Electronic Supplement: Short-term probabilistic hazard assessment in regions of induced seismicity

Ganyu Teng, Jack W. Baker

The electronic supplement contains 1) the detailed estimation of δ_{HF} and a using kriging; 2) a prospective test on the predicted number of earthquakes induced during the injection interval (Equation 1); 3) the deaggregation plot for the IPE from Atkinson *et al.* (2014) (Figure S3); 4) the detailed estimation of $\delta_{A|M_m}$; and 5) the short-term hazard analysis for Oklahoma-Kansas and California based using parameters from Page *et al.* (2016).

Estimation of δ_{HF} and a for hydraulic-fracturing-induced earthquakes

Kriging is used to characterize the spatial distribution of dependent variables based on observed data and their relative locations in the region. The spatial distributions of δ_{HF} and a were estimated using kriging, where the region was divided into 0.01×0.01 grids. In particular, δ_{HF} was estimated using indicator kriging. The dependent binary variable is whether the well is seismogenic (labeled as 1) or not (labeled as 0), which is obtained from Spatiotemporal Association Filter. Ordinary kriging was used to estimate a based on $log_{10}(N)$, where N is the observed number of earthquakes induced with $M \geq 1.5$. 80% of the data was used as the training set and 20% as the test set. Figure S1 shows the semivariograms of δ_{HF} and a, with a cutoff distance of 0.5. They were fitted using exponential models. For δ_{HF}, our model resulted in a test accuracy of 81%. For a, the test root-mean-squared-error was 0.2.

Prospective test

After building the model to predict the number of earthquakes induced during the injection interval (Equation 1), we tested its performance using earthquake and injection data from 2019/01/01 to 2019/07/01. Figures S2a and S2b show the prediction plotted against observation for 2018 and 2019, respectively, where each point is the number of earthquakes induced. For both years, most of the data are within the 95% confidence interval.

Deaggregation

We constructed the deaggregation plot for the IPE from Atkinson *et al.* (2014). We considered magnitudes between 1.5 and 6.0 and a hypocentral distance of 7.2 km. The deaggregation was computed using the following equation:

$$P(m < M|MMI > mmi) = \frac{P(m < M, MMI > mmi)}{P(MMI > mmi)} = \frac{\sum_{m<M} P(MMI > mmi|m)P(m)}{\sum_m P(MMI > mmi|m)P(m)} \quad (1)$$

Where $P(MMI > mmi|m)$ was computed from the IPE and $P(m)$ was estimated from the Gutenberg-Richter distribution with $b = 1$. Figure S3 suggests that earthquakes with $M < 3$ contribute more than 40% to the hazard level of $MMI > 3$.

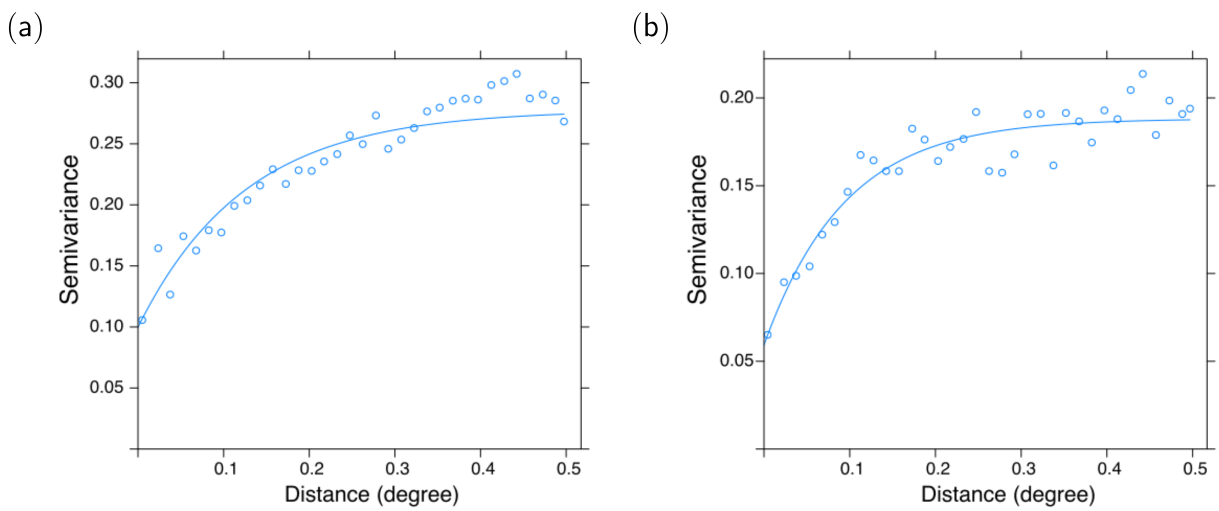

Figure S1: Semivariograms for (a) δ_{HF} and (b) a.

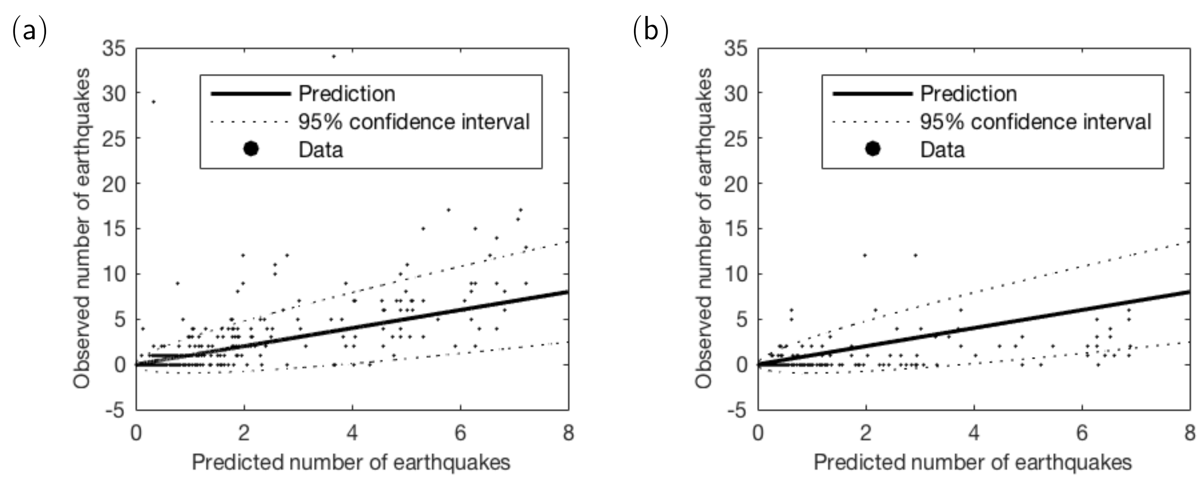

Figure S2: Predicted and observed number of induced earthquakes during the injection interval using (a) 2018 catalog and (b) 2019 catalog.

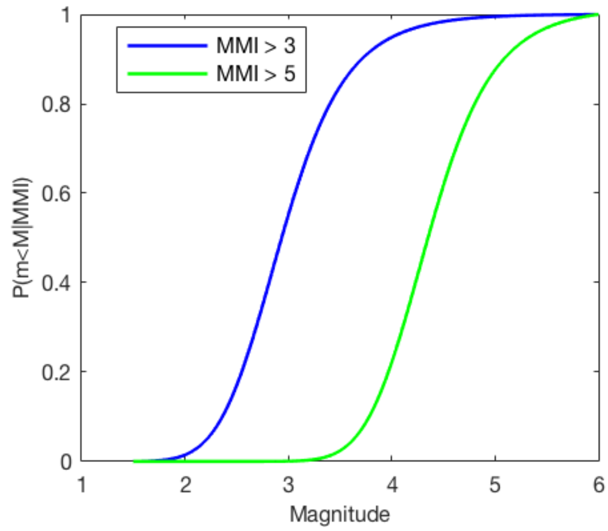

Figure S3: Deaggregation plot for a hypocentral distance of $7.2km$.

	California	Oklahoma-Kansas
a	-2.42	-2.85
p	0.98	0.73
c	0.018	0.018
b	1.00	1.00

Table S1: Aftershock parameters in Page *et al.* (2016)

Estimation of $\delta_{A|M_m}$

$\delta_{A|M_m}$ was estimated using logistic regression as a function of the mainshock magnitude (M_m). In particular, the catalog was first declustered using Reasenberg (1985). Mainshocks were then divided into two groups: mainshocks with dependent events (labeled as 1) or not (labeled as 0). We then performed logistic regression to predict the probability that a mainshock with magnitude Mm triggers aftershocks. The p-values for Oklahoma-Kansas and California catalogs were less than 2×10^{-16}. This method could be improved by considering secondary aftershocks

Hazard analysis for Oklahoma-Kansas and California based on Page *et al.* (2016)

We repeated the hazard analysis for Oklahoma-Kansas and California using parameters developed by Page *et al.* (2016). The parameters are summarized in Table S1. Since their estimation considered sequences with no aftershocks, we set $\delta_{A|M_m}$ in Equation 3 to 1.0. The results are summarized in Figures S4, S5 and S6. The absolute hazard level using the two sets of parameters differs significantly. This is due to the large difference between the two a-values used ($a = -1.62$ and $a = -2.85$). However, both models suggest that for Stillwater, the Poissonian mainshock rate has a more significant contribution to the short-term hazard level compared to natural earthquakes in San Francisco (Figure 14 and S6). Thus we could also conclude that in Oklahoma-Kansas, the mainshock rate could be important to short-term hazard levels.

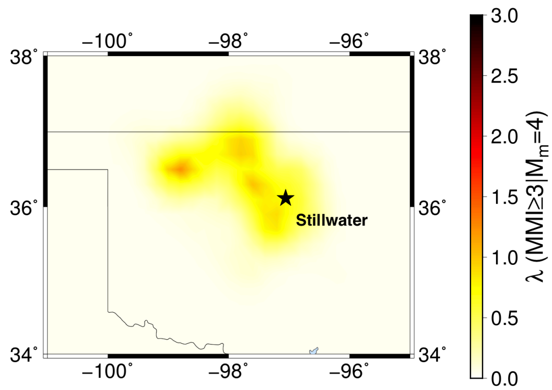

Figure S4: Rate of exceeding $MMI = 3$ over seven days after an $M_m = 4$ mainshock using parameters from Page *et al.* (2016).

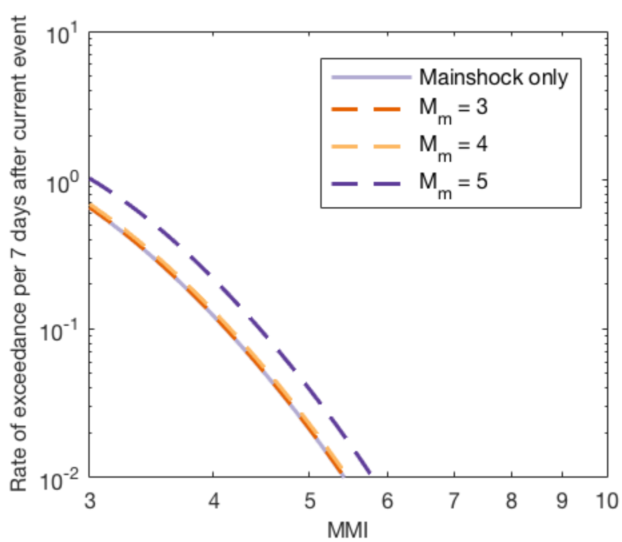

Figure S5: Seven-day hazard curves for Stillwater, conditional on three mainshock magnitude (M_m) values using parameters from Page *et al.* (2016). The hazard is the rate of exceeding an MMI over seven days after an M_m mainshock.

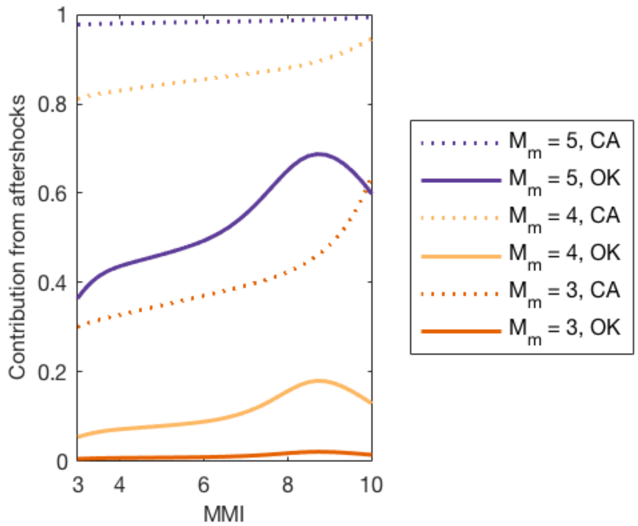

Figure S6: Aftershock contribution to short-term hazard levels for Stillwater in Oklahoma (OK) and San Francisco in California (CA) using parameters from Page *et al.* (2016).

References

Atkinson, G. M., C. B. Worden, and D. J. Wald (2014). Intensity prediction equations for north america, *Bulletin of the Seismological Society of America* **104**(6), 3084–3093.

Page, M. T., N. Van Der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability, *Bulletin of the Seismological Society of America* **106**(5), 2290–2301.

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