Wastewater-Induced Seismicity in California

My Dinh, Michael Jetsupphasuk, Lisa Jian, Frank Mei

December 2017

Abstract

Given evidence of wastewater injection-induced seismicity in the central and eastern United States and the high levels of wastewater injection in California, we develop a generalized statistical method to test the association of wastewater injection with seismicity. Since previous studies have relied on some parametric model to fit the data and since earthquake modeling has been shown to be inaccurate, we propose a completely non-parametric statistical test. The proposed test statistic is an adjustment to the Spearman rank correlation that incorporates information from different lags (under the assumption that wastewater injection causes earthquakes in the future). Though the test statistic is shown to have decent power in simulations, it appears to boast a high false discovery rate in the real data, tested on both California and Oklahoma.

Contents

1 Introduction

When oil is extracted in an oil well, large amounts of water are also produced as a byproduct. Since the water is too saline to be used for consumption or irrigation, it is often injected into an injection well in the ground. In the central and eastern United States, there is a general scientific consensus that wastewater injection has, in several instances, induced earthquakes, some with moment magnitudes greater than 4.0 [3, 5, 7, 8].

The general process for how wastewater may cause seismic activity is generally known [3]. Injected wastewater can cause an increase in pore pressure in the surrounding rock, which, if along a fault, can increase seismic hazard. Additionally, wastewater-induced seismicity can also occur if an injection well changes the loading conditions on a fault. However, measuring the changes in pore pressure, stress along a fault, and the seismic hazard are difficult and are not generalized to all geological formations. Therefore, the observation of wastewater-induced seismicity in the central and eastern U.S. is not necessarily externally valid to other regions.

In California, wastewater injection volumes are higher than in Oklahoma [6]. However, it has not been conclusively shown that wastewater-induced seismicity exists in California; studies have found no general, definitive causal link between wastewater injection and seismicity [6, 9]. However, these studies have relied on some parametric assumptions about the behavior of earthquakes; for example using the Epidemic Type Aftershock Sequence (ETAS) model to decluster the earthquake catalog [6], which has been shown to be an unreliable model for earthquake behavior [2]. The motivation for this paper is to propose a generalized, non-parametric method to measure wastewater-induced seismicity.

Our methodology generally follows (with some exceptions) the methodology outlined by (McClure 2017) [9] but with a different test statistic. The paper is organized as follows: first, in section (2), we outline our methodology and test statistic; in section (3), we test our test statistic on simulated data to check the validity of our statistic and report our results; in section (4) we apply our test on the real data in California and Oklahoma and report our results; in section (5), we acknowledge the limitations of our design and discuss potential future improvements.

2 Methodology

2.1 General Framework

Like (McClure 2017) [9], our design is a longitudinal study rather than a cross-sectional study. With the assumption that confounding variables relating to space are stable over time, the remaining relevant variable is wastewater injection.

Below is a rough outline of our study design:

1. We overlay a grid with blocks of size 0.2° latitude by 0.2° longitude over the states of California and Oklahoma. The size of the blocks is chosen to be consistent with the spatial scales used in (McClure 2017) and in other investigations [9]. The purpose of this is to uniquely associate wells and earthquakes so blocks can be analyzed independently. The drawback of this layout, however, is that earthquakes and wells on the borders of the blocks do not relate to the earthquakes and wells of the neighboring block even if it may make more sense for them to. To examine the sensitivity of our design to this issue, we also run our test on a grid shifted by 0.1°.

2. Within each block, we compute our test statistic. Then, within each block, we run a permutation test by permuting earthquake data with a moving-block method to generate a null distribution of our test statistic, from which we obtain a p-value that measures the significance of our observed test statistic.

3. We aggregate the within-block p-values across blocks using Fisher's combined test [1].

2.2 Data

The time periods analyzed for each state are different due to differences in the availability of data. For California, the time analyzed is between January 1980 and June 2017. For Oklahoma, it is between January 2011 and December 2016. Unlike (McClure 2017) [9], we use monthly data to maintain as much detail as possible, since the time scale of wastewater-induced seismicity is unknown and variable.

Earthquake data from both California and Oklahoma are collected from the United States Geological Survey's earthquake catalog [10]. The minimum magnitude used is 2.5 for California and 3.0 for Oklahoma [6, 9]. Though the magnitude of completeness (M_c) for both regions is actually lower for the relevant time periods, lower magnitudes are less "socially significant" so they are not considered in this analysis. Our analysis only looks at counts of earthquakes and disregards magnitude. Since the available earthquake data are in different magnitude scales and there exists no general way to convert magnitudes from one scale to another, we cannot compare the relative sizes of earthquake events, which our study requires. We do not de-cluster the catalog since, as mentioned previously, the common ETAS [Stark: Is ETAS what others used? Not common to use it that way...] method has been shown to not perform well [2].

Injection well data for California were collected from California's Department of Conservation - Division of Oil, Gas, and Geothermal Resources (DOGGR) [11]. For Oklahoma, they were collected from the Oklahoma Corporate Commission [12].

Geographical map data were collected using the `get_map` function in the R package `ggmap` [4].

2.3 Lag-Adjusted Spearman Rank Correlation

In order to measure the positive dependence structure between seismicity and wastewater injection within a block, we want to construct a statistic that incorporates the following objectives:

- explicitly defines the dependence structure between seismicity and wastewater that we want to test for;
- adjusts for the possibility that the induced seismicity depends on wastewater injection, with an unknown time lag.

2.3.1 Dependence

Positive dependence between seismicity and wastewater injection is difficult to define, as many different dependence structures exist. We instead look for a specific dependence structure that is reasonable: a positive, monotonic relationship between wastewater injection and seismicity, measured by Spearman's rank correlation.

Spearman's correlation measures the monotonic relationship between two variables. It is a more relaxed measure of correlation than Pearson's correlation, which assesses linear relationships. It is also less sensitive to outliers.

Spearman's rank correlation, r_s, of two processes, X and Y, is defined as follows:

$$r_s = 1 - \frac{6D}{N^3 - N},$$

where r_s is the rank correlation, N is the total number of observations in the sample and D is defined as follows:

$$D = \sum [\text{rank}(X_i) - \text{rank}(Y_i)]^2.$$

Equivalently, we can write r_s in terms of the moments of the ranked data $rk(X) = \text{rank}(X)$, $rk(Y) = \text{rank}(Y)$, which is Pearson's correlation coefficient computed on ranked data:

$$r_s = \frac{\gamma_{rk(X),rk(Y)}}{\sigma_{rk(X)}\sigma_{rk(Y)}},$$

where

$$\gamma_{rk(X),rk(Y)} = \text{cov}(rk(X), rk(Y)).$$

More specifically, in our context, to compute the cross-correlation for two series at a particular time lag is to compute the correlation between the two series shifted by the appropriate lag. For instance, if we consider the two processes $X = \{X_1, X_2, ..., X_n\}$ and $Y = \{Y_1, Y_2, ..., Y_n\}$ and we want to compute the cross-correlation at lag τ, this is equivalent to computing Spearman's correlation for $X_\tau = \{X_1, X_2, ..., X_{n-\tau}\}$ and $Y_\tau = \{Y_\tau, Y_{\tau+1}, ..., Y_n\}$, which we define as:

$$r_\tau(X, Y) = \frac{\gamma_{rk(X_\tau),rk(Y_\tau)}}{\sigma_{rk(X_\tau)}\sigma_{rk(Y_\tau)}}.$$

A statistical test on the correlation itself could be defined where the null hypothesis H_0 is $r_\tau \neq 0$ at a fixed τ. If at least one of the series is exchangeable given the other then under the null hypothesis, all permutations of one variable against the other are equally likely to occur and each r_τ generated is equally likely. Therefore a p-value could be generated by computing 1 minus the percentile of the original r_τ in relation to the distribution of r_τ generated by the permutations.

2.3.2 Time Lag Effect of Induced Seismicity

We know injection can cause earthquakes some time later, but the exact time lag of induced seismicity is unknown. Therefore, we want to search for dependence at multiple time lags. Using the time scales by previous studies like (McClure 2017) [9] as a guideline, we search for dependence at time lags up to one year. We want to incorporate the correlations at different lags into one statistic.

Initially, we look at the Spearman's correlations of the data at time lags ranging from one month to a year (12 months) and choose the maximum correlation among them. Then, our test statistics is $T = \max_{\tau \in [1,12]} r_\tau(X, Y)$, where X is wastewater injection and Y is seismicity.

However, that statistic ignores the information provided by other lags, which may also have some sensitivity to the overall relation between water injection and earthquakes. For example, if we have one sequence of correlations $\{0, 0, 1, 0.9, 0\}$, and another $\{0, 1, 0, 0, 0\}$, both have a max correlation of 1. But the first sequence also has a correlation of 0.9 at another lag, which is large. And thus we want to give the first one a higher score by incorporating the correlations at other lags in some way.

We can combine different lags with a p-norm.

The p-norm of a vector $x = (x_1, x_2, ...x_n)$ is

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}.$$

This is a generalization of the method of taking the maximum correlation among lags because as p approaches ∞, the p-norm becomes the maximum, $\|x\|_\infty = \max(|x_1|, |x_2|, ..., |x_n|)$.

However, the p-norm takes each entry at its absolute value, while our objective is to differentiate a positive and negative correlation since positive correlation between data is what is desired. Specifically, there may be cases that the high p-norm value is majorly contributed by a negative correlation taken to its absolute value.

2.3.3 Test Statistic

As a result, we construct our test statistic as:

$$T(X, Y; p) = \|(r_1(X, Y), r_2(X, Y), ..., r_{12}(X, Y))\|_p,$$

where $r_\tau(X, Y) = \frac{\gamma_{rk(X_\tau), rk(Y_\tau)}}{\sigma_{rk(X_\tau)}\sigma_{rk(Y_\tau)}}$ is the Spearman's correlation of wastewater X and seismicity Y at lag τ.

Here, p is a hyper-parameter that we want to tune with simulations.

2.4 Permutation Test

In the previous discussion of Spearman's correlation, we mention that if at least one of the series is exchangeable given the other, then under the null hypothesis, all permutations of one variable are equally likely to be paired with the values of the other. There is no reason to believe wastewater injection is not autocorrelated. Main shocks in the earthquake data, on the other hand, are generally known to be stationary. However, since we do not de-cluster the catalog for main shocks and after-shocks, there is temporal clustering in the data. By permuting moving-blocks, we can control for this clustering pattern. We group the seismicity data by time cells of six months and permute the cells instead. In doing so, the clustering is preserved within a cell and the time cells are exchangeable.

The null hypothesis H_0 implies that the distribution of T is invariant under permutations of time cells. Let $\mathcal{G}$ be the finite group of transformations that permute among time cells of size 6 months, and let its size $|\mathcal{G}| = G$. For permutation $g \in \mathcal{G}$, $T(X, Y) \sim T(X, gY)$, where X is wastewater injection and Y is seismicity.

We take 10,000 random permutations from the whole permutation group $\mathcal{G}$ and look at the distribution of our test statistic T under these permutations. We find a p-value by measuring how extreme is our statistic for the original data. Let G' be the number of permutations $g \in \mathcal{G}$ such that $T(X, Y) \geq T(X, gY)$. We reject H_0 at level α if $P = \frac{G'}{10000} \leq \alpha$.

2.5 Fisher's Combined Test

So far we have only looked at the dependence between wastewater injection and seismicity within a block. In order to aggregate results for different blocks, we use Fisher's combined test.

Let P_i be the p-value of the permutation test within block i. Because under the null, blocks are independent of each other, Fisher's method states that:

$$-2\sum_{i=1}^{k} \ln(P_i) \sim X^2_{2k}$$

where X^2_{2k} is the chi-squared distribution with $2k$ degrees of freedom. Therefore, an overall p-value can be computed from this statistic.

3 Simulation

To examine the power of the test, we test on simulated data to see how it performs.

The simulation has the following steps:

1. randomly choose actual water injection data from one block;
2. simulate a time series of seismicity based on the Poisson model in (McClure 2017) [9];
3. perform the test on the above two time series.

(McClure 2017) constructed a statistical model for natural and induced seismicity [9]:

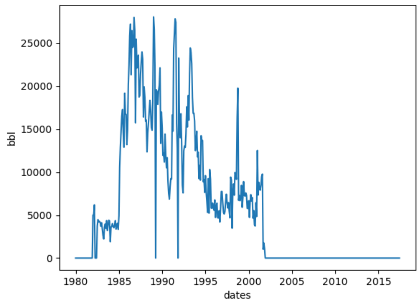
$$y_{ij} = \text{Poi}(e^{\mathcal{N}(0,\sigma_i)}(\mu_i + \beta_i x_{ij}) + a y_{i,j-1} e^{\mathcal{N}(0,\sigma_{II})}),$$

where y_{ij} is the number of earthquakes in block i and month j; x_{ij} is the cumulative water injected; σ_i is a measure for the variance of seismicity; μ_{ij} is the rate of natural seismicity; β_i is the dependence on water volume; a is the degree to which earthquakes cluster; and σ_{II} is the variability of the clustering.

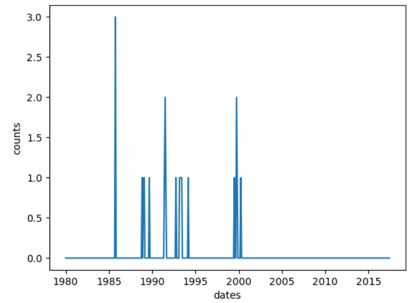
Using this model, we simulate the seismicity data using a slightly adjusted one, which fixed some variables and added multiple historical wastewater terms for a multiple lag effect:

$$y_j = \text{Poi}(e^{\mathcal{N}(0,1)}[\mu + \sum_{k=0}^{5} \text{Unif}(0,1)\beta_k x_{j-k}] + e^{\mathcal{N}(0,1)} a y_{j-1}).$$

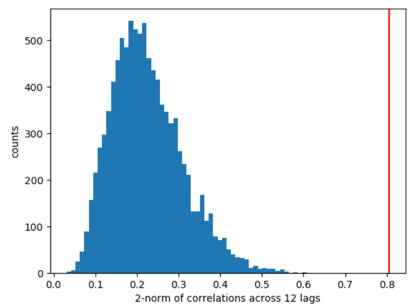
There are still many variables [Stark: What values of the parameters did you use in the simulations?] in this seismicity simulation model, which we use as



(a) real wastewater injection from a block



(b) simulated seismicity



(c) distribution of test statistics under permutation

Figure 1: Simulation Result

parameters to generate different shapes of seismicity data and evaluate the power of our test among these variations. Figure 1 shows an example of a simulation. Through simulation, we see that the test is powerful in detecting dependence between water injection and seismicity despite the randomness in the Poisson generating model.

Furthermore, as stated in previous section, we tune the value p in the p-norm in our test during simulation. Comparing the result given by using a 1-norm, 2-norm, 4-norm, and ∞-norm (max), we observe that when p is smaller, the resulting p-values are smaller and tests are more powerful against this alternative model for induced seismicity.

4 Results

Applying the permutation test to real wastewater and seismicity data within each block and aggregating with Fisher's method, we get a overall p-value of 0.0003. Thus, the null hypothesis is rejected. Out of 87 blocks with both

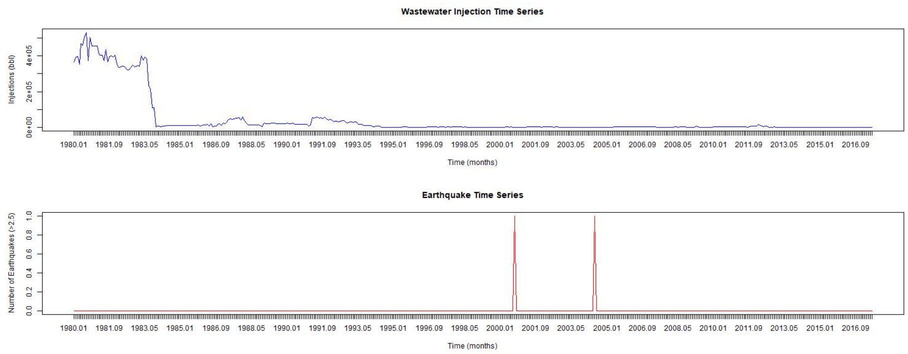


Figure 2: Time Series plots of Grid 1495.

non-zero water injection and non-zero seismicity within the time frame, 10 had p-values under 0.05 (or, 11.5%). See the Appendix for full results.

However, by closely looking at the time series plots of wastewater injection and seismicity in each block that has a significant p-value, we find one as shown in Figure 2.

Visually, seismicity does not have any obvious dependence on wastewater injection in this block, yet the test gives a positive result. A possible explanation is the following: It is easily seen that there is a significant difference between the amount of wastewater injected before 1984 and after. But Spearman's correlation only looks at the ranks of data, which will magnify the small variation in wastewater injection after 1984 and make them much more distinct while they are almost constant in the original data. Thus, in the rank-transformed data, the two spikes in seismicity match some spikes in wastewater injection in the same period, resulting in a small p-value. This example brings forward a possible concern that Spearman's rank correlation may cause false positives. In future work, we should examine and compare the different results using Spearman's and Pearson's correlation.

4.1 Shifted Grid

To examine the sensitivity of our test to the grid layout, we shift the grid by $0.1°$ three times – in the longitudinal direction, in the latitudinal direction, and both – and re-run our test. The percentage of blocks with p-values less than or equal to 0.05 is 16.1%, 16.7%, and 22.3%. The combined p-value in all three cases is 0.000.

4.2 Oklahoma

We also test on Oklahoma wastewater injection and seismicity data, and obtain an overall p-value of $2e-8$. Out of 84 blocks with both non-zero water injection and non-zero seismicity in Oklahoma, 17 had p-values under 0.05 (or, 20.2%). See Appendix for full results.

5 Open Questions and Future Work

From an examination of the real data and our outputted results, there is a clear deficiency in the validity of our test. As mentioned before, perhaps Pearson's correlation is more appropriate for this context. [Stark: I don't think the issue is the statistic: it's the null. The statistic should "wash out" in the simulation.]

Additionally, the decision to choose six months as the size of our time cells to permute is somewhat arbitrary. Six months was chosen as a somewhat conservative estimate given earthquake behavior. However, there may be a more appropriate method to force exchangeability in this case.

In this study, only counts of earthquakes are considered. There are obvious deficiencies in this decision as a 5.0 magnitude earthquake should not be considered equal to a 3.0 earthquake.

As (McClure 2017) [9] mentions in his paper, the grid system could be improved upon since it may not make sense for an earthquake at the border of one block to be associated with an injection well at the other side of the block and not an injection well immediately next it on the other side of the border. Some sort of method that softens the boundaries of the grid may improve upon this model. Further, although McClure does not mention this in his paper, one result of the above is that aggregating the p-values of each block with Fisher's combined test will generate an anti-conservative p-value since the test assumes the p-values are independent.

Additionally, the association we look for may not be the physically appropriate association. We are essentially testing to see if relatively large amounts of water injection line up with large amounts of earthquakes for some lag. However, the absolute values of water injection may be more relevant. Or, perhaps, cumulative injection. Large changes in water volume may also be relevant.

6 Acknowledgments

My Dinh, Michael Jetsupphasuk, Lisa Jian, and Frank Mei each contributed to the design and statistical methods of the study. Although all four members of the team aided in the development in each aspect of the project, the main responsibilities were as follows: Dinh and Jetsupphasuk conducted the literature review; Jetsupphasuk cleaned the data; Jian and Mei coded the simulations; and Jetsupphasuk and Mei wrote this report. Special thanks to Professor Philip Stark for supervising, guiding, and contributing to the project.

7 Appendix

7.1 California Test Results

Table 1: California Results

grid	corr	pvalue
25	0.15	0.62
26	$9.04 \cdot 10^{-2}$	0.94
27	0.26	0.21
28	0.28	$4.5 \cdot 10^{-2}$
29	$6.85 \cdot 10^{-2}$	0.8
66	0.29	0.27
67	0.17	0.44
68	$7.55 \cdot 10^{-2}$	0.79
69	0.13	0.6
70	$6.44 \cdot 10^{-2}$	0.99
71	$8.02 \cdot 10^{-2}$	0.97
72	0.2	0.23
73	$6.7 \cdot 10^{-2}$	0.99
110	$6.04 \cdot 10^{-2}$	0.76
111	0.17	0.65
112	0.25	$7.5 \cdot 10^{-2}$
113	0.33	$6.72 \cdot 10^{-2}$
114	0.16	0.48
115	0.2	0.29
116	0.2	0.33
117	0.1	0.94
118	0.2	0.41
119	0.19	0.24
152	$7.86 \cdot 10^{-2}$	0.96
153	0.29	$3.3 \cdot 10^{-3}$
154	1.02	$1 \cdot 10^{-5}$
155	0.27	0.3
156	$7.06 \cdot 10^{-2}$	1
157	0.14	0.49
158	0.13	0.76
159	$8.24 \cdot 10^{-2}$	0.97
160	0.45	$7 \cdot 10^{-4}$
161	$3.15 \cdot 10^{-2}$	1
162	0.24	$6.45 \cdot 10^{-2}$
163	$9.63 \cdot 10^{-2}$	0.93
164	0.32	$8.39 \cdot 10^{-2}$
165	0.12	0.97
198	0.2	0.33

199	0.12	0.78
200	$3.74 \cdot 10^{-2}$	1
201	0.17	0.49
203	0.19	0.41
204	0.18	0.3
205	$4.42 \cdot 10^{-2}$	1
206	$9.91 \cdot 10^{-2}$	0.9
207	$9.01 \cdot 10^{-2}$	0.94
208	0.11	0.87
209	0.13	0.66
248	0.18	0.37
249	0.14	0.62
250	0.34	$7.9 \cdot 10^{-3}$
251	0.19	0.27
252	0.25	$2.85 \cdot 10^{-2}$
292	0.24	0.12
293	$6.94 \cdot 10^{-2}$	0.99
294	0.14	0.78
295	0.13	0.65
296	$9.56 \cdot 10^{-2}$	0.95
297	0.72	$1 \cdot 10^{-4}$
333	$9.9 \cdot 10^{-2}$	0.9
337	0.15	0.67
338	0.22	0.29
339	0.19	0.57
340	0.12	0.86
381	0.2	0.29
382	0.46	$2.6 \cdot 10^{-3}$
383	0.23	$9.49 \cdot 10^{-2}$
384	0.17	0.43
385	$2.07 \cdot 10^{-2}$	0.7
386	0.19	0.33
431	0.16	0.56
467	$9.42 \cdot 10^{-2}$	0.95
468	0.31	$1.1 \cdot 10^{-2}$
474	0.34	$9.68 \cdot 10^{-2}$
476	0.21	0.42
511	$6.68 \cdot 10^{-2}$	0.98
556	$5.97 \cdot 10^{-2}$	0.99
563	0.42	$4.63 \cdot 10^{-2}$
606	$9.48 \cdot 10^{-2}$	0.93
648	0.52	$1 \cdot 10^{-5}$
649	$5.72 \cdot 10^{-2}$	1
689	0.11	0.88
690	$9.18 \cdot 10^{-2}$	0.86

693	0.26	0.11

7.2 Oklahoma Test Results

Table 2: Oklahoma Results

grid	corr	pvalue
25	0.57	0.15
26	0.67	$5.38 \cdot 10^{-2}$
27	0.58	0.18
28	0.5	$3.83 \cdot 10^{-2}$
29	0.56	0.13
66	0.33	0.89
67	0.45	0.29
68	0.45	0.24
69	0.17	0.94
70	0.31	0.75
71	0.44	0.3
72	0.33	0.66
73	0.78	$5.78 \cdot 10^{-2}$
110	0.34	0.8
111	0.58	$3.36 \cdot 10^{-2}$
112	0.52	$3.95 \cdot 10^{-2}$
113	0.3	0.57
114	1.04	$8 \cdot 10^{-4}$
115	0.3	0.59
116	0.69	$1.5 \cdot 10^{-2}$
117	0.42	0.72
118	0.45	0.15
119	0.87	$6 \cdot 10^{-4}$
152	0.72	$3.18 \cdot 10^{-2}$
153	0.61	$2.35 \cdot 10^{-2}$
154	0.29	0.53
155	0.44	0.49
156	0.39	0.49
157	0.3	0.92
158	0.51	0.3
159	0.21	0.94
160	0.68	$1.96 \cdot 10^{-2}$
161	0.28	0.8
162	0.22	0.95
163	0.21	0.97
164	0.78	$1.7 \cdot 10^{-2}$
165	0.55	0.35

198	0.31	0.95
199	0.27	0.89
200	0.29	0.85
201	0.54	0.15
203	0.39	0.68
204	0.44	$7.25 \cdot 10^{-2}$
205	0.28	0.37
206	0.26	0.76
207	0.46	0.29
208	0.38	0.63
209	0.6	$5.52 \cdot 10^{-2}$
248	0.32	0.66
249	0.27	0.76
250	0.61	$4.15 \cdot 10^{-2}$
251	0.78	$1 \cdot 10^{-5}$
252	0.96	$2 \cdot 10^{-3}$
292	0.37	0.61
293	0.24	0.58
294	0.92	$1.95 \cdot 10^{-2}$
295	0.89	$5 \cdot 10^{-3}$
296	0.49	0.42
297	0.46	0.67
333	0.45	0.11
337	0.31	0.86
338	0.47	0.5
339	0.53	0.35
340	0.56	$8.86 \cdot 10^{-2}$
381	0.48	0.32
382	0.42	0.44
383	0.46	0.4
384	0.64	$9.67 \cdot 10^{-2}$
385	0.44	0.48
386	0.48	0.26
431	0.37	$8.2 \cdot 10^{-2}$
467	0.25	0.97
468	0.37	0.82
474	0.2	0.99
476	0.45	0.59
511	0.28	0.99
556	0.56	0.13
563	0.23	$1.57 \cdot 10^{-2}$
606	0.51	0.24
648	0.36	0.76
649	0.46	0.57
689	0.57	$8.67 \cdot 10^{-2}$
690	0.6	$4.41 \cdot 10^{-2}$

693	0.41	0.56

7.3 Code Availability

Please see our Github repository for code used in this project. R was used primarily for cleaning the data and Python was used for our simulations and test.

References

[1] R.A. Fisher. "Statistical methods for research workers". In: (1925).

[2] Brad Luen and Philip B. Stark. "Testing earthquake predictions". In: *Probability and Statistics: Essays in Honor of David A. Freedman*. Ed. by Deborah Nolan and Terry Speed. Vol. Volume 2. Collections. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2008, pp. 302–315. DOI: 10.1214/193940307000000509. URL: https://doi.org/10.1214/193940307000000509.

[3] William L. Ellsworth. "Injection-Induced Earthquakes". In: *Science* 341.6142 (2013). ISSN: 0036-8075. DOI: 10.1126/science.1225942. eprint: http://science.sciencemag.org/content/341/6142/1225942.full.pdf. URL: http://science.sciencemag.org/content/341/6142/1225942.

[4] David Kahle and Hadley Wickham. "ggmap: Spatial Visualization with ggplot2". In: *The R Journal* 5.1 (2013), pp. 144–161. URL: http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

[5] Richard D. Andrews and Austin Holland. *Statement on Oklahoma Seismicity*. Oklahoma Geological Survey, 2015. URL: http://wichita.ogs.ou.edu/documents/OGS_Statement-Earthquakes-4-21-15.pdf.

[6] Thomas Göebel. "A comparison of seismicity rates and fluid-injection operations in Oklahoma and California: Implications for crustal stresses". In: *The Leading Edge* 34.6 (2015), pp. 640–648. DOI: 10.1190/tle34060640.1. eprint: https://doi.org/10.1190/tle34060640.1. URL: https://doi.org/10.1190/tle34060640.1.

[7] Susan E. Hough and Morgan Page. "A Century of Induced Earthquakes in Oklahoma?A Century of Induced Earthquakes in Oklahoma?" In: *Bulletin of the Seismological Society of America* 105.6 (2015), p. 2863. DOI: 10.1785/0120150109. eprint: /gsw/content_public/journal/bssa/105/6/10.1785_0120150109/3/2863.pdf. URL: +%20http://dx.doi.org/10.1785/0120150109.

[8] Manoochehr Shirzaei et al. "Surface uplift and time-dependent seismic hazard due to fluid injection in eastern Texas". In: *Science* 353.6306 (2016), pp. 1416–1419. ISSN: 0036-8075. DOI: 10.1126/science.aag0262. eprint: http://science.sciencemag.org/content/353/6306/1416.full.pdf. URL: http://science.sciencemag.org/content/353/6306/1416.

[9] Mark McClure et al. "Identifying potentially induced seismicity and assessing statistical significance in Oklahoma and California". In: *Journal of Geophysical Research: Solid Earth* 122.3 (2017). 2016JB013711, pp. 2153–2172. ISSN: 2169-9356. DOI: 10.1002/2016JB013711. URL: http://dx.doi.org/10.1002/2016JB013711.

[10] URL: https://earthquake.usgs.gov/earthquakes/search/.

[11] URL: ftp://ftp.consrv.ca.gov/pub/oil/GIS/.

[12] URL: http://www.occeweb.com/og/ogdatafiles2.htm.