

Evaluation of intensity prediction equations (IPEs) for small-magnitude earthquakes

Ganyu Teng, Jack W. Baker, David J. Wald

August 29, 2022

Abstract

This study assesses existing intensity prediction equations (IPEs) for small unspecified magnitude ($M \leq 3.5$) earthquakes at short hypocentral distances (D_h) and explores such earthquakes’ contribution to the felt shaking hazard. In particular, we consider IPEs by Atkinson and Wald (2007) and Atkinson *et al.* (2014) and evaluate their performance based on “Did You Feel It” (DYFI) reports and recorded peak ground velocities (PGVs) in the central U.S. Both IPEs were developed based on DYFI reports in the central and eastern U.S. with moment magnitudes above $M_w 3.0$. DYFI reports are often used as the ground truth when evaluating and developing IPEs, but they could be less reliable when there are limited responses for small-magnitude earthquakes. We first compare the DYFI reports with intensities interpolated from recorded PGVs. Results suggest a minimal discrepancy between the two when the intensity is large enough to be felt (i.e., $M > 2$ and $D_h < 15$ km). We then compare intensities from 31,617 DYFI reports of 3,049 earthquakes to the two IPEs. Results suggest that both IPEs match well with observed intensities for $2.0 < M < 3.0$ and $D_h < 10$ km, but the IPE by Atkinson *et al.* (2014) matches better for larger distances. We also observe that intensities from DYFI reports attenuate faster compared to the two IPEs, especially for distances greater than 10 km. We then group DYFI reports by inferred V_{S30} , as a proxy for site amplification effects. We observe that intensities at sites with V_{S30} around 300 m/s are consistently higher than at sites with V_{S30} around 700 m/s and are also closer to the two IPEs. Finally, we conduct hazard disaggregation for earthquakes at close distances ($D_h = 7.5$ km) using the observed records. Results suggest that earthquakes with magnitudes below $M 3.0$ contribute more than 40% to the occurrence of felt shaking.

Introduction

The Modified Mercalli Intensity (MMI) scale is used by the U.S. Geological Survey (USGS) for macroseismic assignments and their “Did You Feel It? (DYFI)” system and ShakeMap. The MMI scale was first introduced by Wood and Neumann (1931). Unlike other intensity measures such as spectral acceleration or peak ground acceleration, MMI depends solely on observations, including felt intensities and structural damage (Wood and Neumann, 1931), though Wald *et al.* (1999b) introduced an instrumental version of the scale specifically for use in ShakeMap.

Intensity prediction equations (IPE) have been developed to estimate the MMI anticipated for a given set of earthquake parameters and site distances. Bakun and Wentworth (1997) developed a strategy to bound the epicentral region and moment magnitude (M_w) using MMI observations. The methods were constructed based on 22 earthquakes in California with $M_w > 4.4$. Atkinson and

Wald (2007) introduced IPEs for both western North America (WNA) and central and eastern North America (CENA) using DYFI data mostly based on earthquakes with $M_w \geq 3.0$. DYFI is an online questionnaire that collects data from online users through simple multiple-choice questions about their experiences and observations of an earthquake. The responses are used to infer the MMIs at users' locations through a simple algorithm (Wald *et al.*, 2012). Atkinson and Wald (2007) used the IPEs to infer that the stress drop in CENA was higher than that in WNA. Atkinson *et al.* (2014) revised the IPE by Atkinson and Wald (2007) to improve its performance at large magnitudes and close distances. The model was developed based on DYFI data from earthquakes with $M_w \geq 3.0$. Allen *et al.* (2012) developed a globally applicable IPE based on earthquakes with $M_w > 5$ for active crustal regions. Other intensity attenuation models developed for regions outside the U.S. include Dowrick and Rhoades (2005), Le Goff *et al.* (2014), and Ahmadzadeh *et al.* (2020). Various researchers have evaluated the performances of existing IPEs. Hough (2014) evaluated the IPE by Atkinson and Wald (2007) against DYFI data from 11 moderate earthquakes in CENA and introduced an effective intensity magnitude that, on average, would generate the intensity distribution observed. By comparing the effective intensity magnitude with the event magnitude, she concluded that the stress drops of injection-induced earthquakes are lower than tectonic earthquakes in CENA, though she did not reconcile whether that was due to shallower depths or the nature of induced events. Atkinson *et al.* (2018) compared the IPE from Atkinson *et al.* (2014) to DYFI data in CENA. They found that intensities from induced earthquakes were similar to natural earthquakes within 10 km from the epicenter, but were lower at greater distances. Cremen *et al.* (2017) studied DYFI intensities from induced earthquakes with unspecified magnitude (M) above M3 in the central U.S. and compared them with predictions from Atkinson *et al.* (2014). They found that Atkinson *et al.* (2014) tended to overestimate the intensity at close distances. They also observed that the intensities of induced earthquakes attenuated faster than those of natural earthquakes.

Most of the above IPEs and the corresponding evaluations focused on moderate-to-large earthquakes (i.e., $M > 3.5$). There have been a limited number of studies on small-magnitude earthquakes, mainly because those earthquakes rarely cause structural damage. However, the increasing number of small-magnitude induced earthquakes in the central U.S. can cause felt shaking and raise public concerns (Ellsworth, 2013). Teng and Baker (2020) used the IPE by Atkinson *et al.* (2014) to conduct a hazard analysis for hydraulic-fracturing-induced earthquakes in West Texas and observed that $M \leq 3$ earthquakes could contribute over 40% of the occurrences of $MMI \geq 3$ (i.e., felt shaking). The large contribution is due to the potentially close distance of hydraulic-fracturing-induced earthquakes and the frequency of small-magnitude earthquakes. Moreover, Minson *et al.* (2021) also observed that small-magnitude earthquakes could generate significant shaking, and seismic hazard is often from earthquakes with magnitudes lower than expected. Thus, we aim to evaluate the performance of existing IPEs for small-magnitude earthquakes at close distances and explore the impact of those earthquakes on felt shaking hazard.

In this project, we evaluated the performance of two existing IPEs, AW07 (Atkinson and Wald, 2007) and AWW14 (Atkinson *et al.*, 2014), based on earthquakes in the central U.S. Both models have separate IPEs for the western U.S. and central U.S., and both were developed based on the USGS DYFI database. The largest difference between the two is that Atkinson *et al.* (2014) has a linear relationship between intensity and magnitudes whereas Atkinson and Wald (2007) has a quadratic relationship (Figure 1). As a result, the two models have very similar behavior at moderate magnitudes, but the difference is larger at small and large magnitudes. In particular, Atkinson and Wald (2007) predicts relatively larger intensities for $M < 3$. We also performed hazard disaggregation for earthquakes at close distances based on recorded intensities to study the contribution of small earthquakes on felt shaking hazard.

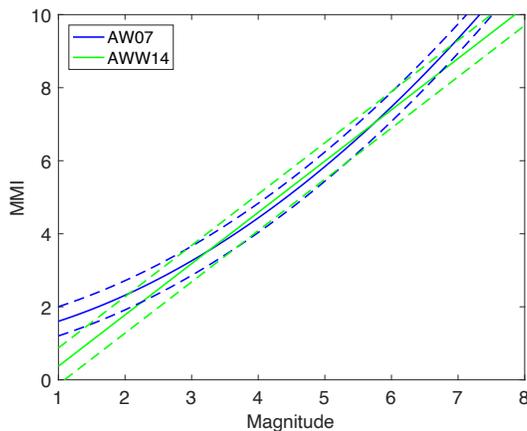


Figure 1: IPEs by Atkinson and Wald (2007) and Atkinson *et al.* (2014), assuming a hypocentral distance of 7.5 km. The solid lines are the expected values and the dashed lines are the one standard deviation ranges.

Data and Processing

We collected DYFI data for earthquakes between 2010 and 2020 with magnitudes between M1.5 and M3.5 and hypocentral distances (D_h) within 30 km in Texas, Oklahoma, and Kansas (Figure 2a). Only records at close distances were considered because of the imperceptible shaking at large distances. Figure 2c summarizes the magnitude distribution of earthquakes with DYFI data. Most of the earthquakes have a magnitude above M2.5, which was expected as earthquakes with stronger shaking are potentially felt by more people. Magnitudes are provided in different scales due to the small magnitudes considered. Figure 3 summarizes the number of earthquakes in various magnitude scales as per the USGS Comprehensive Earthquake Catalog (ComCat, Guy *et al.* 2015). As shown in Figure 3, most of the magnitudes were Ml and M_{bLg} , whereas both IPEs considered were developed based on moment magnitude. According to studies by Rigsby *et al.* (2014) and Trugman *et al.* (2017), the differences among the three magnitude scales were within 0.25 units for small magnitudes considered. Thus, during the evaluation, we used the magnitude scales as per ComCat.

DYFI is an online questionnaire that collects macroseismic data from online users (Wald *et al.*, 2012). The responses are used to generate intensity maps within minutes after earthquakes. In particular, the developers group the responses by ZIP code regions, cities, or predefined rectangular boxes and compute the community weighted sum (CWS) by taking the weighted sum of the average response from each question. The CWS is then related to MMI through a linear relationship and assigned a Community Decimal Intensity (CDI). Previous analyses have shown consistency and comparability of CDIs and MMIs (Wald *et al.*, 2012). Thus in this study, we use CDI to evaluate the performance of two IPEs and study the contribution of small-magnitude earthquakes on felt shaking hazard.

DYFI reports are aggregated within 1 km, 10 km blocks, or ZIP code regions, and one intensity value is provided for every grid. To have a precise distance estimation, we first collected DYFI data in a 1 km grid in Oklahoma, Kansas, and Texas. If there are multiple individual responses within the grid, we treated them as separate responses with identical intensity. Among the data collected, 76% have only one response per grid. There were 3,049 earthquakes with 31,617 DYFI responses having CDI values between 0 and 7.1. There is a potential sampling bias in DYFI data due to “not felt” shaking intensities being less frequently reported (Hough, 2013; Boatwright and Phillips, 2017).

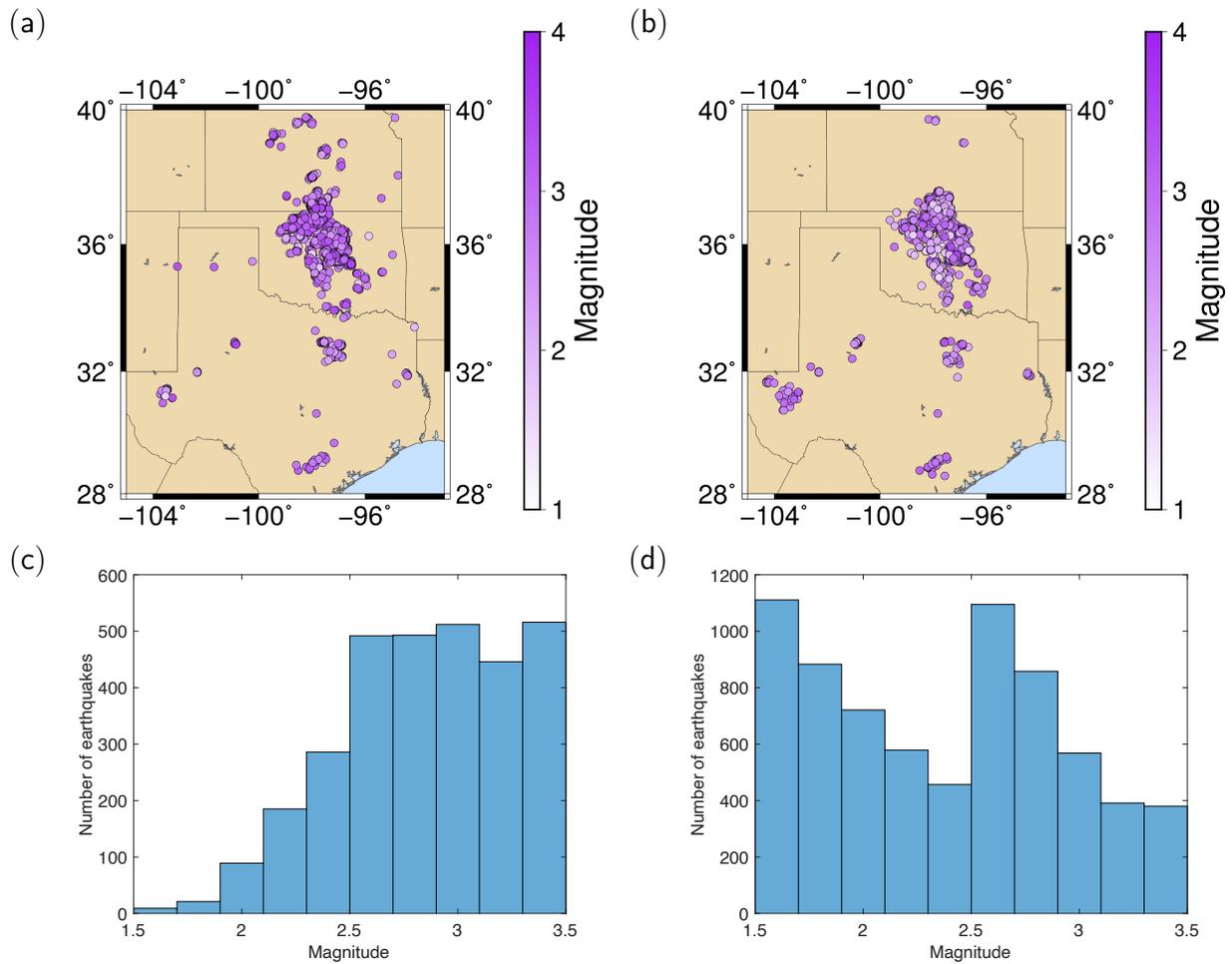


Figure 2: (a) 3,049 earthquakes with DYFI data, (b) 7,043 earthquakes with PGV data. All the earthquakes occurred between 2010 to 2020, with magnitudes from M1.5 to M3.5. Eight hundred and seven of them have both DYFI data and seismic station records available. (c) Earthquake magnitudes for DYFI data, (d) earthquake magnitudes for PGV data.

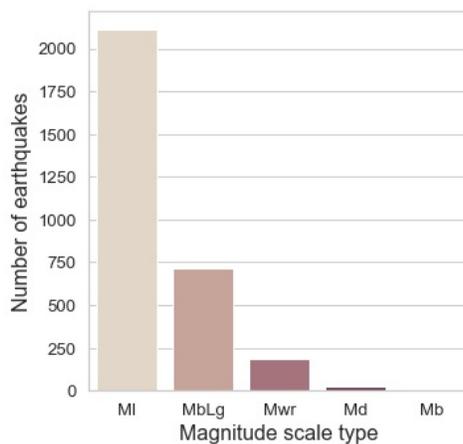


Figure 3: Number of earthquakes in various magnitude scales as per the USGS Comprehensive Earthquake Catalog.

To accurately analyze the contribution of small magnitudes earthquakes on felt shaking hazard, we adjusted CDI values according to ZIP code population using the method by Boatwright and Phillips (2017):

$$CDI^* = CDI + (CDI - 1) \log\left(\frac{7900n}{p}\right) \quad \text{for } p \geq 7900n \quad (1)$$

where CDI is the original CDI value for every individual response, CDI^* is the adjusted intensity, n is the number of DYFI responses in a given ZIP code, and p is the ZIP code population. We collected population data per ZIP code as in 2019 from SimplyAnalytics. The last term, $p/7900$, is to normalize population to DYFI responses. Boatwright and Phillips (2017) used $p/5000$, where $1/5000$ was estimated as the average portion of the population who submit reports from ZIP codes with $CDI \sim 2.0$ based on earthquakes in California. We updated the $1/5000$ term to $1/7900$ according to DYFI responses in Oklahoma, Kansas, and Texas (i.e., the median value in Figure 4a), where we computed the CDI for every ZIP code by averaging the individual reports within the ZIP code. Because the distribution in Figure 4a is highly skewed, we used the median value instead of the mean value to have a more robust representation. Of the DYFI reports, 14% (4,461 out of 31,617) were adjusted due to low reporting percentages, and only 2% of the CDIs were decreased by more than 1 unit (Figure 4b). Since this study focuses on the average intensity trend for small-magnitude earthquakes, we consider that the adjustment has a minor effect on the evaluation.

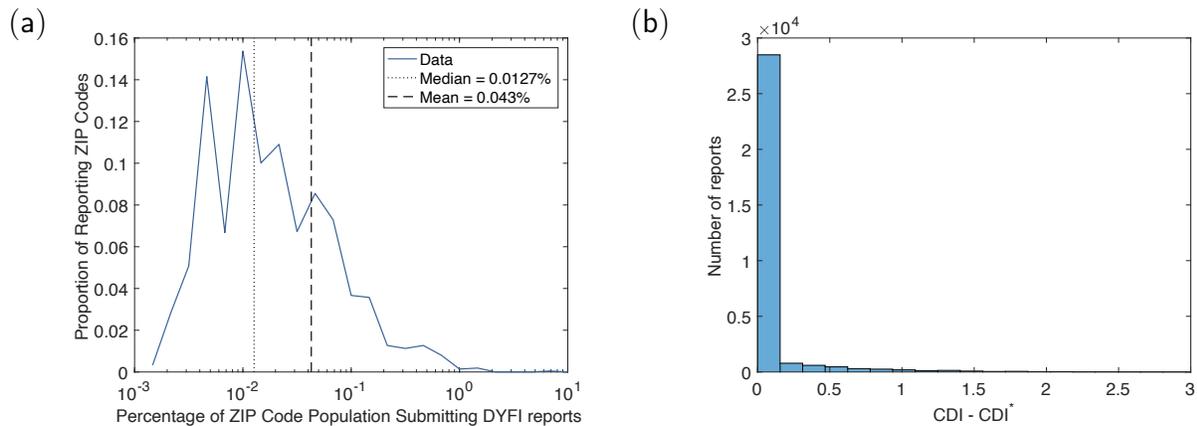


Figure 4: (a) Percentage of ZIP-code population submitting DYFI reports with $1.5 < CDI \leq 2.5$. (b) Distribution of $CDI - CDI^*$.

In addition to DYFI data, we also collected seismic station records for earthquakes in the same region with the same magnitudes and occurrence ranges. We considered earthquakes that have recorded seismograms within 60 km. There are 7,043 earthquakes with nearby seismograms (Figure 2b). Figure 2d summarizes the magnitude distribution. Many records are available for $M < 2.5$, because there are more earthquakes with small magnitudes. A total of 807 of them have both DYFI data and seismic station records available.

CDIs from DYFI data and recorded ground motions

DYFI data are statistically robust due to the large number of responses available (Wald *et al.*, 2012). They are often used as the ground truth when developing IPEs. However, for regions with a small population and situations of small-magnitude earthquakes, DYFI data may be less reliable due to

the limited number of reports. Thus, we first evaluated the CDIs from DYFI reports by studying their consistency with the CDI interpolated using seismic station records.

We considered earthquakes with both seismic records and DYFI reports and computed values of median peak ground velocities over all orientations (PGV_{RotD50}) from the recorded ground motions. Past studies (Wald *et al.*, 1999a; Boatwright *et al.*, 2001; Worden *et al.*, 2012; Caprio *et al.*, 2015) have shown that MMI was better correlated with PGV than other instrumental ground motion parameters, including peak ground acceleration (PGA) and pseudo-spectral acceleration (PSA). Wald *et al.* (1999a) observed that both PGA and PGV correlated well with MMI at low-intensity levels, and PGV correlated better with MMI at high-intensity levels. Worden *et al.* (2012) and Atkinson and Kaka (2007) also showed that PGV performed better when predicting MMI. It has lower standard deviations of residuals compared with PGA and PSA (Atkinson and Kaka, 2007; Worden *et al.*, 2012). Caprio *et al.* (2015) observed that MMI in the central and eastern U.S. correlated best with PGV among other ground motion parameters. Moreover, Wu *et al.* (2003) observed large sharp and impulse-like PGA spikes for small-magnitude earthquakes at close distances. They are localized and disappear quickly from these locations. Given the empirically better performance of PGV, we use it as the ground motion metric to infer MMI.

While USGS ShakeMap uses the larger of the PGVs from the two recorded components to interpolate shaking intensities (Wald *et al.*, 2005), we use the median PGV over all horizontal orientations (PGV_{RotD50}) to represent the shaking intensity. Boore and Kishida (2017) observed that the ratio between the larger horizontal component and PGV_{RotD50} is around 1.1. This results in a difference of around 0.1 MMI units according to the ground motion/intensity conversion equation (GMICE) of Atkinson and Kaka (2007). The difference is insignificant in the context of this study, so we do not worry further about differences amongst PGV metrics.

The 807 earthquakes with both types of data collectively had 12,073 DYFI reports and 2,978 seismic station records. Most of the DYFI reports and seismic station records are at different locations. To compare them, we spatially interpolate PGV at the DYFI report locations based on recorded PGV at seismic station locations. In particular, for every recorded PGV, we computed the total residual with respect to a chosen ground motion model (GMM) and interpolated the residual at the target location (i.e., locations of DYFI reports) according to the correlations among them. The interpolated residual was converted back to the expected PGV values based on the same GMM. Equations 2 to 7 summarize the detailed steps.

First, the actual PGV could be divided into two components:

$$\log(PGV) = \log(PGV_{GMM}) + \epsilon \quad (2)$$

where $\log(PGV)$ is the actual PGV value, $\log(PGV_{GMM})$ is the mean prediction from a GMM, and ϵ is the total residual. We first computed the total residuals at the seismic stations using Equation 2. We used the Atkinson (2015) GMM, which was designed for small-to-moderate earthquakes. This GMM was not developed for $M_w < 3$. We observed that its magnitude scaling is comparable with our PGV data for magnitudes above M3, and it's slightly slower for $M < 3$. Because the PGV estimates are based on a synthesis of both GMM predictions and interpolations from nearby observations, the observations should partially compensate for issues with extrapolating the GMM predictions. Moreover, the data show a potential trend towards larger-than-predicted PGVs for smaller magnitudes, and if this trend is stable and real then it might further increase the importance of $M < 3$ earthquakes for felt shaking. The total residual at the target location (ϵ_Y) given the calculated residuals at the seismic stations (ϵ_X) could be calculated using Equation 3:

$$E(\epsilon_Y | \epsilon_X = x) = \mu_Y + \Sigma_{YX} \Sigma_{XX}^{-1} (x - \mu_X) \quad (3)$$

where ϵ_Y is the interpolated residual at the target location, ϵ_X is a vector of residuals calculated at seismic stations based on recorded PGVs, and μ_Y and μ_X are the means of the residuals (which are zeros). The correlation vector is Σ_{YX} , where every element is the correlation between the residuals at the target location and at one seismic station (i.e., $\Sigma_{YX}(i) = \tilde{\rho}(x_i, y)$). The correlation matrix among seismic stations is Σ_{XX} :

$$\Sigma_{XX}(i, j) = \Sigma_{XX}(j, i) = \tilde{\rho}(x_i, x_j) \quad \text{for } i \neq j \quad (4)$$

$$\Sigma_{XX}(i, i) = 1 \quad (5)$$

According to Park *et al.* (2007), under assumptions of stationarity and isotropy, the correlation between the total residuals at two locations ($\tilde{\rho}(x, y)$) could be computed using the standard deviation of total residuals ($\tilde{\sigma}$), between-event residuals (τ), and within-event residuals (σ) of the chosen GMM:

$$\tilde{\rho}(x, y) = \tilde{\rho}(h) = \frac{\tau^2 + \sigma^2 \rho(h)}{\tilde{\sigma}^2} \quad (6)$$

where $\rho(h)$ is the correlation of within-event residual at two locations with a separation distance of h , and can be described by an exponential function (Sokolov and Wenzel, 2013):

$$\rho(h) = \exp\left(-\frac{h}{R_c}\right) \quad (7)$$

The parameter R_c is the correlation distance and is estimated as 23 km for PGV (Sokolov and Wenzel, 2013). We also tried different R_c values between 20 and 30 km and observed minimal differences in later results.

The interpolated residual at the target location (ϵ_Y) was then converted back to PGV using Equation 2. The interpolated PGV was further transformed to MMI scale using the following GMICE (Atkinson and Kaka, 2007):

$$\begin{aligned} MMI &= 4.37 + 1.32 \log(PGV) & \log(PGV) &\leq 0.48 \\ MMI &= 3.54 + 3.03 \log(PGV) & \log(PGV) &\geq 0.48 \end{aligned} \quad (8)$$

This GMICE was developed using earthquakes in both the western and central U.S., including $M_w < 3$ earthquakes. Atkinson and Kaka (2007) also introduced a GMICE that takes into account magnitude and hypocentral distance. We compared the two formulas and observed that the average difference between the two was only 0.15.

Figure 5 illustrates the above processes for an $M2.0$ earthquake in Oklahoma. In Figure 5a, the PGV values computed using Atkinson (2015) are much lower than the recorded PGVs. Figure 5b shows the spatial distribution of interpolated PGVs using Equations 2 to 7, which are much higher than Figure 5a and more comparable to the recorded PGVs. The interpolated PGVs were then converted to CDI using Equation 8, as shown in Figure 5c. The converted CDIs at the DYFI locations were then compared with the DYFI records.

Figures 6 and 7 summarize the comparison of CDIs interpolated from PGVs (hereafter CDI_{PGV}) with CDIs from DYFI reports (hereafter CDI_{DYFI}). We focused on small magnitude earthquakes ($M \leq 3.5$) at close distances ($D_h \leq 20$ km). Figure 6 shows the CDIs plotted versus hypocentral distance and binned by magnitude, and Figure 7 shows the CDIs plotted versus magnitude and binned by distance. Dots are the individual values, and lines are the weighted average values. We

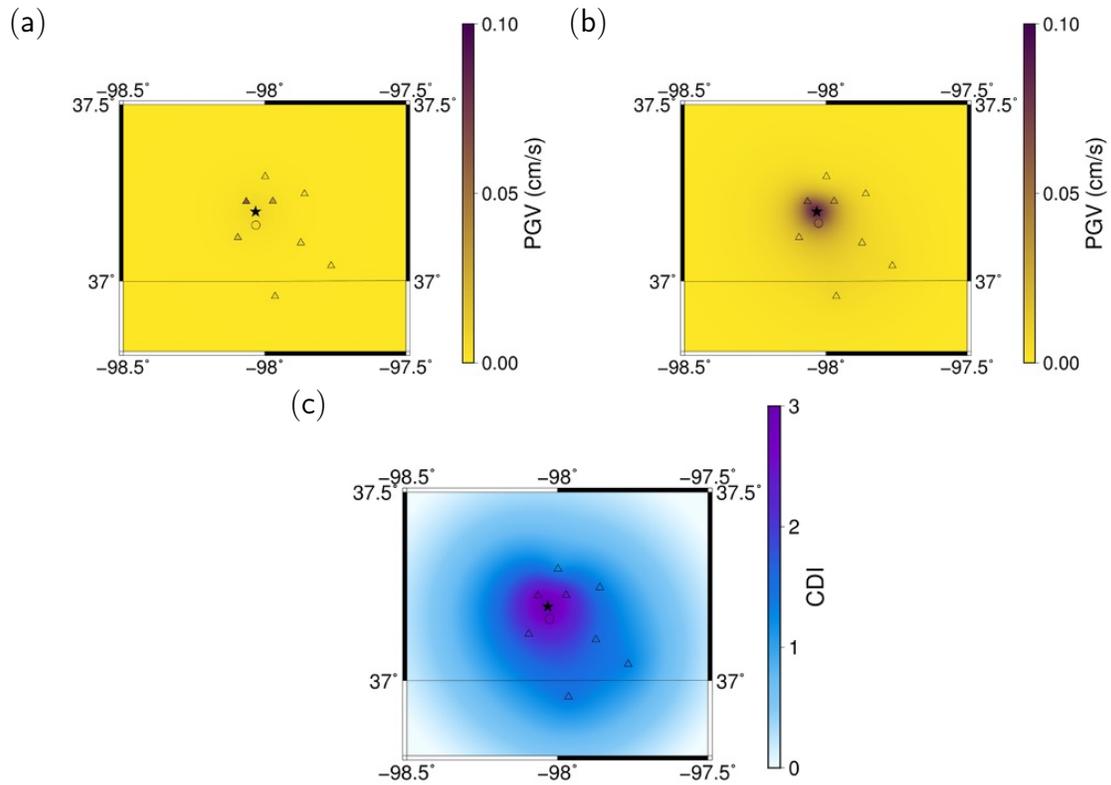


Figure 5: An example of interpolating CDIs at DYFI locations using PGV records. In every subplot, triangles are locations of seismic stations, and circles are DYFI report locations. (a) The background color indicates the PGVs computed using the GMM by Atkinson (2015). Triangle coloring indicates recorded PGVs at seismic stations, and circle coloring indicates the PGVs at DYFI locations using the GMM by Atkinson (2015). (b) The background color and the color of circles indicate the interpolated PGVs using Equations 2 to 7. Triangles coloring indicates recorded PGVs at seismic stations. (c) The spatial distribution of CDIs interpolated from PGVs in (b) using the GMICE by Atkinson and Kaka (2007).

computed the weighted average using Gaussian weighting to smooth the trend:

$$W(x, m) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad w_i = \phi\left(\frac{m_i - m}{\sigma_m}\right)$$

$$W(x, r) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad w_i = \phi\left(\frac{r_i - r}{\sigma_r}\right) \quad (9)$$

where $W()$ is an estimated weighted average value, m and r are the magnitude and distance of interest. x_i is the i^{th} individual observed value, and w_i is the corresponding weight computed from the Gaussian probability density function ($\phi()$) with a mean as the magnitude/distance of interest. For CDI versus distance plots, we used $\sigma_r = 1$ km, and for CDI versus magnitude plots, we used $\sigma_m = 0.1$. The number of data points in every subplot is n . Dashed lines in the figures indicate that the total normalized weights sum to less than 45, indicating that the weighted average is not well constrained.

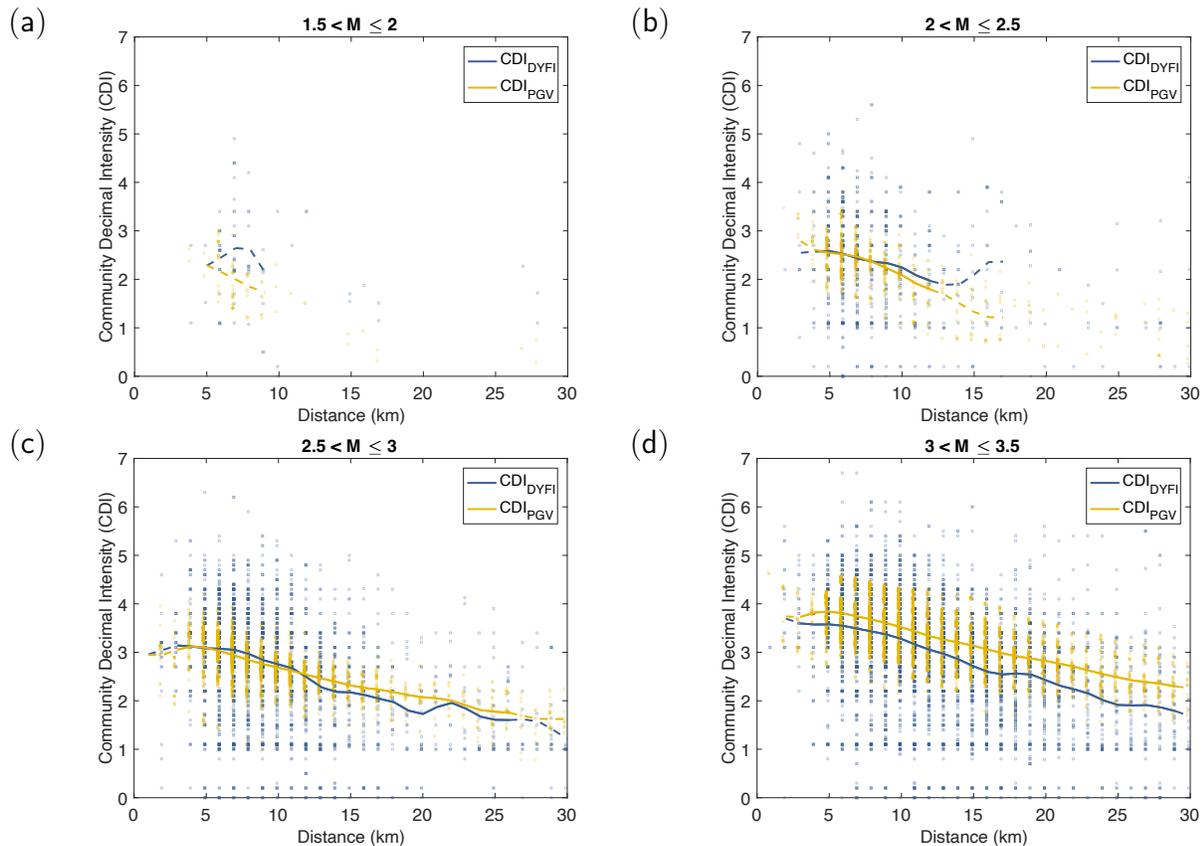


Figure 6: CDIs from DYFI reports and CDIs from PGV records plotted against hypocentral distances. Subplot titles indicate the magnitude ranges considered. Every dot is an individual observation and lines are the corresponding weighted average values. Dashed lines indicate that the average values are not well constrained.

As shown in Figures 6 and 7, CDIs from DYFI reports have much larger variance compared to those interpolated from PGVs. The large variability of CDI_{DYFI} has also been observed by past studies (Atkinson and Wald, 2007). One reason could be that during the computation of CDI_{PGV} , we did not consider the uncertainty involved in GMM and GMICE but focused on the

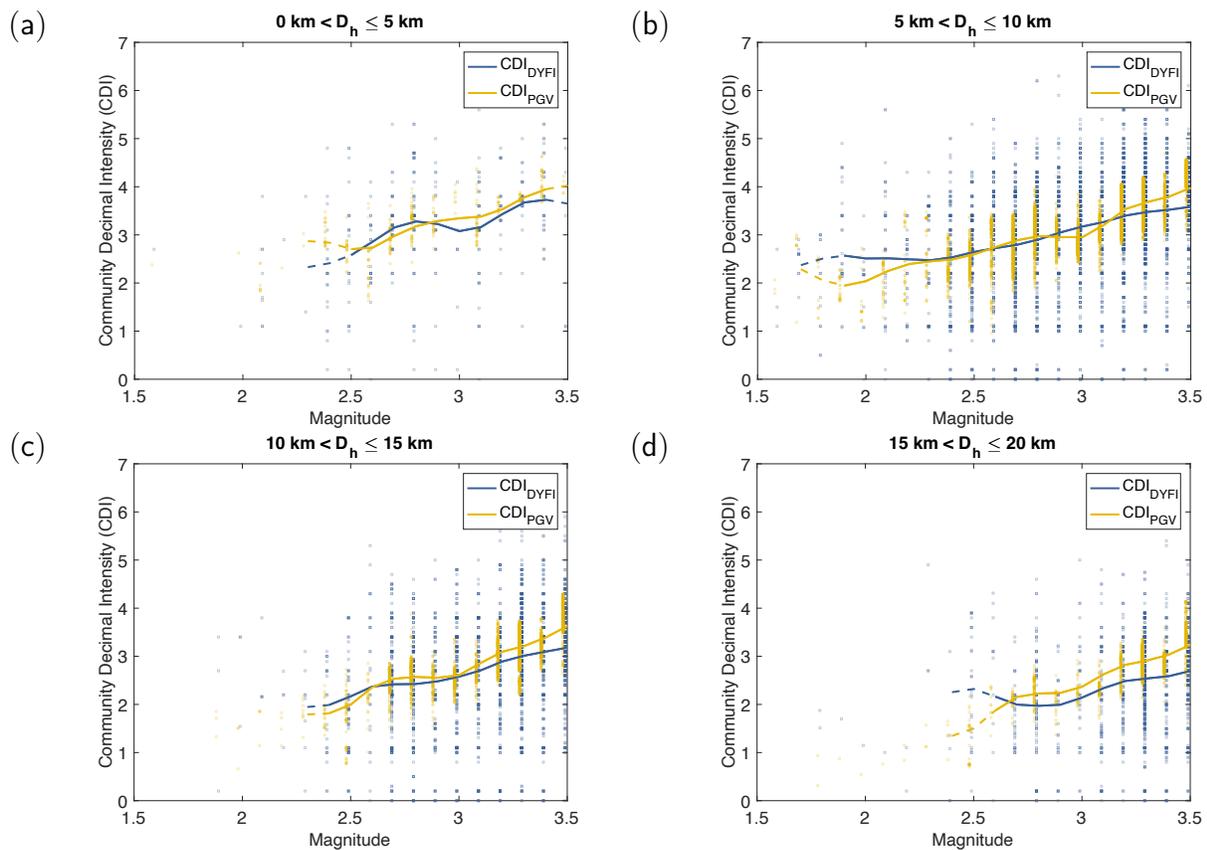


Figure 7: CDIs from DYFI reports and CDIs from PGV records plotted against magnitudes. Subplot titles indicate the hypocentral distance ranges considered. Every dot is an individual observation and lines are the corresponding weighted average values. Dashed lines indicate that the average values are not well constrained.

expected values. Figures 6 and 7 suggest that CDI_{DYFI} matches well with CDI_{PGV} especially at close distances (i.e., $D_h \leq 15$ km) and magnitudes above M2.0. Moreover, CDI_{DYFI} is slightly lower than CDI_{PGV} for $D_h > 15$ km or $M > 3.2$. The differences between CDI_{PGV} and CDI_{DYFI} become larger when the CDI values are too small to be routinely felt (e.g., $M < 2.0$ or $D_h > 20$ km) or when a limited number of DYFI reports are available. These trends are seen in Figure 8, where the difference between the weighted average values (i.e., $W(CDI_{PGV}) - W(CDI_{DYFI})$) are plotted against $W(CDI_{PGV})$ and the number of reports. The solid lines in are the moving averages:

$$\bar{x}_i = \frac{\sum_{j=i-9}^{i+10} x_j}{20} \quad (10)$$

where x_j is the j^{th} individual value in the figure, sorted in order of the x-axis parameter.

In particular, we considered magnitudes from M1.5 to M3.5 with 0.1 interval and distances from 0 to 30 km with 1 km interval. For every combination of magnitude and distance, we computed $W(CDI_{PGV})$ and $W(CDI_{DYFI})$ using Equation 11, where n is the number of individual reports available, and w_i is the weight computed from multivariate Gaussian distribution with means as the magnitude and distance of interest.

$$W(x, m, r) = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad w \sim N(\mu, \Sigma)$$

$$\mu = [m \quad r] \quad \Sigma = \begin{bmatrix} \sigma_m^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix} \quad \sigma_m = 0.1; \quad \sigma_r = 1 \quad (11)$$

We then used Equation 12 to estimate the number of reports:

$$count(m, r) = \frac{\sum_{i=1}^n w_i}{w_0} \quad (12)$$

where w_i is the weight for i^{th} observation computed from the multivariate Gaussian distribution (Equation 11) with mean as the magnitude and distance of interest, and it is normalized by w_0 , which is the weight computed at the mean. As shown in Figure 8a, the difference between CDI_{DYFI} and CDI_{PGV} is close to zero for larger CDI values (i.e., $W(CDI_{PGV}) > 2.0$), and it becomes larger at low intensities (i.e., $W(CDI_{PGV}) < 2.0$). In Figure 8b, the variance is larger when there are fewer reports, and it decreases and stabilizes for cases where more reports are collected. This is consistent with results in Worden *et al.* (2012), where they developed an exponential function to describe the decrease in the standard deviation with respect to the number of DYFI responses.

From the above results, we did not observe any unexpected discrepancy between CDIs from DYFI reports and CDIs interpolated from PGVs. Thus, we concluded that for low-amplitude felt shaking, the DYFI data we collected are reliable for evaluating IPEs.

Evaluation of intensity prediction equations

We evaluated the AW07 and AWW14 IPEs using DYFI reports available in Kansas, Oklahoma, and Texas in the past 10 years. Figures 9 and 10 summarize the comparison of intensities from two IPEs with CDIs from DYFI data. The dots are the individual records, and black lines are the weighted average using Gaussian distribution (Equation 9). Records are grouped into the different magnitude and hypocentral distance ranges, as labeled in the figures. In every subplot, the mean magnitude

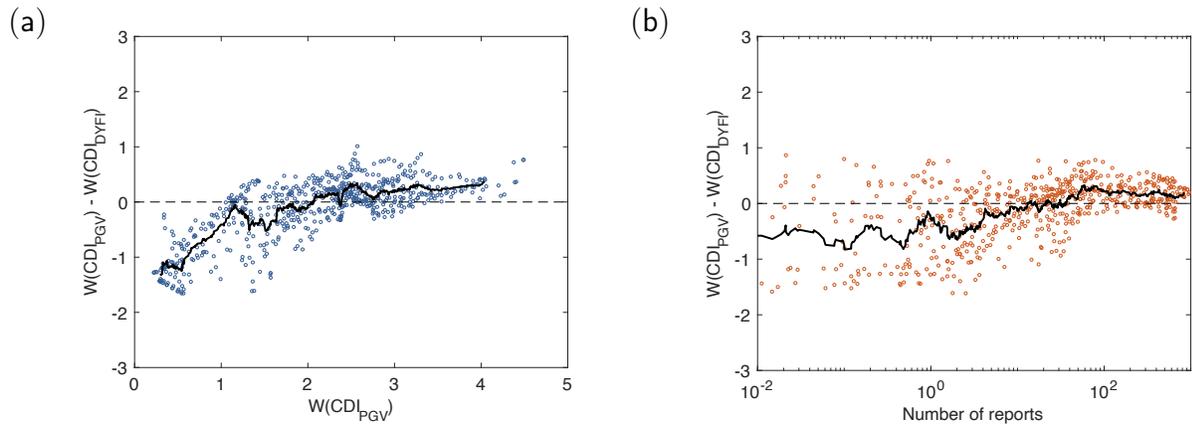


Figure 8: The difference in the weighted average values (i.e., $W(CDI_{PGV}) - W(CDI_{DYFI})$) versus (a) $W(CDI_{PGV})$ using Equation 11 and (b) the number of reports using Equation 12 for the m, r of interest. Every dot corresponds to the value for a pair of m, r of interest. The solid lines are the moving averages computed using Equation 10.

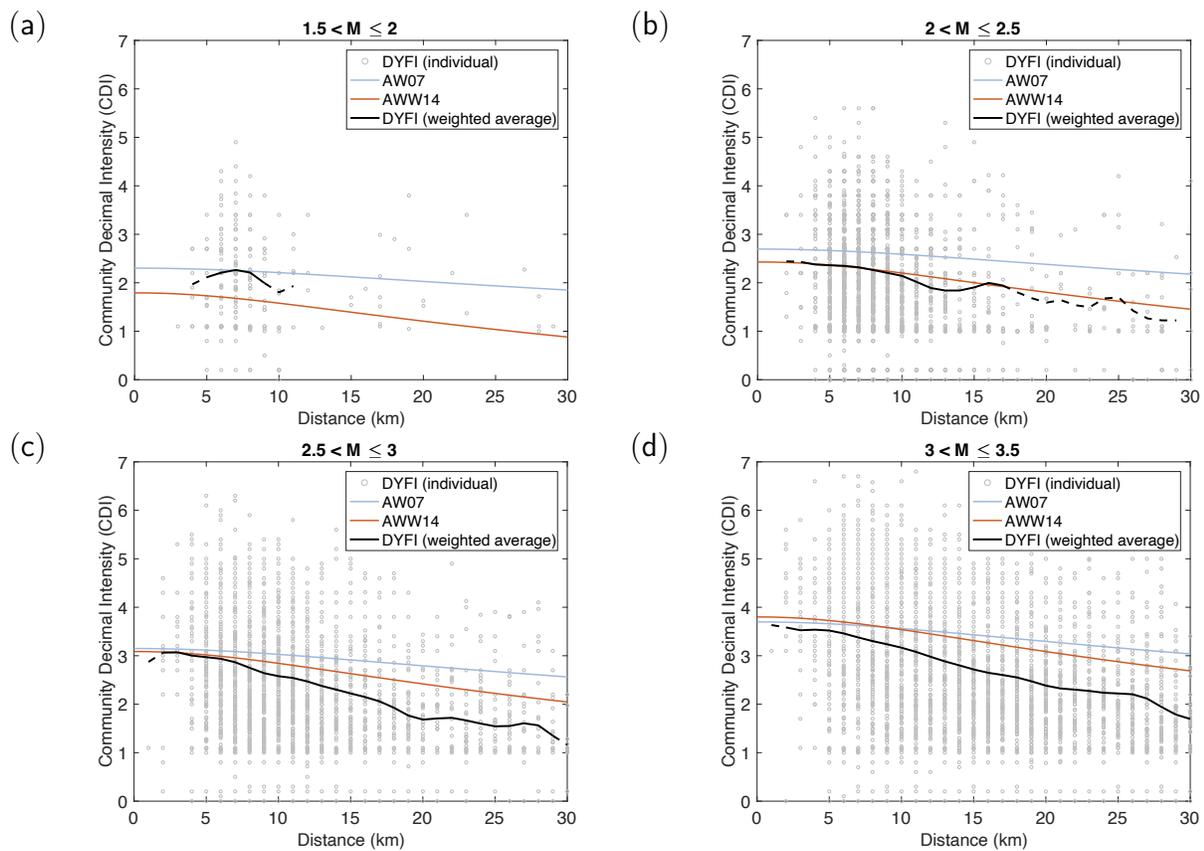


Figure 9: CDIs from DYFI reports versus hypocentral distances, plotted with IPEs (AW07 and AWW14). Every subplot has different magnitude ranges as stated. Grey dots are individual DYFI reports and black lines are the corresponding weighted average values. Black dashed lines indicate that the average values are not well constrained.

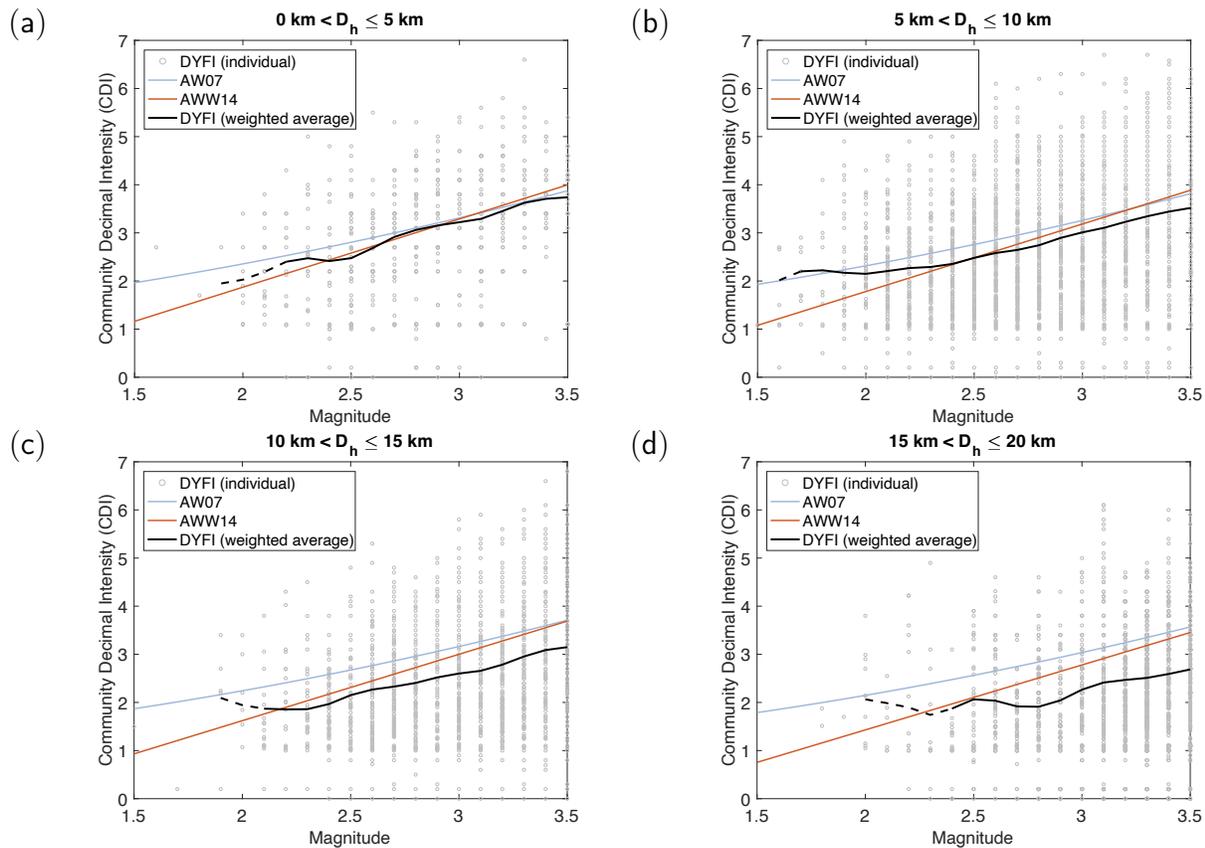


Figure 10: CDIs from DYFI reports versus magnitudes, plotted with IPEs (AW07 and AWW14). Every subplot has different hypocentral distance ranges as stated. Grey dots are individual DYFI reports and black lines are the corresponding weighted average values. Black dashed lines indicate that the average values are not well constrained.

or the mean hypocentral distance of individual reports is used to generate intensities of the two IPEs. Both figures suggest that the two IPEs match well for $D_h < 10$ km and $2.0 < M < 3.0$, and AWW14 performs better for $D_h > 10$ km.

In Figure 10, the gradient of DYFI data to magnitudes is slightly flatter compared to AWW14, closer to that of AW07, especially for $M < 2.5$ and $5 \text{ km} < D_h \leq 10 \text{ km}$. AW07 shows a flattening out at small magnitudes, though the pattern is difficult to compare to DYFI data due to the limited number of reports for $M < 2.0$. Moreover, intensities for $M < 2.0$ are so low that they are less important for the purpose of hazard analysis. In terms of attenuation with distance (Figure 9), the CDIs from DYFI reports attenuate faster than both IPEs, but they are slightly closer to AWW14. The CDIs from DYFI reports are consistently lower than both IPEs at large distances. The fast attenuation of induced earthquakes in the central U.S. has also been observed by others (Hough, 2014; Cremen *et al.*, 2017), potentially due to their shallower depths and lower stress drops compared to natural earthquakes in the western U.S. However, due to the sparse seismic station network in the central and eastern U.S., uncertainties in earthquake locations could be large. We obtained the uncertainty in focal depth, longitude, and latitude from ComCat, and estimated the uncertainty in resulting hypocentral distances. Results show an average hypocentral distance uncertainty of approximately 2 km. The distance uncertainty could affect the attenuation of CDIs, especially for close distances.

Though the IPEs do not consider soil conditions, we explored the effect of soil conditions, such as the time-averaged shear-wave velocity for the upper 30 m depth (V_{S30}) on the DYFI report intensity. We used the V_{S30} values and inverse distance weighting interpolation methods from Zalachoris *et al.* (2017) to estimate the V_{S30} values across the region. We then selected two groups of DYFI reports with well-separated V_{S30} ranges. One group is from central Oklahoma with an average V_{S30} value of 700 m/s, and the other group is from Dallas and northern Oklahoma with an average V_{S30} value of 300 m/s. Figure 11 shows the selected regions where the background color indicates V_{S30} values. Figure 12 shows the comparison of two groups of DYFI data, plotted with two IPEs. The regions with lower V_{S30} values have higher CDIs for a given magnitude and distance than those with higher V_{S30} values. This effect could be due to the amplification of shaking on the softer surficial materials or could be caused by other regional differences. If studies over a broader range of conditions confirm this effect, it suggests that current IPEs could be further refined to account for the effect of soil conditions. Moreover, Figure 12 indicated that intensities of lower V_{S30} s match better with both IPEs compared to the other group as well as mean values of the entire regions (Figures 9 and 10). Though no V_{S30} information was included in the two original IPE models, Atkinson *et al.* (2014) developed the IPE assuming an average V_{S30} of 450 m/s, corresponding to the average amplification of a NEHRP Class C site. They also stated that the predicted intensities would be lower for stiffer sites, consistent with the results in Figure 12.

Fitted IPEs based on DYFI and seismic station data

With the DYFI and seismic station data available, we generated a new IPE to describe the intensities at close distances. We considered observations with hypocentral distances between 5 and 10 km, based on typical hypocentral distances between hydraulic-fracturing-induced earthquakes and the corresponding production site (Teng and Baker, 2020). We refit an IPE to the CDI_{PGV} and CDI_{DYFI} data. Specifically, we considered similar functional forms to Atkinson and Wald (2007) and Atkinson *et al.* (2014) and refit the coefficients of models. Atkinson and Wald (2007) assumes a quadratic relationship between intensities and magnitudes:

$$\hat{y} = f(M, R) = a_0 + a_1M + a_2R + a_3\log R + a_4M\log R + a_5M^2 \quad (13)$$

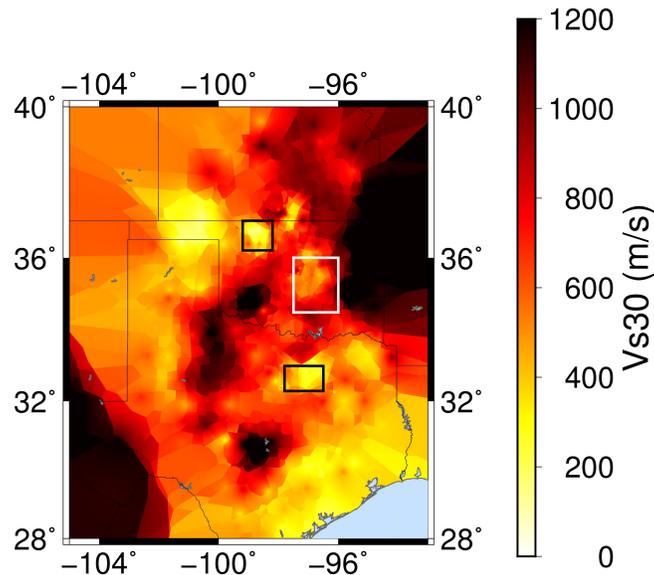


Figure 11: Selected regions with different V_{S30} ranges. DYFI reports in the black boxes have an average V_{S30} of approximately 300 m/s. DYFI reports in the white box have an average V_{S30} of approximately 700 m/s. The mapped V_{S30} values are capped at 1200 m/s for better visualization in the regions of interest.

where $R = \sqrt{D_h^2 + 17^2}$, and D_h is the hypocentral distance. We also considered the functional form from Atkinson *et al.* (2014) for the western U.S., which is same as Equation 13 without the term $a_5 M^2$, and $R = \sqrt{D_h^2 + 14^2}$ instead. Due to the narrow range of hypocentral distances considered, we left the coefficients a_3 and a_4 to be the same as the values in Atkinson and Wald (2007) and Atkinson *et al.* (2014). The remaining coefficients could be estimated by solving an optimization problem:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \text{s.t.} \quad & \frac{\partial f}{\partial M} \geq 0 \\ & \frac{\partial f}{\partial D_h} \leq 0 \end{aligned} \quad (14)$$

where y_i is the i^{th} observation, \hat{y}_i is the corresponding prediction, n is the number of observations, and the partial derivative constraints ensure that the slopes of the predictions are consistent with physical expectations.

Figure 13 summarizes the fitted curves, along with the two original IPEs and the weighted average of the observations. For the DYFI data (Figure 13a), both the linear and quadratic models are closer to AWW14, especially at $M > 2.0$. Though the linear and quadratic models differ for small magnitudes, they have very similar performance in general, with a difference between the root mean square errors less than 0.01. This is because most of the DYFI reports have magnitudes above $M2.0$, and there are limited reports available for small magnitudes. As a result, the trend is not well constrained even though the weighted average values from the observations suggest that the intensities could flatten out at small magnitudes. However, for the PGV data (Figure 13b), there are

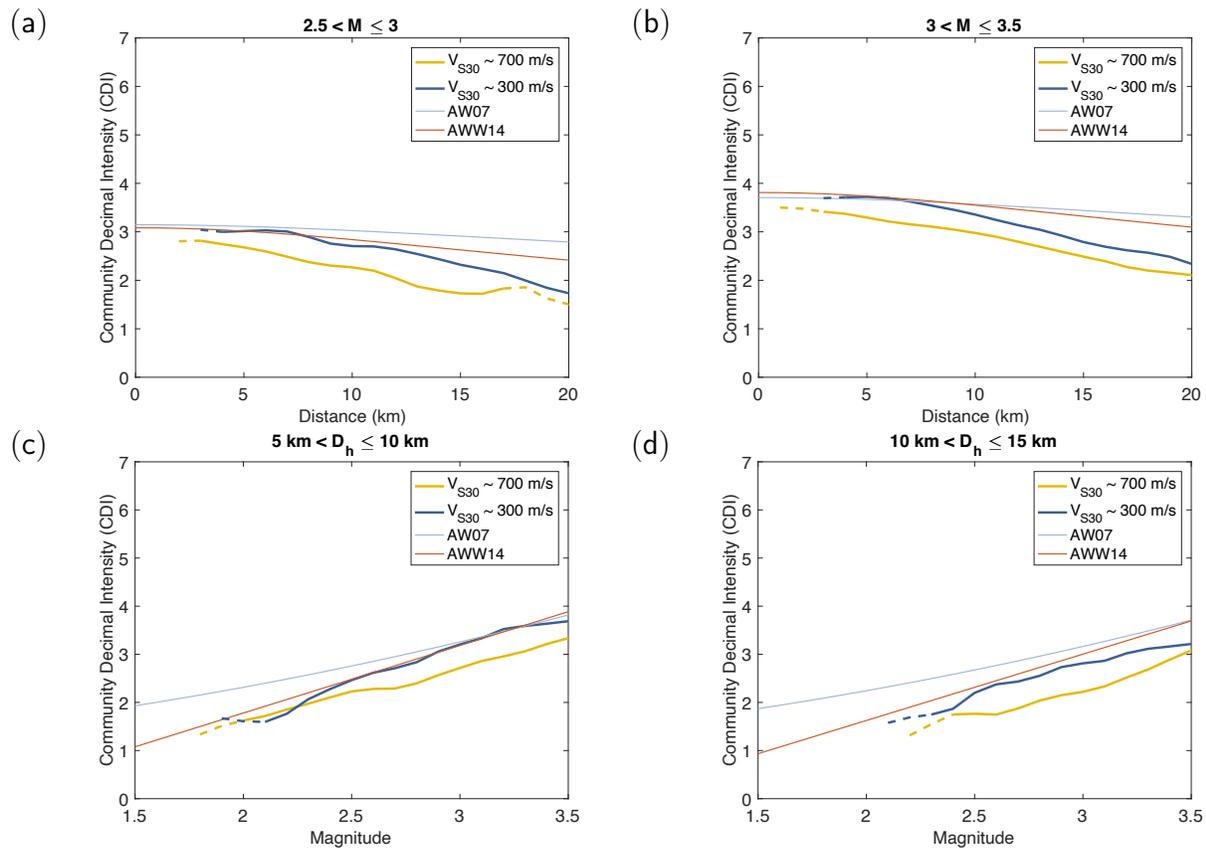


Figure 12: DYFI reports grouped by V_{S30} values, plotted against IPEs (AW07 and AWW14). (a,b) CDIs versus hypocentral distances. (c,d) CDIs versus magnitudes. Every subplot has different magnitude or hypocentral distance ranges as stated. The solid lines are the weighted averages of every group. Dashed lines indicate that the average values are not well constrained.

many more data points available at small magnitudes. Two models are nearly identical, and both suggest that intensities do not flatten out at small magnitudes. For the PGV data, both fitted models are closer to AW07 than AWW14, while the models based on DYFI data are closer to AWW14. One reason is that the DYFI data and PGV data tend to come from different earthquakes. Most of the PGV data are from northern Oklahoma and southern Kansas, whereas most of the DYFI data are from central Oklahoma and Dallas, closer to regions with high populations. Moreover, Atkinson *et al.* (2014) covered earthquakes in Oklahoma and Kansas, whereas Atkinson and Wald (2007) considered fewer earthquakes in this region.

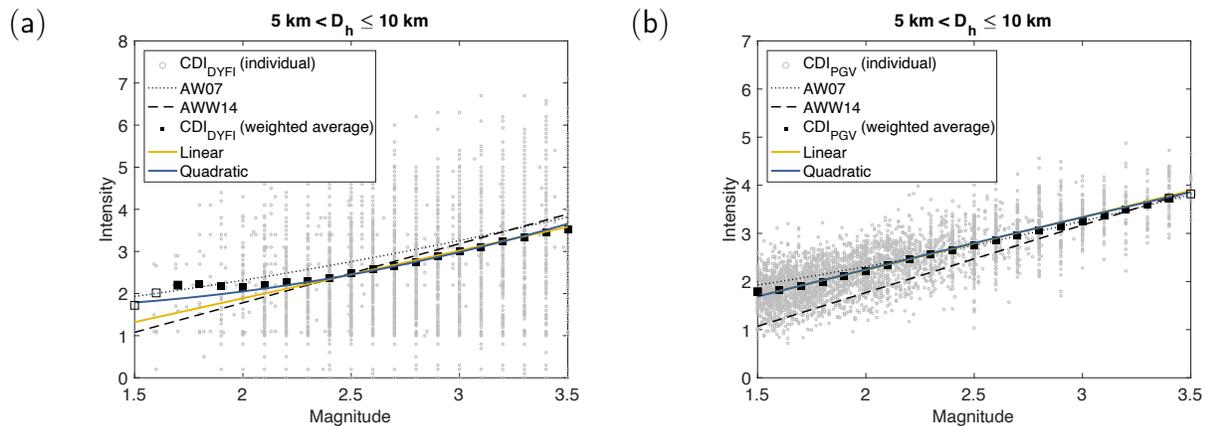


Figure 13: IPEs fitted using (a) DYFI data and (b) PGV data plotted with IPEs (AW07 and AWW14). Open squares indicate that the weighted average values are not well constrained.

Hazard disaggregation

With the fitted IPEs, we conducted hazard analysis and disaggregation for earthquakes at close distances and explored the contribution of small magnitude earthquakes to felt shaking in a situation of typical hydraulic-fracturing induced seismicity. In particular, we considered a site with a Gutenberg-Richter distribution of magnitudes between M1.5 and M6.0 and a b-value of 1.0, occurring at a hypocentral distance of 7.5 km (the same situation as in Teng and Baker, 2020). We compute the probability of a given earthquake causing $MMI > 3$, and then use disaggregation to compute the probability that an $MMI > 3$ ground motion was caused by an earthquake with $M < m$.

The fitted IPEs only cover magnitudes below M3.5, so we use Atkinson *et al.* (2014) for magnitudes above M3.5. As shown in Figure 1, the two IPEs have a negligible difference for magnitudes between M3.5 and M6.0. For the PGV data, we only considered the linear model since the two fitted IPEs are almost identical. For illustration, we assumed a prediction standard deviation of 0.5 for the fitted IPEs, which was the same as Atkinson *et al.* (2014). The disaggregation was computed using the following equation:

$$P(M < m | MMI > mmi) = \frac{P(M < m, MMI > mmi)}{P(MMI > mmi)} = \frac{\sum_{m_i < m} P(MMI > mmi | m_i) P(m_i)}{\sum_{m_i} P(MMI > mmi | m_i) P(m_i)} \quad (15)$$

Figure 14 summarizes disaggregation results from three fitted IPEs together with AW07 and AWW14. In general, the linear IPE from DYFI data matches well with AWW14, whereas the IPE from PGV

data is closer to AW07. The fitted IPE from DYFI data with a quadratic format is between AWW14 and AW07, and it has higher contributions for small magnitudes (i.e., $M < 2$) than all IPEs. Overall, though the contribution of $M < 3$ earthquakes to $MMI > 3$ ground motions could vary by almost 20% for different models, all models suggest a significant contribution of small magnitude earthquakes on the hazard of felt shaking (i.e., $> 40\%$), which is consistent with the observations in Teng and Baker (2020).

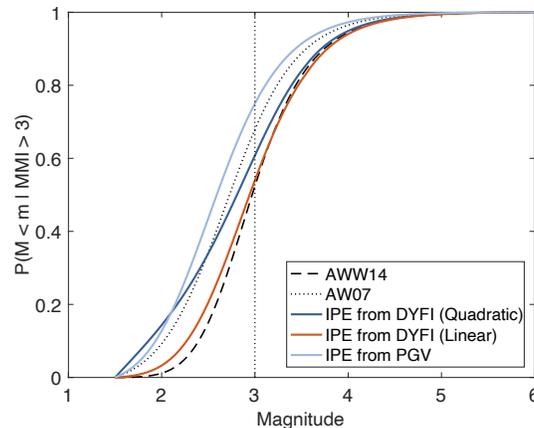


Figure 14: Hazard disaggregation using fitted IPEs, AW07 and AWW14.

Moreover, we observed the significant variation of DYFI data and explored its impact on hazard analysis. We re-conducted hazard disaggregation using the standard deviation of residuals calculated from observations. For the DYFI data, the standard deviation of residuals for both fitted IPEs was 1.0. Figure 15 summarizes the disaggregation result. We also included the results with a standard deviation of 0.5 for comparison. As shown in Figure 15, the contribution of small-magnitude earthquakes increases significantly due to the large variation of DYFI reports. Such large variation, especially at lower intensity levels, has been observed in the past studies (Atkinson and Wald, 2007; Hough, 2013; Cremen *et al.*, 2017). Moreover, individual responses are highly variable, but the resulting trends from large numbers of responses are robust (Wald *et al.*, 2012). For example, the standard deviation for a location with 10 responses is about 0.25 (Wald *et al.*, 2012), in contrast to the standard deviation of about 1 in this study. In this study, we considered DYFI responses aggregated in 1 km grid. Since most grids (76%) have only one response, the standard deviation computed could be an approximation of standard deviation for a single response. When considering the occurrence of felt shaking from induced earthquakes, it remains an open question what metric of shaking is most important. Individual experiences of felt shaking may be important if they cause discomfort, in which case the individual response statistics are important. Alternatively, average felt intensities may be deemed of greater importance, in which case the statistics using lower standard deviations may be appropriate. In the absence of a clear preferred approach, we provide both the Figure 14 and Figure 15 results.

Conclusion

This study evaluated the performance of current IPEs (Atkinson and Wald, 2007; Atkinson *et al.*, 2014) on small-magnitude-induced earthquakes at close distances based on both DYFI reports and PGV recordings. We also performed hazard disaggregation using the observed data to study the

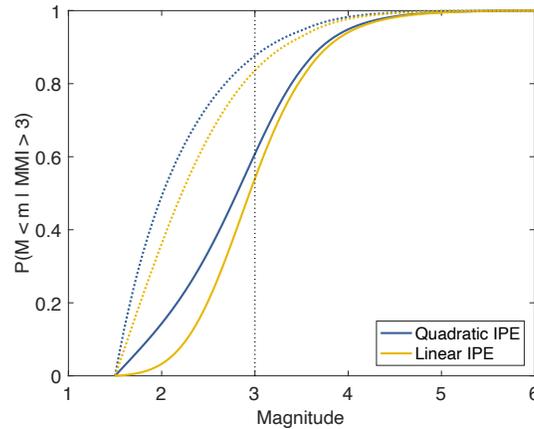


Figure 15: Hazard disaggregation using fitted IPEs, with an assumed standard error of 0.5 (solid line) and a calculated standard error for individual responses of 1.0 (dotted line).

importance of small-magnitude earthquakes on felt shaking.

DYFI data are statistically robust due to the large number of responses available and are often used as the ground truth when evaluating and developing IPEs. However, for regions with low populations and for small-magnitude earthquakes, DYFI data may be less reliable. Thus, we first evaluated the DYFI data by comparing it with intensities interpolated from ground motion records (i.e., PGVs). We considered earthquakes with both DYFI reports and ground motion records available. Results suggest that DYFI data match well with the intensities interpolated from PGVs for $M > 2$ and $D_h < 15$ km. These ranges correspond to situations where the intensity is large enough to be felt (i.e., $MMI > 2$), and thus a larger number of reports are available. The difference between the DYFI and PGV intensities becomes larger when they are too low to be consistently felt, so there are limited data points to constrain the trend. Overall, we did not observe any unexpected discrepancies between intensities from DYFI reports and intensities interpolated from PGVs. For hazard analysis on felt shaking, the DYFI data are reliable as a ground truth for evaluating IPEs.

We evaluated IPEs by Atkinson and Wald (2007) and Atkinson *et al.* (2014), with a focus on $D_h < 15$ km and $2.0 \leq M \leq 3.0$, which correspond to ranges that were not well studied before but could still contribute to felt shaking. Overall, both match well with DYFI data for $D_h < 10$ km and $2.0 \leq M \leq 3.0$, but AWW14 performs better for $D_h > 10$ km. AW07 and AWW14 differ most at small magnitudes ($M < 2$) and larger distances ($D_h > 10$ km), where the former is consistently higher. Results show that intensities from DYFI reports attenuate faster than both IPEs, especially for distances beyond 10 km. The observed intensities are consistently lower than both IPEs at large distances but closer to AWW14. We also explored the effect of V_{S30} on intensities and observed that intensities at sites with $V_{S30} \sim 300$ m/s are consistently higher than intensities at sites with large V_{S30} values (700 m/s). If studies over a broader range of conditions confirm this effect, it suggests that current IPEs could be further refined to account for the effect of soil conditions.

We fitted separate IPEs to CDIs from DYFI reports and CDIs interpolated from PGV records. We used similar functional forms as AW07 and AWW14. In general, the IPEs fitted to PGV data are closer to AW07, whereas the IPEs on DYFI reports match better with AWW14 for $M > 2$. This discrepancy could be due to the difference in the earthquake data used: (1) Most of the PGV records in this study are from northern Oklahoma and southern Kansas, whereas the majority of the DYFI reports are from central Oklahoma and Dallas; (2) Atkinson *et al.* (2014) covered earthquakes in Oklahoma and Kansas, whereas Atkinson and Wald (2007) had fewer data in this region.

We conducted hazard disaggregation for earthquakes at close distances and estimated the contribution of small magnitude earthquakes to felt shaking using the fitted IPEs, AWW14, and AW07. Results show that though the IPEs differ from each other, they all suggest a significant contribution ($\geq 40\%$) from small-magnitude earthquakes to felt shaking hazard. We also observed large variations in individual DYFI reports. When this variation is accounted for, it further increased the contribution of small-magnitude earthquakes on felt shaking.

Data and Resources

The “Did You Feel It” records were provided by Vince Quitarano in April 2020. The corresponding earthquake information is available via the U.S. Geological Survey ComCat database (<https://earthquake.usgs.gov/data/comcat/>; Guy *et al.* 2015). It was collected using the USGS ComCat libcomcat library (<https://github.com/usgs/libcomcat>, last accessed April 2021). Ground motion records were collected from Incorporated Research Institutions for Seismology (IRIS) Data Services (<http://ds.iris.edu/ds/nodes/dmc/>, last accessed May 2020) using the program Standing Order of Data (SOD, <http://www.seis.sc.edu/sod/>, last accessed May 2020). The population data were collected from SimplyAnalytics (<https://simplyanalytics.com/>, last accessed June 2020).

Acknowledgement

We thank Vince Quitarano for help with accessing “Did You Feel It” records. We thank William Ellsworth, Gail Atkinson, the anonymous reviewer and the editor for insightful comments that significantly improved the article. Funding for this work came from the Stanford Center for Induced and Triggered Seismicity. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Ahmadzadeh, S., G. J. Doloei, and H. Zafarani (2020). New intensity prediction equation for Iran, *J. Seismol.* **24**(1), 23–35.
- Allen, T. I., D. J. Wald, and C. B. Worden (2012). Intensity attenuation for active crustal regions, *J. Seismol.* **16**(3), 409–433.
- 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**(2A), 981–992.
- Atkinson, G. M., and S. I. Kaka (2007). Relationships between felt intensity and instrumental ground motion in the central United States and California, *Bull. Seismol. Soc. Am.* **97**(2), 497–510.
- Atkinson, G. M., D. Wald, C. B. Worden, and V. Quitarano (2018). The intensity signature of induced seismicity, *Bull. Seismol. Soc. Am.* **108**(3A), 1080–1086.
- 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**(3), 362–368.

Teng, G., Baker, J. W., and Wald, D.J. (2021). “Evaluation of intensity prediction equations (IPEs) for small-magnitude earthquakes.” *Bulletin of the Seismological Society of America*, 112(1), 316–330. <https://doi.org/10.1785/0120210150>

- Atkinson, G. M., C. B. Worden, and D. J. Wald (2014). Intensity prediction equations for North America, *Bull. Seismol. Soc. Am.* **104**(6), 3084–3093.
- Bakun, W. u., and C. Wentworth (1997). Estimating earthquake location and magnitude from seismic intensity data, *Bull. Seismol. Soc. Am.* **87**(6), 1502–1521.
- Boatwright, J., and E. Phillips (2017). Exploiting the demographics of “Did You Feel It?” responses to estimate the felt area of moderate earthquakes in California, *Seismol. Res. Lett.* **88**(2A), 335–341.
- Boatwright, J., K. Thywissen, and L. C. Seekins (2001). Correlation of ground motion and intensity for the 17 January 1994 Northridge, California, earthquake, *Bull. Seismol. Soc. Am.* **91**(4), 739–752.
- Boore, D. M., and T. Kishida (2017). Relations between some horizontal-component ground-motion intensity measures used in practice, *Bull. Seismol. Soc. Am.* **107**(1), 334–343.
- Caprio, M., B. Tarigan, C. B. Worden, S. Wiemer, and D. J. Wald (2015). Ground motion to intensity conversion equations (GMICEs): A global relationship and evaluation of regional dependency, *Bull. Seismol. Soc. Am.* **105**(3), 1476–1490.
- Cremen, G., A. Gupta, and J. Baker (2017). Evaluation of ground motion intensities from induced earthquakes using “Did You Feel It?” data, in *16th World Conf. on Earthq. Eng.*
- Dowrick, D., and D. Rhoades (2005). Revised models for attenuation of modified Mercalli intensity in New Zealand earthquakes, *Bull. New Zeal. Soc. Earthquake Eng.* **38**(4), 185–214.
- Ellsworth, W. L. (2013). Injection-induced earthquakes, *Science* **341**(6142), 1225942.
- Guy, M. R., J. M. Patton, J. Fee, M. Hearne, E. M. Martinez, D. C. Ketchum, C. B. Worden, V. Quitoriano, E. J. Hunter, G. M. Smoczyk, *et al.* (2015). *National Earthquake Information Center systems overview and integration*, US Department of the Interior, US Geological Survey.
- Hough, S. E. (2013). Spatial variability of “Did You Feel It?” intensity data: Insights into sampling biases in historical earthquake intensity distributions, *Bull. Seismol. Soc. Am.* **103**(5), 2767–2781.
- Hough, S. E. (2014). Shaking from injection-induced earthquakes in the central and eastern United States, *Bull. Seismol. Soc. Am.* **104**(5), 2619–2626.
- Le Goff, B., J. F. Borges, and M. Bezzeghoud (2014). Intensity-distance attenuation laws for the Portugal mainland using intensity data points, *Geophys. J. Int.* **199**(2), 1278–1285.
- Minson, S. E., A. S. Baltay, E. S. Cochran, S. K. McBride, and K. R. Milner (2021). Shaking is almost always a surprise: The earthquakes that produce significant ground motion, *Seismol. Soc. Am.* **92**(1), 460–468.
- Park, J., P. Bazzurro, J. W. Baker, *et al.* (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, *Applications of statistics and probability in civil engineering* pp. 1–8.
- Rigsby, C., R. B. Herrmann, and H. Benz (2014). An Investigation of MbLg Versus Mw for Eastern North America, *Seismol. Res. Lett.* **85**(3), 625–630.

Teng, G., Baker, J. W., and Wald, D.J. (2021). “Evaluation of intensity prediction equations (IPEs) for small-magnitude earthquakes.” *Bulletin of the Seismological Society of America*, 112(1), 316–330. <https://doi.org/10.1785/0120210150>

Sokolov, V., and F. Wenzel (2013). Further analysis of the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA and PGV data from the K-NET and the KiK-net (Japan) networks, *Bull. Earthq. Eng.* **11**(6), 1909–1926.

Teng, G., and J. W. Baker (2020). Short-term probabilistic hazard assessment in regions of induced seismicity, *Bull. Seismol. Soc. Am.* **110**(5), 2441–2453.

Trugman, D. T., S. L. Dougherty, E. S. Cochran, and P. M. Shearer (2017). Source spectral properties of small to moderate earthquakes in southern Kansas, *J. Geophys. Res.* **122**(10), 8021–8034.

Wald, D. J., V. Quitoriano, T. H. Heaton, and H. Kanamori (1999a). Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California, *Earthq. Spectra* **15**(3), 557–564.

Wald, D. J., V. Quitoriano, T. H. Heaton, H. Kanamori, C. W. Scrivner, and C. B. Worden (1999b). TriNet “ShakeMaps”: Rapid generation of peak ground motion and intensity maps for earthquakes in southern California, *Earthq. Spectra* **15**(3), 537–555.

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**(6).

Wald, D. J., B. C. Worden, V. Quitoriano, and K. L. Pankow (2005). ShakeMap manual: technical manual, user’s guide, and software guide, Technical report.

Wood, H. O., and F. Neumann (1931). Modified Mercalli intensity scale of 1931, *Bull. Seismol. Soc. Am.* **21**(4), 277–283.

Worden, C., M. Gerstenberger, D. Rhoades, and D. Wald (2012). Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California, *Bull. Seismol. Soc. Am.* **102**(1), 204–221.

Wu, Y.-M., T.-l. Teng, T.-C. Shin, and N.-C. Hsiao (2003). Relationship between peak ground acceleration, peak ground velocity, and intensity in Taiwan, *Bull. Seismol. Soc. Am.* **93**(1), 386–396.

Zalachoris, G., E. M. Rathje, and J. G. Paine (2017). VS 30 characterization of Texas, Oklahoma, and Kansas using the P-wave seismogram method, *Earthq. Spectra* **33**(3), 943–961.

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.