Assessments of the Performance of the 2017 One-Year Seismic-Hazard Forecast for the Central and Eastern United States via Simulated Earthquake Shaking Data

by Edward M. Brooks, James Neely, Seth Stein, Bruce D. Spencer, and Leah Salditch

ABSTRACT

As a result of wastewater injection from nonconventional oil and gas production, the central and eastern United States experienced a dramatic increase in seismicity. To better characterize the resulting hazard, the U.S. Geological Survey (USGS) began producing one-year seismic-hazard models intended to capture both natural and induced seismicity as of 2016. In its first year, we found that the map performed very well, demonstrating both a good match between the observed and expected number of exceedances, and between observed and predicted shaking. We repeat this analysis for the 2017 map, using "Did You Feel It?" data to explore the map's performance in different regions of the country. We find that the 2017 model performed well, but not as well as the previous year's model. We explore the likelihood of observing the performance seen in 2017, by simulating earthquake shaking realizations using the assumptions of the 2017 hazard model, including *a*- and *b*-values, locations of induced earthquakes, and ground-motion models (GMMs). These simulations indicate a low likelihood of this decrease in performance happening by chance if the assumptions in the hazard model were appropriate. Hence, it is likely that the USGS one-year seismic model's performance reflects a reduction in wastewater injection rates, possibly due to regulatory and economic pressures. Future models could benefit from better modeling how seismic rates may change year-to-year with variations in wastewater injection rates and locations, and improved GMMs.

INTRODUCTION

Increases in nonconventional oil and gas production in the central and eastern United States (CEUS) since 2008 resulted in significantly increased seismicity, most notably in Oklahoma and the surrounding regions (Horton, 2012; Ellsworth, 2013; Keranen *et al.*, 2013, 2014). This region historically has not experienced high shaking and is generally unprepared for the increased seismicity (Liu *et al.*, 2014; Ellsworth *et al.*, 2015).

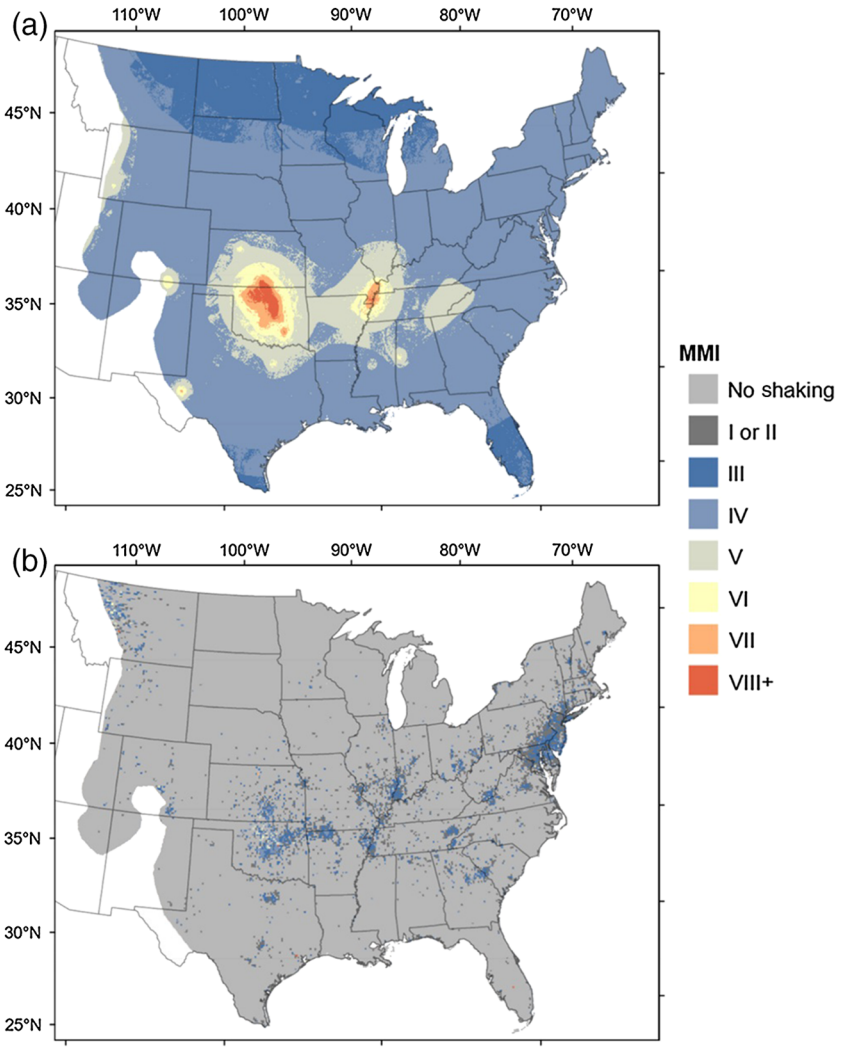
The increased seismicity necessitated reassessment of the resulting hazard. For this purpose, the U.S. Geological Survey (USGS) produced a series of hazard maps intended to be used for one year that focus on the hazard that results from human activity, namely wastewater injection (Petersen *et al.*, 2016, 2017; Petersen, Mueller, Moschetti, Hoover, Rukstales, McNamara, and Llenos, 2018). Developing models for one year of usage, versus a longer window such as 50 yr as in other maps (Petersen *et al.*, 2015), allows responses to the changes that may happen in human activities, a nonsteady state variable.

The USGS's first one-year map (Petersen *et al.*, 2016) accounted for the induced seismicity by defining zones where earthquakes do not appear natural, indicated by a noticeable increase in seismicity near injection wells, both spatially and temporally. They defined separate logic trees for seismicity inside and outside these zones that differ largely in the parameters used to describe catalog duration, smoothing distance, maximum magnitude, and ground-motion models (GMMs). Seismicity rates are inferred from injection rates from the prior 2 yr.

The one-year model has an advantage for assessing the resulting performance of the map; the time necessary to gather data is not so long that one must resort to historic data instead. Although hindcasting, using historic data to assess hazard maps for which catalogs of subsequent shaking do not exist, is useful (Stein *et al.*, 2015; Brooks *et al.*, 2016), gathering data generated entirely after the map was made is preferable for assessing its performance because it is a true test of a model's forecast.

We found that the 2016 model performed better than previous maps studied (Stein *et al.*, 2015; Brooks *et al.*, 2016). Using a similar approach as in this study, Brooks *et al.* (2017) used "Did You Feel It?" (DYFI) data to compare seismic intensity observations to the model's predictions. Both within the entire CEUS, and in the area surrounding Oklahoma where induced seismicity is most prevalent, the data were in good accord with the model's predictions. We thus concluded that the 2016 model was a very good model. Other studies looked at the 2016 model and found general agreement between

doi: 10.1785/0220190007



▲ **Figure 1.** (a) 2017 one-year seismic hazard forecast for the central and eastern United States (CEUS; Petersen *et al.*, 2017). Shaking levels are communicated in modified Mercalli intensity (MMI) units. (b) Maximum "Did You Feel It?" (DYFI) responses in 2017 for the CEUS. Gray regions indicate an absence of DYFI responses, but do not necessarily imply a lack of shaking. The color version of this figure is available only in the electronic edition.

observation and prediction using DYFI and instrumental data (Petersen *et al.*, 2017; Mostafa Mousavi and Beroza, 2018; White *et al.*, 2018).

For the following year, another model for 2017 was developed (Fig. 1a). The 2017 model employs the same logic trees and GMMs as the 2016 model, but uses an updated earthquake catalog for the additional year of seismicity observed (Petersen *et al.*, 2017). A second year of seismic intensity records from DYFI allowed assessment of the 2017 model's performance using the metrics employed in Stein *et al.* (2015) and Brooks *et al.* (2017).

The first metric, the fractional exceedance metric $M0$, is

$$M0(f,p) = |f - p|, \tag{1}$$

in which p is the predicted fraction of sites for which the highest shaking is expected to exceed the model's predictions, and f is the observed fraction of sites for which this actually occurs. The probability p is derived from the fact that probabilistic hazard models seek to predict a level of shaking that should be exceeded only with a certain probability in some time window (Cornell, 1968; Field, 2010). At any point on the map, the probability of exceedance is given by an exponential distribution: $p = 1 - \exp(-\frac{\tau}{T})$. For one year of observation, $\tau = 1$, and the model assumes a return period of $T = 100$ yr. Hence, the probability of exceedance p for the model is roughly $p = 0.01$.

The fractional exceedance metric is implicit in probabilistic seismic-hazard analysis (PSHA). This metric is binary, and only considers whether an observation is over or under the map's prediction. Thus, we also use a squared misfit metric $M1$:

$$M1(s,x) = \frac{1}{N}\sum(x_i - s_i)^2, \tag{2}$$

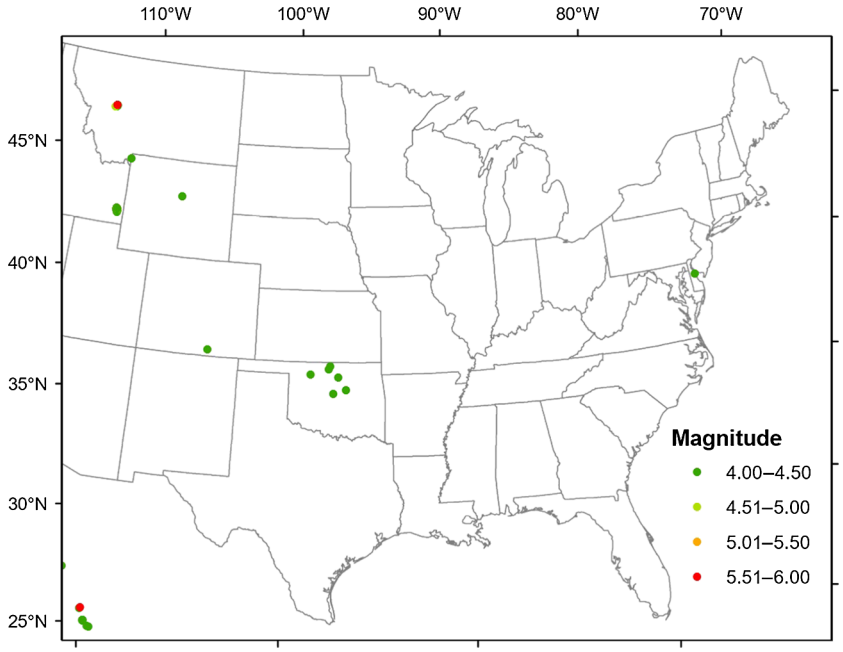
in which x_i and s_i are the maximum observed shaking and predicted shaking at each of the sites $i = 1, 2, ..., N$. Although not the goal of PSHA, this metric captures other important aspects of map performance, notably the spatial match between prediction and observation. Because a map can be successful by $M0$, but less useful overall as a map (Stein *et al.*, 2015; Brooks *et al.*, 2016), we consider both metrics to get a clearer understanding of map behavior. For both metrics, a perfect match between prediction and observation will yield a score of 0; hence, higher scores reflect relatively weaker performance. Uncertainty in the predictions should be reflected in $M0$ and $M1$, so that when uncertainty increases so do the values of $M0$ and $M1$, all other things being equal. Error (or uncertainty) in the observations (e.g., DYFI) would tend to inflate the values of $M0$ and $M1$ relative to error-free observations, and thus would lead to a slight inflation in $M0$ and $M1$.

SEISMICITY AND FELT GROUND SHAKING IN 2017

The 2017 one-year seismic-hazard model emphasizes the hazard most strongly in Oklahoma and surrounding states, mainly Texas and Kansas. This emphasis was also present in 2016. To assess the model's performance, we use the shaking record from the DYFI database (Wald *et al.*, 1999; Atkinson and Wald, 2007). A number of studies compared DYFI data to predicted seismic hazard in recent years (Mak and Schorlemmer, 2016a;

Cremen *et al.*, 2017; White *et al.*, 2018), and noted DYFI's utility for broad areal coverage (Atkinson and Wald, 2007; Hough, 2012; Mak and Schorlemmer, 2016b). Reports for individual events are geocoded by zip code, and annual summaries of maximum observed intensities are compiled on a 10 km grid (Quitoriano *et al.*, 2017). Though there can be issues in quality and completeness in poorly populated areas, we found that the DYFI database, specifically the annual maximum data, is one of the most thorough and robust seismic intensity datasets available, providing the most observations over the largest area (Wald *et al.*, 2012).

Figure 1b shows the maximum shaking reported to DYFI for 2017. The map shows 17,391 sites on the one-year map where at least one report was made. Though the data are sparse, there is a match, broadly speaking, between the expected shaking and the intensity in the reports made, including the highest shaking congregated within Oklahoma. The map also features many reports to the east in the Pennsylvania-Delaware-Jersey tristate area, and to the west in Montana. These reports are geographically consistent with the location of the largest earthquakes observed in the CEUS in 2017 (Fig. 2).

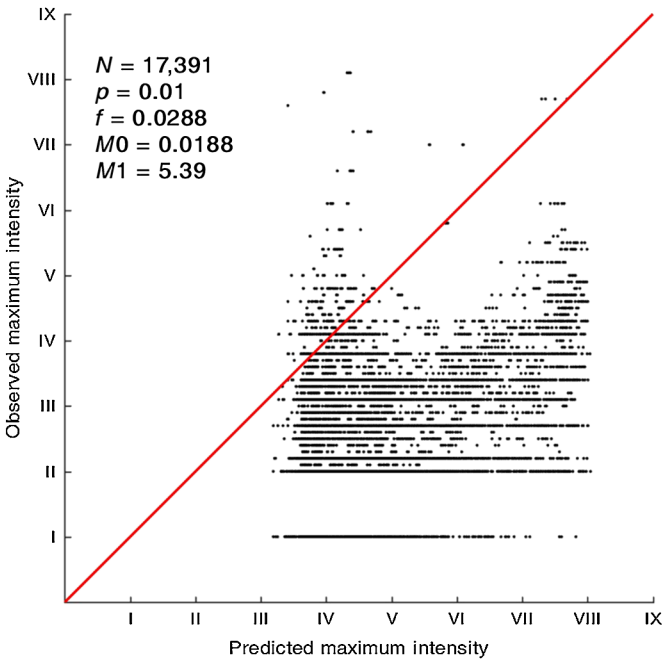


▲ **Figure 2.** Occurrence of **M** 4+ earthquakes near the CEUS in 2017. Earthquakes in Oklahoma, Montana, and Delaware were the primary generators of DYFI responses. The color version of this figure is available only in the electronic edition.

Figure 2 shows that despite the expected high seismicity in the greater Oklahoma area, the largest earthquake in the CEUS in 2017 was an **M** 5.8 event near Lincoln, Montana (McMahon *et al.*, 2017). Similarly, the east coast experienced an **M** 4.1 earthquake in December 2017, located in Dover, Delaware, where seismicity is expected to be low. Conversely, Oklahoma experienced only a few seismic events with magnitude greater than 4, in contrast to the previous year, when it experienced a number of high-shaking events, including the largest recorded in its history, the **M** 5.8 Pawnee event (Yeck *et al.*, 2017). Because the distribution of these 2017 events differs from the previous year's, in which numerous **M** 4+ events occurred in Oklahoma and smaller events were located elsewhere, they provide an opportunity to see how well a map performs when a number of "black swans," rare and unexpected scenarios, occur (Stein *et al.*, 2012).

DYFI AND MAP PERFORMANCE

Figure 3 illustrates the map's overall performance. From the $N = 17{,}391$ responses, we see that 501 sites reported shaking exceeding that predicted. This corresponds to $f = 0.0288$; hence, because $p = 0.01$ the map experienced roughly three times the number of exceedances expected, and thus $M0 = 0.0188$. Furthermore, $M1 = 5.39$, a relatively low score reflecting a reasonable spatial correlation between the maximum shaking at sites that responded to DYFI and the predicted maximum shaking there. Brooks *et al.* (2017) found that for the 2016 model, the CEUS had $M0 = 0.0073$ and $M1 = 4.62$. Although the previous year's lower scores indicate a slightly better performance overall, in general they are similar. Although this is the first study focusing on successive iterations of maps, our prior studies have found maps that yield both $M0$ and $M1$ orders of magnitude higher (Stein *et al.*, 2015).



▲ **Figure 3.** Comparison of predictions from the 2017 hazard map to maximum reported intensity for the entire CEUS. The color version of this figure is available only in the electronic edition.

For areas where predicted intensity is high (e.g., intensity VII+), the range of observed shaking spans intensity II–VIII. A similar trend is present for low-intensity predictions (i.e., intensity III–V). However, between these two regions (e.g., intensity V–VII), observed shaking levels are predominantly lower. As a result, most exceedances come from sites where shaking was predicted to be relatively low. Most of those exceedances cluster around the red 45° line that marks a perfect match between observation and prediction, suggesting that most exceedances are small. Roughly two dozen points are exceedances in which intensity is VII+. Further insight into spatial variations in map performance as a function of the DYFI data can be had by subdividing Figure 1 into smaller geographic regions.

Figure 4a–c shows the predicted and observed shaking, and metric score for the greater Oklahoma area, the region where induced seismicity in the United States is the highest. The greater Oklahoma region contains the majority of locations where observed shaking is well below the model predictions. There are only eight exceedances from 3410 observations that leads to an observed exceedance fraction $f = 0.0023$ and an $M0 = 0.0077$, similar to the 2016 model score (in 2016, $M0_{\mathrm{CEUS}} = 0.0073$, and $M0_{\mathrm{OK}} = 0.0069$). The 2016 model underpredicted shaking and the Oklahoma region experienced more exceedances than expected. In 2017, the opposite occurred and shaking was overpredicted. With so few exceedances in the area, the significance of a lower $M0$ score is lessened (Brooks *et al.*, 2017). In contrast to the similarities in $M0$ between 2016 and 2017, 2017's greater Oklahoma $M1 = 10.46$, substantially higher than for the entire CEUS in 2017 and the CEUS and greater Oklahoma area $M1$ scores for 2016 (in 2016, $M1_{\mathrm{CEUS}} = 4.62$ and $M1_{\mathrm{OK}} = 5.01$).

The effects of the low shaking reported in Oklahoma can be highlighted by examining the CEUS metrics excluding the greater Oklahoma area. Figure 4j shows observed versus predicted shaking for the opposite of the top row, that is, the entire CEUS without the box in the top of Figure 4. The map lacks the exceedances that occurred at higher predicted maximum intensities, because the areas with the highest predicted shaking due to induced seismicity are removed from the map. Hence, there are far more exceedances from areas with lower predicted intensities, largely the Montana and Delaware earthquakes. Oklahoma shaking was heavily overpredicted, but the map as a whole underpredicted shaking, so removing Oklahoma from the dataset yields a larger fraction of site exceedances. Hence, $M0$ increases to 0.0253, a decrease in map performance. However, the large mismatch in the intensity of the shaking predicted in Oklahoma is reflected by the squared misfit $M1$, double the CEUS $M1$ score. Thus, removing the greater Oklahoma area from the data improves the map performance by the squared misfit metric, so $M1 = 4.16$.

Additional subdivisions illustrate other strengths and weaknesses of map performance. Consider responses to the Montana earthquake in the northwest, shown in Figure 4d–f. Because this was an unexpectedly large earthquake for the area, it led to a high number of exceedances. Over 20% of sites reported shaking exceeding the predicted levels, so $M0 = 0.2063$. However, this number may be artificially inflated by a lack of distant responses in areas where shaking may not have exceeded predictions, perhaps due to low population. The few reports from distant areas that do describe shaking are close to the predicted values. As a result of this strong match, this region of the map scores $M1 = 1.98$. This is notably lower than any other $M1$ scores generated for 2017, indicating that the map generally succeeds at matching predictions to observations.

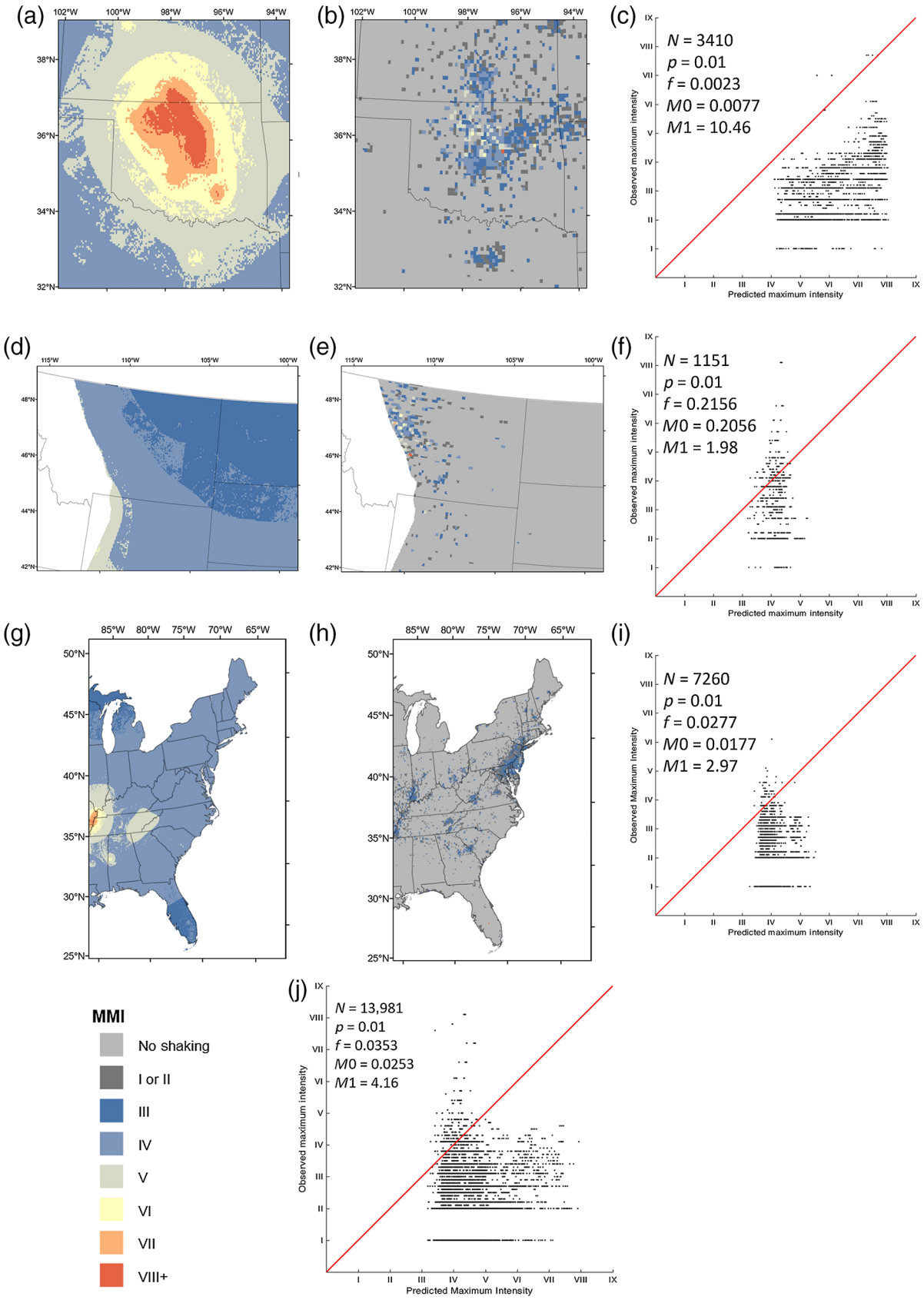
Finally, Figure 4g–i shows the eastern half of the CEUS map, where the seismicity is largely noninduced. The region tends to be very aseismic, though there is a history of events happening along the coast in the past (Hough, 2012; Wolin *et al.*, 2012). This portion of the map contains 40% of the data reported in the entire CEUS but has $M0 = 0.0177$, similar to that of the entire region. The data lack large exceedances, and although there are some instances of overpredicted shaking in Tennessee and the New Madrid region, broadly speaking there is a good visual match between predictions and observations. Hence, a low-squared misfit arises, with $M1 = 2.97$.

The metric calculations indicate that data fit the 2017 map's predictions reasonably well, although not as closely as the 2016 data fit that map. This map performs better than all previous studies we conducted of hazard maps, with one exception, the 2016 one-year map. This difference opens questions of how much variability to expect in performance from year to year, and whether the poorer performance in 2017 is likely to have arisen by chance or instead reflects a meaningful change in seismicity. Furthermore, the decrease in responses that may be a function of lower shaking, low population at the epicenters of events, or earthquake fatigue driving down response rates—or a combination of all three—raises questions regarding data quality and completeness (Mak and Schorlemmer, 2016b).

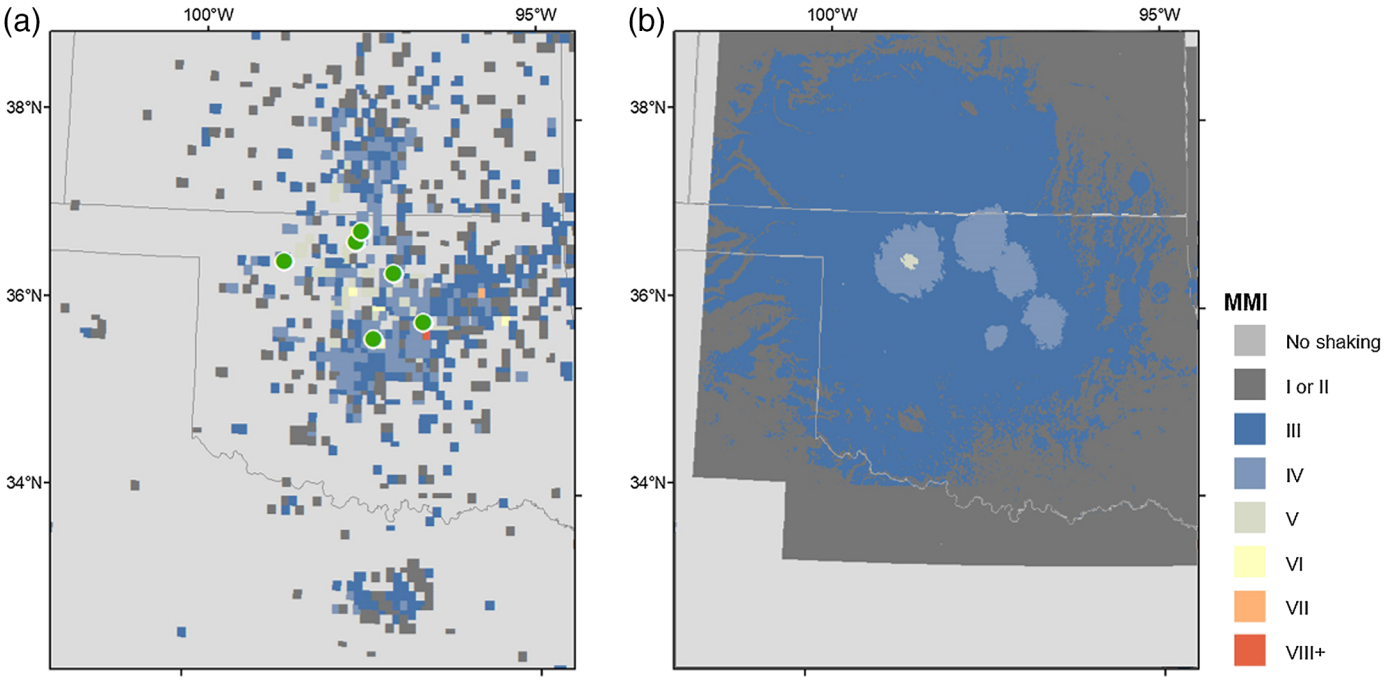
Focusing specifically on Oklahoma, between 2016 and 2017 the number of responses cataloged in the maximum shaking felt per year in the DYFI system dropped by nearly two-thirds. About 15% of the region had shaking reported in 2017, down from 40% in 2016. Figure 5 shows the absence of responses by comparing the DYFI map to what ShakeMap predicted the intensity of shaking would be (Wald *et al.*, 2005; Brooks *et al.*, 2017). In the absence of information, ShakeMap is a suitable approximation of a lower bound for reported shaking. Although there are some regions where DYFI shaking exceeds ShakeMap predictions, there are many more areas devoid of responses that ShakeMap suggests should have felt shaking, particularly toward the western part of Oklahoma. This is likely due to issues in population density compounding issues with decreased response rate. This decrease in reports is also accompanied by a decrease in the intensity of shaking reported. The median maximum annual shaking reported in 2017 is 2.9; down half an intensity unit from the median in 2016 of 3.4.

SHAKING SIMULATIONS

To address the questions of whether low responses are a function of low-shaking levels or low-response rates, we use Monte Carlo simulation to characterize the variability of possible



▲ **Figure 4.** (a,d,g) Predicted and (b,e,h) observed shaking maps, and (a–c) predicted–observed plots for the greater Oklahoma area, (d–f) Montana, and (g–i) the east coast of the United States. (j) Observed–predicted plot for the entire CEUS except the greater Oklahoma area. The color version of this figure is available only in the electronic edition.



▲ **Figure 5.** Comparison between shaking reported to (a) DYFI and (b) predictions from ShakeMap. **M** 4+ earthquakes in Oklahoma are marked with circles on (a). They are excluded on (b) to not obscure contour changes, but are located at the center of each local maximum in the ShakeMap predictions. The color version of this figure is available only in the electronic edition.

shaking histories for the greater Oklahoma area in 2017. Our approach is similar to that of Vanneste *et al.* (2018). The simulations can address the likelihood of the observed decrease in shaking occurring by chance, explore data incompleteness issues, and give insight into how the metrics used for assessment describe the map's performance. For simplicity, we used Oklahoma as the region of interest, and so did not need to account for multiple zones of induced seismicity across the CEUS, or the effect of natural seismicity in areas such as the New Madrid region.

We considered four random processes that define the maximum shaking experienced in a year:

1. where the earthquakes occur;
2. how many earthquakes occur;
3. the magnitude of the earthquakes; and
4. uncertainties in the GMMs that describe shaking.

The first three processes control the distribution of earthquakes in the region and can be described simply. Petersen *et al.* (2016) define regions where wastewater injection is expected to lead to induced seismicity. We consider this area to be uniformly susceptible to induced earthquakes.

In the models, the rate of induced seismicity is defined for an upcoming year as a weighted average of the past 2 yr, with weights of 0.8 and 0.2 for the most recent year and the previous year, respectively (Petersen *et al.*, 2014, 2017). For 2017, given a minimum magnitude of completeness of 2.7, we observe 162 independent (e.g., declustered, with no aftershocks) earthquakes in 2016 and 152 earthquakes in 2015 for the Oklahoma induced zone (Petersen, Mueller, Moschetti, Hoover, Rukstales, McNamara, *et al.*, 2018); hence, we set the expected number of earthquakes $\lambda = 160$ and model the number of possible earthquakes that can occur as a Poisson random variable with this mean.

Earthquake magnitudes are assigned based on the Gutenberg–Richter relationship, assuming $b = 1$, as the 2017 model does (Petersen *et al.*, 2017). To simulate artificial event magnitude using a given b-value, we follow a method given by Zhuang and Touati (2015) using inverse transform sampling. An event's magnitude m is randomly generated, assuming that

$$m = -\frac{1}{b\ln(10)}\ln U + m_0, \tag{3}$$

in which b is the model's b-value, U is a value obtained randomly from a uniform distribution on [0,1], and the minimum magnitude for completeness is $m_0 = 2.7$.

For a more accurate characterization of the 2017 earthquake record, this process of simulating earthquake occurrence is repeated a second time, using $\lambda = 4$, to include earthquakes that happened outside the defined box of induced seismicity.

By simulating these first three random processes, we generate many realizations of seismicity in Oklahoma in 2017. We describe the resulting ground shaking using the GMMs used by Petersen *et al.* (2017). Nine different models with varying weights (Petersen *et al.*, 2014) are used and then aggregated to describe ground shaking (Frankel *et al.*, 1996; Silva *et al.*, 2002; Toro, 2002; Campbell, 2003; Tavakoli and Pezeshk, 2005; Atkinson and Boore, 2006; Atkinson, 2008, 2015; Pezeshk *et al.*, 2011). Because the resulting shaking is given as peak ground acceleration (PGA) and the model we are assessing

predicts shaking as modified Mercalli intensity (MMI), we use techniques in Petersen *et al.* (2017) to convert PGA to MMI (Worden *et al.*, 2012). Each GMM, as well as the conversion from PGA to MMI, has an error term in the form of a Gaussian random variable. The error terms for ground motion are treated as uncorrelated at each site and can be treated as representing the uncertainty in each model, as well as the influence of site effects, directivity, or other modifiers to shaking. The PGA to MMI conversion error is correlated and assumed to be equal at all sites. Using the GMMs, the shaking from each earthquake in a given realization is calculated. After calculation for all events' shaking, the maximum observed shaking at each site, gridded on a 10 km scale, is selected and used to calculate the metrics for map performance.

We calculated 10,000 simulations to explore the full range of outcomes. We call this model, which allows for full variance of all parameters, unconstrained. Figure 6a shows nine of these realizations, and the metric calculations associated with each. Intensity is calculated at each site within the greater Oklahoma region. Hence, the metrics here show performance that would arise with a 100% response rate.

The metrics calculated for the 10,000 unconstrained simulations—in which uncertainties in earthquake count, location, magnitude, GMMs, and PGA to MMI conversion are allowed to vary—are shown in Figure 6b. The results show generally low $M0$ scores, reflecting a tendency for $f \approx p$. A tail drops off for larger $M0$ scores, showing the small possibility of achieving a very large score, indicative of scenarios in which $f \gg p$. Counts of $M0 > 0.02$ decrease substantially and are shown by an extra plot beneath the heatmap. Though the values reach as high as $M0 = 0.8$, most simulations have $M0 \leq 0.02$. About 80% of the simulations fall below this cutoff, 95% of the simulations result in metrics in which $M0 \leq 0.15$. Generally speaking, this distribution agrees with the 2017 DYFI result ($M0 = 0.0077$). The squared misfit metric $M1$ is roughly characterized by a normal curve centered around a mean $M1 = 2.09$, a very low score compared with the 2017 DYFI data ($M1 = 10.46$). The standard deviation for this curve is 0.41. There is an inverse correlation between the fraction of sites that exceed predicted shaking and $M1$. This can be observed by the two trend lines that grow out of the dense grouping of points centered around $(M0, M1) = (0, 2.09)$, which is where $f = p = 0.01$. The sharp upward trend terminates at $M0 = 0.01$, the point where no exceedances are observed ($f = 0$). The shallow downward trend continues for all $M0$ values in the heatmap and outlier plot. Hence, as f increases, the squared misfit decreases. Although only the fractional exceedance metric is implicit in the definition of this hazard map, this result suggests it should be possible to minimize both metrics for a given set of predictions and expected number of exceedances.

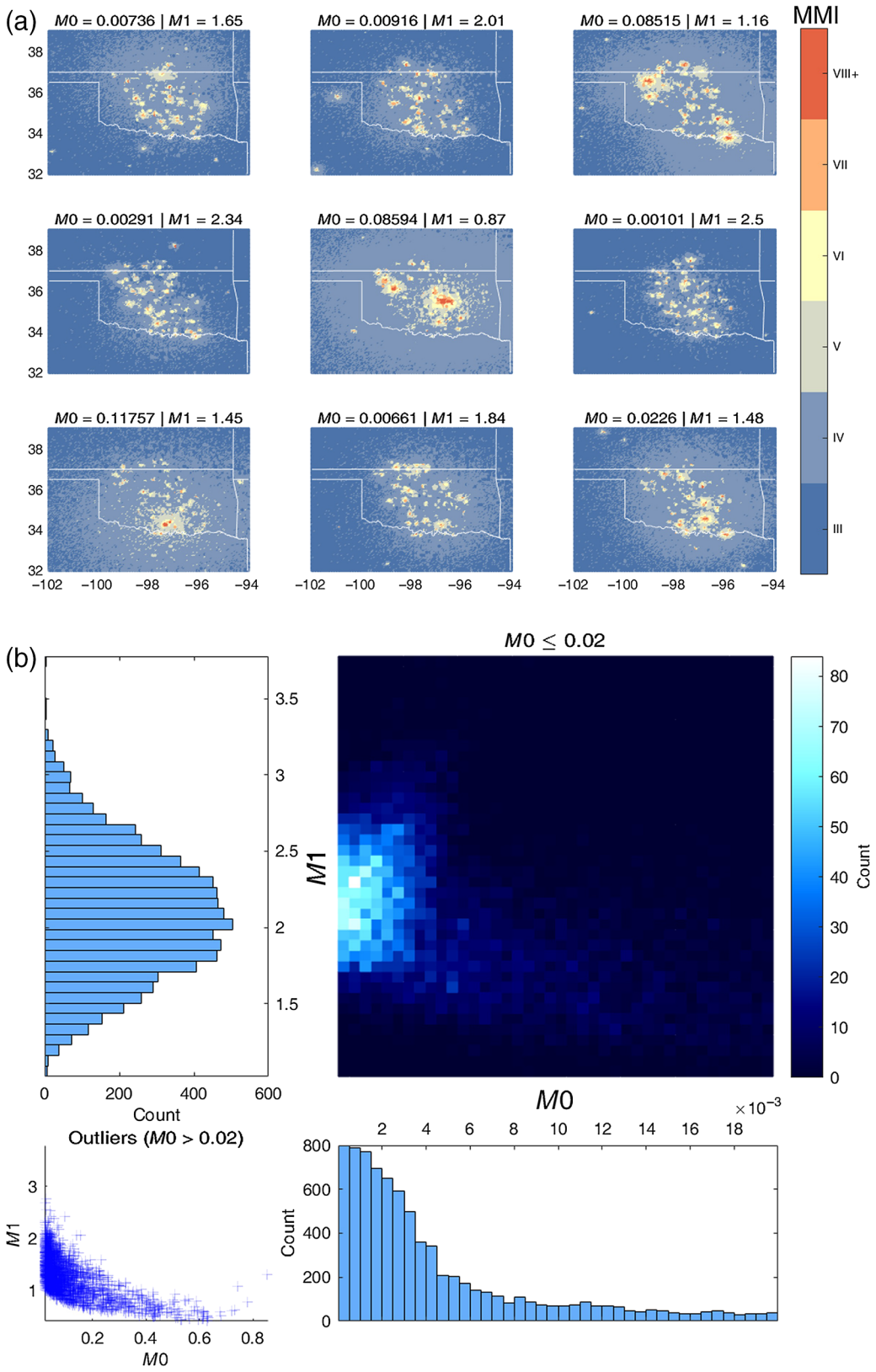
The metrics obtained with DYFI data for 2017 ($M0 = 0.0077$ and $M1 = 10.46$) do not fall within the simulation's distribution due to the high-squared misfit metric $M1$, though they are in reasonable accord with the distribution of the fractional exceedance metric $M0$. However, the discrepancy between the metrics for DYFI and the simulations may result from incompleteness in the DYFI data. Some of the largest earthquakes in Oklahoma in 2017 have few DYFI responses (Fig. 5), and some of the neighboring responses report MMI II intensities, which seem far too low for their proximity to some events' epicenters. It appears there is an issue with the response rate of the DYFI data, resulting in many inconsistent and missing responses.

It is not a fair comparison to compare the metrics for the DYFI data, given the low-response rate, to those for simulated data, in which data coverage is perfect. Brooks *et al.* (2017) found an inverse relationship between response rate and $M1$. Thus, although both DYFI and simulation can be used to assess performance individually, it is unclear whether a comparison between the two is useful for comparing aspects of map performance, or the relationship between $M1$ and the number of responses.

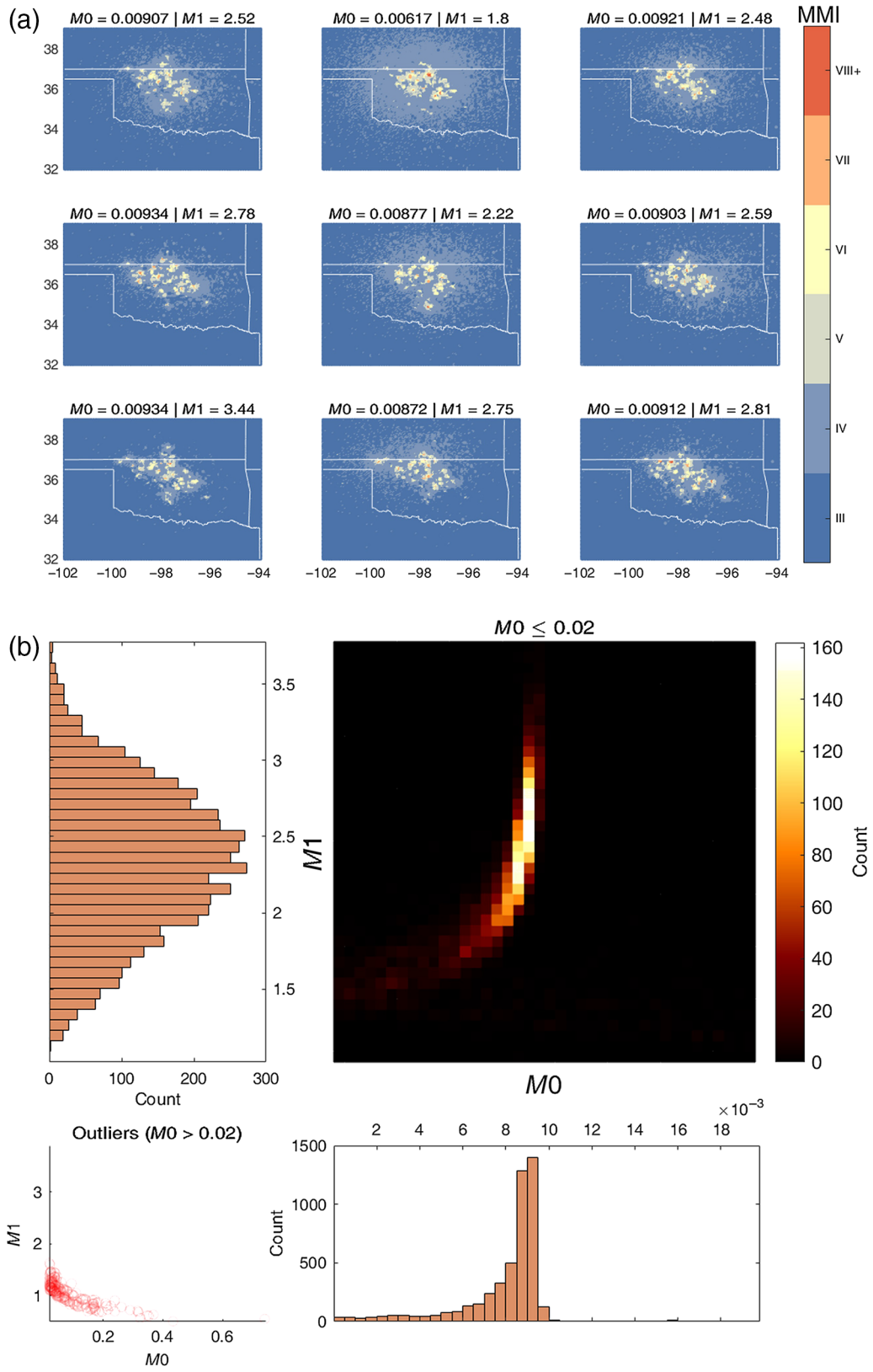
Hence, to better address data discrepancies and generate a more consistent comparison of map performance, we fixed earthquake count, location, and magnitude to the values for declustered events that occurred in 2017, and repeated the simulation procedure considering only the uncertainties in ground motion and PGA to MMI conversion. By repeating this analysis with these fixed parameters, we create a comparable dataset of "observations" to contextualize the results of the simulation. This approach is analogous to using ShakeMap data (Brooks *et al.*, 2017) and approximates having DYFI data with a 100% response rate for the earthquakes observed in 2017, but has the added benefit of also incorporating the uncertainty in the GMMs.

Figure 7a shows some of the 5000 shaking simulations calculated with the observed earthquake catalog in Oklahoma in 2017. We halve the number of simulations to reduce computation time, because we expect substantially smaller variance in the output when there are many fewer input variables. We term these simulations the constrained model, whereas we term the original simulations the unconstrained model. The variance in shaking output between simulations appears to be much smaller, because they simulate the same events so ground motion varies less between simulations. Figure 7b shows the variability in the metrics for the constrained simulations, which is due only to GMM uncertainty. $M0$ is tightly clustered between 0.008 and 0.01, differing strongly from the distribution of $M0$'s for the unconstrained model. About 97% of the simulations have $M0 \leq 0.02$. $M1$ in the constrained simulation has a slightly larger standard deviation than the unconstrained model, with a wider normal curve centered at the mean $M1 = 2.33$. The standard deviation of the $M1$ distribution is 0.48. The outliers $M0 > 0.02$ are far fewer in number, because the lower shaking produces fewer exceedances.

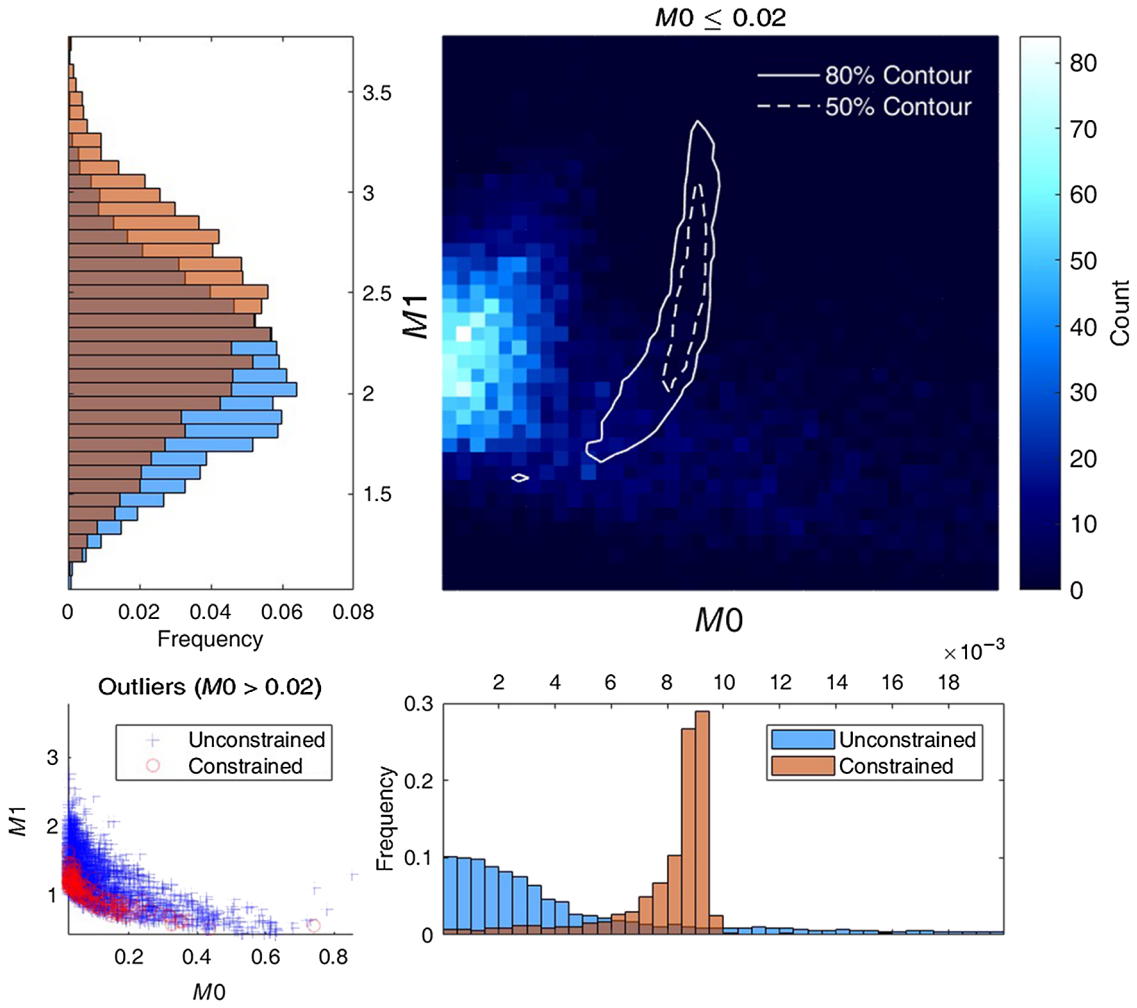
The results from both sets of simulations are superimposed in Figure 8. Counts are normalized to show relative frequency, illustrating the likelihood of getting a specific metric result from the unconstrained and constrained simulations, and comparing the distributions of the two. Through this comparison, we examine the likelihood that the higher misfit observed in 2017 could be attributed to either bad luck, or a flawed assumption in the underlying model (Stein *et al.*, 2011).



▲ **Figure 6.** (a) Simulation outputs, illustrating shaking and metric results, for nine different realizations of 2017. The shaking model is fully unconstrained, with randomness in earthquake count, location, magnitude, and in the ground-motion models (GMMs). Intensity is reported in MMI. (b) Heatmap of distribution of metric scores for 10,000 realizations of 2017 seismicity. The *x* axis is the fractional exceedance metric *M*0, and the *y* axis is the squared misfit metric *M*1. Each axis has a histogram of each metric's distribution, independent of the other. Outliers, defined by $M0 > 0.02$, are plotted in the bottom left. The color version of this figure is available only in the electronic edition.



▲ **Figure 7.** (a) Sample of simulation outputs, illustrating shaking and metric variability for the observed earthquake record in 2017. These constrained simulations have randomness only due to the GMMs. (b) Heatmap of distribution of metric scores from for 5000 realizations of 2017. The *x* axis is the fractional exceedance metric *M*0, and the *y* axis is the squared misfit metric *M*1. Each axis shows a histogram of each metric's distribution, independent of the other. Outliers, defined by $M0 > 0.02$, are plotted in the bottom left. The color version of this figure is available only in the electronic edition.



▲ **Figure 8.** Superposition of unconstrained (background heatmap) and constrained simulations (white contour lines). Contour lines represent where 50% and 80% of the data reside. Histograms show relative frequencies in the distribution of each metric. Outliers ($M0 > 0.02$) are plotted on the bottom left. The color version of this figure is available only in the electronic edition.

We would expect a model that can accurately forecast the shaking for a given year to have a metric score that falls within the range of metric scores achieved from a fully unconstrained simulation.

In essence, we would expect what we observe to be one of the scenarios predicted by the unconstrained model. However, the contours of the constrained model, representing 50% and 80% bounds on the data, are displaced upward for both metrics, giving higher scores, indicating weaker performance than what might be expected from a random scenario from the unconstrained simulation. Though it is possible, based on the distribution of the unconstrained simulation, to have an expected earthquake scenario result in poor map performance (as indicated by the simulations that yielded very high metric scores, indicative of even weaker performance), there is little overlap between the constrained (contours) and the unconstrained (heatmap) simulation results. Hence, from this comparison, we conclude that the poorer performance of the 2017 map arises not from bad luck, but a flawed assumption. Some physical aspect of shaking is not accurately described by the model for the map; otherwise, we would expect to see the constrained simulations overlap with the unconstrained output. Specifically, the tight clustering of the fractional exceedance metric around the predicted number of exceedances ($p = 0.01$) seems to suggest that the earthquakes that occurred in 2017 were insufficient to generate shaking large enough to cause exceedances. This lack of large shaking can also explain the upward shift in the squared misfit metric $M1$. There were too few large events in 2017. Larger events would generate lower scores for both $M0$ and $M1$. This conclusion is reinforced by the issues in the DYFI data, which similarly suffer due to the lack of large shaking events necessary to generate a broad response of quality reports.

Beyond assessing the performance of this specific model, these results have implications for the general issue of how to assess and improve earthquake-hazard maps' performance. The simulations approach is useful for filling in gaps in data and exploring the uncertainty in a map's predictions. Furthermore, the metrics defined in Stein *et al.* (2015) were intended to be

used as a comparative tool to assess the performance of multiple different maps. Through work on many maps, a general understanding of what constitutes high- and low-metric scores may arise, though it is harder to assess a map's performance with no basis for a comparison. Simulation is a tool to address this problem, allowing for comparison of many different shaking realizations. The simulations allow better understanding of map performance if there are no comparisons to be made, and for a better understanding of the likelihood of observing a single outcome (Vanneste *et al.*, 2018). This is important for better using metrics to assess map performance, by moving from assessing relative performance to assessing absolute performance. In other words, we would go beyond asking if a map is better than another map, to asking if a map is good. Such advances will help researchers better understand how and why maps perform the way they do and thus how to better use them for earthquake-hazard mitigation.

CONCLUSIONS

Comparison of shaking observed in 2017 to that predicted by the 2017 hazard model shows an overprediction of shaking. The shaking record for 2017 contained so little shaking that it generated essentially no exceedances. This is in stark contrast with 2016 that had many exceedances throughout the greater Oklahoma area, due to numerous large events, specifically three **M** 5+ episodes, including the **M** 5.8 Pawnee earthquake. These large events, and moderate-to-large shaking episodes in general, dominate the maximum shaking record.

The greatest mismatch between prediction and observation for the 2016 model was in northern Texas (Brooks *et al.*, 2017). Substantial shaking in Dallas did not occur as predicted and maximum DYFI reports in the area were linked to distant larger earthquakes in Oklahoma. Wastewater injection has been found to be linked strongly with Texan seismicity (Hornbach *et al.*, 2016), and though earthquakes may persist after wastewater injection has halted, rates decline following the closure of an injection site (Ogwari *et al.*, 2018). We believe the change in seismicity that follows a reduction in wastewater injection rates cause an increase in metric scores. Hence, future models may benefit from better including information about wastewater injection sites and rates.

This observation suggests that past one-year forecasts may have had caused a reduction in the hazard in Oklahoma, where a similar effect leading to a decrease in large events may be going on. Regulatory efforts capped wastewater injection rates in Oklahoma, leading to a gradual decline in the number of larger earthquakes. Furthermore, oil prices and earthquake rates are correlated (Roach, 2018), so the sharp decline in prices since 2014 (Prest, 2018) may influence rates. Combined economic and regulatory pressures thus led to the decrease in the maximum shaking observed in Oklahoma.

It appears the parameters used to predict seismicity in the 2017 hazard model did not fully account for these changes. The *b*-values in induced zones may appear higher due to the earthquake swarms that occur, which may be occurring presently (Goertz-Allman and Wiemer, 2012). An increase of the *b*-value in this setting would decrease the likelihood of observing higher magnitude events, a possible explanation for what was observed in 2017. Despite this, the *b*-value for the 2017 model, as well as prior years, was 1. Further gains in performance can come from improving GMMs. Many of the GMMs used for the 2017 map —all but Atkinson (2015) that received 75% of the weighting for GMMs—are derived for scenarios that may not apply to the Oklahoma region, including noninduced seismicity and the tectonic setting of the western United States. More localized GMMs may reduce the very large uncertainty that contributes to the misfit in model performance (Moschetti *et al.*, 2018; Novakovic *et al.*, 2018).

Although there is room for improvement in the hazard model, the resulting map is still useful as a whole. The metric scores, as calculated with the DYFI data, tend to be similar to the performance of the previous year's model, with only slightly worse fractional exceedance metrics than for the 2016 model. All regions except Oklahoma, where seismic rates are assumed to be better known and stable, have consistently small squared misfit metric values. As a whole, these results are better than many maps we analyzed by this approach (Stein *et al.*, 2015; Brooks *et al.*, 2016). Hence, despite weaker performance compared with the 2016 map, we believe the 2017 model is a good map. This conclusion is reinforced by the results of our seismicity simulation, which shows performance weaker than the previous year's map, but stronger than performance from our other studies (Stein *et al.*, 2015; Brooks *et al.*, 2016). Although the mismatch in simulated metric distributions appears to reflect assumptions that could be improved, the model's performance is still better, with a much smaller discrepancy between observed and predicted shaking, than that of many of other maps for natural seismicity assessed previously. For the purposes of mitigating risk and anticipating shaking in the future, the 2017 model, and all one-year hazard models, can still inform users about the hazards posed by wastewater injection and other seismically inducing activities.

DATA AND RESOURCES

The 2017 one-year seismic hazard model for the central and eastern United States (CEUS) from induced and natural earthquakes was downloaded from https://www.sciencebase.gov/catalog/item/5a85dc1de4b00f54eb36679d (last accessed December 2018). Maximum intensity "Did You Feel It?" data were provided by David Wald and Gregory Smoczyk, and can be viewed at http://usgs.maps.arcgis.com/apps/webappviewer/index.html?id=9310990e7ce84e3b8567109616b0944d (last accessed January 2019).

ACKNOWLEDGMENTS

The authors would like to thank Gregory Smoczyk for his help acquiring and using the "Did You Feel It?" (DYFI) data, and Dan McNamara and an anonymous reviewer for their insightful and helpful comments. Brooks would like to thank the Northwestern University Institute for Policy Research for

funding his research, and Lewis Brooks for his help in bypassing issues with figure making.

REFERENCES

Atkinson, G. M. (2008). Ground-motion prediction equations for eastern North America from a referenced empirical approach: Implications for epistemic uncertainty, *Bull. Seismol. Soc. Am.* **98,** no. 3, 1304–1318.

Atkinson, G. M. (2015). Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards, *Bull. Seismol. Soc. Am.* **105,** no. 2A, 981–992.

Atkinson, G. M., and D. M. Boore (2006). Earthquake ground-motion prediction equations for eastern North America, *Bull. Seismol. Soc. Am.* **96,** no. 6, 2181–2205.

Atkinson, G. M., and D. J. Wald (2007). "Did You Feel It?" intensity data: A surprisingly good measure of earthquake ground motion, *Seismol. Res. Lett.* **78,** no. 3, 362–368.

Brooks, E. M., S. Stein, and B. D. Spencer (2016). Comparing the performance of Japan's earthquake hazard maps to uniform and randomized maps, *Seismol. Res. Lett.* **87,** no. 1, 90–102.

Brooks, E. M., S. Stein, B. D. Spencer, L. Salditch, M. D. Petersen, and D. E. McNamara (2017). Assessing earthquake hazard map performance for natural and induced seismicity in the central and eastern United States, *Seismol. Res. Lett.* **89,** no. 1, 118–126.

Campbell, K. W. (2003). Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America, *Bull. Seismol. Soc. Am.* **93,** no. 3, 1012–1033.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58,** no. 5, 1583–1606.

Cremen, G., A. Gupta, and J. W. Baker (2017). Evaluation of ground motion intensities from induced earthquakes using "Did You Feel It?" data, *16th World Conf. on Earthquake Engineering*, Santiago, Chile, 12 January 2017.

Ellsworth, W. L. (2013). Injection-induced earthquakes, *Science* **341,** no. 6142, doi: 10.1126/science.1225942.

Ellsworth, W. L., A. L. Llenos, A. F. McGarr, A. J. Michael, J. L. Rubinstein, C. S. Mueller, M. D. Petersen, and E. Calais (2015). Increasing seismicity in the US midcontinent: Implications for earthquake hazard, *The Leading Edge* **34,** no. 6, 618–626.

Field, E. H. (2010). *Probabilistic Seismic Hazard Analysis: A Primer*, available at http://www.opensha.org/ (last accessed June 2018).

Frankel, A. D., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic-hazard maps: Documentation June 1996, *U.S. Geol. Surv. Open-File Rept. 96-532*, Reston, Virginia, 110 pp.

Goertz-Allman, B. P., and S. Wiemer (2012). Geomechanical modeling of induced seismicity source parameters and implications for seismic hazard assessment, *Geophysics* **78,** no. 1, KS25–KS39.

Hornbach, M. J., M. Jones, M. Scales, H. R. DeShon, M. B. Magnani, C. Frohlich, B. Stump, C. Hayward, and M. Layton (2016). Ellenburger wastewater injection and seismicity in North Texas, *Phys. Earth Planet. In.* **261,** 54–68.

Horton, S. (2012). Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake, *Seismol. Res. Lett.* **83,** no. 2, 250–260.

Hough, S. E. (2012). Initial assessment of the intensity distribution of the 2011 M_w 5.8 Mineral, Virginia, earthquake, *Seismol. Res. Lett.* **83,** no. 4, 649–657.

Keranen, K. M., H. M. Savage, G. A. Abers, and E. S. Cochran (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 M_w 5.7 earthquake sequence, *Geology* **41,** no. 6, 699–702.

Keranen, K. M., M. Weingarten, G. A. Abers, B. A. Bekins, and S. Ge (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection, *Science* **345,** no. 6195, 448–451.

Liu, M., G. Luo, H. Wang, and S. Stein (2014). Long aftershock sequences in North China and central US: Implications for hazard assessment in mid-continents, *Earthq. Sci.* **27,** no. 1, 27–35.

Mak, S., and D. Schorlemmer (2016a). A comparison between the forecast by the United States National Seismic Hazard Maps with recent ground-motion records, *Bull. Seismol. Soc. Am.* **106,** no. 4, 1817–1831.

Mak, S., and D. Schorlemmer (2016b). What makes people respond to "Did You Feel It?", *Seismol. Res. Lett.* **87,** no. 1, 119–131.

McMahon, N. D., M. Stickney, R. C. Aster, W. Yeck, H. R. Martens, and H. Benz (2017). Spatiotemporal analysis of the foreshock–mainshock–aftershock sequence of the 6 July 2017 M 5.8 Lincoln, Montana earthquake, *AGU Fall Meeting Abstracts*, New Orleans, Louisiana, 12 December 2017.

Moschetti, M. P., E. M. Thompson, P. M. Powers, S. M. Hoover, and D. E. McNamara (2018). Ground motions from induced earthquakes in Oklahoma and Kansas, *Seismol. Res. Lett.* **90,** no. 1, 160–170.

Mostafa Mousavi, S., and G. C. Beroza (2018). Evaluating the 2016 one-year seismic hazard model for the central and eastern United States using instrumental ground-motion data, *Seismol. Res. Lett.* **89,** no. 3, 1185–1196.

Novakovic, M., G. M. Atkinson, and K. Assatourians (2018). Empirically calibrated ground-motion prediction equation for Oklahoma, *Seismol. Res. Lett.* **108,** no. 5A, 2444–2461.

Ogwari, P. O., H. R. DeShon, and M. J. Hornbach (2018). The Dallas-Fort worth airport earthquake sequence: Seismicity beyond injection period, *J. Geophys. Res.* **123,** no. 1, 553–563.

Petersen, M. D., M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A. D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd, *et al.* (2014). Documentation for the 2014 update of the United States national seismic hazard maps, *U.S. Geol. Surv. Open-File Rept. 2014-1091*, 243 pp., doi: 10.3133/ofr20141091.

Petersen, M. D., M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A. D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd, *et al.* (2015). The 2014 United States national seismic hazard model, *Earthq. Spectra* **31,** no. S1, S1–S30.

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, *Seismol. Res. Lett.* **87,** 1327–1341.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, K. S. Rukstales, D. E. McNamara, and A. L. Llenos (2018). 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes, *Seismol. Res. Lett.* **89,** no. 3, 1049–1061.

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). Data release for 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes, *U.S. Geol. Surv. Data Release*, doi: 10.5066/F7Cf9PC4.

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, *Seismol. Res. Lett.* **88,** no. 3, 772–783.

Pezeshk, S., A. Zandieh, and B. Tavakoli (2011). Hybrid empirical ground-motion prediction equations for eastern North America using NGA models and updated seismological parameters, *Bull. Seismol. Soc. Am.* **101,** no. 4, 1859–1870.

Prest, B. C. (2018). Explanations for the 2014 oil price decline: Supply or demand? *Energy Econ.* **74,** 63–75.

Quitoriano, V., E. M. Thompson, G. M. Smoczyk, and D. J. Wald (2017). Access to "Did You Feel It?" data for induced earthquake studies, *2017 Seismological Society of America Annual Conference*, Denver, Colorado, 20 April 2017.

Roach, T. (2018). Oklahoma earthquakes and the price of oil, *Energy Policy* **121,** 365–373.

Silva, W., N. Gregor, and R. Darragh (2002). *Development of Regional Hard Rock Attenuation Relations for Central and Eastern North America*, Pacific Engineering and Analysis, El Cerrito, California.

Stein, S., R. J. Geller, and M. Liu (2011). Bad assumptions or bad luck: Why earthquake hazard maps need objective testing, *Seismol. Res. Lett.* **82,** no. 5, 623–626.

Stein, S., R. J. Geller, and M. Liu (2012). Why earthquake hazard maps often fail and what to do about it, *Tectonophysics* **562,** 1–25.

Stein, S., B. D. Spencer, and E. M. Brooks (2015). Metrics for assessing earthquake-hazard map performance, *Bull. Seismol. Soc. Am.* **105,** no. 4, 2160–2173.

Tavakoli, B., and S. Pezeshk (2005). Empirical-stochastic ground-motion prediction for eastern North America, *Bull. Seismol. Soc. Am.* **95,** no. 6, 2283–2296.

Toro, G. R. (2002). Modification of the Toro *et al.* (1997) attenuation equations for large magnitudes and short distances, *Risk Eng. Tech. Rept.*, 10 pp.

Vanneste, K., S. Stein, T. Camelbeeck, and B. Vleminckx (2018). Insights into earthquake hazard map performance from shaking history simulations, *Sci. Rep.* **8,** no. 1, 1855.

Wald, D. J., B. C. Worden, V. Quitoriano, and K. L. Pankow (2005). ShakeMap manual: Technical manual, *User's Guide, and Software Guide, Report Number: TM 12–A1*.

Wald, D. J., V. Quitoriano, L. A. Dengler, and J. W. Dewey (1999). Utilization of the internet for rapid community intensity maps, *Seismol. Res. Lett.* **70,** no. 6, 680–697.

Wald, D. J., V. Quitoriano, C. B. Worden, M. Hopper, and J. W. Dewey (2012). USGS "Did You Feel It?" Internet-based macroseismic intensity maps, *Ann. Geophys.* **54,** no. 6, doi: 10.4401/ag-5354.

White, I. J., T. Liu, N. Luco, and A. B. Liel (2018). Considerations in comparing the US Geological Survey one-year induced-seismicity hazard models with "Did You Feel It?" and instrumental data, *Seismol. Res. Lett.* **89,** no. 1, 127–137.

Wolin, E., S. Stein, F. Pazzaglia, A. Meltzer, A. Kafka, and C. Berti (2012). Mineral, Virginia, earthquake illustrates seismicity of a passive-aggressive margin, *Geophys. Res. Lett.* **39,** no. 2, doi: 10.1029/2011GL050310.

Worden, C. B., M. C. Gerstenberger, D. A. Rhoades, and D. J. Wald (2012). Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California, *Bull. Seismol. Soc. Am.* **102,** no. 1, 204–221.

Yeck, W. L., G. P. Hayes, D. E. McNamara, J. L. Rubinstein, W. D. Barnhart, P. S. Earle, and H. M. Benz (2017). Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts, *Geophys. Res. Lett.* **44,** no. 2, 711–717.

Zhuang, J., and S. Touati (2015). Stochastic simulation of earthquake catalogs, *Community Online Resource for Statistical Seismicity Analysis*, available at http://www.corssa.org/export/sites/corssa/.galleries/articles-pdf/zhuang-touati.pdf (last accessed December 2018).

Edward M. Brooks[1]
James Neely
Seth Stein[1]
Leah Salditch
Department of Earth and Planetary Sciences
Technological Institute
Northwestern University
2145 Sheridan Road
Evanston, Illinois 60208-3130 U.S.A.
eddie@earth.northwestern.edu
james@earth.northwestern.edu
seth@earth.northwestern.edu
leah@earth.northwestern.edu

Bruce D. Spencer[1]
Department of Statistics
Northwestern University
2006 Sheridan Road
Evanston, Illinois 60208 U.S.A.
bspencer@northwestern.edu

Published Online 10 April 2019

[1] Also at Institute for Policy Research, Northwestern University, 2040 Sheridan Road, Evanston, Illinois 60208 U.S.A.