Teng, G., and Baker, J. W. (2019). "Seismicity Declustering and Hazard Analysis of the Oklahoma–Kansas Region." *Bulletin of the Seismological Society of America*, 109(6), 2356–2366.

Seismicity Declustering and Hazard Analysis of the Oklahoma–Kansas Region

Ganyu Teng, Jack W. Baker

Abstract

This project evaluates the suitability of several declustering algorithms for induced seismicity and their impacts on hazard analysis in Oklahoma and Kansas. We consider algorithms proposed by Gardner and Knopoff (1974), Reasenberg (1985), Zaliapin and Ben-Zion (2013) and the stochastic declustering method (Zhuang *et al.*, 2002) based on the epidemic type aftershock sequence (ETAS) model (Ogata, 1988, 1998). Results show that the choice of declustering algorithm has a significant impact on the declustered catalog as well as the resulting hazard analysis in Oklahoma and Kansas. The Gardner and Knopoff algorithm, which is currently implemented in the U.S. Geological Survey one-year seismic hazard forecast for the central and eastern U.S., has unexpected features when used for this induced seismicity catalog. It removes 80% of earthquakes and fails to reflect the changes in background rates that have occurred in the past few years. This results in a slight increase in the hazard level from 2016 to 2017, despite a decrease in seismic activities in 2017. The Gardner and Knopoff algorithm also frequently identifies aftershocks with much stronger shaking intensities than their associated mainshocks. These features are mostly due to the window method implemented in the Gardner and Knopoff algorithm. Compared to the Gardner and Knopoff algorithm, the other three methods are able to capture the changing hazard level in the region. However, the ETAS model potentially overestimates the foreshock effect and generates negligible probabilities of large earthquakes being mainshocks. The Reasenberg and Zaliapin and Ben-Zion methods have similar performance on catalog declustering and hazard analysis. Compared to the ETAS method, these two methods are easier to implement and faster to generate the declustered catalog. Results from this study suggest that both Reasenberg and Zaliapin and Ben-Zion declustering algorithms are suitable for declustering and hazard analysis for induced seismicity in Oklahoma and Kansas.

Introduction

The declustered earthquake catalog is often used as input for probabilistic hazard analysis models. Declustering removes dependent earthquakes (i.e., foreshocks and aftershocks) so that the retained earthquakes follow a Poisson distribution, which allows a stable estimation of background rates from short catalogs. However, the identification of independent and dependent earthquakes is not absolute, and there are various declustering algorithms developed based on different assumptions and recorded data. As a result, a single algorithm may not be suitable for all seismic regions. For example, the U.S. Geological Survey (USGS) one-year seismic hazard forecast model used the Gardner and Knopoff declustering algorithm and developed one-year seismic hazard forecasts that accounted for both induced and natural earthquakes in the central and eastern U.S. from 2015 to 2018 (Petersen *et al.*, 2015, 2016, 2017, 2018). The model assumed that the hazard level could be predicted from the declustered earthquake catalog in the previous year. The 2018 forecast showed

an increase in seismic hazard for the Oklahoma and Kansas region despite the reduced number of earthquakes in the prior year (Petersen *et al.*, 2018). We will see that this unexpected hazard increase is largely due to the use of the Gardner and Knopoff declustering algorithm.

In this study, we studied four of the most popular declustering algorithms, the Gardner and Knopoff (1974), Reasenberg (1985), Zaliapin and Ben-Zion (2013), and the stochastic declustering method (Zhuang *et al.*, 2002) based on the epidemic type aftershock sequence (ETAS) model (Ogata, 1988, 1998), for seismic hazard analysis in Oklahoma and Kansas. Gardner and Knopoff (1974) introduced a window method that grouped earthquakes according to the space-time distances among them. For every cluster, the event with the maximum magnitude was indicated as the mainshock, and events within the space-time window of the mainshock were removed. The Gardner and Knopoff algorithm is the simplest and most commonly used algorithm, though it does not identify the higher order aftershocks, which are aftershocks generated by aftershocks (van Stiphout *et al.*, 2012). Another example of window methods is Wooddell and Abrahamson (2014). The second algorithm was a link method developed by Reasenberg (1985) that connected earthquakes according to the spatial and temporal distances among them. Specifically, Reasenberg (1985) developed an interaction zone, specified by space-time distances, centered at each event. Any event within the interaction zone of the previous event was identified as an aftershock. The space and time extent of the interaction zone was defined according to Omori's law and stress redistribution around each earthquake. The cluster was grown by rules of association, and the largest event in the cluster was considered as the mainshock. Other link methods are Frohlich and Davis (1990); Davis and Frohlich (1991). The third method was developed by Zaliapin and Ben-Zion (2013) based on the nearest-neighbor distances of events in space-time-energy domain. The distance between every pair of events was a function of the time difference, spatial separation and magnitude. The nearest-neighbor distances followed a bimodal distribution, where the mode with larger distances corresponded to background events. The last method is the probability-based declustering method (Zhuang *et al.*, 2002) based on the ETAS model (hereafter the ETAS method). The ETAS model was first introduced by Ogata (1988, 1998). It was a stochastic model that uses probabilistic methods to estimate the background seismic rate, which was assumed to be stationary. It did not identify aftershocks, mainshocks, and foreshocks explicitly. Instead, the output of the ETAS model was the probability of an event being a background event, i.e., a mainshock. It considered the probability distributions of the time and location of an offspring event based on Omori's law and other previous studies (Kagan and Knopoff, 1978, 1980). More recent studies also introduced nonstationary ETAS models (Marzocchi and Lombardi, 2008; Kumazawa *et al.*, 2014; Kattamanchi *et al.*, 2017).

This project evaluates the suitability of these four seismic declustering algorithms for induced earthquakes in the Oklahoma and Kansas region. We analyze their effects on declustered catalogs and hazard analyses in three steps: 1) checking how often declustering algorithms take out events that produce stronger shaking than retained events, 2) comparing results from the four algorithms when implemented on an induced catalogs (Oklahoma-Kansas), on natural catalogs (California), and on simulated catalogs where the earthquake occurrence follows the Poisson distribution, 3) conducting one-year hazard analyses for Oklahoma City and evaluating the effect of different algorithms on the ground motion hazard results.

Data and processing

We considered two regions of approximately $1.5 \times 10^5 km^2$ area in the Oklahoma and Kansas region (Figure 1a) and in California (Figure 1b) and obtained earthquake catalogs of the regions from the USGS earthquake catalog website. For the Oklahoma and Kansas region, the magnitude of

completeness (M_c) is 2.5 according to the Goodness of Fit test at 90% level (Wiemer and Wyss, 2000) and past studies in this area (Darold *et al.*, 2015; Vasylkivska and Huerta, 2017). For the selected region in California, M_c is less than 2.0 (Wiemer and Wyss, 2000). We then collected all earthquakes occurring between 2014 and 2018 and with magnitudes greater than 2.7. There were 5,167 and 639 events in the Oklahoma-Kansas and California regions, respectively. We also collected ground motion time series from Incorporated Research Institutions for Seismology (IRIS) Data Services for the same regions and period.

Both catalogs were then declustered using the four algorithms. For the Gardner and Knopoff (1974) algorithm, the spatial and time window lengths are defined as

$$\log_{10} T = \begin{cases} 0.032M + 2.7389 & \text{if } M \geq 6.5 \\ 0.5409M - 0.547 & \text{otherwise} \end{cases} \tag{1}$$

$$\log_{10} L = 0.1238M + 0.983 \tag{2}$$

where $T(days)$ and $L(km)$ are the window lengths in time and space, and M is the magnitude of an earthquake. For the Reasenberg algorithm, we used the default values in Reasenberg (1985) for all the parameters, such as the interaction radius for dependent events ($10\,km$) and the minimum looking ahead time for not clustered events ($1\,day$). However, we adjusted the minimum magnitude of completion to 2.7 based on the selected catalog. For the Zaliapin and Ben-Zion method, the distance between earthquake i and j was defined as

$$\eta_{ij} = \begin{cases} t_{ij}(r_{ij})^d 10^{-bm_i} & \text{if } t_{ij} > 0 \\ \infty & \text{otherwise} \end{cases} \tag{3}$$

where $t_{ij}(years)$ is the event intercurrence time, which is positive if event i occurs before event j; $r_{ij}(km)$ is the spatial distance between two events; d is the dimension of hypocenters or epicenters, which is 1.3 according to Zaliapin and Ben-Zion (2016); b describes the Gutenberg-Richter distribution. The catalog was declustered using the Gaussian mixture model approach on the nearest-neighbor distances. For the ETAS method, we used SEDA, a Matlab package developed by Lombardi (2017), to conduct the Time-Magnitude-Space ETAS modeling developed by Ogata (1988, 1998). This model estimates eight parameters, including the background rate and parameters of Omori's law, using the maximum likelihood method. The output of this model is the probability of each event in the catalog being a background event and its corresponding background rate. We declustered the catalog using a stochastic algorithm (Algorithm 2 in Zhuang *et al.* (2002)) based on the output probability. The results presented in this paper are mean values generated from 50 stochastically declustered catalogs. The ETAS model divides the catalog into a precursory and a target period. It is applied to the target period for parameter estimation and considers the triggering effect of the precursory earthquakes on the target period. Since the ETAS model by Ogata (1988, 1998) assumes that the background rate is independent of time, to capture the changing seismic activity, we divided the Oklahoma-Kansas catalog into four one-year target periods starting from 01/01/2014. Their corresponding precursory periods are the half-years before the target period, which starts from 07/01/2013. We considered the California catalog as a single target period, with its precursory period defined as the time from 07/01/2013 to 12/31/2013.

We plotted the Oklahoma-Kansas catalogs before and after declustering to illustrate the declustering process. Figure 2 shows the spatial distribution of earthquakes with magnitudes above 2.7 in the Oklahoma and Kansas region in 2016 and 2017. Most events were in central and northern Oklahoma. In 2016, there were 1,338 earthquakes with magnitudes above 2.7, and the number

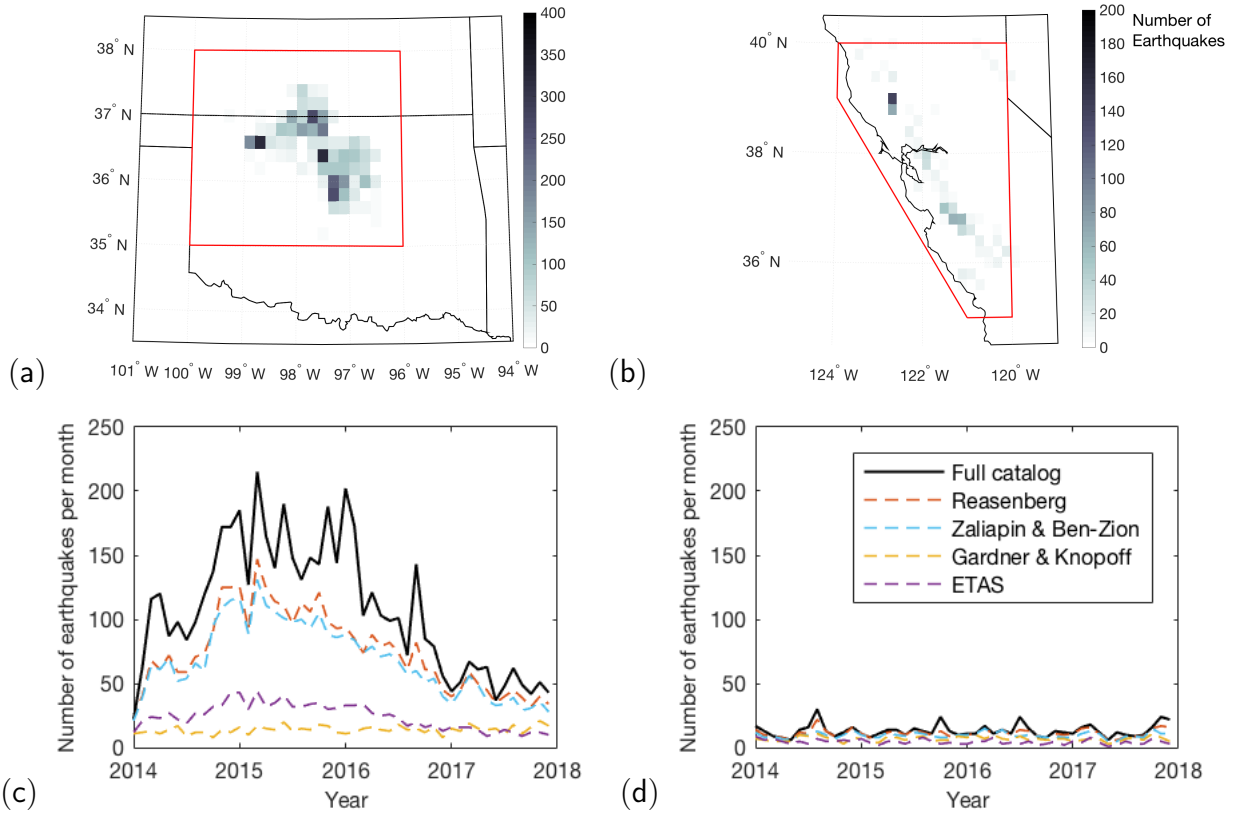


Figure 1: Selected regions (marked in red) with similar sizes in (a) Oklahoma and Kansas and (b) California. The number of earthquakes is plotted in blue. The number of earthquakes per month in (c) Oklahoma and Kansas and (d) California.

decreased to 618 in 2017 due to the reduction in wastewater injection activity in much of the region (Petersen *et al.*, 2018). Figure 3 shows the fraction of earthquakes removed by the four declustering algorithms for the two years. In both years, the Gardner and Knopoff method removed more earthquakes (87% and 67%) compared to the Reasenberg (33% and 18%) and Zaliapin and Ben-Zion methods (39% and 32%). All algorithms removed a smaller fraction of earthquakes in 2017 because the earthquakes were fewer and more spatially dispersed.

Determination of dependent events with stronger ground motions

The typical assumption in hazard analysis is that the declustering process removes dependent events that produce less intense shaking than those retained, so it does not affect the hazard results significantly (Cornell, 1968), though we note that some studies have considered dependent events in the hazard analysis (Boyd, 2012; Marzocchi and Taroni, 2014; Chioccarelli *et al.*, 2018). In this section, we tested the validity of the assumption about declustering when using the selected declustering algorithms in the Oklahoma and Kansas region. In particular, we compared shaking intensities, measured by spectral accelerations of horizontal components at a period of 0.1s ($Sa(0.1s)$), of ground motions of mainshocks and dependent events. The dependent events were defined to be stronger than their mainshock if there was at least one station with $Sa_{dependent\ event}/Sa_{mainshock} > 1.0$. $Sa_{mainshock}$ and $Sa_{dependent\ event}$ are the recorded $Sa(0.1s)$ of the mainshock and dependent event

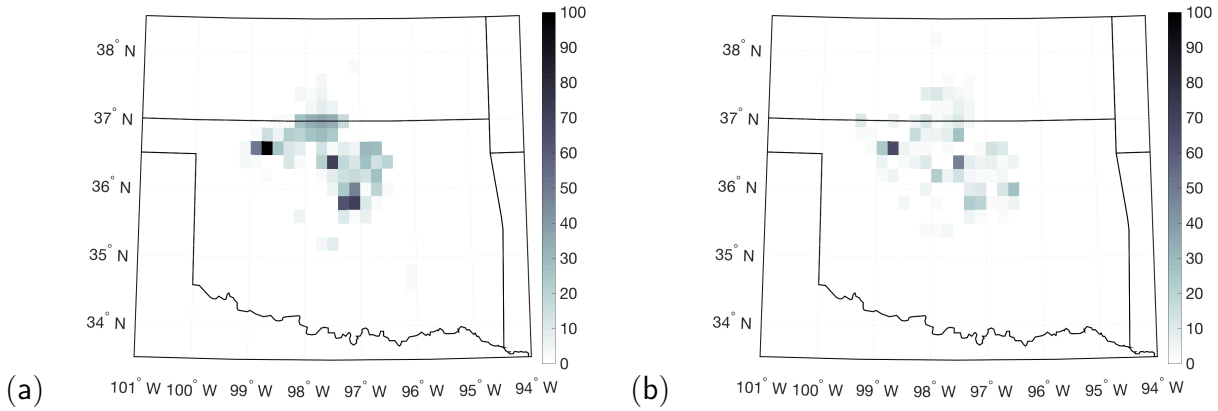


Figure 2: The number of earthquakes with magnitudes above 2.7 in the Oklahoma and Kansas region in (a) 2016 and (b) 2017.

ground motions, respectively. Note that this metric is not intended to define the earthquake itself as stronger, but merely the resulting ground motion at a given station; the motivation for this metric is that if such events and ground motions are removed from the historic catalog, the future catalogs considered by the probabilistic seismic hazard analysis (PSHA) calculations will be missing events with the potential to cause ground shaking of interest. Figure 4 illustrates two examples of response spectra of a mainshock ground motion and its dependent events' ground motions defined by the Gardner and Knopoff algorithm. In Figure 4a, the mainshock ground motion has a higher shaking intensity than its dependent events, with a maximum $Sa_{dependent\ event}/Sa_{mainshock}$ of 0.11. In Figure 4b, the aftershock ground motion is more damaging than the mainshock ground motion, with $Sa_{dependent\ event}/Sa_{mainshock} = 29.3$. We declustered the 2014-2018 Oklahoma-Kansas catalog and computed the number of mainshocks having at least one $Sa_{dependent\ event}/Sa_{mainshock}$ larger than a given ratio. We did not evaluate the ETAS method because it does not match aftershocks and foreshocks to specific mainshocks.

Figure 5 illustrates the cumulative fraction of mainshocks with maximum $Sa_{dependent\ event}/Sa_{mainshock}$ less than values of interest. Clusters without available recorded ground motions were excluded in the calculation. For the Gardner and Knopoff algorithm, approximately 84% of mainshocks had dependent events with higher shaking intensities, and 39% had $Sa_{dependent\ event}/Sa_{mainshock} > 4$ (a somewhat arbitrary threshold chosen for illustration), which indicates that it removed some events that are much stronger than the ones retained. The Reasenberg and Zaliapin and Ben-Zion methods have similar performances. Both identified stronger events as aftershocks or foreshocks, but the $Sa(0.1s)$ ratios were smaller compared to values from the Gardner and Knopoff method, with only around 10% of the mainshocks having $Sa_{dependent\ event}/Sa_{mainshock} > 4$. At all thresholds, the Gardner and Knopoff method had the highest fraction of stronger events (Figure 5). These results illustrate that compared to Reasenberg and Zaliapin and Ben-Zion methods, the Gardner and Knopoff algorithm potentially identifies more dependent events that have stronger shakings than mainshocks. To check the robustness of the results, we repeated the analysis by considering two additional criteria: 1) selecting ground motions with site-event distance $< 40km$, 2) selecting ground motions with least 3 stations having $Sa_{dependent\ event}/Sa_{mainshock} > 1$. The results for all algorithms were similar to the values in Figure 5.

Large $Sa_{dependent\ event}/Sa_{mainshock}$ values indicate the removal of earthquakes far apart but close in magnitude, which suggests that the algorithms remove events causing ground motions that could

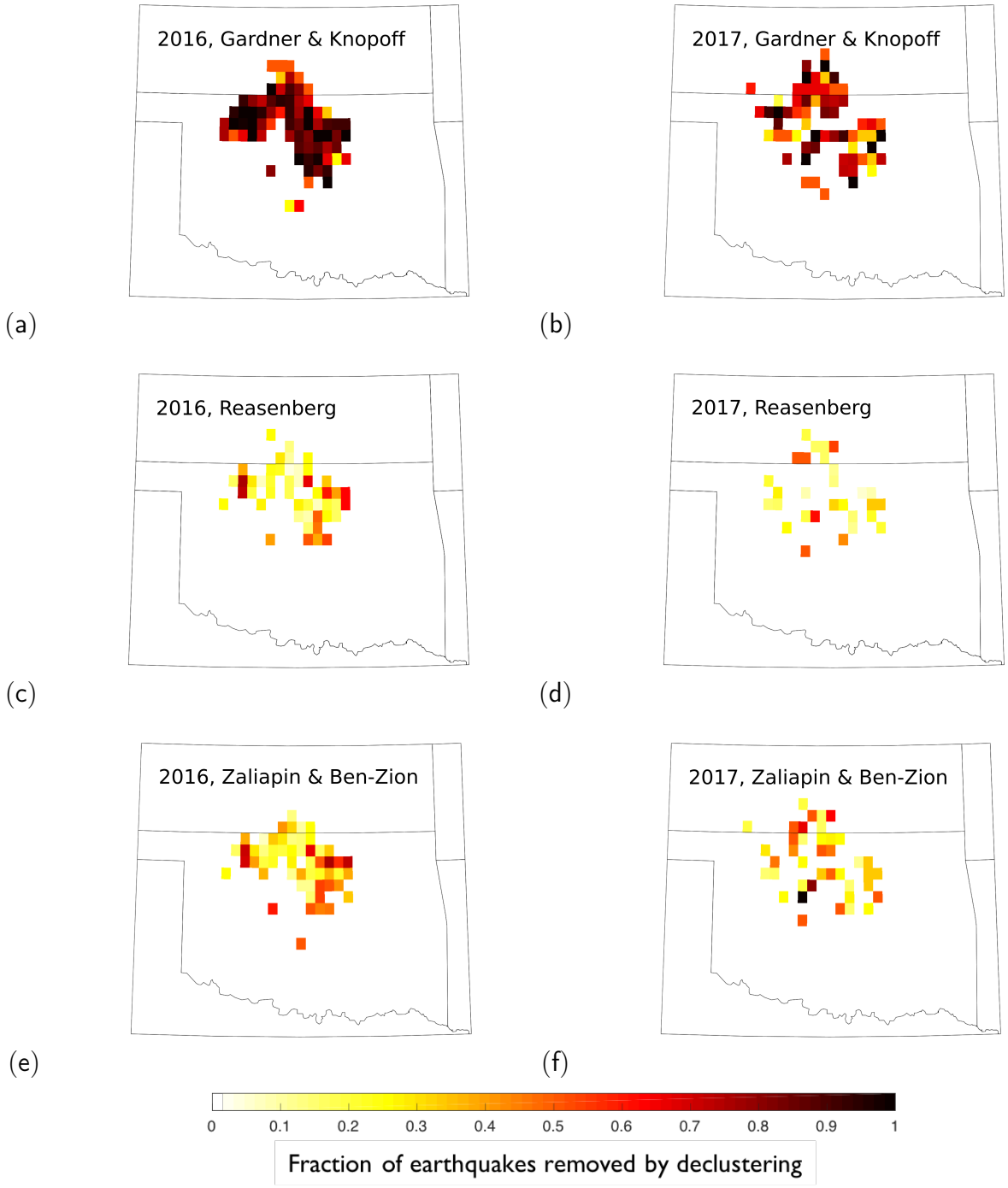


Figure 3: Fraction of earthquakes removed by the Gardner and Knopoff algorithm in (a) 2016 and (b) 2017, the Reasenberg algorithm in (c) 2016 and (d) 2017, the Zaliapin and Ben-Zion algorithm in (e) 2016 and (f) 2017 and the ETAS method in (g) 2016 and (h) 2017.

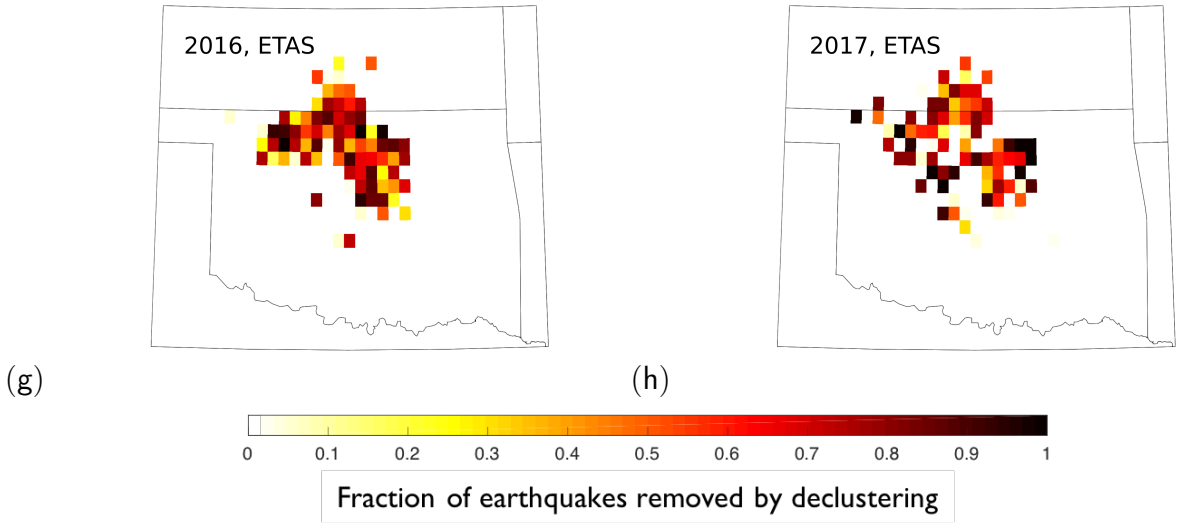


Figure 3: continued

potentially contribute to hazard analysis. The exclusion of many such events distorts the spatial distribution of large events and results in an underestimation of hazard level. The dependent events with large $Sa_{dependent\ event}/Sa_{mainshock}$ values are due to the large spatial window of the Gardner and Knopoff algorithm. A station experiences a stronger shaking from an aftershock than its mainshock when the events are close in magnitude but far apart from each other, with the aftershock closer to the station. For example, Figure 6 is a map of earthquakes from the Figure 4b cluster, in which the aftershock identified by the Gardner and Knopoff algorithm has much stronger shaking than its mainshock. The aftershock is much closer to the station compared to the mainshock. The small difference in magnitudes (the $M3.3$ mainshock and the $M3.0$ aftershock) but large separation distance ($20.2km$) between the aftershock and its mainshock resulted in the large $Sa_{dependent\ event}/Sa_{mainshock}$ value. The large ratio generated by the Gardner and Knopoff method suggests that it potentially groups earthquakes that are far apart but similar in magnitude into one cluster. As a result, this algorithm identifies aftershocks that can be as strong as mainshocks and removes them from the catalog. More earthquakes are declustered by the Gardner and Knopoff algorithm, and this can result in an underestimation of hazard. This large spatial window ($24.6km$ for a $M3.3$ earthquake) was developed based on earthquakes in California and is usually not an issue for natural seismicity as there are seldom two earthquakes, with magnitudes above 3.0 and separated by $20km$, happening within a week. However, this spatial window is not suitable for induced seismicity in Oklahoma due to the high occurrence rate and close time-space separation distances among events. Overall, compared to the Reasenberg and Zaliapin and Ben-Zion methods, the Gardner and Knopoff algorithm removes more events that have much stronger shaking than their mainshocks. Preferred approaches to minimize the removal of stronger dependent events are using the Reasenberg and Zaliapin and Ben-Zion algorithms, or developing region-appropriate parameters for the Gardner and Knopoff algorithm.

Declustered catalogs in Oklahoma-Kansas and California regions

We next compared the selected algorithms' performances for the California and Oklahoma-Kansas catalogs defined in the Data and Processing section. After declustering the catalogs, we first tested

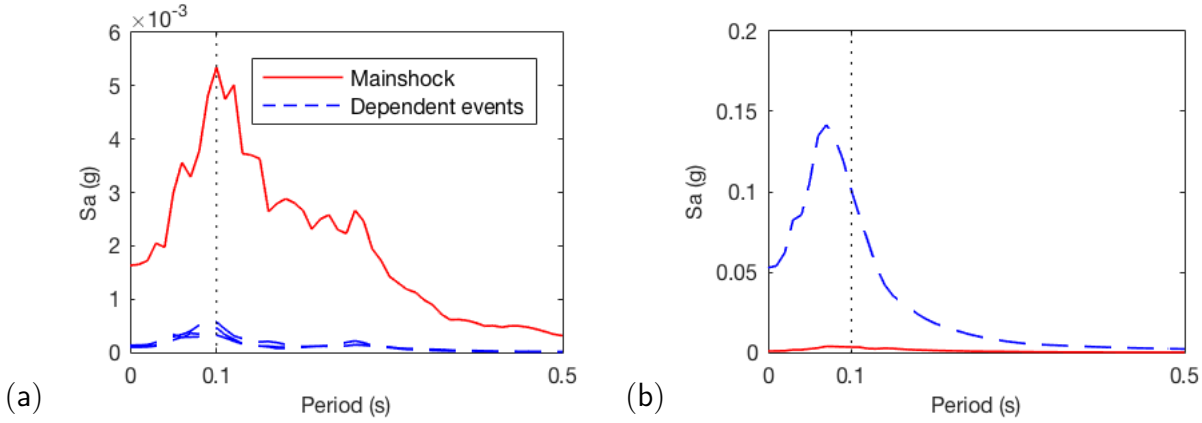


Figure 4: Response spectra of two sample clusters where (a) the mainshock has stronger shaking than its dependent events; (b) the dependent event has stronger shaking than the mainshock.

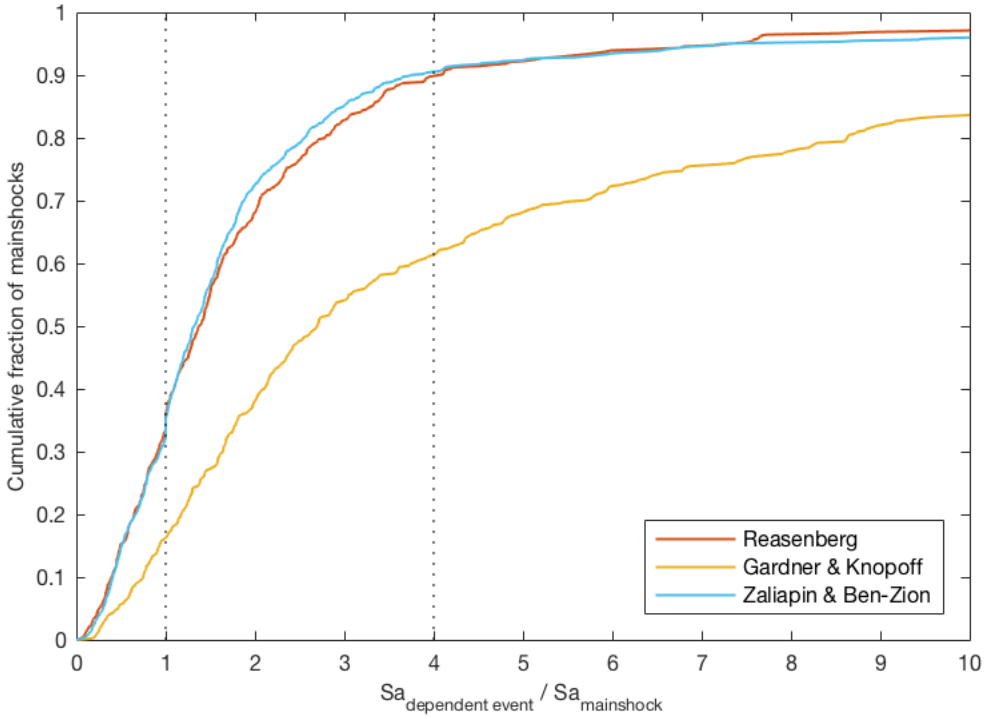


Figure 5: Fraction of mainshocks with maximum $Sa_{dependent\ event}/Sa_{mainshock}$ less than a given ratio.

the temporally homogeneous Poisson hypothesis of the declustered catalogs using the conditional chi-square test and the Kolmogorov–Smirnov test at a significance level of 0.1 (Luen and Stark, 2012). The non-Poissonian sequences of the induced seismic catalogs are due to the presence of dependent earthquakes and the temporally in-homogeneous base rate caused by human activity. An effective declustering catalog should remove dependent earthquakes while capturing the temporally in-homogeneous rate. For four-year declustered catalogs, only the catalog from the Gardner and Knopoff method did not reject the null hypothesis. The Gardner and Knopoff method removed

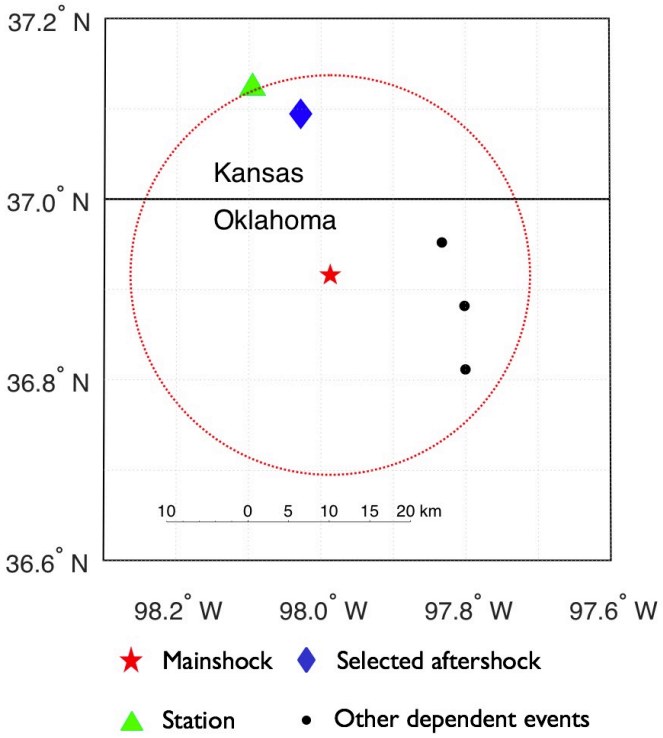


Figure 6: Maps of a cluster where the aftershock is stronger than its mainshock. The response spectra of the mainshock and the selected aftershock are shown in Figure 4b.

many earthquakes and produced a catalog with a constant rate in time (Figure 1c). This suggests that this method removes both dependent earthquake sequences and the changing base rate. The other three declustered catalogs were nearly Poissonian over four-month time intervals. For the Reasenberg and Zaliapin and Ben-Zion algorithms, the null hypothesis was not rejected for more than 70% of the four-month declustered catalogs. For the ETAS model, more than 90% of the four-month declustered catalogs were not rejected. The results suggest that these three methods are able to remove dependent events and generate Poissonian catalogs for a short time interval in which the base rate is almost constant. They also preserve the non-stationary base rate when considering a longer time interval (Figure 1c). For induced seismicity, we often conduct the hazard analysis for a short time interval (no longer than a one-year period), we thus conclude that the Reasenberg, Zaliapin and Ben-Zion and ETAS methods are suitable even though their declustered 4-year catalogs were not consistent with the null hypothesis.

We then studied the magnitude distribution of the declustered catalog by fitting it with a truncated Gutenberg-Richter distribution (Cornell and Vanmarcke, 1969).

$$log_{10}N(M \geq m) = a - b(m - m_{min}) \tag{4}$$

where m is the magnitude of interest, $m_{min} = 2.7$ is the predefined minimum magnitude, a is the total number of earthquakes with $m \geq m_{min}$, and b describes the decrease in the number of earthquakes as magnitudes increase. The b was estimated using the maximum likelihood method proposed by Utsu (1965):

$$\hat{b} = \frac{nlog_{10}e}{\sum_{i=1}^{n}(m_i - m_{min} + m_{bin}/2)} \tag{5}$$

where m_i is the magnitude of event i, n is the number of events in the declustered catalog and m_{bin} is the binning width of the catalog.

The results are summarized in Figure 7. All $\hat{b}$ for the California catalog are close to 1.0, which suggests that the selected methods have similar performances on natural seismicity. However, in Oklahoma and Kansas, the $\hat{b}$ from different declustering algorithms vary significantly. The Gardner and Knopoff method removed more small earthquakes ($M < 3.5$) compared to others, producing the smallest fitted $\hat{b} = 0.79$. This resulted in lower predicted occurrence rates for small events but higher rates for large events. The values computed from the Reasenberg and the Zaliapin and Ben-Zion algorithms are close to each other as well as the one obtained from the undeclustered catalog. The ETAS method resulted in the largest $\hat{b}$ as it produced very small probabilities of large earthquakes (e.g., $M > 4.5$) being mainshocks, which resulted in low occurrence rates for those earthquakes as well as an underestimation in the hazard level. This was also observed by Console *et al.* (2010), in which they suggested that the ETAS model tends to overestimate the foreshock effect. Moreover, for the Oklahoma and Kansas catalog (Figure 7a), the fit of the Gutenberg-Richter distribution is not perfect for magnitudes greater than 4. This can be addressed by using a larger assumed magnitude of completeness. We repeated the analysis using $M_c = 3.5$ and observed that the effect of declustering algorithms on $\hat{b}$ was still significant, with The ETAS model having $\hat{b}$ of 2.22 and the Gardner and Knopoff method having the smallest $\hat{b}$ of 1.24.

We also compared the reduction in the number of earthquakes after declustering by the selected methods. Figure 1c and 1d illustrate the number of mainshocks per month from 2014 to 2018 in the two regions. All declustering algorithms worked well in California as they took out the sporadic aftershock sequences and resulted in relatively constant background rates. In Oklahoma and Kansas, the Gardner and Knopoff algorithm removed a large number of earthquakes and failed to capture the changing background seismic rate. This resulted in relatively uniform rates throughout the four years, which was inconsistent with the changing seismic activity in the region. The ETAS method reflected the changing seismicity, but the change is minor compared to the full catalog. This is because the ETAS model assumes that the background intensity at a given location remains constant for the time range considered (i.e., one year in this study). This can be improved by considering nonstationary ETAS models (Marzocchi and Lombardi, 2008; Kumazawa *et al.*, 2014; Kattamanchi *et al.*, 2017). Both the Reasenberg and Zaliapin and Ben-Zion algorithms preserved the changing seismic rate while removing aftershock sequences. The results address the importance of selecting a suitable declustering algorithm for induced earthquakes in the Oklahoma and Kansas region.

Declustering for a simulated Poisson catalog

In addition to testing the declustering algorithms on recorded earthquake catalogs, we explored how they behaved on Monte-Carlo-simulated catalogs where the earthquake occurrence follows a Poisson distribution (i.e., with no foreshocks, aftershocks or clusters). Since declustering is designed to remove dependent (i.e., non-Poissonian) earthquakes, an effective declustering algorithm should remove a negligible number of earthquakes in the simulated catalogs.

We defined a $200km \times 200km$ region and generated 5,000 earthquakes within the region. We made three assumptions: 1) the earthquake occurrence was a stationary Poisson process with a predefined mean of λ ($count/year/km^2$). The catalog length was set such that it enveloped 5,000 earthquakes and had a coefficient of variation of 0.0141. 2) the earthquake location was uniformly distributed in the $200km \times 200km$ region, 3) the earthquake magnitude followed a truncated Gutenberg–Richter distribution between 2.7 and 7.0, with $b = 1.0$. Figure 8 visualizes 4 years (out of 69 years) of

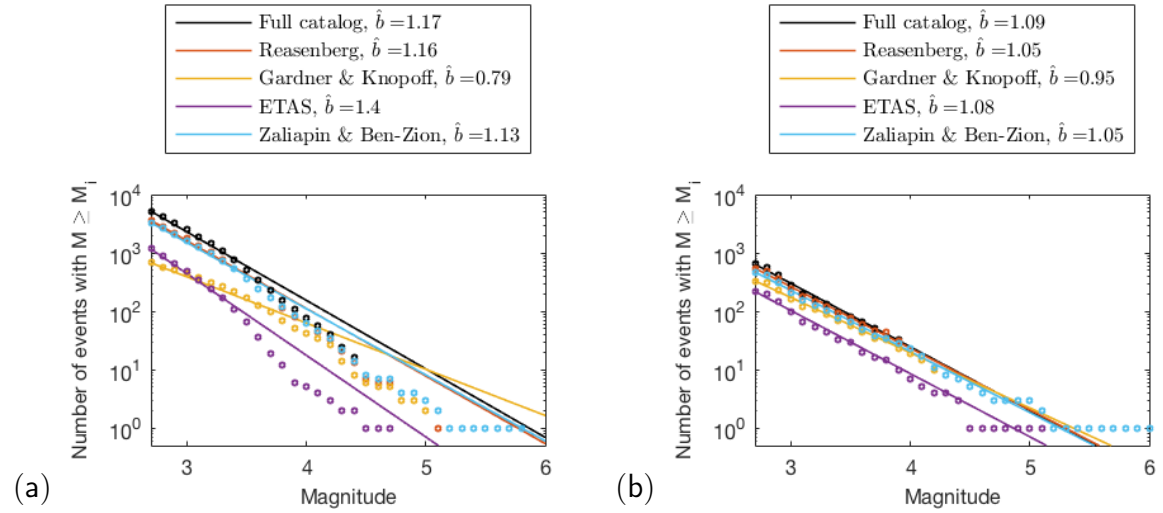
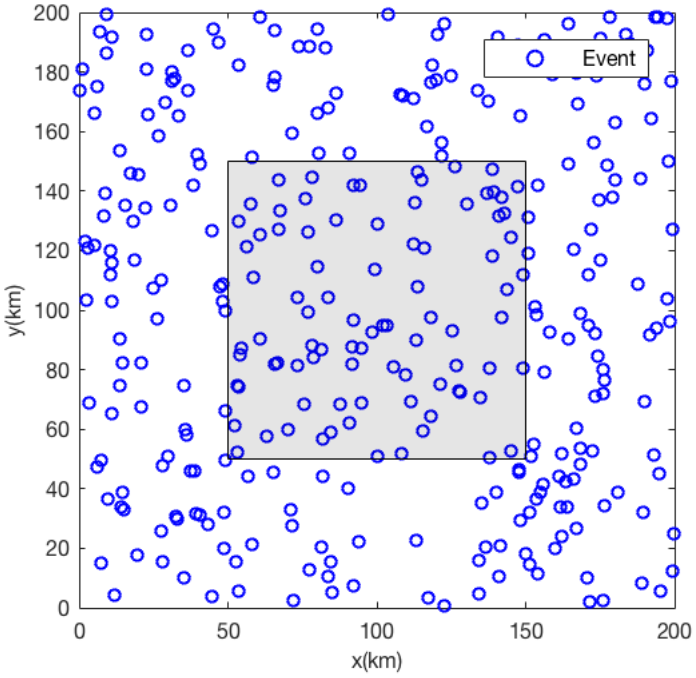


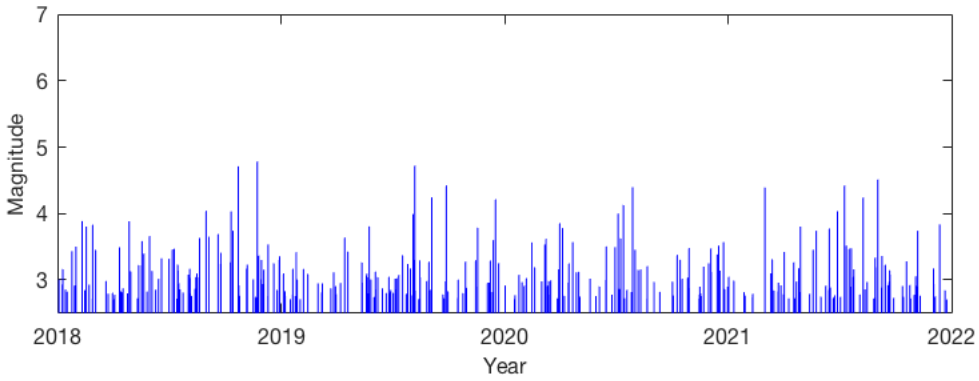
Figure 7: The number of earthquakes with magnitudes greater than values of interest in (a) Oklahoma and Kansas and (b) California. The dots are data from the declustered catalog, and the line describes the truncated Gutenberg-Richter distribution using the fitted b (Equation 5).

one simulated earthquake catalog with $\lambda = 1.825 \times 10^{-3}(count/year/km^2)$. Figure 8a shows the locations of simulated earthquakes and Figure 8b illustrates the earthquake occurrence versus time. We then used the four algorithms to decluster the simulated catalogs. To avoid any inaccuracy caused by clusters near the boundary, we computed the removal ratio as the ratio between the number of events removed and the total number of simulated events in the $100km \times 100km$ region at the center (the shaded area in Figure 8a).

Figure 9 summarizes the removal ratios plotted versus λ for simulated catalogs with a range of λ, for each of the selected declustering algorithms. The removal ratios from the California and Oklahoma-Kansas catalogs are also plotted for reference. In Figure 9, the Gardner and Knopoff curve shows an unexpected feature — the removal ratio increases with λ. It reaches above 0.5 when $\lambda > 5 \times 10^{-3} count/year/km^2$. Moreover, the Gardner and Knopoff curve is slightly below the ratios computed from the declustered California and Oklahoma-Kansas catalogs, suggesting that the Gardner and Knopoff method removed independent earthquakes while declustering the recorded catalogs. This unexpected behavior is due to the large time and space windows implemented in the Gardner and Knopoff algorithm. The removal ratio from the Zaliapin and Ben-Zion method stays around 32% for different λ. This is because, for a Poisson process, the distribution of nearest-neighbor distances with different λ is unimodal and can be approximated by the Weibull distribution (Zaliapin and Ben-Zion, 2013). As a result, the non-zero removal ratio is caused by the implementation of the Gaussian mixture model in this method. Removal ratios computed using the Reasenberg and ETAS methods remain at almost zero for all λ, which is the desired behavior and is mostly due to the implementation of Omori's law in both methods. Omori's law states that the frequency of aftershocks decays with the time after the mainshock and is proportional to the reciprocal of the time difference. This prevents declustering Poissonian earthquake catalogs with large λ. Overall, an effective declustering algorithm should remove dependent earthquakes only and thus a negligible number of events in the simulated Poissonian catalogs. Thus we conclude that the Reasenberg and ETAS methods are more effective compared to the other two algorithms.



(a)



(b)

Figure 8: An example of 4 years (out of 69 years) of one simulated Poissonian catalog with $\lambda = 1.825 \times 10^{-3}(count/year/km^2)$. (a) Locations of earthquakes, where circles are simulated earthquakes. We computed the removal ratios of the declustering algorithms using the earthquakes in the shaded $100km \times 100km$ region. (b) Magnitudes and occurrence times of earthquakes, where every vertical line represents one occurrence.

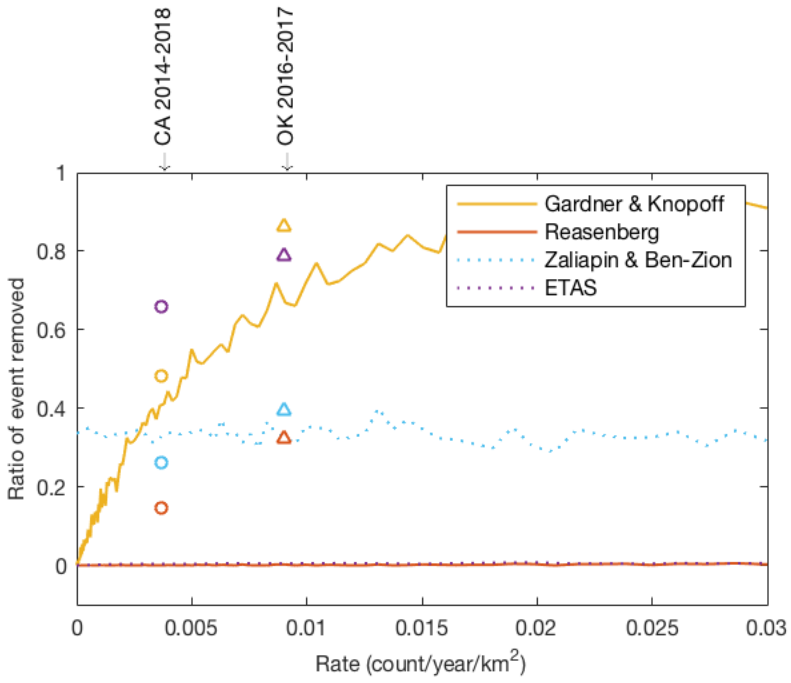


Figure 9: Removal ratios when testing four declustering algorithms on simulated Poissonian catalogs with different mean occurrence rates.

Hazard analysis for Oklahoma City

After studying the performance of the four declustering methods on the declustered catalogs, we explored their effects on the induced seismic hazard analysis by generating 2016 and 2017 hazard curves for Oklahoma City. The hazard curve described the annual rate of exceeding a given intensity measure (IM) and was computed using the PSHA method:

$$\lambda(IM > im) = \sum_{i=1}^{n} \nu_i \int_{m_{min}}^{m_{max}} P(IM > im | r_i, m) f_M(m) dm \tag{6}$$

The PSHA integrates hazard generated by earthquakes with magnitudes between m_{min} and m_{max} from all seismic sources. In this study, we divided the Oklahoma and Kansas region into $0.2^\circ \times 0.2^\circ$ grids and considered each grid as a seismic point source, with the site-source distance defined as the distance from the center of the grid to the site of interest. ν_i is the annual rate of earthquakes for a seismic source i and is computed from a declustered catalog. We estimated ν_i as the number of mainshocks in a given year. $P(IM > im | r_i, m)$ is the probability of an earthquake, with a magnitude of m and $r_i\,km$ away from the site of interest, producing an IM higher than a predefined value. Its value is determined using a ground motion prediction equation (GMPE). In this study, the IM was defined as $Sa(0.1s)$, which was determined using the GMPE introduced by Atkinson

(2015). For every seismic source, f_M was the probability density function of the Gutenberg-Richter distribution truncated between magnitudes of 2.7 and 7.0. Though b often varies with regions, we used $b = 1.0$ to be consistent with the USGS seismic hazard forecast (Petersen *et al.*, 2015, 2016, 2017, 2018).

Figure 10 illustrates one-year hazard curves generated using the 2016 and 2017 Oklahoma-Kansas catalogs. The Gardner and Knopoff algorithm with $b = 1.0$, the method used in the USGS seismic hazard forecast, generated the lowest hazard level in 2016. This was because the Gardner and Knopoff method removed more earthquakes compared to other declustering algorithms, as shown in Figure 3, which resulted in lower earthquake rates (ν_i in Equation 6) and thus lower hazard levels. The hazard curves from the Reasenberg and Zaliapin and Ben-Zion algorithms are similar and can be four times higher than the Gardner and Knopoff case. Moreover, besides the Gardner and Knopoff curve, all other curves show a significant drop in 2017, which is consistent with the expectation as the number of events was reduced by half in 2017. However, the Gardner and Knopoff method produced an increased hazard level in 2017 due to the use of window method together with the sparser earthquake distribution in 2017, which resulted in the increased number of mainshocks identified. This can explain the increased hazard level in the 2018 USGS one-year seismic hazard forecast, despite the reduced number of recorded earthquakes in 2017. Overall, the Gardner and Knopoff and ETAS methods remove more events and generate lower hazard levels compared to the other two algorithms. The Gardner and Knopoff algorithm also fails to capture the change in the earthquake rate and thus seismic hazard.

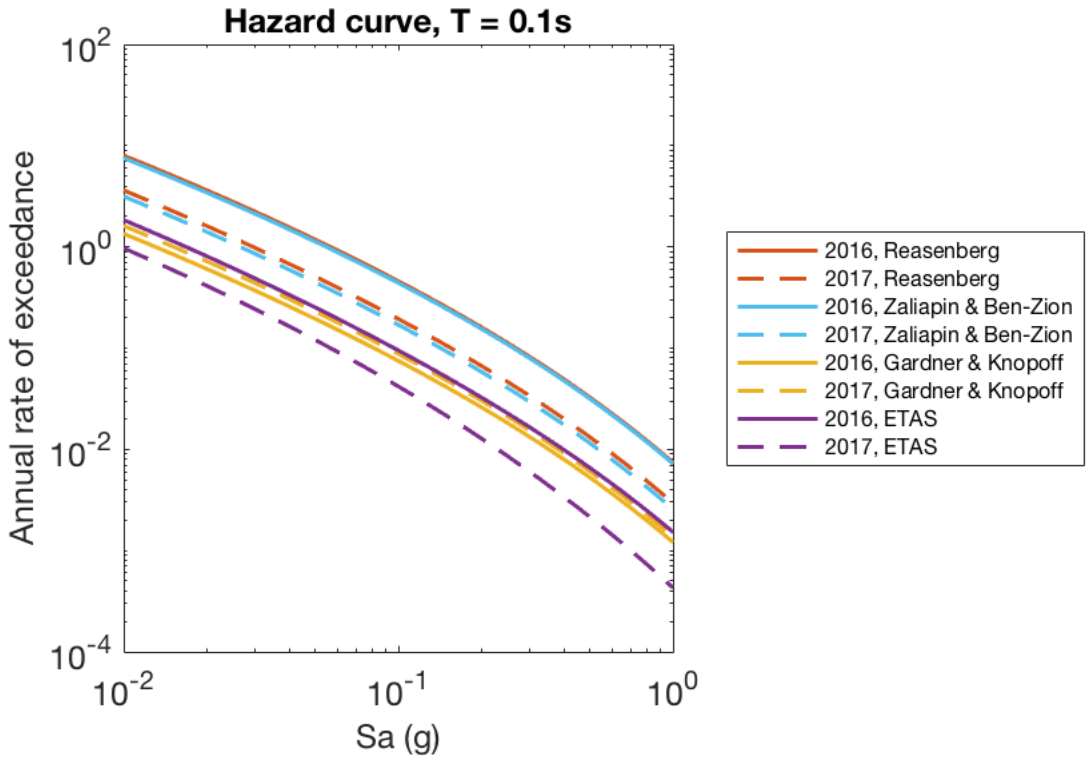


Figure 10: One-year hazard curves of $Sa(0.1s)$ for Oklahoma City in 2016 and 2017 using the four declustering methods.

Conclusions

This study evaluated the suitability of using four declustering algorithms proposed by Gardner and Knopoff (1974), Reasenberg (1985), Zaliapin and Ben-Zion (2013) and the stochastic declustering method (Zhuang *et al.*, 2002) based on the ETAS model (Ogata, 1988, 1998). We explored three aspects of the algorithms' behaviors: 1) the removal of events with stronger shaking intensities than the events retained, 2) their performances on declustering induced, natural, and simulated Poissonian catalogs, 3) their impacts on the hazard analysis for Oklahoma City.

We evaluated how often the Gardner and Knopoff, Reasenberg and Zaliapin and Ben-Zion algorithms took out dependent events with ground motions stronger than their mainshocks. We compared the $Sa(0.1s)$ from ground motions of mainshocks and their corresponding dependent events from the Oklahoma-Kansas catalog. The Gardner and Knopoff algorithm often identified events with stronger shaking as aftershocks — 84% of mainshocks had $Sa_{dependent\,event}/Sa_{mainshock} > 1$ and 39% had $Sa_{dependent\,event}/Sa_{mainshock} > 4$. The other two algorithms also removed dependent events with stronger shaking, but less often — only around 10% of the mainshocks had $Sa_{dependent\,event}/Sa_{mainshock} > 4$. Large $Sa_{dependent\,event}/Sa_{mainshock}$ values indicate that the algorithm identifies aftershocks that are distant from mainshocks but have similar magnitudes, which suggests the removal of events that potentially contribute to hazard level. The large ratios produced by the Gardner and Knopoff algorithm were due to the large spatial window defined in this algorithm, which was developed based on California earthquakes. Preferred approaches to minimize the removal of stronger dependent events are using the Reasenberg and Zaliapin and Ben-Zion algorithms, or developing region-appropriate parameters for the Gardner and Knopoff algorithm.

We used the four methods to decluster California and Oklahoma-Kansas earthquake catalogs and studied their behaviors. For the California catalog, all algorithms worked well as they removed dependent events and resulted in a stable estimation of background rates. Moreover, they all generated $\hat{b}$ close to 1.0. For Oklahoma and Kansas, the effect of different declustering algorithms on the declustered catalog was significant. The Gardner and Knopoff algorithm removed many more events and resulted in almost constant rates along the time. It also generated the smallest $\hat{b} = 0.79$. Other three algorithms removed some events and were able to preserve the changing background rate. However, the ETAS method generated the largest $\hat{b} = 1.40$, which suggested that it potentially overestimated the foreshock effect and produced small probabilities of large earthquakes ($M > 4.5$) being mainshocks. This can result in an underestimation of the hazard level. These observations address the importance of selecting a suitable declustering algorithm for induced seismicity.

We then tested algorithms on simulated Poissonian catalogs where earthquake occurrences were independent of each other. The Reasenberg and ETAS methods had the desired behavior of removing negligible numbers of earthquakes from simulated catalogs. However, the Gardner and Knopoff method removed a considerable number of events, and the removal ratio increased with the mean earthquake occurrence rate. It exceeded 0.5 when the occurrence rate was approximately $5 \times 10^{-3} count/year/km^2$. Both values were close to the declustered 2014-2018 California catalog. These observations suggest that the Gardner and Knopoff method identifies independent events as aftershocks or foreshocks when the background earthquake rate is high. The Zaliapin and Ben-Zion method removed almost constant numbers of earthquakes (32%) from simulated catalogs with different λ. This is because, for a Poisson process, the distribution of nearest-neighbor distances with different λ is unimodal but not Gaussian.

Finally, we computed the 2016 and 2017 one-year hazard curves for Oklahoma City using the four declustering algorithms. The Gardner and Knopoff algorithm, with $b = 1.0$, generated the lowest hazard level in 2016, which was almost one-fourth of that from the Reasenberg and Zaliapin and Ben-Zion algorithms. This was because the Gardner and Knopoff method removed more earthquakes

(80%) compared to the latter two methods (30%). For the Gardner and Knopoff method, there was a slight increase in the hazard level in 2017, though the number of recorded earthquakes decreased by half. This was due to the use of the window method and the sparser earthquake distribution in 2017, which resulted in an increased number of mainshocks identified, thus an increased hazard level. These observations suggest that the Gardner and Knopoff algorithm fails to capture the changing seismic rate and the hazard levels in different years. The use of this algorithm partly explains the increased hazard in the 2018 USGS seismic hazard forecast despite a reduced number of earthquakes the prior year.

Data and Resources

The earthquake occurrences in California and the Oklahoma and Kansas region were obtained from the USGS earthquake catalog website (https://earthquake.usgs.gov/earthquakes/search, last accessed March 2019). Ground motion records were collected from Incorporated Research Institutions for Seismology (IRIS) Data Services (http://ds.iris.edu/ds/nodes/dmc/, last accessed April 2019) using the program Standing Order of Data (SOD, http://www.seis.sc.edu/sod/, last accessed April 2019).

Acknowledgement

We thank the anonymous reviewers for insightful comments that significantly improved the manuscript. Funding for this work came from the Stanford Center for Induced and Triggered Seismicity.

References

Atkinson, G. M. (2015). Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards, *Bulletin of the Seismological Society of America* **105**(2A), 981–992.

Boyd, O. S. (2012). Including foreshocks and aftershocks in time-independent probabilistic seismic-hazard analyses, *Bulletin of the Seismological Society of America* **102**(3), 909–917.

Chioccarelli, E., P. Cito, and I. Iervolino (2018). Disaggregation of sequence-based seismic hazard, in *Proceedings of the 16th European conference on earthquake engineering (16ECEE), Thessaloniki, Greece*.

Console, R., D. Jackson, and Y. Kagan (2010). Using the etas model for catalog declustering and seismic background assessment, *Pure and applied geophysics* **167**(6-7), 819–830.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bulletin of the seismological society of America* **58**(5), 1583–1606.

Cornell, C. A., and E. H. Vanmarcke (1969). The major influences on seismic risk, in *The fourth world conference on earthquake engineering*, volume 1, pp. 69–83.

Darold, A. P., A. A. Holland, J. K. Morris, and A. R. Gibson (2015). Oklahoma earthquake summary report 2014, *Okla. Geol. Surv. Open-File Rept. OF1-2015* pp. 1–46.

Davis, S. D., and C. Frohlich (1991). Single-link cluster analysis of earthquake aftershocks: Decay laws and regional variations, *Journal of Geophysical Research: Solid Earth* **96**(B4), 6335–6350.

Frohlich, C., and S. D. Davis (1990). Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues, *Geophysical Journal International* **100**(1), 19–32.

Gardner, J., and L. Knopoff (1974). Is the sequence of earthquakes in southern california, with aftershocks removed, poissonian?, *Bulletin of the Seismological Society of America* **64**(5), 1363–1367.

Kagan, Y., and L. Knopoff (1978). Statistical study of the occurrence of shallow earthquakes, *Geophysical Journal International* **55**(1), 67–86.

Kagan, Y., and L. Knopoff (1980). Spatial distribution of earthquakes: the two-point correlation function, *Geophysical Journal International* **62**(2), 303–320.

Kattamanchi, S., R. K. Tiwari, and D. S. Ramesh (2017). Non-stationary etas to model earthquake occurrences affected by episodic aseismic transients, *Earth, Planets and Space* **69**(1), 157.

Kumazawa, T., Y. Ogata, *et al.* (2014). Nonstationary etas models for nonstandard earthquakes, *The Annals of Applied Statistics* **8**(3), 1825–1852.

Lombardi, A. M. (2017). Seda: A software package for the statistical earthquake data analysis, *Scientific reports* **7**, 44171.

Luen, B., and P. B. Stark (2012). Poisson tests of declustered catalogues, *Geophysical journal international* **189**(1), 691–700.

Marzocchi, W., and A. M. Lombardi (2008). A double branching model for earthquake occurrence, *Journal of Geophysical Research: Solid Earth* **113**(B8).

Marzocchi, W., and M. Taroni (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis, *Bulletin of the Seismological Society of America* **104**(4), 1838–1845.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical association* **83**(401), 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics* **50**(2), 379–402.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016). Seismic-hazard forecast for 2016 including induced and natural earthquakes in the central and eastern united states, *Seismological Research Letters* **87**(6), 1327–1341.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, J. L. Rubinstein, A. L. Llenos, A. J. Michael, W. L. Ellsworth, A. F. McGarr, A. A. Holland, *et al.* (2015). Incorporating induced seismicity in the 2014 united states national seismic hazard model: Results of 2014 workshop and sensitivity studies .

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, K. S. Rukstales, D. E. McNamara, R. A. Williams, A. M. Shumway, P. M. Powers, P. S. Earle, *et al.* (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes, *Seismological Research Letters* **89**(3), 1049–1061.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. M. Shumway, D. E. McNamara, R. A. Williams, A. L. Llenos, W. L. Ellsworth, A. J. Michael, *et al.* (2017). 2017 one-year seismic-hazard forecast for the central and eastern united states from induced and natural earthquakes, *Seismological Research Letters* **88**(3), 772–783.

Reasenberg, P. (1985). Second-order moment of central california seismicity, 1969–1982, *Journal of Geophysical Research: Solid Earth* **90**(B7), 5479–5495.

Utsu, T. (1965). A method for determining the value of" b" in a formula log n= a-bm showing the magnitude-frequency relation for earthquakes, *Geophys. Bull. Hokkaido Univ.* **13**, 99–103.

van Stiphout, T., J. Zhuang, and D. Marsan (2012). Seismicity declustering, *Community Online Resource for Statistical Seismicity Analysis* **10**, 1–25.

Vasylkivska, V. S., and N. J. Huerta (2017). Spatiotemporal distribution of oklahoma earthquakes: Exploring relationships using a nearest-neighbor approach, *Journal of Geophysical Research: Solid Earth* **122**(7), 5395–5416.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan, *Bulletin of the Seismological Society of America* **90**(4), 859–869.

Wooddell, K. E., and N. A. Abrahamson (2014). Classification of main shocks and aftershocks in the nga-west2 database, *Earthquake Spectra* **30**(3), 1257–1267.

Zaliapin, I., and Y. Ben-Zion (2013). Earthquake clusters in southern california i: Identification and stability, *Journal of Geophysical Research: Solid Earth* **118**(6), 2847–2864.

Zaliapin, I., and Y. Ben-Zion (2016). A global classification and characterization of earthquake clusters, *Geophysical Journal International* **207**(1), 608–634.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association* **97**(458), 369–380.