

Applying Spatial Causal Inference on Induced Seismicity

Yuchen Xiao¹, Corwin Zigler¹, Peter Hennings², Alexandros Savvaidis¹, and Michael Pyrcz³

¹University of Texas at Austin

²Bureau of Economic Geology, University of Texas at Austin

³Chevron Energy Technology Company

November 22, 2022

Abstract

Saltwater disposal has been identified as the dominant causal factor that contribute to induced seismicity. Physical models rely on mechanistic understanding to infer causality where they evaluate various conditions for fault slips albeit with a high degree of uncertainty due to sparse data and subsurface heterogeneity. Given these uncertainties, statistical analysis is designed to measure statistical associations in the observed data with parametric regression models and interpret the significance of specific coefficient as evidence of causation. However, it is often difficult to interrogate the coefficients between different statistical models as the coefficients hold different implications. We propose a causal inference framework with the potential outcomes perspective to explicitly define what we meant by causal effect and declare necessary assumptions to ensure consistency between models for model comparison. The proposed workflow is applied to the Fort-Worth Basin of North Central Texas with the area of interest is discretized into non-overlapping grid blocks. 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. In addition, our analysis is repeated for different grid configurations to directly assess the sensitivity of statistical results. We have identified a stable and statistically significant causal relationship between the presence of saltwater disposals and the number of earthquakes and have estimated there are, on average, 13 more earthquakes occurring in grids with saltwater disposals.

Applying Spatial Causal Inference on Induced Seismicity

Yuchen Xiao¹, Corwin Zigler³, Peter H. Hennings⁴, Alexandros Savvaidis⁴,
Michael J. Pyrcz^{1,2}

¹Hildebrand Department of Petroleum and Geosystems Engineering, University of Texas at Austin,
Austin, Texas, USA

²Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA

³Department of Statistics and Data Sciences, University of Texas at Austin, Austin, Texas, USA

⁴Bureau of Economic Geology, University of Texas at Austin, Austin, Texas, USA

Key Points:

- The separation of causal and statistical conditions force consideration on important assumptions like the strong ignorability
- Sensitivity analysis of grid configuration on statistical results is necessary for raster-based spatial problems
- There is a stable and significant causal relationship between saltwater disposal and induced seismicity in the Fort-Worth Basin

Corresponding author: Yuchen Xiao, xiao.jack@utexas.edu

Corresponding author: Corwin Zigler, cory.zigler@austin.utexas.edu

Abstract

Saltwater disposal has been identified as the dominant causal factor that contribute to induced seismicity. Physical models rely on mechanistic understanding to infer causality where they evaluate various conditions for fault slips albeit with a high degree of uncertainty due to sparse data and subsurface heterogeneity. Given these uncertainties, statistical analysis is designed to measure statistical associations in the observed data with parametric regression models and interpret the significance of specific coefficient as evidence of causation. However, it is often difficult to interrogate the coefficients between different statistical models as the coefficients hold different implications. We propose a causal inference framework with the potential outcomes perspective to explicitly define what we meant by causal effect and declare necessary assumptions to ensure consistency between models for model comparison. The proposed workflow is applied to the Fort-Worth Basin of North Central Texas with the area of interest is discretized into non-overlapping grid blocks. 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. In addition, our analysis is repeated for different grid configurations to directly assess the sensitivity of statistical results. We have identified a stable and statistically significant causal relationship between the presence of saltwater disposals and the number of earthquakes and have estimated there are, on average, 13 more earthquakes occurring in grids with saltwater disposals.

Plain Language Summary

Causal inference, a sub-field of statistics, has gained popularity across other quantitative fields of medicine, epidemiology, and social sciences to provide evidence of causality but has not been previously explored in geoscience. We apply a causal framework with the potential outcomes perspective, the outcomes we would observe under a counterfactual scenario, to analyze the effect of saltwater disposal on earthquakes. We found there is a statistically significant causal relationship between saltwater disposal and the number of earthquakes and estimated, on average, there are 13 more earthquakes occurring in grid with saltwater disposal. We performed sensitivity analysis on the effect of grid configuration on statistical results that is unique in raster-based spatial analysis.

1 Introduction**1.1 Background**

Saltwater disposal (SWDs) has been linked to the recent increase of earthquakes in various regions of the United States (Ellsworth, 2013; Frohlich et al., 2016a; Grigoratos et al., 2020b; Hennings et al., 2019; Justinic et al., 2013; Keranen et al., 2013; Langenbruch & Zoback, 2017; McClure et al., 2017; Walsh & Zoback, 2015; Weingarten et al., 2015). In Texas, the development of shale hosted hydrocarbon resources in the Permian Basin, Eagle Ford Basin and Barnett Basin has resulted in a rapid expansion in both the number of SWDs and the cumulative injection volume, along with an abrupt increase in the number of earthquakes in respective basins (Hennings et al., 2019; Hornbach et al., 2015; Ogwari et al., 2018; L. Quinones et al., 2019; Scales et al., 2017; Zhai & Shirzaei, 2018). Of particular importance is the Fort-Worth Basin which hosts Barnett Shale in the North Texas that include most of the Dallas-Fort Worth (DFW) metropolitan area. Although the rate of earthquake activity in the DFW region has decreased since its peak in 2015, the potential linkages to oil and gas activity continuous to be a concern and put the social license of developing oil and gas resources in Texas at stake.

In response to this concern, the TexNet Seismological Observatory and the Center for Integrated Seismicity Research (CISR) at The University of Texas at Austin were

66 established to monitor potentially induced seismicity and to better understand the earth-
67 quake activities across the State of Texas (Hennings et al., 2019; Savvaidis et al., 2019).
68 One of the overarching goals of TexNet-CISR is to improve causative understanding of
69 the relationship between SWDs and onset earthquakes and the quantification of any iden-
70 tified causal relationship.

71 Advanced physics-based modeling has indicated the significant increases in pore
72 pressure from large-scale SWD activities, which reduces frictional resistance of critically
73 stressed faults, can induce fault slips (Fan et al., 2019; Zhai & Shirzaei, 2018; Keranen
74 et al., 2014; Lund Snee & Zoback, 2016). However, the physical models do not provide
75 direct evidence of whether an instance of earthquake is coincidental or whether there ex-
76 ist a clear causal relationship between larger number of earthquake and large number
77 of SWDs (Hornbach et al., 2016; Fan et al., 2019; Langenbruch & Zoback, 2016; McClure
78 et al., 2017).

79 To complement deterministic physical models, statistical analyses can provide ad-
80 ditional evidence of the causal relationship between SWD activity and earthquakes which
81 has practical and policy related to SWD regulation. In particular, causal inference, a sub-
82 field of statistics, has gained popularity across other quantitative fields of medicine, epi-
83 demiology, and social sciences to provide evidence of causality but has not been previ-
84 ously explored in geoscience (C. M. Zigler & Dominici, 2014; C. M. Zigler et al., 2018;
85 Dominici & Zigler, 2017; Papadogeorgou et al., 2019; C. M. Zigler & Papadogeorgou, 2021;
86 Reich et al., 2020; Imbens & Rubin, 2015; Hahn et al., 2020). An integral component
87 of causal inference is the notion of potential outcomes where we conceive of different out-
88 comes for a unit (e.g., a particular location in the study region) under different treat-
89 ment options, noting that only one outcome can be ultimately observed for that unit (Imbens
90 & Rubin, 2015). Using the notion of potential outcomes, causal inference methodology
91 allows practitioner to explicitly define the causal effect at the unit-level as the difference
92 between the potential outcomes (Imbens & Rubin, 2015). More specifically, the causal
93 effect of interest in this work, formalized with potential outcomes, is the difference be-
94 tween what earthquake activity would potentially be at a location if SWDs were present
95 and what earthquake activity would potentially be at the same location absent SWDs.
96 The “fundamental problem of causal inference”, where only one outcome can be observed
97 at a given location, motivates the use of average comparison across multiple locations
98 within the area of interest where SWDs are and are not present to compute the aver-
99 age causal effect (Holland, 1986). Importantly, the assumption of strong ignorability that
100 there are unmeasured confounding features is essential in causal inference (Imbens & Ru-
101 bin, 2015). The strong ignorability assumption, in this context, clarifies that the con-
102 founding factors would be unmeasured variables that jointly dictate SWD activity and
103 earthquakes.

104 Although an existing body of work has used parametric regression models to es-
105 tablish spatiotemporal correlations between SWD fluid injection and earthquakes and
106 has interpreted the statistical significance of specific coefficients as evidence of causation,
107 the caveat here is “correlation does not imply causation” (Hornbach et al., 2015; Fasola
108 et al., 2019; McClure et al., 2017; Grigoratos et al., 2020a; Aldrich, 1995; Langenbruch
109 & Zoback, 2016). More specifically, we argue the causal validity from statistical anal-
110 ysis is not completely determined by statistical model specification, but rather related
111 to explicit or implicit assumptions about the study design (C. M. Zigler & Dominici, 2014;
112 Dominici & Zigler, 2017). For example, McClure et al. (2017) first describe their model
113 specifications with modeling assumptions (i.e., assume the number of earthquakes is gen-
114 erated from a Poisson distribution) and then discuss the causal assumptions, strong ig-
115 norability assumption, to ensure the associations can be formally interpreted as demon-
116 strating causality within their longitudinal study design (McClure et al., 2017).

117 We expand the workflow in McClure et al. (2017) and propose a new spatial causal
118 inference workflow that integrates the notion of potential outcomes and relevant assump-

119 tions for the assessment of causality for induced seismicity. We apply two statistical meth-
 120 ods for two specific aspects of the average causal effect of interest. First, we offer a randomization-
 121 based test of the null hypothesis of no causal effect of SWD placement on earthquakes,
 122 tailoring the null distribution of the test to the specifics of the study design. Second, we
 123 estimate the average causal effect and its uncertainty of SWDs on earthquakes with the
 124 average difference between the potential outcomes across multiple grids within the area
 125 of interest. Our focus on the effects of presence or absence of SWDs versus, for exam-
 126 ple, other work’s focus on the effects of distributed SWD volume, is meant to simplify
 127 the problem and focus on key features of the causal framework. With new developments
 128 in spatial causal inference, exposure models that link the influence of particular SWD
 129 to earthquakes through distributed volume will be incorporated in future analysis (C. Zigler
 130 et al., 2020).

131 In addition to explicit causal considerations, we also offer an assessment of sensi-
 132 tivity of the statistical results to decisions about how to process the spatial data into a
 133 raster layer for analysis, specifically, the size and offset of spatial grids. Studies typically
 134 select a single grid configuration, chosen based on underlying knowledge or convenience,
 135 and then condition all inference on the chosen configuration (McClure et al., 2017; Grig-
 136 oratos et al., 2020a). Lack of a universally accepted grid configuration, even for the same
 137 study area, invites an assessment of how sensitive a given study’s results are to a cho-
 138 sen configuration to determine the possibility of analysis artifacts that are attributable
 139 to different grid configurations. Rather than condition inference on a single configura-
 140 tion, we conduct statistical analyses and summarise the statistical results under a vari-
 141 ety of configurations to gauge sensitivity to the grid size and placement of the raster
 142 layer. Results point towards the potential for sensitivity to grid configuration that war-
 143 rants careful consideration in raster-based spatial analysis.

144 In Section 2, we expound our design decisions and describe the specifics of two desig-
 145 nated statistical methods. More specifically, we detail the implications of two causal
 146 conditions and highlight how formulating the problem with potential outcomes forces
 147 deliberate considerations on the placement of SWDs to approximate a randomized ex-
 148 periment (Section 2.2 and Section 2.3). We argue more emphasis should be placed on
 149 the proper construction of the null distribution for hypothesis testing and demonstrate
 150 how our approach arrives at proper null distribution for our study and for previous stud-
 151 ies (Section 2.4.2 and Appendix Appendix E). We further differentiate between the marginal
 152 interpretation and the conditional interpretation in raster-based spatial problems (Sec-
 153 tion 2.4.3). Lastly, we perform sensitivity analysis to directly assess the impacts of grid
 154 configuration on the statistical results (Section 2.6). We discuss our results in Section
 155 3 and motivate future researches in Section 4.

156 2 Methods

157 2.1 Data Assembly and Parameterization

158 Our study area is the Dallas Fort-Worth (DFW) Basin in North-Central Texas. Nu-
 159 merous studies have documented the evolution of earthquake sequences, collected exten-
 160 sive compilations of mapped faults, and conducted numerical simulations of hydrolog-
 161 ical modeling and fault activation in the area of interest (Hennings et al., 2019; Frohlich
 162 et al., 2016a, 2020; Fan et al., 2019; Hornbach et al., 2016; Scales et al., 2017; L. A. Quinones
 163 et al., 2018; Lund Snee & Zoback, 2016; Gao et al., 2019). We refer to above references
 164 for complete background information for the study area.

165 We use the North Texas Earthquake Study (NTXES) catalog (2008-2018), collected
 166 at the South Methodist University (SMU), in this study (L. Quinones et al., 2019; DeShon
 167 et al., 2019). We have not screened the earthquake catalog because there are significantly
 168 less earthquakes that have magnitudes above 2.0 compared to those in Oklahoma. There

169 are only about 103 earthquakes after declustering assuming the magnitude of complete-
 170 ness is 2. We are aware that the SMU had few earthquake monitoring stations back in
 171 2008 and the temporary stations have mostly captured the aftershocks, not the main shocks.
 172 We aim to demonstrate the merits of our statistical framework and avoid being ham-
 173 pered by data-related issues. We use the operator reported SWDs injection volume data
 174 in DFW area from 2000s to 2017. It is available to download from Texas Railroad Com-
 175 mission website. The study area is within 32.07 degree to 33.68 degree latitude and -98.38
 176 degree to -96.74 degree longitude. A specific coordinate reference system is used to con-
 177 vert from latitude and longitude coordinates to Cartesian coordinates. The perimeter
 178 of the study area is selected to best encompass all available SWDs and earthquakes while
 179 constraining the total area.

180 2.2 Causal Quantities of Interest

181 We discretize the study area into non-overlapping grid blocks, each block represent-
 182 ing an observational unit of analysis. Each grid block is indexed i , taking on values $1, \dots, N$.
 183 The presence or absence of SWDs in grid block i is denoted as W_i , taking on value 0 if
 184 grid i does not have SWDs and 1 if the grid i does have SWDs. Hence \mathbf{W} indicates the
 185 presence or absence of SWDs across all grids within the area of interest and it is a rep-
 186 resentation of the spatial placement of SWDs in the study area. Let $Y_i(0)$ denote the po-
 187 tential outcome, that is, the number of earthquakes that would occur at grid i if there
 188 were no SWDs. Define $Y_i(1)$ analogously to be the potential outcome for grid i if there
 189 were SWDs present in that grid block. The individual-level causal effect of the presence
 190 of SWDs on the number of earthquakes in block i is defined as $Y_i(1) - Y_i(0)$. The av-
 191 erage causal effect over the study area is defined as $\bar{Y}(1) - \bar{Y}(0)$, which is the average
 192 over the sample of the individual-level effects. The above potential outcomes notation
 193 implicitly assumes that there is “no interference” between grids, where the presence or
 194 the absence of SWDs in one grid block does not impact the number of earthquakes in
 195 other grid blocks and vice versa. This assumption is also employed in McClure et al. (2017)
 196 and Grigoratos et al. (2020a) where injection volume in one grid is assumed to not im-
 197 pact the modeled outcome of the number of earthquakes in other grids. The validity of
 198 this assumption may warrant more careful consideration in studies of induced seismicity,
 199 a point to which we return in Section 4.

200 With the above definition of causal effect, we can explicitly state a sharp null hypo-
 201 thesis of no causal effect of the presence of SWDs on earthquakes in any grid block
 202 as $Y_i(1) = Y_i(0)$ for all i , corresponding to the hypothesis that the presence of SWDs
 203 does not causally affect the number of earthquakes in any grid of the study area. We de-
 204 velop an appropriate null distribution and test for this null hypothesis using a random-
 205 ization distribution that considers all plausible values of W . In addition to a statistical
 206 test of the sharp null hypothesis of no causal effect, we also estimate the magnitude of
 207 the average causal effect across the study area, $\bar{Y}(1) - \bar{Y}(0)$. Ideally, if SWDs were ran-
 208 domly allocated in the area of interest, then testing the null hypothesis and estimating
 209 the average causal effect would be trivial (Imbens & Rubin, 2015; McClure et al., 2017).
 210 In reality, using observations across a study area where some locations have SWDs re-
 211 quires careful considerations of why SWDs are placed in their observed locations (Imbens
 212 & Rubin, 2015), so one can judge the extent to which this placement could be reason-
 213 ably assumed to be random with respect to earthquakes. This judgment will be dictated
 214 in large part by a) assumption about the mechanism determining the placement of the
 215 SWDs, which we elaborate as the strong ignorability assumption in Section 2.3; and b)
 216 assumption about the spatial distribution of SWD placement, which will dictate the con-
 217 struction of an appropriate null distribution for a hypothesis test of no causal effect in
 218 Section 2.4.

219

2.3 Strong Ignorability and the “Assignment” of SWDs

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

The main assumption dictating the extent to which the study can reasonably approximate the design of a randomized experiment is that of strongly ignorable treatment assignment, or strong ignorability (Imbens & Rubin, 2015). This assumption states that, whatever the mechanism dictating the presence or absence of SWDs across the study area, it can be regarded as “random” in the sense that it is unrelated to the potential outcomes of the number of earthquakes for any grid. Formally, this assumption specifies conditional independence between W_i and $Y_i(0), Y_i(1)$, conditional on other grid features. In other words, there are no unobserved confounding factors, such as human attribution or geologic factors, that dictate both the placement of SWDs and the occurrence of earthquakes (McClure et al., 2017). One potential threat to the validity of this assumption is confounding due to the location of geologic faults. It is reasonable to suggest that locating SWDs closer to or farther from geologic faults might make it more or less likely their fluid injection triggers fault slips (McClure et al., 2017; Keranen et al., 2013; Hincks et al., 2018; Gao et al., 2019). Intentional placement of SWDs in relation to fault locations would violate the ignorability assumption and indicate poor approximation of a controlled experiment that randomly place SWDs. We expect this threat in our analysis to be minimal, since operators typically did not have complete information on fault locations, which are typically mapped after the occurrence of earthquakes (Hennings et al., 2019; Horne et al., 2020), which themselves may be induced at long time lags following the initiation of SWD (McClure et al., 2017; Fasola et al., 2019; Schoenball & Ellsworth, 2017).

240

2.4 Randomization-Based Hypothesis Test

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

Beyond the assumption of ignorability, the notion of approximating a randomized experiment also points towards consideration of alternative values of \mathbf{W} that might have arisen from a similar design to serve as the basis of a null distribution for the test of the sharp null hypothesis. In our proposed workflow, the plausible \mathbf{W} correspond to plausible arrangement of SWDs in the study area, which should correspond to the unique spatial characteristics evident in the observed placement of SWDs. 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 values of \mathbf{W} corresponding to alternative random assignments in a randomized experiment. To construct such a null distribution of plausible alternative values of \mathbf{W} , we need to model the mechanism that simulate the randomization of \mathbf{W} which matches with the unique spatial characteristics in the observed placement of SWDs. In particular, the randomization of \mathbf{W} grants every grid (i.e., even those grids without observed SWDs or observed earthquakes) to be eligible for having SWDs and constructs just one null distribution corresponding to plausible SWD placement across the entire study area.

257

258

259

260

261

262

263

264

265

266

267

268

269

270

To reflect the spatial structure inherent to the observed placement of SWDs, we select the Log-Gaussian Cox Process, (LGCP), to reproduce different SWD point patterns that assemble the observed SWD point pattern, Figure 1, where the number of SWDs is fixed for every reproduction. The comparison between the set of first- and second-order spatial summary functions of the observed SWD point pattern and those of the fitted LGCP model are displayed in the right columns of Figure 2 and Figure 3, respectively. We observe all empirical summary functions (i.e., black lines) are within the confidence intervals of the fitted LGCP model (i.e., shaded grey regions) and are near the expectations of the fitted LGCP model (i.e., red dotted lines). For illustrations, eight simulated SWD point patterns are shown in Figure 4, where some are indistinguishable from the observed SWD point pattern (i.e., Figure 1) in terms of inter-distances and spatial correlations. In short, while one might envision a controlled experiment where each grid is randomly assigned to either have or have not an SWD or where a fixed number of SWDs are placed in a manner that reflects complete spatial randomness, we advocate instead

271 for the approximation of an experiment where a fixed number of SWDs are placed across
 272 the entire study region in a manner that reflects basic spatial features of the observed
 273 SWD distribution.

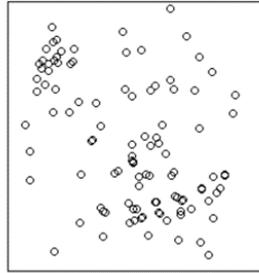


Figure 1. The observed SWD point pattern is shown where the black bounding box is the perimeter of the study region.

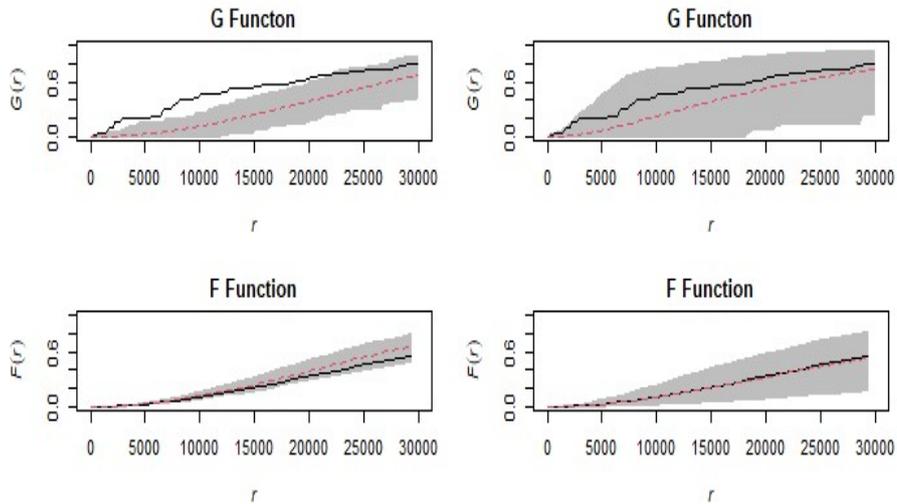


Figure 2. The first-order empirical summary functions (i.e., the cumulative nearest-neighbor distances, the G function, and the cumulative empty-space distances, the F function), shown in black line, are compared to that of the Complete Spatial Random (CSR) (left column) and the LGCP (right column), respectively. The grey intervals are the confidence intervals constructed from 3000 Monte Carlo simulations and the red dotted lines are the expectations with respect to each point process models (Baddeley et al., 2015; Illian et al., 2008). It is obvious the empirical summary functions run outside the confidence intervals of CSR, indicating the observed SWD point pattern is not of CSR origin. In comparison, the empirical summary functions nearly match the expectation of the summary functions of the fitted LGCP model, indicating the LGCP model fits well with the observed SWD point pattern in term of inter-distances.

274 To test the sharp null hypothesis of no causal effect, we repeat the generation of
 275 **W** with the LGCP model 1000 times and calculate 1000 average causal effects under dif-
 276 ferent realized **W** to constitute a reliable null distribution for hypothesis testing. To test
 277 the null hypothesis, we calculate the test statistics described below for each randomly-

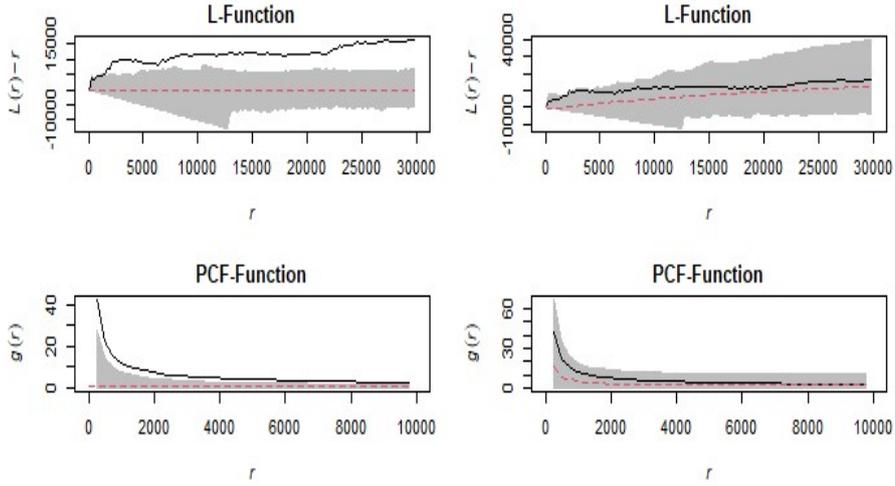


Figure 3. The second-order empirical summary functions (i.e., the L-function and the Pair Correlation Function (PCF)) are compared to that of the CSR (left column) and the LGCP (right column), respectively. Again, the left column shows the empirical summary functions run outside the confidence intervals of CSR, indicating the observed SWD point pattern is not of CSR origin. In comparison, the empirical summary functions nearly match the expectations of the summary functions of the LGCP, indicating the LGCP fits well with the observed SWD point pattern in term of correlations.

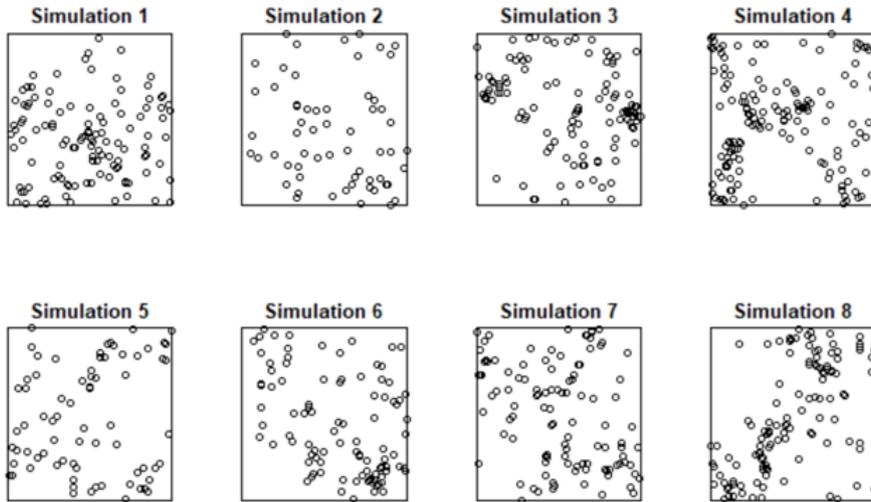


Figure 4. Eight simulated SWD point patterns are compared to the observed SWD point pattern in Figure 1. Some are very similar to the observed SWD point pattern in terms of inter-distance and spatial correlation.

278 generated value of \mathbf{W} , calculating a p-value to describe the observed value of the test
 279 statistic relative to the distribution of simulated values under the assumed sharp null hy-
 280 pothesis. A large p-value indicates an observed relationship between SWD placement
 281 and earthquakes that is consistent with no causal effect and ignorable SWD placement;

282 a low p-value indicates evidence to reject the sharp null and is interpreted as evidence
 283 of a causal effect.

284 **2.4.1 Rank Transformation**

285 A unique feature in raster-based spatial analysis is there might be excessive zeros.
 286 For example, consider a 30 by 30 discretization scheme where there are 900 grid blocks
 287 but only less than 80 grid blocks have non-zero counts of earthquakes. Consequently, a
 288 standard t-statistic may not be appropriate because of the excessive zero counts; in any
 289 particular discretization, the vast majority of grids do not have earthquakes. To address
 290 this, we implement a ranked T statistic to directly minimize the impacts of the zero-counts
 291 (Imbens & Rubin, 2015).

292 **2.4.2 The Proper Null Distribution for Hypothesis Testing**

293 Following above, we highlight it is the explicit definition of the causal effect and
 294 the transparent characterization of the sharp null hypothesis prompt us to utilize the
 295 LGCP model to approximate a randomized experiment and to construct proper null dis-
 296 tribution that best reflect the intended null hypothesis. Furthermore, we naturally cal-
 297 culate one p-value for the entire area of interest following the study design and the causal
 298 assumptions instead of for every grid. We now distinguish our approach with the pre-
 299 vious studies where we focus on the construction of the null distribution and the impli-
 300 cations of the marginal interpretation and the conditional interpretation.

301 Approximating a randomized experiment where the SWDs are assigned across the
 302 study area clarifies two points that distinguish with previous work. First, the null dis-
 303 tribution is constructed with respect to the entire study region instead of specific to each
 304 grid (McClure et al., 2017; Grigoratos et al., 2020a). Second, the inclusion of all grid cells
 305 within the area of interest focuses on marginal interpretation instead of conditional in-
 306 terpretation (McClure et al., 2017; Grigoratos et al., 2020a). McClure et al. (2017) pro-
 307 pose to use the resampling method to construct null distribution for hypothesis testing
 308 specific to each grid. The resampling method requires the identically and independently
 309 distributed (*iid*) assumption to guarantee the resampling is done in a way that reflects
 310 the intended null hypothesis (Carsey & Harden, 2013; Hall & Wilson, 1991; McClure et
 311 al., 2017). However, the sampled distributed volumes for each grid are not *iid* because
 312 they are spatially correlated which produce narrower null distribution and lead to overly-
 313 optimistic p-value for each grid. In Appendix Appendix E, we further illustrate that our
 314 approach even constructs proper null distribution with distributed volumes which is di-
 315 rectly applicable to previous studies (McClure et al., 2017; Grigoratos et al., 2020a).

316 **2.4.3 The Inclusion of Zero-Counts**

317 Both McClure et al. (2017) and Grigoratos et al. (2020a) have only analyzed grids
 318 that had hosted at least one earthquake and have made an implicit statistical condition
 319 that grids with zero observed earthquakes are implausible to have nonzero predicted earth-
 320 quakes. In fact, knowing locations with low predicted earthquakes that actually have zero
 321 earthquakes would provide very useful information on the causal relationship we want
 322 to investigate, as would knowing locations with zero observed earthquakes but have pre-
 323 dicted to have many (Panzeri et al., 2008). Our general causal formulation of the prob-
 324 lem with the potential outcomes perspective focus considerations on the randomization
 325 of \mathbf{W} . For every randomization, we allow the SWDs to be allocate to any locations within
 326 the area of interest as long as they match the spatial characteristics of the observed SWD
 327 point pattern. Our definition of the causal effect and the hypothesis testing procedure
 328 would be ill-conceived if the SWDs are only allowed to be allocate in the grids with ob-
 329 served earthquakes. We underline that the selection of the study area should be made
 330 before any statistical analysis and not dictated by the statistical analysis to avoid selec-

331 tion biases where the latter does not provide population-level summary (i.e., conditional
332 interpretation vs. marginal interpretations).

333 2.5 Estimating the Average Causal Effect

334 Testing the causal effect between the presence or absence of SWDs and the num-
335 ber of earthquakes is the first step towards understanding causality, quantifying such causal
336 effect is the second step. Given the explicit definition of the causal effect with the po-
337 tential outcomes perspective, we are interested to know what would the average num-
338 ber of earthquakes be if all grids were to have SWDs, $\bar{Y}(1)$ in notation format. Similarly,
339 we ask what would the average number of earthquakes be if all grids were to have no SWDs,
340 $\bar{Y}(0)$ in notation format. More importantly, what is the difference between the average
341 potential outcomes? We define such difference between the average potential outcomes
342 as the average causal effect and denote it as τ_{fs} where:

$$343 \tau_{fs} = \bar{Y}(1) - \bar{Y}(0) = \sum_{i=1}^N (Y_i(1) - Y_i(0)) / N \quad (1)$$

344 We interpret τ_{fs} as the average increase in the number of earthquakes for any grid
345 in the area of interest with SWDs compared to without SWDs and it serves as an ap-
346 proximation to the average causal effect (Imbens & Rubin, 2015). Provided with the causal
347 assumptions, we employ the LGCP model in a similar fashion to compute the average
348 causal effects for every randomization of \mathbf{W} . We further derive the expectation and vari-
349 ance of the average causal effect. Stepping away from hypothesis testing marks an im-
350 portant milestone to deepen our understanding of the causal relationship between SWDs
351 and the number of earthquakes. It is often trivial to prove, in terms of statistically sig-
352 nificant p-values, that the onset earthquakes are linked to SWDs. Perhaps, it is more
353 consequential to quantify the effects of SWDs on the number of earthquakes. The spec-
354 ifications of the method are provided in Appendix D.

355 2.6 Assessing the Sensitivity of Grid Configuration

356 In previous studies, McClure et al. (2017) divided the State of California and Ok-
357 lahoma into uniform grid blocks of 0.2 latitude and 0.2 longitude (roughly 22.5 km by
358 18 km) and performed one grid offset which found no significant difference in results. Grigoratos
359 et al. (2020a) calculated p-values in 20 km grid blocks and took the median value from
360 the sixteen 20 km grid blocks as the p-value for a 5 km grid block. The above approaches
361 either somewhat disregard the impacts of grid sizes and grid offsets or failed to properly
362 capture the variations resulting in erratic behaviors of p-values (McClure et al., 2017;
363 Grigoratos et al., 2020a). Consider a study area in a 4 by 4 discretization scheme (left
364 of Figure 5), where the pivot, defined as the bottom left corner, is allowed to move within
365 in a grid block of the same size as the grid blocks in the study area and the center of that
366 grid block coincides with the original pivot. The pivot is randomly shifted 100 times (right
367 of Figure 5) to generate 100 slightly different raster layers. For every unique raster layer,
368 we repeat the statistical analyses. In addition, we calculate the average across the 100
369 raster layers and repeat the statistical analyses. This process is repeated for a range of
370 grid sizes where the statistical analyses are repeated 101 times for every grid size. Al-
371 though grid configurations have been partially informed by domain expertise in previ-
372 ous studies, they are still arbitrary and provoke instability in statistical results. Sensi-
373 tivity analysis of grid configuration is therefore critical because different areas of inter-
374 est with different data availability might require different grid sizes. For example, the
375 State of Oklahoma has hosted thousands of $M \geq 3$ earthquakes where DFW has hosted
376 significantly less. The average number of earthquakes in 5 km grid blocks could vary dras-
377 tically depending on the residing States. To our knowledge, there has not been compre-
378 hensive sensitivity analysis on grid configuration and we aim to bridge this gap in the

379 literature. We repeat our analyses described in Section 2.4 and Section 2.5 for 3,131 times
 380 (i.e., 101 grid offsets for every grid size with total 31 different grid sizes), respectively,
 381 and summarize the results across all grid offsets for every grid size to gauge sensitivity
 382 of the analysis about grid configuration.

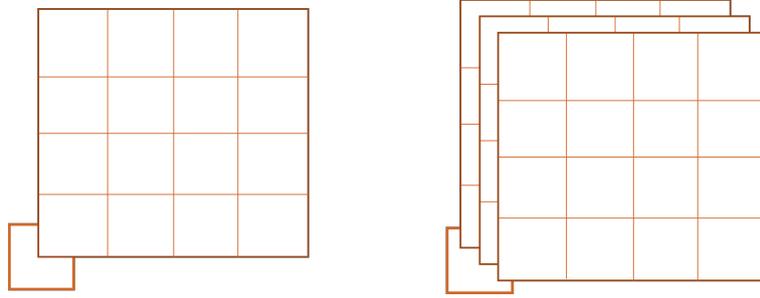


Figure 5. Illustration of Grid Offsets

383 **3 Results and Discussion**

384 **3.1 Results from the Randomization-Based Hypothesis Test**

385 Figure 6 displays the p-values from the sharp null hypothesis test where the y-axis
 386 shows the p-values transformed into log-scale to help better examining clusters near zero
 387 and the x-axis shows all the grid sizes implemented in the study. Every p-value repre-
 388 sents a unique combination of grid size and grid offset. Each green point is a statistically
 389 significant p-value calculated from performing the hypothesis test on one grid offset for
 390 a particular grid size and the blue point is the p-value calculated from performing the
 391 hypothesis test on the average data values of 100 raster layers from grid offsets. Because
 392 only the statistically significant p-values are marked with dots, the overall distribution
 393 of the p-values for every grid size is rendered with a violin plot. Our results should be
 394 differentiated from Grigoratos et al. (2020a), where they summarized median p-values
 395 from the 20 *km* grids for the 5 *km* grids and did not perform independent hypothesis
 396 test on the 5 *km* grids (Grigoratos et al., 2020a).

397 Overall, Figure 6 indicates there is a stable and statistically significant causal rela-
 398 tionship between the presence of SWDs and the number of earthquakes over the en-
 399 tire study area across a range of grid sizes. There is a trend from larger grid sizes to smaller
 400 grid sizes where there are less extremely small p-values that are below 0.0010. This is
 401 potentially caused by the reduction in the ratio of non-zero counts and zero-counts as
 402 the grid size diminishes, when there are many grids and the overwhelming majority are
 403 without earthquakes.

404 To recognize the impacts of grid offset, we focus on grid size 10.2 *km* by 12.3 *km*.
 405 This grid size has a wider and a more uniform distribution of p-values as displayed in
 406 the violin plot, which indicates grid offsets can have a large impact on the p-values for
 407 this specific grid size. The wide distribution of p-values suggests small shifts in the seg-
 408 mentation of the study area might separate critical clusters differently and result in a
 409 divergence of statistical results.

410 Following above, there is another observed trend where larger grid sizes have flatter
 411 distribution of p-values and smaller grid sizes have more concentrated distribution
 412 of p-values. Because the pivot is allowed to move within a grid block of the same size
 413 as the grid blocks in the study area, larger grid blocks are more likely to experience more
 414 distinct placements of the raster layer due to different initial grid offsets and have more

415 divergent results. In comparison, smaller grid blocks are more constrained to obtain di-
 416 vergent results.

417 Through the compelling visualization, we conclude there is a stable and statisti-
 418 cally significant causal relationship between the presence of SWDs and the number of
 419 earthquakes for the entire study area. In addition, we argue comprehensive sensitivity
 420 analysis of grid configuration is necessary in raster-based spatial analysis to demonstrate
 421 the stability of statistical results and to arrive at objective conclusions.

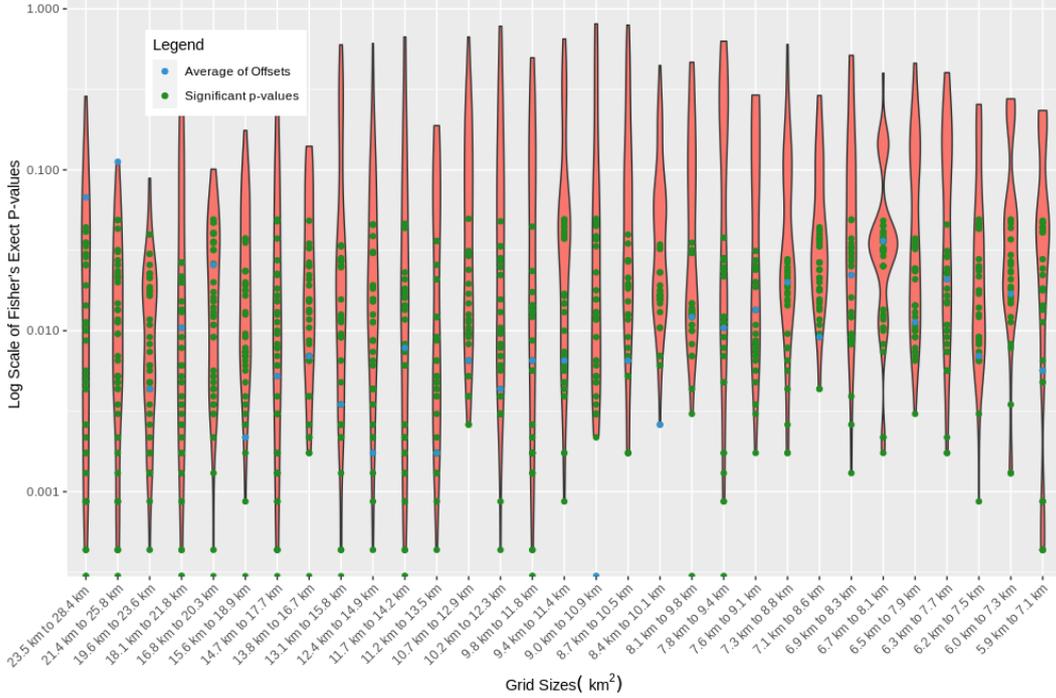


Figure 6. The p-values from hypothesis test are displayed. Only the statistically significant p-values are marked with green dots and the p-value computed from the average of 100 raster layers is marked with blue dot.

422 **3.2 Results from Estimating the Average Causal Effect**

423 Figure 7 shows the average causal effects in boxplot for different grid configurations.
 424 The solid bar is the median of the average causal effects summarised over all grid off-
 425 sets for every grid size. There are two noteworthy observations. First, there is a decreas-
 426 ing trend in the average causal effect as the grid size diminishes. This is an unique arti-
 427 fact for raster-based spatial problems and it is expected since both the SWDs and earth-
 428 quakes are measured at points, so increasingly finer grids will eventually separate each
 429 earthquake and each SWD into its own grid cell, resulting in a null effect estimate. Sec-
 430 ond, there is a general trend that larger grid sizes exhibit greater variations in the av-
 431 erage causal effect from grid offsets. Larger grid sizes experience more distinctive dis-
 432 cretization schemes from grid offsets where they have more ways to divide critical clus-
 433 ters and undoubtedly result in more diverse statistical results.

434 Table 1 shows the average height of the confidence intervals of the average causal
 435 effects and the ratio of the 90% confidence intervals that overlap zero. Because we gen-
 436 erate a pair of the average causal effect and the corresponding confidence interval for ev-

437 every grid offset, it is hard to visualize all the confidence intervals for every grid size. Al-
 438 ternatively, we calculate the height of every confidence interval (i.e., subtracting the lower
 439 bound from the upper bound) and summarize the average height over all grid offsets for
 440 every grid size. The average confidence interval height is providing some sense of the typ-
 441 ical uncertainty around a point estimate for a given grid size.

442 Table 1 illustrates yet another unique artifact to raster-based spatial problems where
 443 the uncertainty around the point estimates goes down with decreasing grid size, primar-
 444 ily because of the increasing number of observations. The rightmost column in Table 1
 445 shows the ratio of the 90% confidence intervals that overlaps 0 for every grid size. If a
 446 confidence interval overlaps 0, it serves as evidence that it is not significantly different
 447 from 0. For the largest grid size, 50% of the confidence intervals include zero. This per-
 448 centage is increasing for smaller grid sizes where the confidence intervals overlap 0 for
 449 all grid offsets for grid sizes smaller than 10.7 km by 12.9 km. Furthermore, all 95% con-
 450 fidence intervals overlap 0 for all grid offsets for every grid size thus they are not shown.

451 We conclude using the average difference between the potential outcomes across
 452 grids as an approximation to the average causal effect is very sensitive to grid configu-
 453 rations. Unless there are specific grid sizes of interest, it is difficult to make any inter-
 454 pretation. We select 19.6 km by 23.6 km and 18.1 km by 21.8 km as the closet grid sizes
 455 to the grid size used in McClure et al. (2017) (i.e., 18 km by 22.5 km) where we find the
 456 expectation of the average causal effects for the two grid sizes to be 13. In other words,
 457 we expect there are, on average, 13 more earthquakes occurring in any grid with SWDs
 458 versus without SWDs within the area of interest provided the selected grid size is sound
 459 from domain expertise (McClure et al., 2017). Importantly, we note the choice of grid
 460 size(s) of interest should be made based on physical understanding of the problem, and
 461 then sensitivity to grid offsets within those relevant grid sizes should be gauged. We pro-
 462 vide the wide range of grid sizes for illustration only - some grid sizes could presumably
 463 be ruled out as irrelevant for our analysis, and we underline the choice of grid size should
 464 not be based on the convenience of statistical methods.

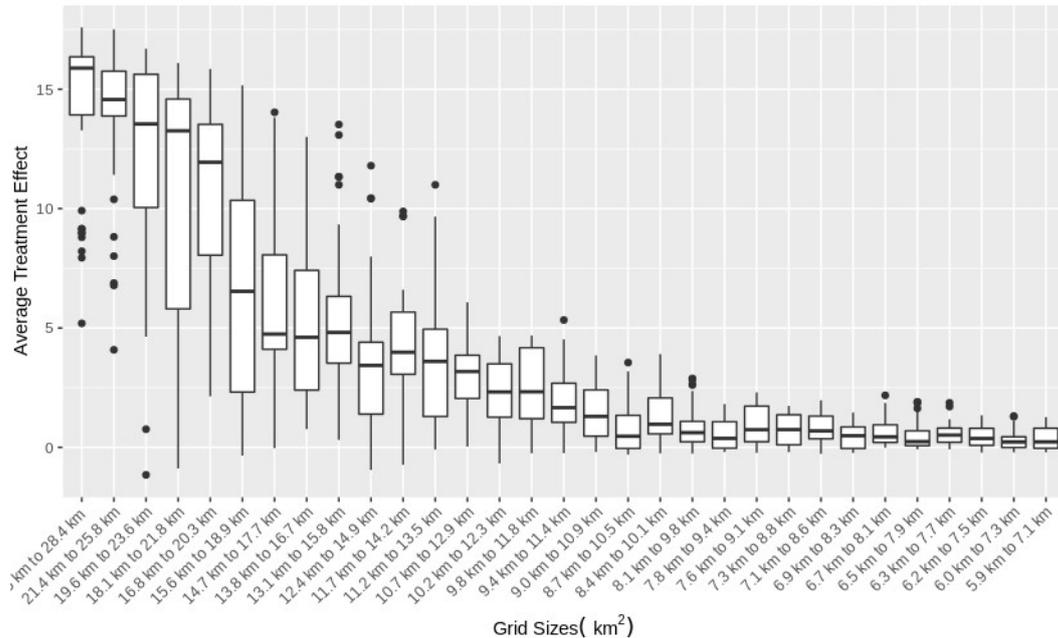


Figure 7. The boxplot displays the average causal effects across all grid offsets for every grid size. The average over the grid offsets is indicated by the solid horizontal bar.

Table 1. Summary of the Average Magnitude of the Average Causal Effect and the Ratio of the Confidence Intervals (CI) that Overlaps Zero

Gridsize	Average90CI	Average95CI	Ratio of 90%CIs that Overlaps Zero
23.5 km to 28.4 km	30.28	34.54	0.50
21.4 km to 25.8 km	28.92	33.21	0.50
19.6 km to 23.6 km	25.89	29.52	0.56
18.1 km to 21.8 km	23.36	26.30	0.72
16.8 km to 20.3 km	22.92	26.08	0.66
15.6 km to 18.9 km	17.13	18.68	0.92
14.7 km to 17.7 km	16.04	17.76	0.92
13.8 km to 16.7 km	13.81	15.37	0.94
13.1 km to 15.8 km	14.15	15.71	0.82
12.4 km to 14.9 km	9.78	11.10	0.98
11.7 km to 14.2 km	11.32	12.80	0.88
11.2 km to 13.5 km	9.61	11.04	0.94
10.7 km to 12.9 km	8.67	10.13	1.00
10.2 km to 12.3 km	7.13	8.36	1.00
9.8 km to 11.8 km	7.53	8.70	1.00
9.4 km to 11.4 km	6.05	7.36	1.00
9.0 km to 10.9 km	5.04	6.38	1.00
8.7 km to 10.5 km	3.28	4.64	1.00
8.4 km to 10.1 km	4.21	5.61	1.00
8.1 km to 9.8 km	3.24	4.63	1.00
7.8 km to 9.4 km	2.51	3.90	1.00
7.6 km to 9.1 km	3.26	4.75	1.00
7.3 km to 8.8 km	2.77	4.25	1.00
7.1 km to 8.6 km	3.12	4.57	1.00
6.9 km to 8.3 km	2.13	3.63	1.00
6.7 km to 8.1 km	2.46	3.98	1.00
6.5 km to 7.9 km	1.93	3.48	1.00
6.3 km to 7.7 km	2.20	3.74	1.00
6.2 km to 7.5 km	1.79	3.37	1.00
6.0 km to 7.3 km	1.40	2.98	1.00
5.9 km to 7.1 km	1.58	3.19	1.00

4 Conclusion and Future Work

Improvements in the understanding of the causal relationship between SWD and induced seismicity, more importantly, the quantification of such relationship, require advancement in statistical analysis that bypass certain limitations in deterministic approaches. Traditional parametric regression models presume the specified regression models accurately reflect the true relationship between the variables of interest which, when coupled with the (often implicit) assumption of strong ignorability, can provide evidence of causation. In contrast, we propose a general causal formulation of the spatial problem where we explicitly define the causal estimand with the potential outcomes perspective and implement appropriate statistical methods for subsequent testing and estimation. The causal conditions are deliberately separated from the statistical conditions so that the causal estimand is purposefully chosen rather than inherited from the specified parametric models with the expectation that the chosen estimand is more directly relevant to address the scientific question. Note that this perspective does not preclude the usefulness of re-

479 gression modeling strategies, it only serves to separate key determinations of causal va-
480 lidity from the specification of such models.

481 Using the potential outcomes perspective, we explicitly define what is meant by a
482 causal effect, and then use the framing relative to the approximate design of a random-
483 ized experiment as a benchmark to guide the analysis and interpretation of threats to
484 validity. In particular, this led to different choices about grid configuration to include
485 in the analysis and the construction of an appropriate null distribution. We perform in-
486 ferences on two specific aspects of the average causal effect. First, we perform a sharp
487 null hypothesis to test the statistical significance of the causal effect between the pres-
488 ence of SWDs and the number of earthquakes. We find a stable and statistically signif-
489 icant causal relationship between the presence or absence of SWDs and the number of
490 earthquakes for the entire study area across a range of grid sizes. This result is consis-
491 tent with the results from other studies which found strong evidence of wastewater-induced
492 seismicity in the DFW region of North-Central Texas. Second, we estimate the average
493 causal effect and observe there are, on average, 13 more earthquakes for any grid with
494 SWDs versus without SWDs for grid sizes 19.6 *km* by 23.6 *km* and 18.1 *km* by 21.8 *km*.
495 We emphasize grid configuration has a material consequence on the statistical results.
496 Grid configuration should be studied empirically because different areas of interest have
497 different data availability and acquire different grid configurations. Domain knowledge
498 should guide the choice of grid configuration but it can not replace empirical experimen-
499 tation.

500 We highlight the statistical analyses that adopt causal inference framework with
501 causal inference terminology are not superior by default (C. M. Zigler & Dominici, 2014).
502 For example, the work from McClure et al. (2017) demonstrates causality under the strong
503 ignorability assumption in a longitudinal design. The causal inference methodology only
504 provides a formal structure to frame the question, whether our analysis demonstrates
505 causality depend on how much we believe the LGCP model accurately reproduce the ob-
506 served placement of SWDs conditional on all potential confounding variables - namely,
507 the exclusion of geologic faults. We view the abovementioned two approaches are more
508 different in styles and less different in substances. Observational studies face confound-
509 ing problems that are difficult to fully account for: oversimplification of the complexity
510 of the problem (e.g. adding the strong ignorability assumption and no interference as-
511 sumption) can potentially invalidate the causal portion of the analysis (Carone et al.,
512 2020). Nevertheless, we argue a more general causal formulation of the problem with the
513 potential outcomes perspective could continuously reflect the complexity of the problem
514 and improve the clarity and transparency regarding the most important tenets for dis-
515 cerning whether empirical statistical analyses provide evidence of causality between SWD
516 and seismicity (Imbens & Rubin, 2015; Carone et al., 2020).

517 Causal inference methodology has become popular and led to important contribu-
518 tions in a variety of other disciplines including education, psychology, economics, epi-
519 demiology, medicine, sociology (Friedrich & Friede, 2020; Glass et al., 2013; Imbens &
520 Rubin, 2015). It has been mostly unexplored in the areas of geoscience and engineering
521 where the objective is to infer causality between spatial variables. Spatial causal infer-
522 ence is a fast-growing field and has contributed to air pollution epidemiology that sup-
523 port important regulatory policies. A major obstacle of applying spatial causal inference
524 in raster-based spatial problems is the assumption of no interference. For example, it is
525 reasonable to suggest that the presence or absence of SWDs in some neighboring grids
526 could all contribute to the occurrence of earthquakes for a grid. There is an inherent many-
527 to-one and one-to-many relationship where many SWDs reside in different grids all con-
528 tribute to the occurrence of earthquakes for a grid and one SWD might affect the oc-
529 currence of earthquakes in many different grids. Similarly, it is logical to postulate that
530 distributed volume in some neighboring grids affect the occurrence of earthquakes for
531 a reference grid. Difficulties arise when extending the causal inference framework to ac-

532 knowledge the interference that would arise when earthquakes at a given location might
 533 depend on the distributed volume at the location from multiple SWDs. These are sub-
 534 ject to future works with a recently developed bipartite interference network in spatial
 535 causal inference where the target is to investigate the causal effect while relaxing the in-
 536 dependence assumption between grids (C. M. Zigler & Papadogeorgou, 2021; C. Zigler
 537 et al., 2020; Giffin et al., 2020; Marrett et al., 2018).

538 **Appendix A A Brief Introduction to Causal Inference**

539 As Rubin (1974) points out, the problem of causal inference is a missing data prob-
 540 lem: given any treatment assigned to an individual unit, the potential outcome associ-
 541 ated with any alternate treatment is missing (Rubin, 1974; Imbens & Rubin, 2015). The
 542 assignment mechanism, therefore, plays a key role and answers questions such as: how
 543 is it determined which units get which treatments or, equivalently, which potential out-
 544 comes are realized and which are not.

545 We will now allude necessary notations in a general case. Let us index units in a
 546 population of size N by i , taking on values $1, \dots, N$, and denote W_i as the treatment in-
 547 dicator for unit i , taking on values 0 (control treatment) and 1 (active treatment). Let
 548 $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes of unit i for the control and active treat-
 549 ments, respectively. Recall only one of the potential outcomes will ultimately be real-
 550 ized and therefore possibly observed. Let Y_i^{obs} denotes this realized and possibly observed
 551 outcome:

$$Y_i^{obs} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases} \quad (\text{A1})$$

Analogously, let Y_i^{mis} denotes the missing potential outcome:

$$Y_i^{mis} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{if } W_i = 0, \\ Y_i(0) & \text{if } W_i = 1. \end{cases} \quad (\text{A2})$$

By the same token, \mathbf{Y}^{obs} and \mathbf{Y}^{mis} are the corresponding N -vectors. Usually we
 want to characterize the potential outcomes in terms of the observed and missing out-
 comes therefore we invert these notations:

$$Y_i(0) = \begin{cases} Y_i^{mis} & \text{if } W_i = 1, \\ Y_i^{obs} & \text{if } W_i = 0. \end{cases} \quad \text{and} \quad Y_i(1) = \begin{cases} Y_i^{mis} & \text{if } W_i = 0, \\ Y_i^{obs} & \text{if } W_i = 1. \end{cases} \quad (\text{A3})$$

We define the assignment mechanism to be the function that assigns probabilities
 to all 2^N possible values for the N -vector of assignments \mathbf{W} , given the N -vectors of po-
 tential outcomes $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ (Imbens & Rubin, 2015). The assignment mechanism
 is then a row-exchangeable function $Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1))$, taking on values in $[0, 1]$, sat-
 isfying

$$\sum_{\mathbf{w} \in [0,1]^N} Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = 1 \quad (\text{A4})$$

552 for all $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$.

553 **A1 The Stable Unit Treatment Value Assumption (SUTVA)**

554 Besides potential outcomes and assignment mechanism, additional assumptions are
 555 needed for causal validity under RCM. Here, we only touch one component of the Sta-
 556 ble Unit Treatment Value Assumption (SUTVA), the No Interference assumption, which
 557 will be sufficient for our purposes. The No Interference assumption states the treatments

558 applied to one unit do not affect the outcome for another unit (Imbens & Rubin, 2015).
 559 Put simply, the potential outcomes and assigned treatments for any unit do not vary with
 560 the treatments assigned to other units.

561 Appendix B Log-Gaussian Cox Point Process

562 The LGCP is chosen because it scored lowest Akaike information criterion (AIC)
 563 and Bayesian information criterion (BIC) values compared to other alternatives, shown
 564 in Table B1 (Baddeley et al., 2015; Illian et al., 2008).

Table B1. Comparison of AIC and BIC Values Between Some Point Process Models

	LGCP	Thomas	MatClust	Cauchy
AIC	45500.29	45604.65	45601.63	45630.82
BIC	45601.97	45607.34	45604.32	45633.50

565 Appendix C Fisher’s Exact Test with Sharp Null Hypothesis

566 Given data from a completely randomized experiments with SUTVA, Fisher’s Ex-
 567 act Test (FET) is assessing the sharp null hypothesis of no effect of the treatment ver-
 568 sus no treatment, that is, the null hypothesis under which, for each unit in the exper-
 569 iment, both values of the potential outcomes are identical (Mehta & Patel, 1983; Imbens
 570 & Rubin, 2015). Consider any test statistic T : a function of the stochastic assignment
 571 vector, \mathbf{W} ; the observed outcomes, \mathbf{Y}^{obs} . The sharp null hypothesis allows us to deter-
 572 mine the distribution of T , generated by the complete randomization of units across treat-
 573 ments. The test statistic is stochastic solely through the stochastic nature of the assign-
 574 ment. We refer to the distribution of the statistic determined by the randomization as
 575 the randomization distribution of the test statistic T . Using this distribution, we can com-
 576 pare the actually observed value of the test statistic, T^{obs} , against the distribution of T
 577 under the null hypothesis. An observed value that is “very unlikely”, given the null hy-
 578 pothesis will be taken as evidence against the null hypothesis using p-value (Imbens &
 579 Rubin, 2015). Hence, the FET approach entails the two steps: (i) the choice of a sharp
 580 null hypothesis, and (ii) the choice of test statistic.

581 C1 Ranked Statistics

582 Ranked Statistic is an important class of test statistics involves transforming the
 583 data to ranks before calculating the test statistics. Such a transformation is attractive
 584 when the data have a distribution with a substantial number of outliers (Imbens & Ru-
 585 bin, 2015). We consider the large-portion of zero-counts as outliers and use normalized
 586 rank to reduce the impacts of zero-counts in the test statistics. The definition for the
 587 normalized rank with ties is:

$$588 R_i = R_i(Y_1^{obs}, \dots, Y_N^{obs}) = \sum_{j=1}^N \mathbf{1}_{Y_j^{obs} < Y_i^{obs}} + \frac{1}{2} \left(1 + \sum_{j=1}^N \mathbf{1}_{Y_j^{obs} = Y_i^{obs}} - \frac{N+1}{2} \right) \quad (C1)$$

589 Given the N ranks R_i , $i = 1, \dots, N$, an obvious test statistic is the absolute value
 590 of the difference in average ranks for the treated and control units:

$$T^{rank} = |\bar{R}_t - \bar{R}_c| = \left| \frac{\sum_{i:W_i=1} R_i}{N_t} - \frac{\sum_{i:W_i=0} R_i}{N_c} \right| \quad (C2)$$

591 where we denote \bar{R}_t to be the ranked statistic in the active treatment group (Imbens &
 592 Rubin, 2015).

593 Figure C1 shows the improvements in the statistical results after the rank trans-
 594 formation. The p-value under the regular T-statistic distribution (right) is drastically
 595 reduced by an order of magnitude after taking normalized rank transformation (left),
 596 minimizing the impacts of zero-counts.

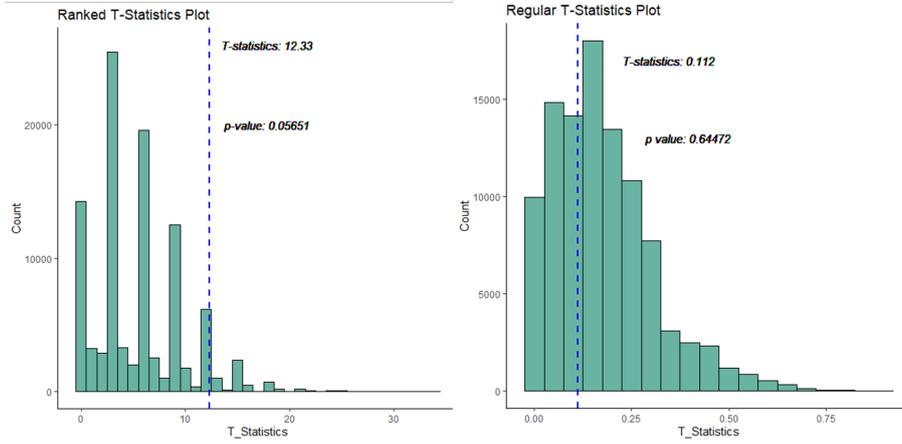


Figure C1. Comparison Between Ranked T-Statistics and Regular T-Statistics

597 Appendix D Neyman’s Repeated Sampling

598 The Fisher’s Exact Test provides limited information besides establishing the causal
 599 link between earthquakes and the presence of SWDs which is trivial to most researchers
 600 and policy makers (Grigoratos et al., 2020a; McClure et al., 2017; Frohlich et al., 2016b;
 601 Fan et al., 2019; Hennings et al., 2019). To broaden the scope of interpretation, Neyman’s
 602 Repeated Sampling is employed to find the average causal effects of the presence of SWDs
 603 on the number of earthquake in any grid over the entire study area. Neyman’s two basic
 604 questions are: (1) what would the average outcome be if all units were exposed to
 605 the treatment, $\bar{Y}(1)$? (2) How did that compare to the average outcome if all units were
 606 exposed to the control treatment, $\bar{Y}(0)$? Most importantly, what is the difference between
 607 these averages? Neyman’s approach was to develop an estimator of the average causal
 608 effect and derive its expectation and variance under repeated sampling.

609 D1 Unbiased Estimation of the Average Causal Effect

610 The population average causal effect τ_{fs} has the form:

$$611 \tau_{fs} = \bar{Y}(1) - \bar{Y}(0) = \sum_{i=1}^N (Y_i(1) - Y_i(0)) / N \quad (D1)$$

612 where $\bar{Y}(0)$ and $\bar{Y}(1)$ are the averages of the potential control and treated outcomes, re-
 613 spectively:

$$614 \bar{Y}(0) = \frac{1}{N} \sum_{i=1}^N Y_i(0) \quad (D2)$$

$$\bar{Y}(1) = \frac{1}{N} \sum_{i=1}^N Y_i(1) \quad (D3)$$

The LGCP enables generating thousands of random assignment vectors while preserving spatial correlation, where $N_t = \sum_{i=1}^N W_i$ are the number of grid blocks have SWDs and the remaining $N_c = \sum_{i=1}^N (1 - W_i)$ are the number of grid blocks absent SWDs. Because of the randomization, a natural estimator for the average causal effect is the difference in the average outcomes between those assigned to treatment and those assigned to control:

$$\hat{\tau}^{dif} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs} \quad (D4)$$

where:

$$\bar{Y}_c^{obs} = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs} \quad (D5)$$

$$\bar{Y}_t^{obs} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} \quad (D6)$$

To prove our estimator $\hat{\tau}^{dif}$ is unbiased for τ_{fs} , we use the fact that $Y_i^{obs} = Y_i(1)$ if $W_i = 1$, and $Y_i^{obs} = Y_i(0)$ if $W_i = 0$, to rewrite the estimator $\hat{\tau}^{dif}$ as:

$$\hat{\tau}^{dif} = \frac{1}{N} \sum_{i=1}^N \left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1 - W_i) \cdot Y_i(0)}{N_c/N} \right) \quad (D7)$$

The potential outcomes are treated as fixed, the only component in this statistic that is random are the treatment assignments obtained from our simulated realizations that preserve spatial correlation (Imbens & Rubin, 2015). Thus, $Pr(W_i = 1 | \mathbf{Y}(0), \mathbf{Y}(1)) = \mathbb{E}_W[W_i | \mathbf{Y}(0), \mathbf{Y}(1)] = N_t/N$ and $\hat{\tau}^{dif}$ is unbiased for the average causal effect τ_{fs} :

$$\begin{aligned} \mathbb{E}_W \left[\hat{\tau}^{dif} \mid \mathbf{Y}(0), \mathbf{Y}(1) \right] &= \frac{1}{N} \sum_{i=1}^N \left(\frac{\mathbb{E}_W[W_i] \cdot Y_i(1)}{N_t/N} - \frac{\mathbb{E}_W[1 - W_i] \cdot Y_i(0)}{N_c/N} \right) \\ &= \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \tau_{fs} \end{aligned} \quad (D8)$$

D2 The Sampling Variance of the Neyman Estimator

We develop an estimator for the sampling variance and appeal to a central limit argument for the large sample normality of $\hat{\tau}$ over its randomization distribution and use its estimated sampling variance to create a large-sample confidence interval for the average causal effect τ_{fs} .

$$\mathbb{V}_W(\hat{\tau}^{dif}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N} \quad (D9)$$

where

$$S_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{obs} - \bar{Y}_c^{obs} \right)^2 \quad (D10)$$

and

$$S_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=0} (Y_i^{obs} - \bar{Y}_t^{obs})^2 \quad (D11)$$

The third term, S_{ct}^2 , is the population variance of the unit-level treatment effects and is generally impossible to estimate empirically. Recall potential outcomes $Y_i(1)$ and $Y_i(0)$ cannot be both observed. Assuming the treatment effects are constant and additive ($Y_i(1) - Y_i(0) = \tau_{fs}$ for all units), then the third term is equal to zero and we have the reduced version of an unbiased estimator for the sampling variance:

$$\hat{V}^{neyman} = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} \quad (D12)$$

This estimator is widely used for two reasons. First, by implicitly setting the third term equal to zero, the expected value of the \hat{V}^{neyman} is at least as large as the true sampling variance equal to $\bar{Y}_t^{obs} - \bar{Y}_c^{obs}$, irrespective of the heterogeneity in the treatment effect, because the third term is non-negative. Hence, the confidence interval generated using this estimator in large sample would be greater or equal to the nominal coverage and is statistically conservative. Second, using \hat{V}^{neyman} as an estimator for the sampling variance of $\bar{Y}_t^{obs} - \bar{Y}_c^{obs}$ is that it is always unbiased for the sampling variance of $\hat{\tau}^{dif}$ as an estimator of the infinite super-population average causal effect. From above derived estimator for sampling variance, we construct 90% and 95% confidence intervals below, respectively (Imbens & Rubin, 2015).

$$CI^{0.90}(\tau_{fs}) = \left(\hat{\tau}^{dif} - 1.645 \cdot \sqrt{\hat{V}}, \hat{\tau}^{dif} + 1.645 \cdot \sqrt{\hat{V}} \right) \quad (D13)$$

$$CI^{0.95}(\tau_{fs}) = \left(\hat{\tau}^{dif} - 1.96 \cdot \sqrt{\hat{V}}, \hat{\tau}^{dif} + 1.96 \cdot \sqrt{\hat{V}} \right) \quad (D14)$$

Appendix E Extending to Studies of Distributed Volume

We highlight that the spatial point process model can be applied to previous studies to construct proper null distribution specific to each grid. More specifically, we generate a permutation of the distributed volume over the entire area of interest for every reproduction of the SWD point pattern. In particular, the permutation of the distributed volume is governed by a deterministic equation where the cumulative injection volume for all SWDs are sampled from the empirical histogram of the cumulative injection volume, shown in Figure E1. Given enough repetitions, there is an empirical distribution of distributed volume specific to each grid. We argue the resulting empirical distribution, specific to each grid, is a more appropriate null distribution than the one obtained from resampling for the following reasons:

1. Every permutation of distributed volume is simulated using the same deterministic physical model under realistic SWD allocation scheme where the cumulative injection volume for every SWD is drawn from the empirical distribution. Every permutation of distributed volume is realistic, more specifically, the distributed volume specific to each grid is realistic.
2. The null distribution of distributed volume for a particular grid block is constituted from different permutations of the distributed volume of the same grid. Because every permutation of the distributed volume is independent, every permutation of distributed volume of that grid is independent. This conforms with the *iid* sampling assumption.
3. Given enough permutation of distributed volume, those *iid* samples of distributed volume construct non-spatial and unbiased null distribution for every grid block.

685 4. The null distributions for all grids are converging under the Law of Large Num-
 686 bers (LLN) with enough permutations (Casella & Berger, 2001). This is crucial
 687 if p-value comparisons are needed between grid blocks, since p-values are not com-
 688 parable across different null distributions. Because the resampling method does
 689 not guarantee *iid* samples, it is less straightforward to apply LLN for convergence.

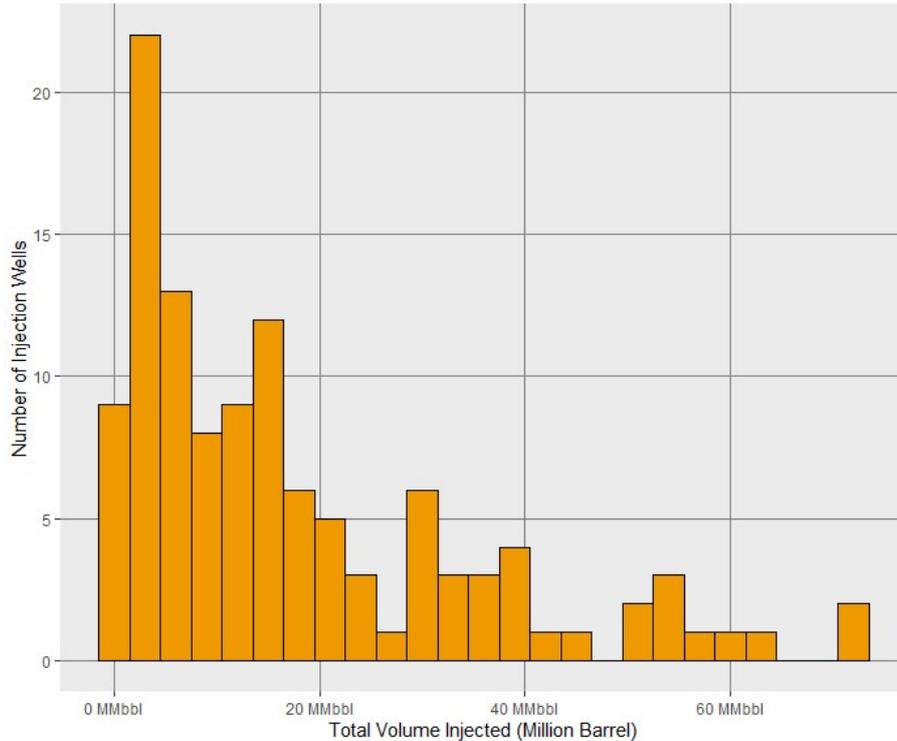


Figure E1. The empirical histogram of cumulative injection volume

690 Figure E2 shows the entire workflow from a random placement of SWDs to a per-
 691 mutation of distributed volume. We use inverse-distance method to calculate distributed
 692 volume for the DFW region. Grigoratos et al. (2020a) implemented a more appropriate
 693 physics-based diffusion model that includes time to calculate distributed volume across
 694 12 years period. We are precluding time in our current analysis, thus the inverse-distance
 695 method is sufficient to produce modest distributed volume across some time period

696 Acknowledgments

697 Thank you to Louis Quinones and Heather DeShon for providing their databases of the
 698 Dallas Forth-Worth region. Thank you to Iason Grigoratos for helping to decluster the
 699 earthquake catalog. The financial support of the TexNet Research and the Center for
 700 Integrated Seismicity Research is gratefully acknowledged. The earthquake data reported
 701 in this paper is available from DeShon et al. (2019) and the saltwater disposal data is
 702 available from Texas Data Repository (<https://dataverse.tdl.org/dataset.xhtml?persistentId=doi:10.18738/T8/CEQEDF>)
 703

704 References

705 Aldrich, J. (1995, 11). Correlations genuine and spurious in pearson and yule.
 706 *Statist. Sci.*, 10(4), 364–376. Retrieved from <https://doi.org/10.1214/ss/>

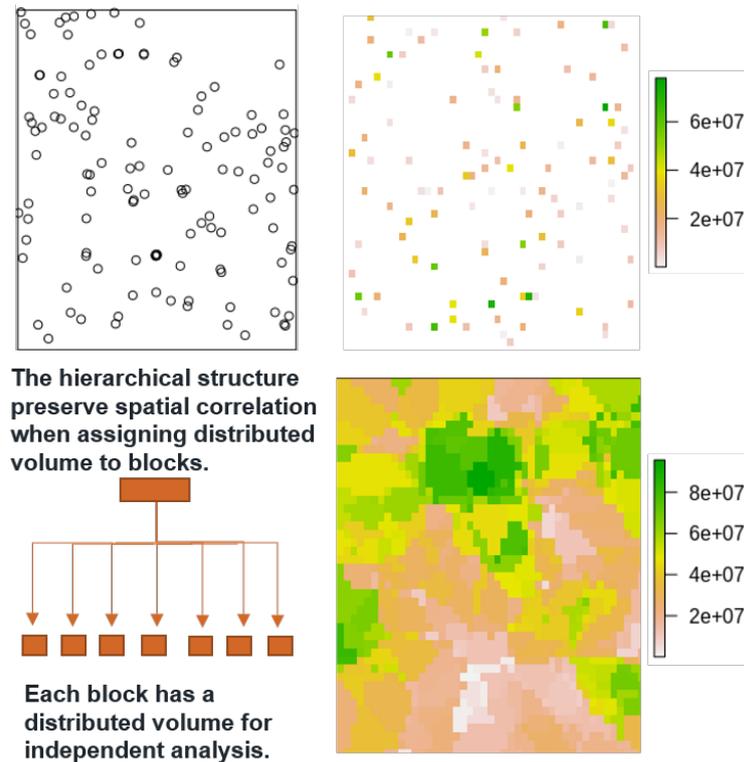


Figure E2. A SWD point pattern (top left panel) is simulated from the fitted LGCP and each SWD is assigned with a cumulative injection volume (top right panel) from the empirical histogram (Figure E2). Inverse-distance method is applied to calculate the distributed volume for the entire study area (bottom right panel). The hierarchical structure (bottom left) preserves the spatial correlation of every permutation of distributed volume at the high level where every permutation is independent. Under LLN, the empirical distribution of distributed volume specific to each grid converges to a proper null distribution.

- 1177009870 doi: 10.1214/ss/1177009870
- 707 Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns methodology*
708 *and applications with r*. CRC Press.
- 709 Carone, M., Dominici, F., & Sheppard, L. (2020). In pursuit of evidence in air
710 pollution epidemiology: the role of causally driven data science. *Epidemiology*,
711 *31*(1), 1–6.
- 712 Carsey, T. M., & Harden, J. J. (2013). *Monte carlo simulation and resampling meth-*
713 *ods for social science*. Sage Publications.
- 714 Casella, G., & Berger, R. (2001). *Statistical inference*. Duxbury Resource Center.
715 Textbook Binding.
- 716 DeShon, H. R., Hayward, C. T., Ogwari, P. O., Quinones, L., Sufri, O., Stump, B.,
717 & Beatrice Magnani, M. (2019). Summary of the North Texas Earthquake
718 Study Seismic Networks, 2013–2018. *Seismological Research Letters*, *90*(1),
719 387–394. doi: 10.1785/0220180269
- 720 Dominici, F., & Zigler, C. (2017). Best practices for gauging evidence of causality in
721 air pollution epidemiology. *American journal of epidemiology*, *186*(12), 1303–
722 1309.
- 723 Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142). Re-
724 trieved from <https://science.sciencemag.org/content/341/6142/1225942>
725

- 726 doi: 10.1126/science.1225942
- 727 Fan, Z., Eichhubl, P., & Newell, P. (2019). Basement Fault Reactivation by Fluid In-
- 728 jection Into Sedimentary Reservoirs: Poroelastic Effects. *Journal of Geophysi-*
- 729 *cal Research: Solid Earth*, *124*(7), 7354–7369. doi: 10.1029/2018JB017062
- 730 Fasola, S. L., Brudzinski, M. R., Skoumal, R. J., Langenkamp, T., Currie, B. S.,
- 731 & Smart, K. J. (2019). Hydraulic fracture injection strategy influences the
- 732 probability of earthquakes in the eagle ford shale play of south texas. *Geo-*
- 733 *physical Research Letters*, *46*(22), 12958-12967. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085167)
- 734 agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085167 doi:
- 735 <https://doi.org/10.1029/2019GL085167>
- 736 Friedrich, S., & Friede, T. (2020). Causal inference methods for small non-
- 737 randomized studies: Methods and an application in covid-19. *Contemporary*
- 738 *clinical trials*, *99*, 106213.
- 739 Frohlich, C., De Shon, H., Stump, B., Hayward, C., Hornbach, M., & Walter, J. I.
- 740 (2016a). A historical review of induced Earthquakes in Texas. *Seismological*
- 741 *Research Letters*, *87*(4), 1022–1038. doi: 10.1785/0220160016
- 742 Frohlich, C., De Shon, H., Stump, B., Hayward, C., Hornbach, M., & Walter, J. I.
- 743 (2016b). A Historical Review of Induced Earthquakes in Texas. *Seismological*
- 744 *Research Letters*, *87*(4), 1022–1038. doi: 10.1785/0220160016
- 745 Frohlich, C., Hayward, C., Rosenblit, J., Aiken, C., Hennings, P., Savvaidis, A.,
- 746 ... DeShon, H. R. (2020). Onset and Cause of Increased Seismic Activ-
- 747 ity Near Pecos, West Texas, United States, From Observations at the Laji-
- 748 tas TXAR Seismic Array. *Journal of Geophysical Research: Solid Earth*,
- 749 *125*(1), 1–14. Retrieved from <https://doi.org/10.1029/2019JB017737> doi:
- 750 [10.1029/2019JB017737](https://doi.org/10.1029/2019JB017737)
- 751 Gao, R., Pelletier, I., Horne, E., Nicot, J.-P., & Hennings, P. (2019). Basin-scale
- 752 hydrogeological modeling of the fort worth basin ellenburger group for pore
- 753 pressure characterization. In *Agu fall meeting abstracts* (Vol. 2019, pp. H53C–
- 754 01).
- 755 Giffin, A., Reich, B., Yang, S., & Rappold, A. (2020). *Generalized propensity score*
- 756 *approach to causal inference with spatial interference*.
- 757 Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J. M. (2013). Causal infer-
- 758 ence in public health. *Annual Review of Public Health*, *34*(1), 61-75. Retrieved
- 759 from <https://doi.org/10.1146/annurev-publhealth-031811-124606>
- 760 (PMID: 23297653) doi: 10.1146/annurev-publhealth-031811-124606
- 761 Grigoratos, I., Rathje, E., Bazzurro, P., & Savvaidis, A. (2020a). Earthquakes
- 762 Induced by Wastewater Injection, Part II: Statistical Evaluation of Causal Fac-
- 763 tors and Seismicity Rate Forecasting. *Bulletin of the Seismological Society of*
- 764 *America*, *110*(5), 2483–2497. doi: 10.1785/0120200079
- 765 Grigoratos, I., Rathje, E., Bazzurro, P., & Savvaidis, A. (2020b). Earthquakes
- 766 Induced by Wastewater Injection, Part I: Model Development and Hindcast-
- 767 ing. *Bulletin of the Seismological Society of America*, *110*(5), 2466–2482. doi:
- 768 [10.1785/0120200078](https://doi.org/10.1785/0120200078)
- 769 Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian Regression Tree
- 770 Models for Causal Inference: Regularization, Confounding, and Heterogeneous
- 771 Effects (with Discussion). *Bayesian Analysis*, *15*(3), 965 – 1056. Retrieved
- 772 from <https://doi.org/10.1214/19-BA1195> doi: 10.1214/19-BA1195
- 773 Hall, P., & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing.
- 774 *Biometrics*, *47*(2), 757–762. Retrieved from [http://www.jstor.org/stable/](http://www.jstor.org/stable/2532163)
- 775 [2532163](http://www.jstor.org/stable/2532163)
- 776 Hennings, P. H., Snee, J. E. L., Osmond, J. L., Deshon, H. R., Dommissie, R.,
- 777 Horne, E., ... Zoback, M. D. (2019). Injection-Induced Seismicity and Fault-
- 778 Slip Potential in the Fort Worth Basin, Texas. *Bulletin of the Seismological*
- 779 *Society of America*, *109*(5), 1615–1634. doi: 10.1785/0120190017
- 780 Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma’s induced seis-

- 781 micity strongly linked to wastewater injection depth. *Science*, 1255(March),
782 1251–1255.
- 783 Holland, P. W. (1986). Statistics and causal inference. *Journal of the American*
784 *Statistical Association*, 81(396), 945–960. Retrieved from [http://www.jstor](http://www.jstor.org/stable/2289064)
785 [.org/stable/2289064](http://www.jstor.org/stable/2289064)
- 786 Hornbach, M. J., Deshon, H. R., Ellsworth, W. L., Stump, B. W., Hayward,
787 C., Frohlich, C., ... Luetgert, J. H. (2015). Causal Factors for Seismicity
788 near Azle, Texas. *Nature Communications*, 6, 1–11. Retrieved from
789 <http://dx.doi.org/10.1038/ncomms7728> doi: 10.1038/ncomms7728
- 790 Hornbach, M. J., Jones, M., Scales, M., DeShon, H. R., Magnani, M. B., Frohlich,
791 C., ... Layton, M. (2016). Ellenburger wastewater injection and seismicity
792 in North Texas. *Physics of the Earth and Planetary Interiors*, 261, 54–68.
793 Retrieved from <http://dx.doi.org/10.1016/j.pepi.2016.06.012> doi:
794 10.1016/j.pepi.2016.06.012
- 795 Horne, E., Hennings, P., Osmond, J., & Deshon, H. (2020). Structural characteriza-
796 tion of potentially seismogenic faults in the Fort Worth Basin. *J*, 8(2). doi: 10
797 .1190/INT-2019-0188.1
- 798 Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). (Vol. Vol. 70). John Wiley
799 & Sons.
- 800 Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and*
801 *biomedical sciences: An introduction*. Cambridge University Press. doi: 10
802 .1017/CBO9781139025751
- 803 Justinic, A. H., Stump, B., Hayward, C., & Frohlich, C. (2013). Analysis of
804 the Cleburne, Texas, Earthquake Sequence from June 2009 to June 2010.
805 *Bulletin of the Seismological Society of America*, 103(6), 3083–3093. doi:
806 10.1785/0120120336
- 807 Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially
808 induced earthquakes in Oklahoma, USA: Links between wastewater injection and
809 the 2011 Mw 5.7 earthquake sequence. *Geology*, 41(6), 699–702.
- 810 Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014).
811 Sharp increase in central Oklahoma seismicity since 2008 induced by mas-
812 sive wastewater injection. *Science*, 345(6195), 448–451. Retrieved
813 from <https://science.sciencemag.org/content/345/6195/448> doi:
814 10.1126/science.1255802
- 815 Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in Oklahoma
816 respond to decreased saltwater injection rates? *Science Advances*, 2(11). doi:
817 10.1126/sciadv.1601542
- 818 Langenbruch, C., & Zoback, M. D. (2017). Response to Comment on “How will in-
819 duced seismicity in Oklahoma respond to decreased saltwater injection rates?”.
820 *Science Advances*, 3(8), 1–10. doi: 10.1126/sciadv.aao2277
- 821 Lund Snee, J.-E., & Zoback, M. D. (2016). State of stress in Texas: Implications
822 for induced seismicity. *Geophysical Research Letters*, 43(19), 10,208–10,214.
823 Retrieved from [https://agupubs.onlinelibrary.wiley.com/doi/abs/](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070974)
824 [10.1002/2016GL070974](https://doi.org/10.1002/2016GL070974) doi: <https://doi.org/10.1002/2016GL070974>
- 825 Marrett, R., Gale, J. F., Gómez, L. A., & Laubach, S. E. (2018). Correlation Anal-
826 ysis of Fracture Arrangement in Space. *Journal of Structural Geology*, 108, 16–
827 33. doi: 10.1016/j.jsg.2017.06.012
- 828 McClure, M., Gibson, R., Chiu, K.-K., & Ranganath, R. (2017). Identifying poten-
829 tially induced seismicity and assessing statistical significance in Oklahoma and
830 California. *Journal of Geophysical Research: Solid Earth*, 122(3), 2153–2172.
- 831 Mehta, C. R., & Patel, N. R. (1983). A network algorithm for performing Fisher’s
832 exact test in $r \times c$ contingency tables. *Journal of the American Statistical*
833 *Association*, 78(382), 427–434. Retrieved from [https://doi.org/10.1080/](https://doi.org/10.1080/01621459.1983.10477989)
834 [01621459.1983.10477989](https://doi.org/10.1080/01621459.1983.10477989) doi: 10.1080/01621459.1983.10477989
- 835 Ogwari, P. O., DeShon, H. R., & Hornbach, M. J. (2018). The Dallas-Fort Worth

- 836 airport earthquake sequence: Seismicity beyond injection period. *Journal of*
837 *Geophysical Research: Solid Earth*, 123(1), 553-563. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB015003)
838 agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB015003 doi:
839 <https://doi.org/10.1002/2017JB015003>
- 840 Panzeri, S., Magri, C., & Carraro, L. (2008). Sampling bias. *Scholarpedia*, 3(9),
841 4258. (revision #148550) doi: 10.4249/scholarpedia.4258
- 842 Papadogeorgou, G., Mealli, F., & Zigler, C. M. (2019). Causal inference with in-
843 terfering units for cluster and population level treatment allocation programs.
844 *Biometrics*, 75(3), 778–787.
- 845 Quinones, L., Deshon, H. R., Jeong, S., Ogwari, P., Sufri, O., Holt, M. M., &
846 Kwong, K. B. (2019). Tracking Induced Seismicity in the Fort Worth
847 Basin: A Summary of the 2008–2018 North Texas Earthquake Study Cata-
848 log. *Bulletin of the Seismological Society of America*, 109(4), 1203–1216. doi:
849 10.1785/0120190057
- 850 Quinones, L. A., DeShon, H. R., Magnani, M. B., & Frohlich, C. (2018, 04). Stress
851 orientations in the fort worth basin, texas, determined from earthquake fo-
852 cal mechanisms. *Bulletin of the Seismological Society of America*, 108(3A),
853 1124-1132. Retrieved from <https://doi.org/10.1785/0120170337> doi:
854 10.1785/0120170337
- 855 Reich, B., Yang, S., Guan, Y., Giffin, A., Miller, M. J., & Rappold, A. (2020). A re-
856 view of spatial causal inference methods for environmental and epidemiological
857 applications. *arXiv: Methodology*.
- 858 Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-
859 randomized studies. *Journal of educational Psychology*, 66(5), 688.
- 860 Savvaidis, A., Young, B., Huang, G.-c. D., & Lomax, A. (2019). Texnet: A statewide
861 seismological network in texas. *Seismological Research Letters*, 90(4), 1702–
862 1715.
- 863 Scales, M. M., DeShon, H. R., Magnani, M. B., Walter, J. I., Quinones, L., Pratt,
864 T. L., & Hornbach, M. J. (2017). A Decade of Induced Slip on the Causative
865 Fault of the 2015 Mw 4.0 Venus Earthquake, Northeast Johnson County,
866 Texas. *Journal of Geophysical Research: Solid Earth*, 122(10), 7879–7894. doi:
867 10.1002/2017JB014460
- 868 Schoenball, M., & Ellsworth, W. L. (2017). A systematic assessment of
869 the spatiotemporal evolution of fault activation through induced seis-
870 micity in oklahoma and southern kansas. *Journal of Geophysical Re-*
871 *search: Solid Earth*, 122(12), 10,189-10,206. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB014850)
872 agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB014850 doi:
873 <https://doi.org/10.1002/2017JB014850>
- 874 Walsh, F. R., & Zoback, M. D. (2015). Oklahoma’s recent earthquakes and saltwater
875 disposal. *Science Advances*, 1(5), 1–10. doi: 10.1126/sciadv.1500195
- 876 Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., & Rubinstein, J. L. (2015).
877 High-rate injection is associated with the increase in U.S. mid-continent seis-
878 micity. *Science*, 348(6241), 1336–1340. doi: 10.1126/science.aab1345
- 879 Zhai, G., & Shirzaei, M. (2018). Fluid Injection and Time-Dependent Seismic Haz-
880 ard in the Barnett Shale, Texas. *Geophysical Research Letters*, 45(10), 4743–
881 4753. doi: 10.1029/2018GL077696
- 882 Zigler, C., Forastiere, L., & Mealli, F. (2020). *Bipartite interference and air pollution*
883 *transport: Estimating health effects of power plant interventions*.
- 884 Zigler, C. M., Choirat, C., & Dominici, F. (2018). Impact of national ambient air
885 quality standards nonattainment designations on particulate pollution and
886 health. *Epidemiology (Cambridge, Mass.)*, 29(2), 165.
- 887 Zigler, C. M., & Dominici, F. (2014). Point: clarifying policy evidence with
888 potential-outcomes thinking—beyond exposure-response estimation in air
889 pollution epidemiology. *American journal of epidemiology*, 180(12), 1133–
890 1140.

891 Zigler, C. M., & Papadogeorgou, G. (2021). Bipartite Causal Inference with Interfer-
892 ence. *Statistical Science*, *36*(1), 109 – 123. Retrieved from [https://doi.org/](https://doi.org/10.1214/19-STS749)
893 [10.1214/19-STS749](https://doi.org/10.1214/19-STS749) doi: [10.1214/19-STS749](https://doi.org/10.1214/19-STS749)