



## EVALUATION OF GROUND MOTION INTENSITIES FROM INDUCED EARTHQUAKES USING “DID YOU FEEL IT?” DATA

G. Cremen<sup>(1)</sup>, A. Gupta<sup>(2)</sup>, J. Baker<sup>(3)</sup>

<sup>(1)</sup> Graduate Student, Stanford University, gcremen@stanford.edu

<sup>(2)</sup> Graduate Student, Stanford University, abhineet@stanford.edu

<sup>(3)</sup> Associate Professor, Stanford University, bakerjw@stanford.edu

### **Abstract**

The Central and Eastern United States has recently experienced a large number of earthquakes that are suspected of being induced by anthropogenic activities. Seismic risk assessment is known to be sensitive to ground motion predictions, so it is important to understand whether the intensity of ground shaking produced by those earthquakes differs relative to motions from comparable natural earthquakes. Unfortunately, due to sparse instrumentation in this area, we have limited recorded strong motion data and thus the question has not been resolved definitively. Here we attempt to address this question using U.S. Geological Survey “Did You Feel It?” (DYFI) data. Using a large set of DYFI survey responses (each with an inferred Macroseismic Intensity and a corresponding earthquake magnitude and distance), we evaluate differences between responses to natural and induced earthquakes. We find a trend that induced earthquakes produce comparable or possibly larger intensities at close distances to the causal earthquake, but that these intensities attenuate faster than natural earthquakes. This finding is consistent with previous literature on the topic, which infers that this effect may be due to induced earthquakes being shallow but having relatively low stress drops. Further we find that the deviations cannot be explained by underlying factors such as differences in exposed populations, survey response rates, or deviations in responses after a sequence of felt earthquakes. This work lends further credibility to the hypothesis that induced earthquakes are capable of producing strong near-fault ground shaking. Future work will investigate the impact of this phenomenon on seismic risk in the Central and Eastern United States.

**Keywords:** *Induced earthquakes; Central and Eastern United States; “Did You Feel It?” data; Macroseismic intensities*

## 1. Introduction

There has recently been a significant increase in the occurrence of earthquakes in the Central and Eastern United States (CEUS), which is believed to be a result of anthropogenic activities. This is a cause for concern due to the potential damage that shaking from these earthquakes could cause. With this as motivation, we investigate whether the intensity of the ground shaking produced by such earthquakes differs from that produced by comparable naturally occurring earthquakes. Understanding potential differences in intensity is essential for assessing the relative seismic risk that accompanies these earthquakes.

There are theoretical reasons why ground shaking from induced seismicity may differ from naturally occurring events. The source depths of induced earthquakes are generally shallower than those of comparable naturally occurring earthquakes. In addition, potential differences may be linked to induced earthquakes having lower stress drop values, which has been suggested by previous studies [1].

Due to sparse instrumentation in the area of interest, we have limited recorded strong motion data (e.g., the PEER NGA-East database, [ngawest2.berkeley.edu](http://ngawest2.berkeley.edu), contains only 40 recordings that fit the selection criteria of this study). Thus strong ground motion data have limited ability to discriminate differences in ground motion intensity between potentially induced earthquakes and those that are naturally occurring.

However, an abundance of U.S. Geological Survey “Did You Feel It?” (DYFI) survey data has been collected in the region of interest and these have been shown to provide accurate characterization of ground motion intensities [e.g. 2]. In this study, we will employ these data (which includes earthquake magnitudes and response distances in addition to macroseismic intensities) to attempt to overcome the challenge mentioned above. Our use of DYFI data will be twofold.

Firstly, we aim to detect differences in the intensity of the two types of earthquake via a linear regression model for intensity from these data that accounts for both earthquake magnitude and distance from the epicenter. Distinction between the earthquake types (and hence the detection of intensity differences) will be achieved by altering the predictor variables of the model if necessary. Furthermore, we will use these data to investigate if any differences in intensity observed from the aforementioned model can be explained by underlying factors, such as disparities in survey response rates or populations exposed to the earthquakes. If underlying factors are not found to explain the intensity differences in the model, this is an indication that the intensities associated with potentially induced earthquakes in the CEUS are apparently different from those associated with comparable naturally occurring earthquakes in the same region.

## 2. DYFI Database

The source of data for this study is the U.S. Geological Survey (USGS) “Did You Feel It?” (DYFI, [earthquake.usgs.gov/earthquakes/dyfi/](http://earthquake.usgs.gov/earthquakes/dyfi/)) system. This online questionnaire rapidly collects macroseismic intensity data and localized damage reports on earthquakes recently felt (or not felt) by internet users [3]. Questions are worded such that, via an algorithm, the Modified Mercalli Intensity (MMI) may be computed at the observer’s location [4-7]. MMI values assigned to individual responses are subsequently averaged by the DYFI program across postal zip codes or geocoded boxes (typically at either 10- or 1-km spacing [3]) and thus variability across the affected region can be mapped [2]. Different indicated categories are weighted to give a “community weighted sum” (CWS) for the area, which is related to MMI through a linear regression and is assigned a Community Decimal Intensity (CDI) [3, 8].

DYFI data can be surprisingly robust, making up in quantity what they may lack in quality [2]. Previous studies have found that DYFI MMI observations provide reliable information on ground motion amplitudes and thus can be used to make inferences about earthquake ground motions. For example, they have provided convincing evidence that earthquake stress drops are higher in the CEUS than in California [2].

For this study, we considered non-suspect individual response (non-aggregated) DYFI data from all earthquakes between 2000 and November 2015 with magnitude  $\geq 3$  and with at least five associated individual DYFI responses, in the states listed in Table 1. We extracted all responses within 50 km of the reference earthquake, given our focus on strong shaking. We excluded the 5.4% of responses that had poor location



precision (i.e., at city level or less). This process resulted in 104,023 responses to 972 earthquakes, which were then considered in the following analysis.

Table 1 – States in which earthquakes of interest to this study occurred

Alabama	Illinois	Missouri	Pennsylvania
Arkansas	Indiana	New Jersey	South Carolina
Colorado <sup>1</sup>	Kansas	New Mexico <sup>1</sup>	Tennessee
Connecticut	Maryland	New York <sup>2</sup>	Texas
Georgia	Michigan	North Carolina	Virginia
	Mississippi	Oklahoma	West Virginia

<sup>1</sup> Only earthquakes that occurred within, or east of, the state's region of induced seismicity specified by USGS 2014 Hazard Maps (see Fig.1) were considered (since only the CEUS was of interest in this study).

<sup>2</sup> Earthquakes at latitudes exceeding 44° were ignored due to their proximity with the Canadian border (only U.S. population data were used in this study).

In addition to CDI values calculated from reports by survey respondents, the following database fields were utilized:

- Location of each earthquake (both geographic coordinates and a supplementary description of location on either a state, regional or city level are included).
- Time of occurrence of each earthquake.
- Magnitude of each earthquake.
- Respondents' locations (in geographic coordinates) and the degree of precision of these locations (e.g. house-, block-, or city-level).
- Respondents' distances from respective events.

Distinction between earthquake types was made on the basis of geographic location. Earthquakes that fell within the locations and dates set out in USGS 2014 Hazard Maps documentation [9], as well as events occurring in Kansas in 2015, were classified as potentially induced. The remainder were classified as natural (Fig.1). In all, 103 earthquakes are classified as naturally occurring, with the number of respondents within 50 km for each of these events ranging from 1 to 16,972. 869 earthquakes are classified as potentially induced, with the number of respondents within 50 km for these events ranging from 1 to 6,761. Magnitudes of the naturally occurring earthquakes range from 3 to 5.8, while those of the potentially induced earthquakes range from 3 to 5.6.

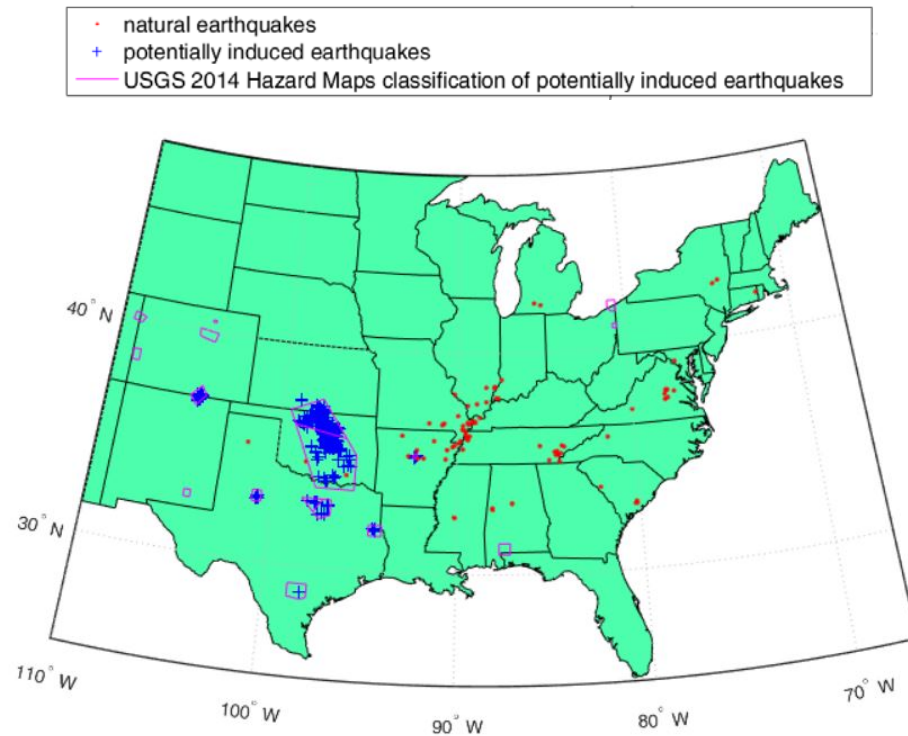


Fig. 1 – Map of considered earthquakes and their classifications as natural or potentially induced. Note that earthquakes plotted in red within polygons occurred outside of the dates specified in the documentation for induced seismicity in the region.

### 3. Investigating potential differences in intensity between natural and potentially induced earthquakes

In order to quantify possible differences in intensity between potentially induced and naturally occurring earthquakes in the CEUS, a regression model was developed using responses from the database outlined in the previous section, describing intensity as a function of both magnitude and epicentral distance (epicentral distance was chosen over hypocentral distance since the available depth data were not well constrained). Linear regression analysis was used to develop the model, with the data being split equally into one training and one test set. A forward selection scheme was implemented for the purposes of choosing parameters, with a  $p$ -value  $< 0.05$  supporting inclusion of a parameter [10]. Potential differentiation between the two types of earthquake and their associated intensities was achieved using indicator terms for certain predictors. To keep the regression analysis focused on relatively strong motions of engineering interest, only responses within 30 km of  $3 \leq M \leq 3.5$  were considered. All responses within 50 km of larger earthquakes were considered.

The main response distance measure included in the forward selection scheme contained an additive term to capture near-distance saturation at close response distances, in keeping with previously published equations that relate intensity and distance [11]. This term was larger in the case of naturally occurring earthquakes, which accounts for such earthquakes having typically greater focal depths than their potentially induced counterparts. An additional interaction term between magnitude and response distance was considered in the scheme for naturally occurring earthquakes since it was shown to influence intensity in preliminary models. Basic regression analyses indicated that higher order forms of both response distance and magnitude had negligible impact on intensity in the presence of other predictors, so these terms were not considered. Inclusion of type-dependent indicators on individual magnitude and distance terms were not found to significantly impact intensity, and were omitted from the final scheme.

The equation obtained from regression analysis for the forward selection scheme corresponding to the lowest test error, may be expressed as follows:

$$CDI = 0.4466 + 0.8809m - 0.02945\sqrt{(2.0 + 3.0I)^2 + d^2} + 0.002100m \times d \times I \quad (1)$$

where  $d$  is the epicentral distance of the response in km,  $m$  is the magnitude and  $I$  is an indicator variable equal to 1 for naturally occurring earthquakes and 0 for potentially induced earthquakes. The coefficients in the “(2.0 + 3.0  $I$ )” term were estimated from a manual search of parameters that led to a good fit with observed data, and the remaining coefficients were computed using linear least-squares regression.

The model is compared with the intensity prediction equation for Eastern United States by [11] in Fig.2 below. Note that we assumed a focal depth of 8 km when calculating the hypocentral distance term of the [11] model (AWW14), since that is also the value of depth used in [11]. AWW14 tends to predict significantly larger intensities than our model, especially at close distances, with differences in intensity as large as 1.15 for magnitude 4.8. The exception is magnitude 3, which is in good agreement with intensities predicted for natural and potentially induced earthquakes. One possible explanation for the difference between the models may be the fact that Eq. (1) is directly fit to CEUS data, whereas the AWW14 model was originally calculated with data from the Western United States and adjusted using Eastern United States intensity residuals.

Fig.2 also shows rolling average plots of the intensity data, with error bars of two standard deviations in length, over 10 km epicentral response distance widths for fixed magnitude bins of width 0.6. The central magnitudes for each of the four figures are 3.3, 4, 4.5 and 4.8 respectively. The model developed in this study is in significantly better agreement with the data than AWW14.

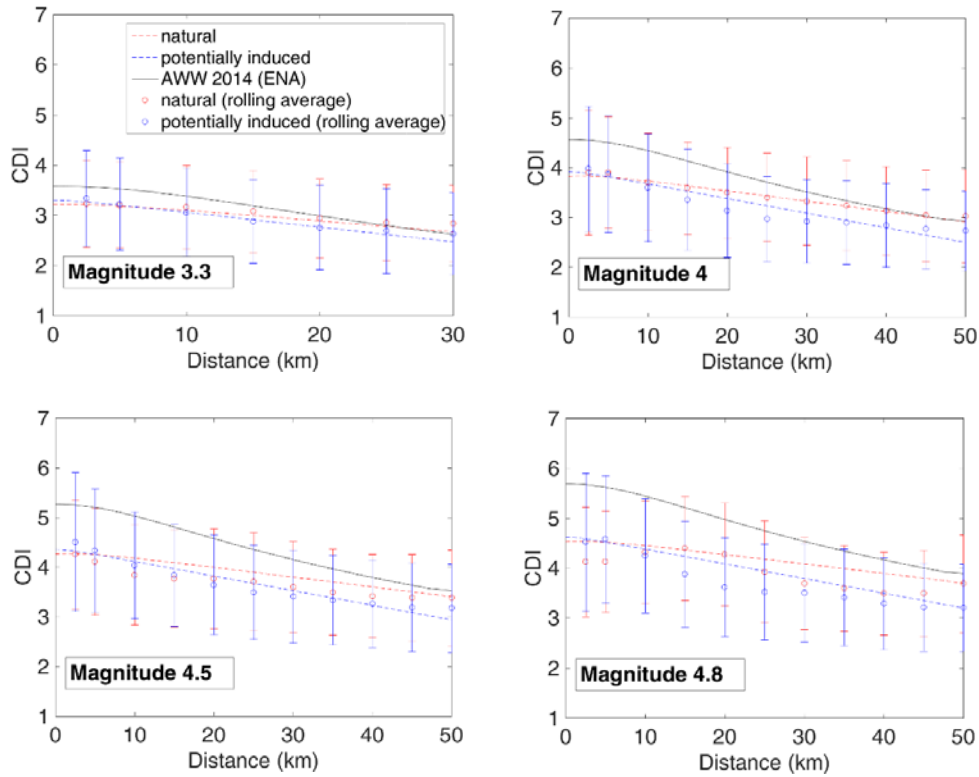


Fig. 2 – Rolling average and standard deviation plots of intensity data (over 10 km wide epicentral response distance bins and 0.6-wide magnitude bins) superimposed on the intensity-distance relationships for our model and that of AWW14. An additional set of data points centered on 2.5 km, with epicentral response distance width of 5 km, is included in each plot to highlight near-distance intensities. Note that the distance range considered for each magnitude is in line with the criteria set out in Section 3.

Fig.3 shows rolling average plots of the data, with error bars of two standard deviations in length, over 0.4-wide magnitude bins for fixed epicentral distance bins of width 3 km. The central distances for the two plots are 1.5 km and 25 km. (Note that error bars are not shown for magnitudes of 5 or greater in the case of the smaller distance due to a sparsity of available data for these magnitudes). Our model can be seen to be a considerably better fit for the data than AWW14 over the majority of magnitudes investigated in this study (the fact that AWW14 is a better fit of the data in the case of the aforementioned larger magnitudes for the smaller distance plot is not surprising, given that the amount of data available to constrain our model at such magnitudes is only 0.2% of the total relevant data). In addition, it is apparent that our model correctly predicts both potentially induced intensities to be similar to those of natural earthquakes at close distances, but smaller at farther distances.

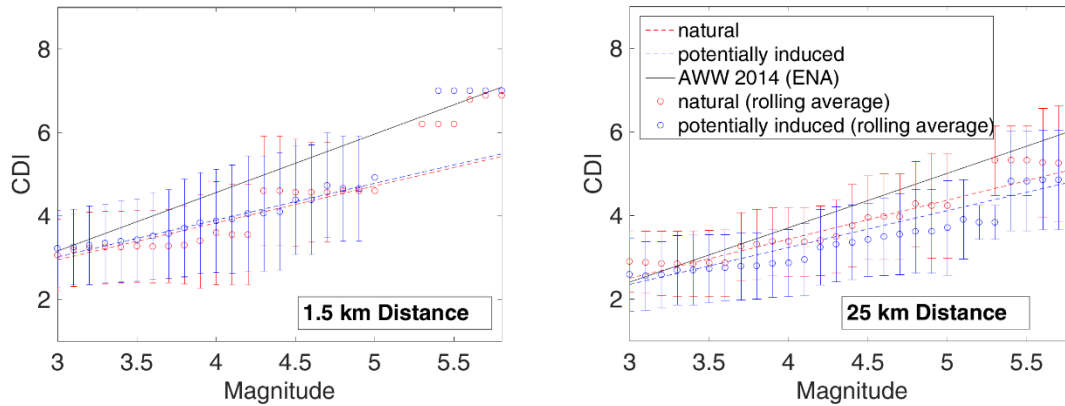


Fig. 3 – Rolling average and standard deviation plots of intensity data (over 0.4-wide magnitude bins and 3 km wide epicentral response distance bins).

Further confirmation of the superior fit of our model to the data of interest can be found by comparing the distance-dependent intensity residuals of the models over all magnitudes.

Residuals at a given epicentral distance are calculated as follows [10]:

$$\text{Residual} = I_o - I_p \quad (2)$$

where  $I_o$  is the observed intensity for the distance/magnitude combination and  $I_p$  is the predicted intensity from the model for the distance/ magnitude combination.

Rolling average plots of the residuals, computed over 2 km epicentral response distance widths, are included (Fig.4) to give an indication of whether the residuals tend to be zero mean. Such plots indicate that the residuals of our model typically tend to be significantly closer to zero mean than those of AWW14. Furthermore, the maximum deviation from zero of the rolling average plot associated with AWW14 is 3.54 times larger than that of the rolling average plot associated with our model for naturally occurring earthquakes and 3.06 times larger for potentially induced earthquakes. Similar rolling average plots for magnitude-dependent residuals, computed over 0.8-wide magnitude bins, are smaller in absolute value for our model than for AWW14 over 87.1% of binned magnitudes in the case of naturally occurring earthquakes and over 100% of binned magnitudes in the case of potentially induced earthquakes. The maximum deviation of these plots from zero is 1.08 times larger for AWW14 in the case of naturally occurring earthquakes (0.59 compared to 0.55) and 2.76 times larger for AWW14 in the case of potentially induced earthquakes (0.96 compared to 0.35).



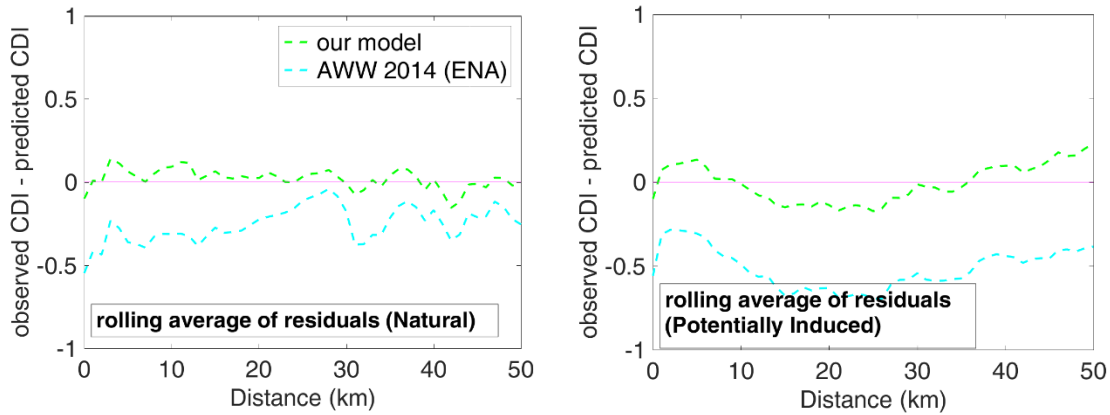


Fig. 4 – Comparison of residuals for our model and those of AWW14.

Hence, the notable deviations of our model from AWW14 do not threaten its validity within the context of this study and we can use our model to draw conclusions about the data of interest.

It appears from Fig.2 above that intensity associated with potentially induced earthquakes is similar to or potentially larger than that of naturally occurring earthquakes of the same magnitude for very close distances (up to 5 km approximately in epicentral distance), but smaller for further distances because of more rapid attenuation. This is consistent with previous literature on the topic [1], which infers that the faster attenuation of induced earthquake intensities may be the result of such earthquakes having relatively lower stress-drops, which are offset at close distances by their shallower depths. For the magnitudes investigated, differences in CDI do not appear to exceed approximately 0.5.

#### 4. Factors that may impact intensity differences

We study some underlying factors, like differences in response rates between induced and natural earthquakes, to assess whether they may explain the intensity differences observed in our model.

##### 4.1 Differences in response rates

Given the frequency of earthquakes and public awareness in some regions with potentially induced earthquakes, we considered whether people feeling potentially induced earthquakes might be more or less likely to respond in the DYFI system, and whether that might influence the mean reported CDI values. This section provides some analysis to explore that question.

Differences in response rates between the two earthquake types were determined for given magnitude and distance combinations, which involved a three-step process. Firstly, response rates for individual earthquakes under a given epicentral distance condition were calculated as follows:

$$R = \frac{n}{P} \times 100 \quad (3)$$

where  $R$  is the response rate expressed in percentage,  $n$  is the number of DYFI responses received and  $P$  is the size of the population [12], under the given distance condition for the earthquake.

Secondly, all such non-zero response rates satisfying a further magnitude condition were grouped by earthquake type and the means of both groups of response rates were determined. Finally, the ratio of the mean conditional percentage response rates for both earthquake types was computed using the following equation:

$$\text{Ratio of response rates} = \frac{E[R_{ind}]}{E[R_{nat}]} \quad (4)$$

where  $E[.]$  is the expectation operator,  $R_{ind}$  is the conditional percentage response rate associated with potentially induced earthquakes and  $R_{nat}$  is that associated with naturally occurring earthquakes. It is important to note that a ratio was only computed if it included data from at least two induced and two natural earthquakes.

Differences in percentage response rates do exist between the two types of earthquake, for various choices of distance and magnitude binning. For example, ratios as large as 4.13 and as small as 0.03 are observed for a distance bin width of 5 km and a magnitude binning of 0.1. Across the range of magnitudes and distances, people are on average approximately 10% more likely to respond to a natural earthquake than a potentially induced earthquake. Since the two earthquake types generally occur in different parts of the CEUS, this suggests that there are regional differences in response behavior. Our conclusions differ somewhat from those of [13], who found DYFI response behavior to be similar in California and the CEUS, regardless of the seismicity at hand. However, [13] compared response rates for regions with different seismicity, whereas we examined response rates for different earthquake types across regions within the CEUS only. Additionally, [13] compared response rates under the assumption that additional factors, such as educated population size, average household size and median population age, remain constant across the different regions. This extra information was not accounted for in our analysis.

To determine if the differences in percentage response rate between the two earthquake types found above may influence the differences in intensity observed, the relationship between percentage response rate and mean reported intensity was examined. Fig.5 shows percentage response rate (from Eq. (3)) versus mean CDI for both types of earthquake over a range of epicentral distance bins for magnitude 3.

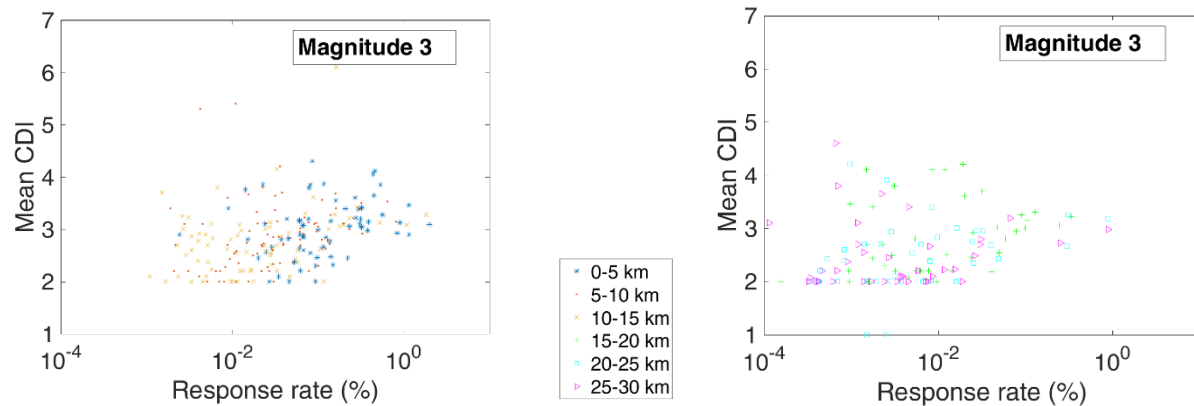


Fig. 5 – Mean intensities versus percentage response rates. Note that the distance bins plotted for each magnitude are in line with the criteria set out in Section 3.

From Fig.5, there appears to be a slight positive trend between percentage response rates and mean CDI, given a particular magnitude and distance bin. (Varying the distance bin width or the magnitude does not alter the nature of the relationship between the two variables.) This is consistent with [13], who found that CDI had a positive effect on DYFI response rate in the CEUS within a regression model. The trend is not strong enough in our data for us to deem it significant, however (as suggested by the p-values of percentage response rate when regression lines are fit between percentage response rate and mean CDI over the different binned distances).

While there are differences in percentage response rates between the two earthquake types, with there being a tendency for more frequent DYFI responses to natural earthquakes than potentially induced earthquakes, these differences cannot explain the differences in intensity observed between the two types of earthquake since there is not a strong relationship between percentage response rate and mean intensity.

#### 4.2 Differences in exposed population

Given that the average response intensity associated with a small exposed population for an earthquake may be more strongly influenced by the presence of outlier responses than the average intensity associated with a larger exposed population for the same percentage response, we considered whether the sizes of populations exposed to the two types of earthquake might affect the observed intensities. The exposed population [12] for an earthquake is defined as the population present within an epicentral distance radius equivalent to the magnitude-dependent limiting response distance specified in Section 3.



The ratio of exposed population for natural and potentially induced earthquakes under a given magnitude and epicentral distance condition was computed using the following equation:

$$\text{Ratio of exposed population} = \frac{E[EP_{ind}]}{E[EP_{nat}]} \quad (5)$$

where  $E[.]$  is the expectation operator,  $EP_{ind}$  is the (non-zero) exposed population associated with the given magnitude and distance condition from potentially induced earthquakes and  $EP_{nat}$  is that from naturally occurring earthquakes.

Large differences in exposed population are apparent between the two types of earthquake, for various combinations of distance and magnitude bin widths. For example, ratios as big as 5.08 and as small as 0.14 are observed for a distance bin width of 5 km and a magnitude binning of 0.2. However, such differences in exposed population are irrelevant if no relationship exists between exposed population and intensity. Included in Fig.6 below are plots of exposed population versus mean CDI for all types of earthquake, within epicentral distance bins of 5 km width, for magnitude 3.

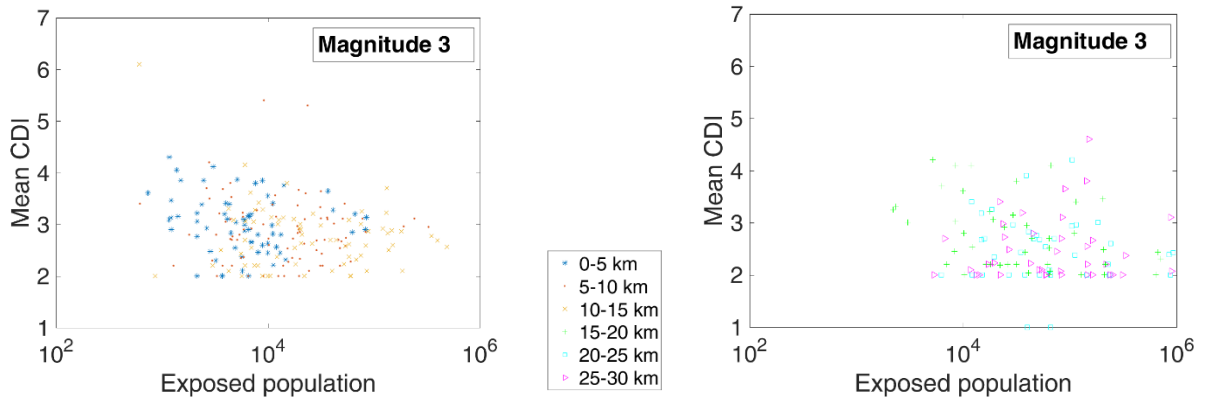


Fig. 6 – Mean intensities versus exposed population. Note that the distance bins plotted for each magnitude are in line with the criteria set out in Section 3.

Fig.6 indicates no strong trend between exposed population and mean CDI, given a particular magnitude and distance bin. Varying the distance bin width and the magnitude did not cause the nature of the relationship between the two variables to change. We conclude that, while differences exist between the sizes of the populations exposed to the two types of earthquake, they do not explain the differences in intensity observed between the types since no strong relationship exists between exposed population and mean intensity.

#### 4.3 Behavior towards repeated earthquakes

Differences in response behavior within a series of felt earthquakes may arise if a population grows more or less sensitive to the occurrence of earthquakes in a sequence. Since deviations in response behavior within sequences of felt earthquakes could have an influence on the CDI-distance relationships obtained previously, it is conceivable that such deviations may explain the differences in intensity observed between the two types of earthquakes. Another potential cause of deviations in response behavior is the variability of the time of day at which earthquakes occur in a sequence.

For this analysis, sequences of earthquakes with identical magnitudes, occurring within 8 km of each other within a period of 8 months and associated with at least 15 responses within 30 km of their epicenter, were examined. 22 sequences that satisfied these criteria and contained at least 2 earthquakes were analyzed. Magnitudes of these sequences ranged from 3 to 3.6. Fig.7 shows the relationship between time since the sequence began and the comparison of mean intensity of individual earthquakes with the mean intensity of the first event in the sequence for all 22 sequences and an epicentral response distance limit of 30 km. Points plotted in black indicate earthquakes that occurred during the night (i.e. between midnight and 7 AM).

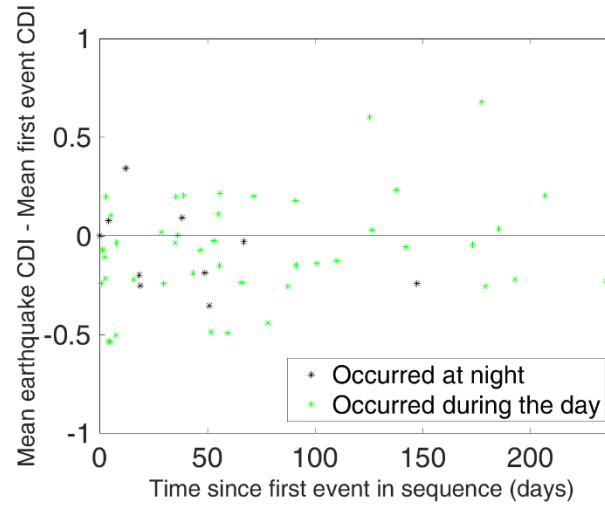


Fig. 7 – Relative intensity of sequential earthquakes versus time since sequence began.

As seen in Fig.7, there is no distinct relationship between earthquake intensity reported and position in the sequence. Furthermore, there is no meaningful trend evident between timing of earthquake (i.e. daytime or nighttime) and the mean earthquake intensity obtained from responses.

The response distance limit was changed to both 20 km and 50 km. In both cases, no distinct relationship between earthquake intensity and either sequence position or earthquake timing was observed. This implies that the response distance radius employed does not influence the nature of the trends associated with earthquake intensities during sequences.

In addition, responses from nearby cities with specific population thresholds were studied independently due to the possibility of city-specific trends existing over sequences. Cities located within 60 km of the earthquakes, with populations or greater than 1000 or 50,000, were investigated for each sequence. For each case, the closest city associated with every earthquake in a sequence was identified and the mean intensity computed for an area of 0.1 degree in longitude and longitude around the city. Again, it was found that no clear trend exists between earthquake intensity and either sequence position or earthquake timing.

In summary, there are no firm deviations in response behavior within the sequences of felt earthquakes studied, so response to earthquake sequences should not cause the intensity differences observed between potentially induced and naturally occurring earthquakes in the CEUS.

#### 4.4 Differences in response accuracy

The accuracy of reported intensity values at a given distance for a particular magnitude may reflect, for example, the amount of user experience with the DYFI reporting system. We evaluated standard deviations of reported intensity values, in line with previous studies on the accuracy of DYFI responses [14], to ensure that the level of accuracy of responses to one type of earthquake did not affect estimates of mean CDI (e.g., via an increased presence of outliers).

We examined the standard deviations of responses across 1 km epicentral distance bins for magnitudes 3, 4, 4.5 and 4.8. To avoid complications from small sample sizes, only bins with more than 5 responses were considered. To account for uncertainties in estimated standard deviations due to finite sample sizes, the standard error (s.e.) of each standard deviation was computed as follows [15]:

$$\text{s.e.} = \frac{\sigma}{\sqrt{2(n-1)}} \quad (6)$$

where  $\sigma$  is the standard deviation of responses in a magnitude-and-distance bin and  $n$  is the number of observations in that bin. Fig.8 shows standard deviations of responses for natural and potentially induced earthquakes having magnitudes 3 and 4.

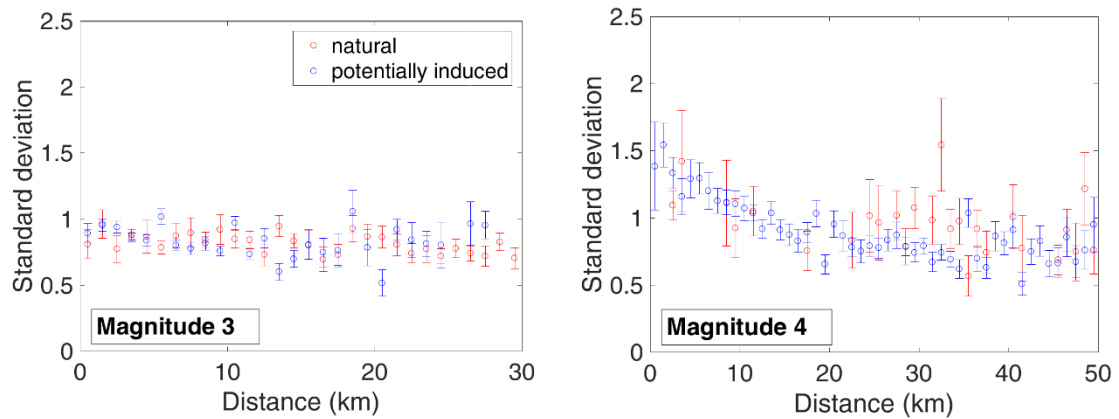


Fig. 8 – Standard deviations (best estimates and  $\pm$  one standard error) for the two types of earthquakes, considering magnitude 3 and 4 earthquakes, and 1 km wide epicentral distance bins. The distance bins plotted for each magnitude match the Section 3 criteria.

There are a number of magnitude-and-distance bins in which the standard deviations of CDIs from the two types of earthquakes differ by more than the standard errors. The differences show no systematic trend, however, and some number of differing values are expected due to statistical sampling. We thus conclude that there are no systematic differences in the accuracy of reported CDI values, and that the level of accuracy does not influence the differences in intensity observed between the two types of earthquake.

## 5. Conclusions

We examined 15 years of “Did You Feel It?” data from the Central and Eastern United States, to evaluate potential differences in shaking intensity resulting from potentially induced and natural earthquakes (with distinction between the two earthquake types being made on the basis of geographic location). Consistent with previous literature on the topic, we find that the intensities associated with potentially induced earthquakes in the region are similar to or possibly larger than those for naturally occurring earthquakes of the same magnitude for close distances (up to approximately 5 km in epicentral distance) but that the intensities attenuate faster with distance. Furthermore, we find that the differences in intensity observed cannot be explained by differences in exposed population, survey response rates, or deviations in response behavior during a sequence of felt earthquakes. This work lends further credibility to the hypothesis that induced earthquakes are capable of producing strong near-fault ground shaking. Because the classification of induced versus natural earthquakes was made on the basis of location, it is possible that this finding is due in part to regional differences in ground motion attenuation. But the wide range of locations of the data, and the fact that the observed trends are consistent with expectations regarding induced earthquakes, suggests that these differences are likely attributable to the induced nature of many of the earthquakes. This work lays the foundations for further work that will investigate the impact of this phenomenon on seismic risk in the Central and Eastern United States.

## 6. Acknowledgements

We thank David Wald and Vincent Quitoriano for helping to provide the DYFI data used here, and for helpful comments that improved this manuscript. Partial support for this work was provided by the Stanford Center for Induced and Triggered Seismicity.

## 7. References

- [1] Hough SE (2014): Shaking from Injection-Induced Earthquakes in the Central and Eastern United States. *Bulletin of the Seismological Society of America*, **104** (5), 2619–2626.



- [2] Atkinson GM, Wald DJ (2007): “Did You Feel It?” Intensity Data: A Surprisingly Good Measure of Earthquake Ground Motion. *Seismological Research Letters*, **78** (3), 362–368.
- [3] Wald DJ, Quitoriano V, Worden CB, Hopper M, Dewey JW (2011): USGS “Did You Feel It?” Internet-based macroseismic intensity maps. *Annals of Geophysics*, **54** (6), 688–707.
- [4] Wood HO, Neumann F (1931): Modified Mercalli intensity scale of 1931. *Bulletin of the Seismological Society of America*, **21** (4), 277-283.
- [5] Dewey JW, Reagor BG, Dengler LA, Moley K (1995): Intensity distribution and isoseismal maps for the Northridge, California, earthquake of January 17, 1994. *U.S. Geological Survey Open-File Report 95-92*, United States Geological Survey, Reston, USA.
- [6] Wald DJ, Quitoriano V, Heaton T, Kanamori H, Scrivner CW, Worden CB (1999a): TriNet “ShakeMaps”: Rapid Generation of Instrumental Ground Motion and Intensity Maps for Earthquakes in Southern California. *Earthquake Spectra*, **15** (3), 537-555.
- [7] Dewey JW, Wald DJ, Dengler LA (2000): Relating conventional USGS Modified Mercalli Intensities to intensities assigned with data collected via the Internet. *Seismological Research Letters*, **71** (2), 264.
- [8] Dengler LA, Dewey JW (1998): An Intensity Survey of Households Affected by the Northridge, California, Earthquake of 17 January, 1994. *Bulletin of the Seismological Society of America*, **88** (2), 441-462.
- [9] Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field N, Chen R, Rukstales KS, Luco N, Wheeler RL, Williams RA, Olsen A (2014): *Documentation for the 2014 Update of the United States National Seismic Hazard Maps*. *U.S. Geological Survey Open-File Report 2014/1091*, United States Geological Survey, Reston, USA.
- [10] James G, Witten D, Hastie T, Tibshirani R (2013): *An Introduction to Statistical Learning with Applications in R*. Springer-Verlag, 1<sup>st</sup> edition.
- [11] Atkinson GM, Worden CB, Wald DJ (2014): Intensity Prediction Equations for North America. *Bulletin of the Seismological Society of America*, **104**(6), 3084–3093.
- [12] Minnesota Population Center (2011): *National Historical Geographic Information System: Version 2.0 2011*, University of Minnesota, Minneapolis, USA.
- [13] Mak S, Schorlemmer D (2016): What Makes People Respond to “Did You Feel It?”. *Seismological Research Letters*, **87** (1), 119–131.
- [14] Wald DJ, Quitoriano V, Worden CB, Hopper M, Dewey JW (2011): USGS “Did You Feel It?” Internet-based macroseismic intensity maps. *Annals of Geophysics*, **54**(6), 688–707.
- [15] Ahn S, Fessler JA (2003): Standard Errors of Mean, Variance, and Standard Deviation Estimators. *Technical Report 413*, Communications and Signal Processing Laboratory, EECS Department, University of Michigan, Ann Arbor, USA.