Spatial Causal Inference on Induced Seismicity

Yuchen Xiao[1], Corwin Zigler[2], Alexandros Savvaidis[1], and Peter Hennings[1]

[1]University of Texas at Austin
[2]The University Of Texas at Austin

November 24, 2022

Abstract

Saltwater disposal (SWDs) has been linked to the recent increase of earthquakes in various regions of the United States. In some cases, the strong temporal and spatial associations have provided unequivocal evidences to the scientific community that wastewater injection is one of the dominant causal factors to the onset seismicity. In addition, numerous physical models have suggested that the increase in pore pressure from wastewater injection is capable to induce fault slips, providing further physical evidences. Another growing body of literature sorts to rigorously prove causality with statistical analysis where they propose statistical frameworks with parametric regression models to evaluate whether the observed earthquakes were occurring more often than by random chances and tested the statistical significance of the observed occurrences of earthquake to arrive at causal interpretations. We propose causal inference frameworks with the potential outcomes perspective to explicitly define what we meant by causal effect with mathematical formulations and declare necessary assumptions to ensure consistency between models for model comparison. In particular, we put considerations on two common difficulties in raster-based spatial statistical analysis, the spatial correlation, which can be described by Tobler's first law of geography where near things are more related than distant things, and interference, a causal inference term, where treatments applied to some spatially indexed units affect the outcomes at other spatially indexed units, mostly due to complex physical processes. The study region, the Fort-Worth Basin of North Central Texas, is discretized into non-overlapping grid blocks. The first proposed workflow adopts a cross-sectional study design on aggregated earthquake catalog and injection data where two statistical methods are employed to test the significance of the causal effect between the presence or absence of saltwater disposals and the number of the earthquakes and to estimate the magnitude of the average causal effect. The second proposed workflow incorporates the temporal domain which holds more scientific interests. Finally, the analysis is repeated for different grid configurations to directly assess the sensitivity of statistical results.

Spatial Causal Inference in Induced Seismicity

Yuchen Xiao[1], Corwin Zigler [2], Peter Hennings [3], Alexandros Savvaidis [3], Michael Pyrcz [1]

[1]Hildebrand Department of Petroleum and Geosystems Engineering
The University of Texas at Austin

[2]Department of Statistics and Data Sciences
The University of Texas at Austin

[3]Bureau of Economic Geology
The University of Texas at Austin

December, 14th 2021

Table of Contents

Table of Contents

Motivation Problem

Induced Seismicity in Fort-Worth Basin

- The rate of seismicity increased markedly from 2008 through 2015 and it coincides spatiotemporally with injection of 2 billion barrels of wastewater into deep aquifers.
- Physical models evaluate conditions of fault slips, but do not provide direct evidence of whether exist a clear causal relationship between large number of saltwater disposals (SWDs) and large number of earthquakes.
- Providing evidence of the causal relationship has practical and policy-relevant implications.
- Framing the problem with potential outcomes framework and evaluating common threats to causal validity.

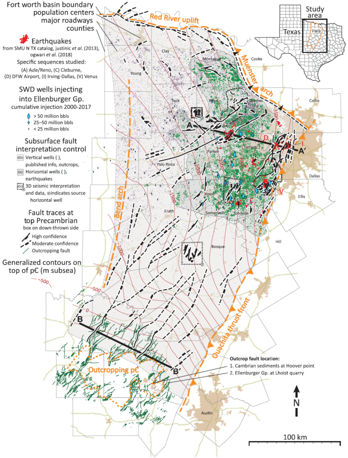


Figure: Map showing the north-central Texas, the traces of faults, saltwater disposal wells and earthquakes from Hennings et al. (2019).

Previous Studies

McClure et al. (2017)

- Adopt the raster-based design and seek associations between injection volume and seismicity on a time scale of order 1 year and on a spatial scale of order 18 km to 22.5 km,
- **Causal question:** Is there a causal relationship between injection volume and the number of earthquakes over time (i.e., annually) in discrete grid blocks.
- Only interested in grid blocks with both injection and seismicity.
- Assessing whether observed associations between injection and seismicity occur more often than would be expected by chance against some benchmark.

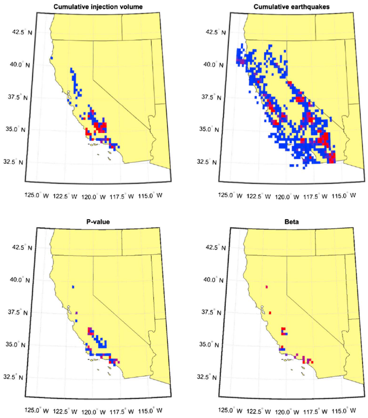


Figure: Figure 2 from McClure et al. (2017)

Causal Question

- A general question such as "do SWDs cause earthquakes" is consistent with a broad range of more finely specified quantitative research questions.
- Each with possibly distinct challenges and threats to causal validity.
- Causal inference methodology is designed to support the **mathematical precision** with respect to the definition of specific causal hypothesis.
- Isolate threats to validity that are specific to the causal hypothesis of interest and point toward ways to resolve those threats.

Causal Question

Broad Question: Do SWDs cause earthquakes?

- **Our Causal Question**: Does the presence of an SWD operating during 2010-2014 cause earthquakes around that location during 2014-2015 relative to what would have occurred in the absence of an SWD operating during that time?
- **Main threats to validity**: Confounding due to spatial differences (e.g., location of hydrocarbon deposits, location of faults).
- **McClure's Causal Question**: Does a XXXX barrels increase in annual wastewater injection volume at a location in a given year cause **more** earthquakes at that location during that year, relative to what would have occurred if wastewater injection volume had not been increased?
- **Main threats to validity**: Time-varying confounding due to well operating decisions that depend on previous treatments (i.e., previous annual injection volume) and previous outcomes (i.e., previous annual seismicity rate).

Table of Contents

The Opportunity and Goals

- Propose statistical analysis with in-depth presentation of **causal inference methodology**.
- The potential outcomes frame defines the causal effect for an individual as the difference between the outcomes that would be observed for that individual with versus without the exposure or intervention.
- Explicitly define the causal effects of interest in a **formal mathematical** way, and thus can make formal statistical statements about them.
- Critically evaluate assumptions and simplifications needed to conduct either a cross-sectional study or a temporal analysis.
- Improve the clarity and transparency regarding the most important tenets for discerning whether empirical statistical analyses provide evidence of causality between SWD and seismicity.

Common Threats to Causal Validity - Spatial Confounding

- The **spatial confounders** are variables that jointly dictate the placement of SWDs and the earthquake activity (e.g., **hydrocarbon deposits**, **geologic faults**, **geologic elevation**).
- For example, there could be possible **geologic reasons** that SWDs are placed near **hydrocarbon deposits** and the **hydrocarbon deposits** also influence the location of earthquakes, then the hydrocarbon deposits is a spatial confounder.
- The **intensity of gas producing wells** is used as first-order proxy to the location of hydrocarbon deposits where that intensity is used as a covariate for fitting the spatial point process model.

Table of Contents

Data Preprocessing

- Declustered the (NTXES) catalog, (2008-2018), from SMU with Reasenberg's algorithm assuming a magnitude of completeness of 2.
- Discard recorded earthquakes before 2013, since the systematic operation of network started in 2013.
- Use the operator reported SWDs injection volume data in DFW area from 2000s to 2017.
- Screen the SWDs operating between 2010 to 2014 with average monthly injection volumes (AMIV) above 250,000 bbl.
- Select earthquakes occurred in 2014 and 2015 from the declustered earthquake catalog.
- A total of 91 SWDs and 60 earthquakes are selected for the analysis.
- Consequently, the specifc causal question is **whether the presence of SWDs operating with AMIV above 250,000 from 2010 to 2014 causally affect the number of earthquakes at that location in 2014 and 2015.**

Sharp Null Hypothesis

- **Causal question**: whether the presence of operating SWDs during 2010 to 2014 causally affect the number of earthquakes occurred in 2014 and 2015 over the entire study region.
- Denote $Y_i(0)$ to be the potential outcome of the number of earthquakes that would occur at grid block i in 2014 and 2015 combined, if there were no SWDs operating from 2010 to 2014.
- Denote $Y_i(1)$ analogously to be the potential outcome for grid block i in 2014 and 2015 combined, if there were SWDs operating in that grid block from 2010 to 2014.
- $$H_o : \underbrace{Y_i(1)}_{\text{The potential outcome of unit i under treatment}} = \underbrace{Y_i(0)}_{\text{The potential outcome of unit i under control}}$$
- $$H_a : \underbrace{Y_i(1)}_{\text{The potential outcome of unit i under treatment}} \neq \underbrace{Y_i(0)}_{\text{The potential outcome of unit i under control}}$$
- Causal effect is explicitly defined as $Y_i(1) - Y_i(0)$, which is 0 under the H_o.
- All grid blocks are intervened and the causal estimand pertains to the entire study region.

Randomization-Based Hypothesis Testing

- The key idea of a randomization-based test of the null hypothesis is to compare the **observed relationship** between the presence of SWDs and earthquakes against what would be observed under the observed distribution of earthquakes but under **various probabilistically-generated alternative random assignments**.
- The Log-Gaussian Cox point process (LGCP) model is used to simulate random point patterns that mimic the unique spatial characteristics in the observed placement of SWDs.

Grid Sizes and Grid Offsets

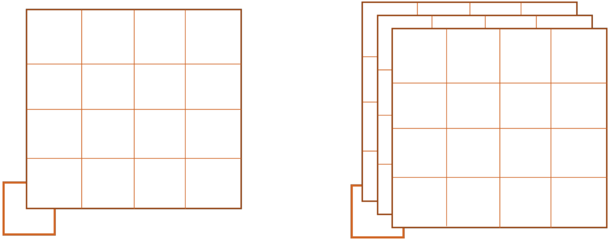
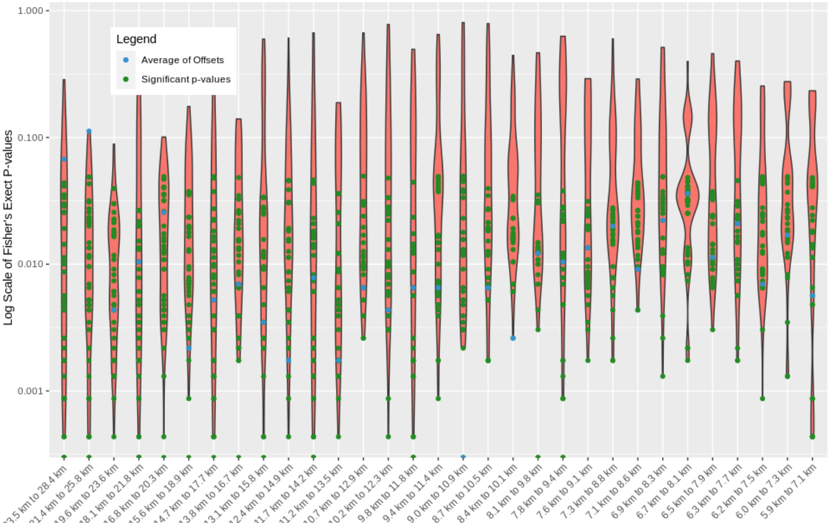


Figure: A schematic showing the variation of grid offsets

1. Consider a 4 by 4 discretization scheme, where the pivot defined as the bottom left corner.
2. The pivot moves within a fixed grid block of the same size.
3. This is repeated for a range of grid sizes.
4. Sensitivity analysis of grid configurations (i.e., grid sizes and grid offsets) is critical because different areas of interest with different data availability might require different grid sizes.

Table of Contents

Results


Legend
Average of Offsets
Significant p-values
Log Scale of Fisher's Exect P-values
1.000
0.100
0.010
0.001
3.5 km to 28.4 km
21.4 km to 25.8 km
19.6 km to 23.6 km
18.1 km to 21.8 km
16.8 km to 20.3 km
15.6 km to 18.9 km
14.7 km to 17.7 km
13.8 km to 16.7 km
13.1 km to 15.8 km
12.4 km to 14.9 km
11.7 km to 14.2 km
11.2 km to 13.5 km
10.7 km to 12.9 km
10.2 km to 12.3 km
9.8 km to 11.8 km
9.4 km to 11.4 km
9.0 km to 10.9 km
8.7 km to 10.5 km
8.4 km to 10.1 km
8.1 km to 9.8 km
7.8 km to 9.4 km
7.6 km to 9.1 km
7.3 km to 8.8 km
7.1 km to 8.6 km
6.9 km to 8.3 km
6.7 km to 8.1 km
6.5 km to 7.9 km
6.3 km to 7.7 km
6.2 km to 7.5 km
6.0 km to 7.3 km
5.9 km to 7.1 km

The "No Interference" Assumptions

- The no interference assumption, in our context states, the presence of SWD operating in one grid block does not affect the number of earthquakes in other grid blocks.
- The no interference assumption, in the context of McClure et al. (2017) states, the injection volume in one grid block does not affect the number of earthquakes in other grid blocks.
- The no interference assumption is obviously violated in light of the true underlying physical diffusion process.
- There might be few earthquakes in an isolated area with only one injection well, and there might be more earthquakes in an area with many injection wells scattered around the area.

Common Threats to Causal Validity - Time-Varying Confounding

- **Time-varying confounding due to well operating decisions that depend on previous treatments and outcomes.**
- For example, if the operator notices that there have been no earthquakes after operating SWDs at a location for some years, the operator might increase the injection volume. The opposite might also happen where the operator reduces the injection volume of some SWDs after an uptick in earthquakes in the immediate surroundings.
- Annual number of earthquakes for grid block i at year t **only** depends on the annual injection volume at year t and does not depend on injection volume from previous years.
- A more general formulation would correspond to the consensus that significant pore pressure build-up takes from several months to a couple years.

Conclusion

1. A broad question, "do SWDs cause earthquakes", is consistent with many more finely specified quantitative research questions.
2. **Causal inference methodology** support **mathematical precision** by explicitly define what is meant by a causal effect using the potential outcomes framework.
3. The specific causal question under investigation is a direct consequence of data preprocessing.
4. Data collection is critical to statistical analysis. Collecting spatial covariates (i.e., potential confounders) and incorporating them appropriately in analysis is challenging.
5. Detail the common threats to causal validity (i.e., confounding and interference) and point toward ways to resolve those threats.
6. Sensitivity analysis of grid configuration is needed to gauge the variations in statistical results.
7. Future work will relax the "no interference" assumption and estimate the causal effects in presence of interference.
8. If interested, our manuscript "**Applying Spatial Causal Inference on Induced Seismicity**" can be found at https://www.essoar.org/doi/10.1002/essoar.10507179.1

Acknowledgement

- Thanks to Louis Quinones and Heather DeShon for providing SMU earthquake catalog.
- Thanks to Iason Grigoratos for helping to decluster the earthquake catalog and helpful discussion.
- Thanks the financial support from the Center for Integrated Seismicity Research.
- Thanks to all the co-authors for their mentorship and contributions.

The University of Texas at Austin
Department of Statistics and Data Sciences
College of Natural Sciences

Bureau of Economic Geology
TexNet CISR

The University of Texas at Austin
Hildebrand Department of Petroleum and Geosystems Engineering
Cockrell School of Engineering

References

- Grigoratos, I., Rathje, E., Bazzurro, P., Savvaidis, A.(2020a).Earthquakes782Induced by Wastewater Injection, Part II: Statistical Evaluation of Causal Fac-783tors and Seismicity Rate Forecasting.Bulletin of the Seismological Society of784America,110(5), 2483–2497. doi: 10.1785/0120200079
- Hennings, P. H., Snee, J. E. L., Osmond, J. L., Deshon, H. R., Dommisse, R.,794Horne, E., . . . Zoback, M. D. (2019). Injection-Induced Seismicity and Fault-795Slip Potential in the Fort Worth Basin, Texas.Bulletin of the Seismological796Society of America,109(5), 1615–1634. doi: 10.1785/0120190017

Thank you!

"Appropriate methods do not (yet) exist, so researchers hoping to learn about contagion from observational social network data are sometimes faced with a dilemma: they can abandon their research program, or they can use inappropriate methods. This chapter will focus on the challenges and the open problems and will not weigh in on that dilemma, except to mention here that the most responsible way to use any statistical method, especially when it is well-known that the assumption on which it rests do not hold, is with a healthy dose of skepticism, with honest acknowledgment and deep understanding of the limitations, and with copious caveats about how to interpret the results" - **Elizabeth L. Ogburn**