracturing stage = 100.

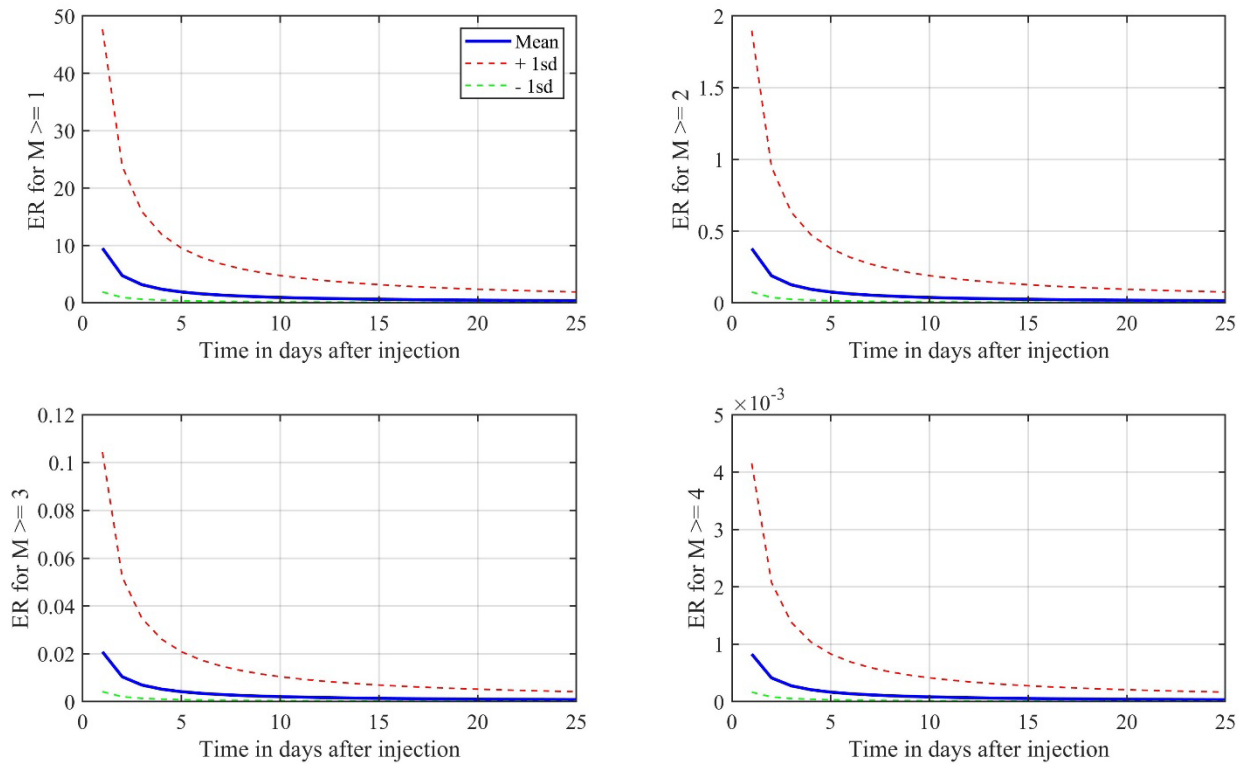


Figure 2. Exceedance rate for magnitudes equal to or greater than different values (1, 2, 3, and 4) versus time after hydraulic fracturing. Dashed lines show mean \pm 1 standard deviation (sd).

Machine Learning Approach

To predict parameters of IIS based on machine learning approaches, we use the Extreme Gradient Boosting algorithm (Chen and Guestrin, 2016), which provides an optimized implementation of Gradient Boosting Trees. Our focus is on predicting two parameters of IIS: the number of earthquakes and earthquake magnitude. As with our analytical approach, the area of interest is in NEBC, particularly the Montney unconventional resource play with thousands of hydraulic fracturing stages and associated IIS. We consider the problem of predicting the number of earthquakes as a regression problem with our target variable (number of earthquakes) taking on any integer values ≥ 0 . We compile extensive input datasets for our machine learning modeling that corresponds to various spatial (d_g) and temporal (d_t) grids with d_g between 0.01 and 0.3 degree and d_t between 1 and 30 days. In total, 168 datasets are assembled. Furthermore, we employ cross validation in our approach, which measures how well a machine learning model can be generalized to unseen data. In cross validation, a single parameter k (a value of 10 used in this study) is used, which refers to the number of times that a given dataset is split into training and test sets. To evaluate the results, we use the R^2 metric. Our best result is obtained for the

dataset with $d_g = 0.1$ degree and $d_t = 10$ days with an average R^2 of 0.65 (over the 10-fold cross validation splits). We consider prediction of earthquake magnitudes as both a regression and a classification problem. Here, instead of predicting a numeric value, we re-formulate the problem to predict whether earthquake magnitude is smaller or larger than a threshold ($M3$). Our output variable are two classes depending on whether the predicted magnitude is < 3 (class 0) or ≥ 3 (class 1). Our results show that considering magnitude prediction as a regression problem does not result in an acceptable R^2 with the best R^2 of 0.28 among all the 10 folds of cross validation splits. Figure 3 shows the results of predicting the number of earthquakes and earthquake magnitude considering the regression scheme for the best case; $d_g = 0.1$ degree, $d_t = 10$ days, $k = 6$, and $R^2 = 0.94$ when predicting the number of earthquakes (Figure 3a and b) and $d_g = 0.1$ degree, $d_t = 10$ days, fold = 7, and $R^2 = 0.28$ when predicting the earthquake magnitude (Figure 3c).

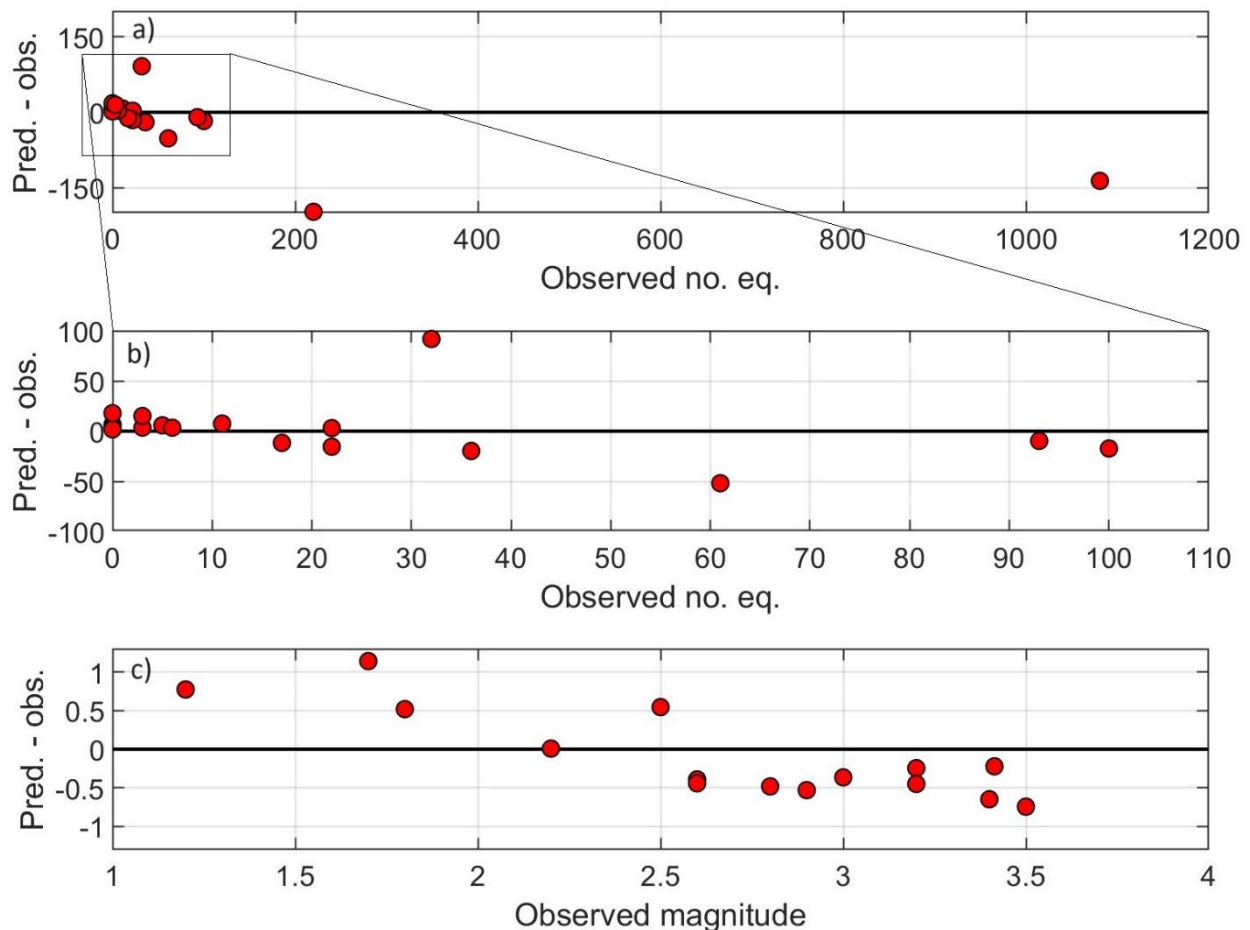


Figure 3. Residuals (prediction minus observed) versus observed number of earthquakes (a and b) and earthquake magnitude (c).

However, we achieve a much better results when considering the prediction of earthquake magnitude as a classification problem. In terms of performance metric, we use the area under the curve of precision and recall (AUC PR). Our results show AUC PR of 0.31 to 0.89 across the 10-

fold cross validation splits. Table 1 shows class predictions for the best case with $d_g = 0.1$ degree, $d_t = 10$ days, $k = 5$, and AUC PR = 0.89.

Table 1. Prediction of the magnitude class (0 = magnitude < 3; 1 = magnitude \geq 3). TP = true positive, TN = true negative, FP = false positive, FN = false negative.

Observed class	Predicted class	TP/TN/FP/FN
0	0	TN
0	0	TN
0	1	FP
1	1	TP
1	0	FN
0	0	TN
1	1	TP
1	1	TP
1	1	TP
0	1	FP
0	0	TN
0	0	TN
0	0	TN
0	0	TN
0	0	TN

Acknowledgements

This research is partially supported by Natural Resources Canada's Induced Seismicity Research Project, Environmental Geoscience Program, and an OERD research grant for the seismic risk assessment of CCUS.

References

- Gutenberg, B., and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of Seismological Society of America*, 34, 185–188.
- Sieh, K., Junes, L., Hauksson, E., Hudnut, K., et al. (1993). Near-field investigations of the Landers earthquake sequence, April to July 1992. *Science*, 260, 171–176.
- Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophysics Magazine*, 30, 521–605.
- Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco, California, U.S.A., 785-794, <https://doi.org/10.1145/2939672.2939785>.