Using Simulated Ground Motions to Constrain Near-Source Ground Motion Prediction Equations in Areas Experiencing Induced Seismicity

Samuel A. Bydlon[1], Abhineet Gupta[2], and Eric M. Dunham[1]

1. Department of Geophysics, Stanford University, Stanford, CA

2. Department of Civil and Environmental Engineering, Stanford University, Stanford, CA

Abstract

Recent increases in seismic activity in historically quiescent areas such as Oklahoma, Texas, and Arkansas have spurred the need for investigation into expected ground motions associated with these seismic sources. The neoteric nature of this seismicity increase corresponds to a scarcity of ground motion recordings within ∼20 km of earthquakes M_w 3.0 and greater. To aid the effort of constraining near-source ground motion prediction equations (GMPEs) associated with induced seismicity, we develop a framework for integration of synthetic ground motion data from simulated earthquakes into the GMPE development process. We demonstrate this framework by developing a GMPE for a target region encompassing north-central Oklahoma and south-central Kansas. We first gather a catalog of recorded ground mo-

tions from M_w 3-4 earthquakes that occurred in the target region. Using constraints on the region's material structure, including well log data that provides insight into the characteristics of shallow sedimentary layers, we perform point-source simulations intended to mimic a selection of recorded earthquakes from the target region. Simulated earthquake sources are constrained by available moment tensors and locations. Once we determine that our simulations produce realistic ground motions, we combine recorded and synthetic ground motion data to produce a composite ground motion catalog. We use this composite catalog to develop a regionally-specific GMPE for our target region. This framework can be exported to other regions where near-source ground motion data are sparse and can be used to improve constraints on near-source GMPEs, which could directly benefit seismic hazard estimates.
Keywords: Ground motions, induced seismicity, simulation

Introduction

Over the last 10-15 years, parts of the central and eastern United States, such as Oklahoma, Kansas, and Texas, have experienced a dramatic increase in earthquake rates (Ellsworth, 2013). A number of authors have linked this increased seismic activity to wastewater disposal associated with oil and gas

operations (Walsh and Zoback, 2015; Kim, 2013; Keranen et al., 2013, 2014; Frohlich et al., 2014; Weingarten et al., 2015). Such earthquakes are often referred to as "induced" or "triggered" events, depending on the human-associated change in stress levels as a fraction of the ambient shear stress level acting on a fault, and in the central US have been as large as M_w 5.8 (near Pawnee, OK, on 3 September 2016). Since earthquakes in these areas occur at shallow depths, ground motions at the surface could be greater than those of deeper events for similar epicenter distances and magnitudes. Some induced events occur near populated areas (such as Oklahoma City) or critical infrastructure (such as the large and complex pipeline crossroads and crude oil storage facilities in Cushing, OK), prompting demand for well-constrained ground motion prediction equations (GMPEs), which express expected ground motion intensity measures, such as peak ground velocity (PGV), peak ground acceleration (PGA), and spectral accelerations (SA) as functions of magnitude and source-to-site distance. GMPEs are critical elements of seismic hazard evaluation and prediction, including in the national seismic hazard maps produced by the USGS (Petersen et al., 2014).

The most robust regional GMPEs are developed via regression of recorded ground motion data. However, for magnitudes and source-to-site distances

of engineering interest, only well-instrumented areas such as California have enough data to properly constrain GMPEs, especially at short source-to-site distances. Even in these areas, data scarcity is a problem for large events, such as earthquakes M_w 7+. As described in Yenier and Atkinson (2015), there are several different approaches to GMPE development in data-poor areas. One approach is to use ground motion simulations in combination with regional parameters that describe site, source, and path effects to generate synthetic ground motion data. The complexity of these simulations can vary widely, from simple stochastic point sources in a homogeneous material structure to finite-fault sources in regionally-specific heterogeneous material structures (Atkinson and Boore, 1995; Toro et al., 1997; Atkinson and Silva, 2000; Silva et al., 2002; Atkinson and Boore, 2006; Frankel, 2009). This paper presents a method for constraining GMPEs in data-poor areas using ground motion simulations. A second approach, called the hybrid empirical method, uses GMPEs from data-rich (host) regions tuned for use in a data-poor (target) region by relationships between stochastic simulations of the host and target regions (Campbell, 2003; Pezeshk et al., 2011). A third approach, called the referenced empirical approach, uses GMPEs from host regions tuned for use in a target region by relationships between predictions

of empirically-derived GMPEs of the host and observed motions in the target regions (Atkinson, 2008, 2010; Atkinson and Motazedian, 2013; Hassani and Atkinson, 2015). Most recently, Yenier and Atkinson (2015) strategically combined elements of all three of the aforementioned approaches to generate GMPEs that can be regionally adjusted by tuning just a handful of parameters and applied the methodology to a target region that covered the central and eastern US.

The neoteric nature of the seismicity increase in the central US corresponds to a scarcity of ground motion recordings within ∼50 km of earthquakes M_w 3 and greater, with increasing scarcity at larger magnitudes (Gupta et al., 2017). As expected, data scarcity is most extreme at the closest hypocentral distances, namely less than 10 km. Therefore, GMPEs for the central US derived from real ground motion recordings suffer from poor constraints. Atkinson (2015) produced GMPEs for small-to-moderate earthquakes at short hypocentral distances (defined as less than 40 km) for application to induced seismicity hazards. These GMPEs were constructed using data from the Next Generation Attenuation-West project (Ancheta et al., 2014), which is primarily comprised of ground motions from natural earthquakes in active tectonic settings, and the assumption that ground mo-

tions from induced events will be comparable to those of natural earthquakes of the same magnitude and hypocentral distance (Atkinson, 2015). This assumption has been shown to be reasonable in areas experiencing induced seismicity related to geothermal activity and gas extraction (Douglas et al., 2013; Edwards and Douglas, 2014).

In this study, we develop ground motion prediction equations for M_w 3-4 earthquakes for the northern Oklahoma and southern Kansas target area using a combination of recorded and simulated ground motion data. We focus on small earthquakes in this study, which is the first part of a larger effort to construct GMPEs for a wide range of magnitudes. The small events considered in this study can be described as point moment tensor sources. This allows us to focus mainly on wave propagation effects and the role of the assumed material structure in controlling ground motions. This assumption has been made in other studies that employ point moment tensor sources in ground motion simulations (Yenier and Atkinson, 2014; Atkinson, 2015). Specifically, Atkinson (2015) found that assuming an effective depth term that accounts for near-source distance saturation effects between 1 and 3 km for earthquakes as small as M_w 4 did not significantly affect the fit of a GMPE. The next step in our overall effort will extend our ground motion

simulations and GMPEs to earthquakes M_w >4; however, we will employ finite-fault sources to simulate such earthquakes.

To establish confidence that our simulations are producing realistic ground motions, we perform validation exercises in which we simulate earthquakes by mimicking (to the best of our abilities) the conditions of real, recorded events (i.e., depth, focal mechanism, magnitude, native material structure, etc.) and comparing the synthetic data to actual recordings. We then simulate a suite of hypothetical earthquakes in the same material structure used in the validation exercises to generate a synthetic ground motion catalog. We catalog publicly available recorded ground motion data for the target region and verify that the properties of our synthetic catalog reflect those of actual recorded data at distances where there is sufficient recorded data.

After combining the recorded and synthetic catalogs, we use a regression equation inspired by and similar to that of Shahjouei and Pezeshk (2016) to generate ground motion prediction equations for small (M_w 3-4) earthquakes in the Oklahoma/Kansas area for several key intensity measures.

Target Region and Associated Ground Motion Data

This study focuses on an area encompassing central Oklahoma, north-central Oklahoma, and south-central Kansas. This target region has been identified as an area of seismic significance due to a marked increase in the rate of earthquakes starting around 2009 (Walsh and Zoback, 2015). We start with the ground motion database for the central and eastern United States collected and processed as per Gupta et al. (2017). This data was collected via IRIS data services (http://ds.347iris.edu/ds/nodes/dmc/) using the SOD interface. Earthquake magnitudes, locations, and depths were updated to those contained in the USGS ANSS Composite Catalog (http://earthquake.usgs.gov/data/comcat/) The instruments that collected these measurements had a range of sampling rates, the lowest being 40 Hz. Therefore, the data was filtered using a 4th-order acausal Butterworth filter with low and high-pass frequencies of 0.3 and 20 Hz, respectively. This filtering could lead to an underestimation of PGA, but for the periods of interest in this study (up to $\sim$5 Hz) spectral values of the filtered and unfiltered data do not differ significantly (Gupta et al., 2017)

We extract the subset of ground motion records associated with earthquakes with epicenters between 35°-38° N latitude, 96°-99.5° W longitude,

and magnitudes M_w 3-4. The earliest ground motion record in this subset is 22 November 2004, although the majority of records are for earthquakes occurring during 2009 and later. The most recent ground motion record in this subset is from 31 December 2015. The catalog consists of ground motion intensity measures from 22,374 ground motion records associated with 1,692 unique earthquakes. Figure 1 shows a map of the earthquake epicenters in the target region catalog. Figure 2 shows ground motion intensity data (PGV, PSA(T=0.2s)) from the target region catalog as a function of hypocentral distance. For reference, M_w 3.5 GMPEs from Atkinson (2015) are also shown. The notable decrease in the density of ground motion recordings at near-source (<10 km) distances motivates our use of ground motion simulations as a proxy for actual ground motion recordings.

Target Area Material Structure

We simulate ground motions by propagating seismic waves through heterogeneous material structures. In this study, we attempt to build realistic material structures (i.e., P and S-wave speeds, density) by integrating data from well logs from central and northern Oklahoma. Well logs were provided to us by oil and gas companies with operations in the area. The well logs are

composite measurements, meaning that more than 20 individual wells spanning central OK were gathered and combined to generate statistics on the velocity structure of the region. We were not provided separate data from each well. Figure 3 shows an example of statistics from a single formation in the form the data was received. We believe that this composite log provides high quality constraints on formation thicknesses and average wave speeds and densities. We are less comfortable with the implied constraints on standard deviations of fluctuations provided by the composite logs, and suspect that by combining multiple logs into a single set of univariate statistics for each formation, there is a good chance of overestimating standard deviations. We do perform simulations including small-scale heterogeneity constrained by these well logs, but more for the general purpose of understanding the effects of small-scale heterogeneity and wave scattering on ground motions rather than as a mechanism to improve our ability to produce realistic ground motions for earthquakes in the target region.

The primary feature of the target region's geology is the contact between the igneous basement and overlying sedimentary layers. The depth of this contact over the target region ranges is often estimated at ∼2-3 km (Keranen et al., 2013, 2014). Our obtained well logs indicate the depth of this contact at

154 ∼2.5 km in our target region. The sedimentary layers include the Arbuckle
155 formation and comprise the target formations for much of the wastewater
156 injection in the area.

157 We incorporate attenuation in our simulations via the S- and P-wave
158 quality factors Q_S and Q_P through the relationships $Q_S = 100V_S$ and $Q_P =$
159 $2Q_S$, where V_S is S-wave velocity in km/s. Over the shallowest 10 km in our
160 target region velocity structure, Q_S ranges from ∼130-360 and Q_P ranges
161 from ∼260-720. This method of linking attenuation to local S-wave speeds
162 has been used in ground motion simulations such as Olsen (2000) and Olsen
163 et al. (2009), although in these studies the relationship between Q_S and V_S
164 was chosen to be $Q_S = 50V_S$ for simulations in the high-attenuation Los
165 Angeles basin. Estimates of Q for the central US tend to be higher than
166 for the western US (detailed below), indicating that attenuation is lower in
167 the central US than in the western US. We therefore choose to represent the
168 attenuation structure of our target region using higher quality factors. We
169 find that using $Q_S = 100V_S$ does a good job producing simulated ground
170 motions that attenuate in a manner similar to actual data from our target
171 region.

172 Benz et al. (1997) reported a frequency-independent value of $Q = 1291$

for the central US based on observations between 1.5 and 7 Hz. Erickson et al. (2004) described Q in the central US as $Q(f) = 640(\pm 225)f^{0.344(\pm 0.22)}$, where f is the frequency of interest. We note that these studies based Q estimates using ground motion data at distances >150 km and report values of crustal Q, whereas we attempt to separate Q_S and Q_P and are concerned with shallow depths as compared to the entire crust below out target region. Since attenuation tends to increase at shallower depths, we infer that for our purposes it is reasonable to assume values of Q_S and Q_P lower than the values of crustal Q reported in these studies, but that they can be viewed as an upper bound for our target region.

Our material structures have two primary components, a 1D "backbone" material structure and a 3D small-scale heterogeneities described by von Karman power spectral density functions. We aim to assess the sensitivity of our ground motion simulations to choices made when constructing material structures by focusing on the effects of using different 1D structures and the inclusion or exclusion of 3D small-scale heterogeneities.

We focus on two 1D structures in our ground motion simulations (Figure 4), the first of which is the material structure used to locate earthquakes in Keranen et al. (2014). As noted in Keranen et al. (2014), this material struc-

ture is well-constrained between ∼2.7 and ∼15 km depth. In this material structure, ∼2.7 km depth is treated as the boundary between fast-velocity igneous basement below and a low-velocity sedimentary layer above. Analysis of well logs from central/north Oklahoma indicate that the structure of the upper ∼2.7 km is more complicated. To account for this complexity, we construct a material structure that starts with the 1D profile from Keranen et al. (2014), but replaces the upper ∼2.7 km with velocity data obtained from well logs. A notable feature of the well log data is that the Arbuckle formation, one of the formations into which large volumes of high-salinity produced water are disposed (Walsh and Zoback, 2015), has a larger average P- and S-wave speed than the basement rock.

Small-scale heterogeneities in material properties contribute to wave scattering in a manner dependent on the wavelengths of the waves under consideration and the characteristics of the heterogeneities themselves, such as their size, anisotropy ratio, and amplitude contrast relative to surrounding material (Frankel and Clayton, 1986; Hartzell et al., 2010; Imperatori and Mai, 2013; Bydlon and Dunham, 2015). Measurements of small-scale heterogeneity at resolutions of typical wave propagation simulations (<100 m) over large areas are uncommon, especially in areas with a relatively short

history of regular seismic activity such as the central US; therefore, such heterogeneity is often represented in a stochastic fashion. We generate stochastic small-scale heterogeneity using the von Karman power spectral density function (PSDF) since it has been shown to be well-suited for statistically characterizing crustal material property fluctuations (Frankel and Clayton, 1986). The von Karman PSDF has the form

$$P(k) = \frac{4\pi\sigma^2\nu\sqrt{a_x^2 + a_y^2 + a_z^2}}{(1 + \sqrt{k_x^2 a_x^2 + k_y^2 a_y^2 + k_z^2 a_z^2})^{\nu+1}}. \quad (1)$$

where k_x, k_y, and k_z are wavenumber components for each dimension, a_x, a_y, a_z are the correlation lengths in each dimension, ν is the Hurst exponent, and σ is the standard deviation of the fluctuations, normalized by the mean value of that property. The von Karman PSDF is commonly used in ground motion simulations to describe small-scale fluctuations in material properties (Frankel and Clayton, 1986; Hartzell et al., 2010; Imperatori and Mai, 2013; Bydlon and Dunham, 2015). These parameters can be constrained by inverting sonic logs (Holliger, 1996, 1997; Savran and Olsen, 2016; Kruiver et al., 2017), but typically a large number of logs is needed to estimate the lateral variation of these fluctuations. Often, these parameters are estimated by parameter space studies of wave propagation in heterogeneous media where synthetic ground motions are generated that attempt to match characteris-

tic of real ground motions such as teleseismic travel time anomalies and the decay rates of coda waves (Frankel and Clayton, 1986; Imperatori and Mai, 2013).

We use the univariate statistics from well log data to constrain the standard deviation of fluctuations in V_P, V_S, and density in our target region. In this study, we perform simulations of earthquakes in material structures that have a 1D backbone and small-scale heterogeneity using the statistics obtained from the composite logs. We find that the addition of small-scale heterogeneity does not improve our ground motion simulations, indicating that using the composite log data to constrain small-scale heterogeneity is not an appropriate choice and that inclusion of realistic small-scale heterogeneity into these simulations would require more granular information on the material structure. Table 1 shows the average V_P, V_S, and density values and associated normalized standard deviations for each layer in the well log data.

Although data from well logs provide constraints on V_P, V_S, and density, we must make assumptions about the other parameters in the von Karman PSDF for Oklahoma, namely, correlation length and Hurst exponent. Using sonic logs from the Los Angeles basin, CA, Savran and Olsen (2016) found

that Hurst exponents range between 0 and 0.2 and vertical correlation lengths range between 15 and 150 m. In this study, we generate small-scale heterogeneity with Hurst exponent of 0.1 and correlation length of 100 m. Due to computational limitations on grid spacing we resolve in our simulations, we set a minimum wavelength cutoff of 200 m.

In the section *Experiments with Different 1D Material Structures and Small-Scale Heterogeneity*, we show the results of our experiments testing whether small-scale heterogeneity strongly affect simulated ground motions. It is first important to explain how we validate events and determine whether a simulated event is producing ground motion data that reasonably represents true ground motions. This procedure is explained in the section *Validation with Recorded Events*. We note in advance that our findings indicate that we are best able to simulate realistic ground motions using a 1D material structure where the top 2.7 km of the material structure is constrained by the composite well log data. We consider this to be our preferred material structure and is what we use to produce simulated ground motion data for inclusion into our final GMPE.

Validation with Recorded Events

Since ground motion simulations are able to generate synthetic ground motion data with spatial resolution greater than actual ground motion recordings, synthetic ground motion data offers an opportunity to augment ground motion databases, particularly at near-source distances. It is important to demonstrate that simulations produce synthetic ground motion data that respect recorded ground motion data when such recordings are available. To demonstrate this, we extract ground motion intensity measures from the recorded and synthetic ground motion data (as a function of hypocentral distance and magnitude) and perform a 2D Kolmogorov-Smirnov test on the distributions to determine whether it is reasonable to assume that the synthetic and recorded ground motion datasets were drawn from the same distribution. If we find that this assumption is reasonable for any individual event we simulate, we add the synthetic ground motion data to the target region ground motion dataset. The central idea of this synthetic/recorded ground motion data integration process is that if simulations can produce ground motion data that reasonably match recorded data (especially in areas where recordings tend to be sparse), we can use the synthetic ground motion data to fill out ground motion datasets and constrain GMPEs. This

approach could be used to generate GMPEs for regions more localized that those for which GMPEs are typically designed. For instance, in this study we show that this process can be used to generate a GMPE for central Oklahoma and southern Kansas, as compared to the entire central and eastern United States.

We simulate ground motions using the ground motion simulation code waveqlab3d, developed by Duru and Dunham (2016). Our ground motion simulations have two customizable components we use to mimic real events. One of these components is the material structure. In simulations of earthquakes in central Oklahoma and southern Kansas, we use the preferred material structure described in the section *Target Area Material Structure*. The second component is the description of the earthquake source. For M_w 3-4 earthquakes, we describe sources as point moment tensors following the method of Petersson et al. (2016). We use the moment rate function proposed by Brune (1970),

$$\dot{M}_0(t) = M_0 \omega_0^2 t e^{-\omega_0 t} H(t), \tag{2}$$

where M_0 is the moment of the earthquake and $\omega_0 = 2\pi f_c$ for corner frequency f_c. We find that the choice of corner frequency can have a significant effect on ground motions. We perform multiple simulations of individual

events while varying corner frequency and choose the simulated data that best fits the recorded data. For the Oklahoma/Kansas target region, we aim to choose corner frequencies such that we respect stress drop estimates of earthquakes from intraplate regions experiencing induced seismic events. Goertz-Allmann et al. (2011) estimated stress drops from events in the Basel geothermal field of 0.1 to 100 MPa with median 2.32 MPa. Huang and Beroza (2015) calculated P and S-wave averaged stress drops of earthquakes in the 2010-2011 Guy-Greenbrier, Arkansas, sequence using a Brune spectral model and reported a range of 1.02 to 42.50 MPa with a median of 10.57 MPa.

Based on the rule-of-thumb for fourth order numerical schemes (e.g. Levander (1988), the scheme used in this study is sixth order) used to estimate maximum resolved frequencies (5 nodes per wavelength times the minimum shear-wave speed), we estimate that by using 25 m grid spacing in our simulations we resolve frequencies up to a maximum of $\sim$10.5 Hz. By comparing our frequency limit to the frequency bands that contribute signal power to oscillator response at various periods described in Bora et al. (2016), we believe that we are adequately resolving frequencies such that we can estimate peak spectral accelerations up to 5 Hz (T=0.2s). We appear to be close to resolving the frequencies necessary to drive a 10 Hz oscillator,

but we err on the side of caution and report simulated intensity measures only up to PSA(T=0.2s). We report values of PGA, estimated to be simply the maximum acceleration value obtained from velocity time series. Due to the frequency limitation of our simulations, we note that we are possibly underestimating PGA using bandlimited time series.

M_w3.4 Event near Anthony, Kansas

We demonstrate the validation process by example. As an example event, we choose a M_w 3.4 earthquake that occurred on 17 October 2015 ∼13 km southeast of Anthony, Kansas, at a depth of 4 km. The moment of the earthquake was 1.58×10^{14} N-m. We describe the source as a point moment tensor with strike = 280°, dip = 35° and rake -55°, which was obtained by cross-referencing the online USGS earthquake archive and the St. Louis University Earthquake Center's catalog. We perform a series of simulations with moment rate functions corresponding to Brune spectra with varying corner frequencies. We use our preferred material model, a 1D structure where depths below 2.7 km follow that of the material model used in Keranen et al. (2014) and depths less than 2.7 km are the average properties taken from the composite well log shown in Table 1. The corner frequency that generates synthetic ground motion data with the best match to recorded data

has $f_c \sim 6.4$ Hz. We compute stress drop by assuming a circular rupture (Eshelby, 1957),

$$\Delta\sigma = \frac{7}{16} M_0 \left(\frac{f_c}{k V_S} \right)^3, \tag{3}$$

where M_0 is moment, f_c is corner frequency, V_S=3430 m/s is the S-wave velocity near the source, and k= 0.372 for S waves (Brune, 1971). We compute a stress drop for this event (using f_c= 6.4 Hz) of ∼8.6 MPa. It is important to note that estimates of stress drop can vary widely depending on the model and assumptions that are used to make the estimate. We do not aim to provide precise constraints on stress drops using this simulation approach. Instead, our choice of stress drop is used to set a corner frequency that defines the spectral content of our earthquakes sources. However, the estimated stress drops are quite consistent with values inferred for earthquakes in this region.

We generate synthetic ground motion data at the free surface with station spacing of 1 km over a 40×40 km grid such that the epicenter of the simulated event is at the center of the grid. For each synthetic ground motion, we extract ground motion intensity measures such as peak ground velocity and peak ground acceleration. Figure 5 shows ground motion intensity measures, as a function of hypocentral distance, of the entire target region ground

motion dataset, the synthetic M_w3.4 Anthony, KS, ground motion data, and the recorded M_w3.4 Anthony, KS, ground motion data. In our simulations we do not directly account for lateral variability in site effects that could translate to variability in simulated ground motions. The variability in our ground motion simulations (as can be seen in Figure 5) comes from the sampling of ground motions at many different azimuths for each distance. In each individual simulation, the variability is therefore a function of the interaction of the radiation pattern of the earthquake being simulated and the velocity structure though which the ensuing seismic waves propagate.

To assess whether our simulations are producing synthetic ground motion data that reasonably represent true ground motions, we perform a 2D Kolmogorov-Smirnov (K-S) test on the synthetic and recorded ground motion data following the algorithm presented in Peacock (1983). The two-sample K-S test is a non-parametric statistical test used to determine whether two samples are drawn from the same or different distributions. The null hypothesis is that two samples are drawn from the same distribution. The samples tested include the synthetic ground motion data and all recorded ground motion data with a source-to-receiver distance less than or equal to the maximum source-to-receiver distance of the synthetic ground motion data. When

we perform the 2D K-S test on the ground motion data from the M_w 3.4 Anthony, KS, event with a significance level of α=0.05, we do not reject the null hypothesis that the samples were drawn from the same parent distribution. The P-value of this test is 0.2781.

In most cases, there are few data points in our recorded ground motion catalog within 20 km, limiting the power of statistical testing. Due to computational limitations, however, we cannot simulate to farther distances at our current spatial resolution in a reasonable amount of time. Therefore, part of determining whether a simulated event reasonably mimics the characteristics of recorded ground motions from the event includes visual inspection. We extrapolate the trend of our simulated ground motion data and compare to the recorded ground motion data at distances farther than we simulate to determine whether the simulated ground motion data respect the trends observed in the recorded ground motion data.

Since the synthetic ground motion data from our M_w 3.4 Anthony, KS, earthquake simulation passes both the 2D K-S test and visual inspection, we conclude that our simulation produces ground motion data that reasonably represent true ground motions and we add the synthetic ground motion data to the target region ground motion dataset.

Experiments with Different 1D Material Structures and Small-Scale Heterogeneity

In this section we explore the effects of different 1D material structures and the introduction of small-scale heterogeneity with parameters constrained by the aforementioned composite well log data. For each test, we use a source description that attempts to mimic the previously described M_w 3.4 event near Anthony, KS. We then change the material structure depending on which specific feature we want to isolate.

Effects of Well Log Constrained 1D Material Structure

We perform an experiment where we simulate an earthquake (the M_w 3.4 event near Anthony, KS) using two different 1D material structures. One of those material structures is taken from the study by Keranen et al. (2014). The second material structure is constructed such that for depths below 2.7 km, the material structure used is that of Keranen et al. (2014) and for depths less than 2.7 km, the average properties are taken from the composite well log shown in Table 1. We note that the upper 2.7 km of the Keranen et al. (2014) material structure is described in their study as "not well-constrained by available data." Results for PGV and PSA(T=0.2s) values are shown

in Figure 6. We find that using only the Keranen et al. (2014) structure, simulated ground motions appear to represent the range of ground motion intensities fairly well, but at the farthest distances we simulate there appears to be a slight trend toward ground motion intensities higher than are seen in the recorded data, in particular for PGV, an intensity measure typically associated with low-frequency motions. This trend could be the beginning of a change in decay rate that is not consistent with recorded data, but simulations out to further distances are needed to confirm this result. We currently do not have the computational resources to extend the simulations to the distances needed to confirm that the decay rate changes significantly at distances greater than ∼20 km.

Effects of Small-Scale Heterogeneity Constrained by Well Logs

Using the univariate statistics of the individual formations from the well log data, we generate random perturbations in material properties (via a von Karman power spectral density function) that we add to the our preferred 1D "backbone" material structure. Parameter choices and generation methodology for these perturbations are explained in the section *Target Area Material Structure*. Figure 7 shows an example of two simulations performed mimicking the aforementioned M_w 3.4 Anthony, KS, event. One of these simulations

employs the preferred 1D material structure constrained by composite well log data in the upper 2.7 km, which has been shown in several figures in this study. The second simulation has the same 1D backbone, but includes isotropic small-scale heterogeneity described using a correlation length of 100 m, Hurst exponent of 0.1, and standard deviations corresponding to the values in Table 1. The perturbations are scaled to the average velocities and densities of each individual formation.

We find that the addition of small-scale heterogeneity reduces the range of ground motion intensities we observe in simulations, an effect that becomes particularly pronounced at our farthest simulated distances (>15 km). Additionally, the rate of decay of ground motion intensities as a function of distance is smaller when small-scale heterogeneity is added to the material structure. The simulation performed using only the 1D structure produces a trend of ground motion intensity decay with distance with much greater agreement to the recorded data than does the simulation with additional small-scale heterogeneity. We believe that this effect could be a result of the composite nature of the well log data. Combining multiple wells likely overestimates the standard deviations reported in Table 1. This notion is further evidenced by the observation that many standard deviations in Table

1 are much higher than typically estimated in the upper crust (high estimates tend to be around 10% (Frankel and Clayton, 1986; Imperatori and Mai, 2013; Bydlon and Dunham, 2015; Savran and Olsen, 2016)). Without more granular data to constrain von Karman parameters for the target region, we conclude that for the purposes of this study, it is best to generate a final GMPE that includes simulated data using our preferred 1D material structure without small-scale heterogeneity.

Construction of Composite Recorded/Simulated GMPE

We simulate a set of earthquakes that occurred in the target region such that we can add the simulated ground motion data to our catalog of recorded ground motions from northern Oklahoma and southern Kansas. We then use this composite simulated/recorded ground motion catalog to construct a GMPE for M_w 3-4 earthquakes for this target region. Details of the simulated events are shown in Table 2. The earthquakes were selected because they were reasonably well-recorded, meaning there are at least $\sim$10 ground motion recordings in our catalog, including recordings with hypocentral distances less than 25 km (the farthest hypocentral distances we simulate), and we could locate an associated moment tensor for the event. We select moment tensors

by searching the earthquake catalog provided by the USGS Earthquake Hazards Program and the Moment Tensor Solution Database provided by Saint Louis University's Earthquake Center. If both catalogs have a moment tensor for any individual earthquake and their preferred values differ (this occurrence is most commonly associated with differences in earthquake depths), we choose the solution provided by Saint Louis University. Since we do not have precise information on stress drops from these earthquakes at this time, we run multiple simulations of each event while varying corner frequency and report the corresponding stress drop that produced the best-fitting ground motion data according to the procedure described in the section *Validation with Recorded Events*. All of our inferred stress drops fall between 1 and 42.5 MPa, in line with observations of stress drops of induced intraplate earthquakes (Goertz-Allmann et al., 2011; Huang and Beroza, 2015).

Once the individual events have been validated, we take the simulated data and add it to the catalog of ground motions recorded in the target area. Figure 8 shows the data in this composite catalog, with a color scheme indicating which data points are recorded and which are simulated. We fit the data to produce a GMPE for the target area using a slightly modified version of the functional form presented in Shahjouei and Pezeshk (2016).

The only difference in our GMPE is that the distance range where the rate of attenuation is different is 40-120 km, instead of 60-120 km as in Shahjouei and Pezeshk (2016). This change reflects the characteristics of the recorded ground motions in our target region. We use nonlinear regression to fit a GMPE to our composite catalog to predict median ground motion intensities using the functional form:

$$\begin{aligned}\log(\bar{Y}) = c_1 + c_2\boldsymbol{M} + c_3\boldsymbol{M}^2 + (c_4 + c_5\boldsymbol{M}) \times \min\{\log(R), \log(40)\} \\ + (c_6 + c_7\boldsymbol{M}) \times \max[\min\{\log(R/40), \log(120/40)\}, 0] \\ + (c_8 + c_9\boldsymbol{M}) \times \max\{\log(R/120), 0\} + c_{10}R,\end{aligned} \tag{4}$$

with

$$R = \sqrt{R_{hyp}^2 + c_{11}^2}, \tag{5}$$

where $\bar{Y}$ is the median value of the specified ground motion intensity measure (in units of cm/s^2 for PGA and PSA, cm/s for PGV), $\boldsymbol{M}$ is the moment magnitude, R_{hyp} is the hypocentral distance in km, and $c_1 - c_{11}$ are the coefficients obtained when fitting the ground motion data using Equation (4). PSA are pseudospectral accelerations computed with a 5% damping parameter. The coefficients computed for PGA, PGV, and PSA at T= 1, 0.5, and 0.2s (1, 2, and 5 Hz, respectively) are reported in Table 3. Figure

9 shows GMPEs for PGV and PSA(T=0.2s) for a M_w3.5 earthquake in the target region constructed via fitting the composite catalog with Equation 4.

We interpret the coefficient c_{11} as the "effective depth" parameter reported in studies such as Atkinson and Silva (2000); Yenier and Atkinson (2014); Atkinson (2015); Yenier and Atkinson (2015), and Atkinson et al. (2016) that aims to capture near-source distance saturation effects. We find that this effective depth term varies between ground motion intensity measures, but is consistently less than 3 km, which is less than but similar to observations from ground motion data of earthquakes up to M_w 4 in the Geysers region of California (Atkinson et al., 2016), where this term was found to be near 3 km for earthquakes of M_w 4.

Figure 10 shows the residuals, defined as the difference (in log units) between the observed and predicted ground motion intensities, for PGV and PSA(T=0.2s) with means and standard deviations of the data binned every 10 km. We compute means and standard deviations for the combined dataset, the simulated data only, and the recorded data only. The means for all of those permutations are near zero or well within 1 standard deviation. The results from the residual computations indicate that our simulated data has amplitude and decay characteristics nearly identical to the recorded data.

Therefore, we conclude that our simulations are producing realistic ground motions.

Discussion and Conclusions

The key development presented in this study is a framework for incorporating realistic ground motion simulations (validated against recorded data) into ground motion catalogs for the purposes of augmenting ground motion datasets and developing GMPEs in data-poor regions. This framework involves the collection of best available constraints on the material structure of the target region and source parameters of individual events. These constraints are then incorporated into ground motion simulations intended to mimic events in the target region. We compare the synthetic data produced via simulation of an individual event to recorded ground motions from the same event to ensure that our simulations are producing realistic ground motion data. Once we are confident that our synthetic data is realistic we add the synthetic data to our target region ground motion catalog. This increases the number of near-source ground motion recordings we can use to constrain GMPEs. Finally, we fit GMPEs for several key ground motion intensity measures using an equation that describes the characteristics of the ground

motion dataset. This framework can be used to develop GMPEs tailored specifically to regions much more refined than typical scales of applicability, which can be areas as large as the central and eastern US. Since ground motions likely exhibit intra-region variability (i.e., ground motions from a M_w 3.5 earthquake in north Oklahoma are likely different than a M_w 3.5 earthquake in the New Madrid Seismic Zone), the development of regionally-specific GMPEs could lead to improvements in seismic hazard forecasting, in particular in areas experiencing induced seismicity.

Using this framework, we have constructed a GMPE for M_w 3-4 earthquakes for north-central Oklahoma and south-central Kansas using a composite catalog consisting of recorded ground motions and synthetic ground motions from simulations of events that occurred in the target region. We compare our GMPE to available GMPEs for M_w 3-4 earthquakes in the central and eastern US, most notably the GMPE presented for this magnitude range in Atkinson (2015), and find that these GMPEs generally predict similar values for ground motions as does our target region GMPE. There are, however, some notable differences that we believe result from 1) the recorded/simulated nature of our ground motion dataset and the amount of near-source ground motion data with which we constrain our GMPE, and

2) the fact that our GMPE is designed for a subregion of the central US that likely has earthquake and/or wave propagation characteristics distinct from the average properties of the entire central and eastern US. Our results indicate that the rate of decay of ground motion intensities in our target region varies as a function of distance similar to the findings of Pezeshk et al. (2011),Boore and Thompson (2015), and Shahjouei and Pezeshk (2016), except that our data indicate changes in the rate of decay at 40 and 120 km. Although we use a similar functional form to construct GMPEs as does Shahjouei and Pezeshk (2016), we note in that study GMPEs were constructed for earthquakes M_w 5-8, so comparisons of such GMPEs to those presented in this study are not valid.

We also present results of several experiments designed to understand sensitivity of 3D ground motion simulations to material structure choice. We find that data from our target region does not sufficiently constrain the statistical parameters of small-scale heterogeneity for use in the development of our target region GMPE. We believe that better constraints on such parameters could lead to the inclusion of small-scale heterogeneity in similar ground motion simulations.

We believe that physics-based synthetic ground motion data, such as the

data generated in this study, could be extended by combination with synthetic data produced via stochastic approaches, such as in Graves and Pitarka (2004); Frankel (2009); Graves and Pitarka (2010); Bommer et al. (2017), to produce broadband synthetics. By including stochastically generated synthetics at high frequencies, the useful frequency band of synthetic ground motions could be extended. This would be particularly useful for estimated intensity measures such as peak ground acceleration and peak spectral accelerations above 10 Hz, which require information on frequencies that are currently beyond what can be accurately resolved by our simulations.

In future work, we aim to use this framework to develop GMPEs for the north-central Oklahoma and south-central Kansas region for earthquakes M_w >4. This will require a change in our approach to earthquake sources, since it will be important to account for finite source effects when simulating large earthquakes.

Data and Resources

Ground motion data used in this study was provided by coauthor Abhineet Gupta and was collected from IRIS Data Services (http://ds.347iris.edu/ds/nodes/dmc/). Processing methods are detailed in upcoming manuscript Gupta et al. (2017).

The well log data was obtained from a member of the Stanford Center for Induced and Triggered Seismicity.

Moment tensors were obtained from the USGS Earthquake Archive (https://earthquake.usgs.gov/ea and the Moment Tensor Solution Database provided by Saint Louis University's Earthquake Center (http://www.eas.slu.edu/eqc/eqcmt.html). Last accessed October 2016.

Acknowledgments

Funding was provided by the Stanford Center for Induced and Triggered Seismicity (SCITS). We would like to thank the directors and participants of SCITS for their feedback during development. Special thanks goes out to Bill Ellsworth, who facilitated the acquisition of the well log data we used in this study and for teaching Sam important lessons about GMPEs. Thank you to Kyle Withers, who always gives an honest attempt to answers Sam's questions when they are asked.

References

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore,

613 D. M., et al. (2014). Nga-west2 database. *Earthquake Spectra*, 30(3):989–
614 1005.

615 Atkinson, G. M. (2008). Ground-motion prediction equations for eastern
616 north america from a referenced empirical approach: Implications for
617 epistemic uncertainty. *Bulletin of the Seismological Society of America*,
618 98(3):1304–1318.

619 Atkinson, G. M. (2010). Ground-motion prediction equations for hawaii from
620 a referenced empirical approach. *Bulletin of the Seismological Society of*
621 *America*, 100(2):751–761.

622 Atkinson, G. M. (2015). Ground-motion prediction equation for small-
623 to-moderate events at short hypocentral distances, with application to
624 induced-seismicity hazards. *Bulletin of the Seismological Society of Amer-*
625 *ica*.

626 Atkinson, G. M. and Boore, D. M. (1995). Ground-motion relations for
627 eastern north america. *Bulletin of the Seismological Society of America*,
628 85(1):17–30.

629 Atkinson, G. M. and Boore, D. M. (2006). Earthquake ground-motion pre-

630 diction equations for eastern north america. *Bulletin of the Seismological*
631 *Society of America*, 96(6):2181–2205.

632 Atkinson, G. M. and Motazedian, D. (2013). Ground-motion amplitudes
633 for earthquakes in puerto rico. *Bulletin of the Seismological Society of*
634 *America*, 103(3):1846–1859.

635 Atkinson, G. M. and Silva, W. (2000). Stochastic modeling of califor-
636 nia ground motions. *Bulletin of the Seismological Society of America*,
637 90(2):255–274.

638 Atkinson, G. M., Yenier, E., Sharma, N., and Convertito, V. (2016). Con-
639 straints on the near-distance saturation of ground-motion amplitudes for
640 small-to-moderate induced earthquakes. *Bulletin of the Seismological So-*
641 *ciety of America*.

642 Benz, H. M., Frankel, A., and Boore, D. M. (1997). Regional lg attenuation
643 for the continental united states. *Bulletin of the Seismological Society of*
644 *America*, 87(3):606–619.

645 Bommer, J. J., Stafford, P. J., Edwards, B., Dost, B., van Dedem, E.,
646 Rodriguez-Marek, A., Kruiver, P., van Elk, J., Doornhof, D., and Nti-
647 nalexis, M. (2017). Framework for a ground-motion model for induced

seismic hazard and risk analysis in the groningen gas field, the netherlands. *Earthquake Spectra*.

Boore, D. M. and Thompson, E. M. (2015). Revisions to some parameters used in stochastic-method simulations of ground motion. *Bulletin of the Seismological Society of America*.

Bora, S. S., Scherbaum, F., Kuehn, N., and Stafford, P. (2016). On the relationship between fourier and response spectra: Implications for the adjustment of empirical ground-motion prediction equations (gmpes). *Bulletin of the Seismological Society of America*.

Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of geophysical research*, 75(26):4997–5009.

Brune, J. N. (1971). Correction to "tectonic stress and the spectra of seismic shear waves from earthquakes". *Journal of Geophysical Research*, 76:5002.

Bydlon, S. A. and Dunham, E. M. (2015). Rupture dynamics and ground motions from earthquakes in 2-d heterogeneous media. *Geophysical Research Letters*, 42(6):1701–1709.

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. *Bulletin of the Seismological Society of America*, 93(3):1012–1033.

Day, S., Bielak, J., Dreger, D., Larsen, S., Graves, R., Pitarka, A., and Olsen, K. (2003). Tests of 3d elastodynamic codes. *Final Report to Pacific Earthquake Engineering Research Center, Lifelines Program TASK 1A02*, pages 1–32.

Douglas, J., Edwards, B., Convertito, V., Sharma, N., Tramelli, A., Kraaijpoel, D., Cabrera, B. M., Maercklin, N., and Troise, C. (2013). Predicting ground motion from induced earthquakes in geothermal areas. *Bulletin of the Seismological Society of America*, 103(3):1875–1897.

Duru, K. and Dunham, E. M. (2016). Dynamic earthquake rupture simulations on nonplanar faults embedded in 3d geometrically complex, heterogeneous elastic solids. *Journal of Computational Physics*, 305:185–207.

Edwards, B. and Douglas, J. (2014). Magnitude scaling of induced earthquakes. *Geothermics*, 52:132–139.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, 341(6142):1225942.

683 Erickson, D., McNamara, D. E., and Benz, H. M. (2004). Frequency-
684 dependent lg q within the continental united states. *Bulletin of the Seis-*
685 *mological Society of America*, 94(5):1630–1643.

686 Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal
687 inclusion, and related problems. In *Proceedings of the Royal Society of*
688 *London A: Mathematical, Physical and Engineering Sciences*, volume 241,
689 pages 376–396. The Royal Society.

690 Frankel, A. (2009). A constant stress-drop model for producing broadband
691 synthetic seismograms: Comparison with the next generation attenuation
692 relations. *Bulletin of the Seismological Society of America*, 99(2A):664–
693 680.

694 Frankel, A. and Clayton, R. W. (1986). Finite difference simulations of seis-
695 mic scattering: Implications for the propagation of short-period seismic
696 waves in the crust and models of crustal heterogeneity. *Journal of Geo-*
697 *physical Research: Solid Earth*, 91(B6):6465–6489.

698 Frohlich, C., Ellsworth, W., Brown, W. A., Brunt, M., Luetgert, J., Mac-
699 Donald, T., and Walter, S. (2014). The 17 may 2012 m4. 8 earthquake

near timpson, east texas: An event possibly triggered by fluid injection. *Journal of Geophysical Research: Solid Earth*, 119(1):581–593.

Goertz-Allmann, B. P., Goertz, A., and Wiemer, S. (2011). Stress drop variations of induced earthquakes at the basel geothermal site. *Geophysical Research Letters*, 38(9).

Graves, R. and Pitarka, A. (2004). Broadband time history simulation using a hybrid approach.

Graves, R. W. and Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, 100(5A):2095–2123.

Gupta, A., Baker, J. W., and Ellsworth, W. L. (2017). Assessing ground motion amplitudes and attenuation for small to moderate induced and tectonic earthquakes in the central and eastern united states. *Seismological Research Letters*.

Hartzell, S., Harmsen, S., and Frankel, A. (2010). Effects of 3d random correlated velocity perturbations on predicted ground motions. *Bulletin of the Seismological Society of America*, 100(4):1415–1426.

Hassani, B. and Atkinson, G. M. (2015). Referenced empirical ground-motion model for eastern north america. *Seismological Research Letters*, 86(2A):477–491.

Holliger, K. (1996). Upper-crustal seismic velocity heterogeneity as derived from a variety of p-wave sonic logs. *Geophysical Journal International*, 125(3):813–829.

Holliger, K. (1997). Seismic scattering in the upper crystalline crust based on evidence from sonic logs. *Geophysical Journal International*, 128(1):65–72.

Huang, Y. and Beroza, G. C. (2015). Temporal variation in the magnitude-frequency distribution during the guy-greenbrier earthquake sequence. *Geophysical Research Letters*, 42(16):6639–6646.

Imperatori, W. and Mai, P. M. (2013). Broad-band near-field ground motion simulations in 3-dimensional scattering media. *Geophysical Journal International*, 192(2):725–744.

Keranen, K. M., Savage, H. M., Abers, G. A., and Cochran, E. S. (2013). Potentially induced earthquakes in oklahoma, usa: Links between wastewater injection and the 2011 mw 5.7 earthquake sequence. *Geology*, 41(6):699–702.

735 Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., and Ge, S.
736 (2014). Sharp increase in central oklahoma seismicity since 2008 induced
737 by massive wastewater injection. *Science*, 345(6195):448–451.

738 Kim, W.-Y. (2013). Induced seismicity associated with fluid injection into
739 a deep well in youngstown, ohio. *Journal of Geophysical Research: Solid*
740 *Earth*, 118(7):3506–3518.

741 Kruiver, P. P., van Dedem, E., Romijn, R., de Lange, G., Korff, M., Stafleu,
742 J., Gunnink, J. L., Rodriguez-Marek, A., Bommer, J. J., van Elk, J., et al.
743 (2017). An integrated shear-wave velocity model for the groningen gas
744 field, the netherlands. *Bulletin of Earthquake Engineering*, pages 1–26.

745 Levander, A. R. (1988). Fourth-order finite-difference p-sv seismograms. *Geo-*
746 *physics*, 53(11):1425–1436.

747 Olsen, K. (2000). Site amplification in the los angeles basin from three-
748 dimensional modeling of ground motion. *Bulletin of the Seismological So-*
749 *ciety of America*, 90(6B):S77–S94.

750 Olsen, K., Day, S., Dalguer, L., Mayhew, J., Cui, Y., Zhu, J., Cruz-Atienza,
751 V., Roten, D., Maechling, P., Jordan, T., et al. (2009). Shakeout-d:
752 Ground motion estimates using an ensemble of large earthquakes on the

753 southern san andreas fault with spontaneous rupture propagation. *Geo-*
754 *physical Research Letters*, 36(4).

755 Olsen, K. B. and Mayhew, J. E. (2010). Goodness-of-fit criteria for broad-
756 band synthetic seismograms, with application to the 2008 mw 5.4 chino
757 hills, california, earthquake. *Seismological Research Letters*, 81(5):715–723.

758 Peacock, J. (1983). Two-dimensional goodness-of-fit testing in astronomy.
759 *Monthly Notices of the Royal Astronomical Society*, 202(3):615–627.

760 Petersen, M. D., Moschetti, M. P., Powers, P. M., Mueller, C. S., Haller,
761 K. M., Frankel, A. D., Zeng, Y., Rezaeian, S., Harmsen, S. C., Boyd,
762 O. S., et al. (2014). Documentation for the 2014 update of the united states
763 national seismic hazard maps. Technical report, US Geological Survey.

764 Petersson, N. A., O'Reilly, O., Sjögreen, B., and Bydlon, S. (2016). Dis-
765 cretizing singular point sources in hyperbolic wave propagation problems.
766 *Journal of Computational Physics*, 321:532–555.

767 Pezeshk, S., Zandieh, A., and Tavakoli, B. (2011). Hybrid empirical ground-
768 motion prediction equations for eastern north america using nga models
769 and updated seismological parameters. *Bulletin of the Seismological Soci-*
770 *ety of America*, 101(4):1859–1870.

771 Savran, W. and Olsen, K. (2016). Model for small-scale crustal heterogeneity
772 in los angeles basin based on inversion of sonic log data. *Geophysical*
773 *Journal International*, 205(2):856–863.

774 Shahjouei, A. and Pezeshk, S. (2016). Alternative hybrid empirical ground-
775 motion model for central and eastern north america using hybrid sim-
776 ulations and nga-west2 models. *Bulletin of the Seismological Society of*
777 *America*.

778 Silva, W., Gregor, N., and Darragh, R. (2002). Development of regional hard
779 rock attenuation relations for central and eastern north america. *Pacific*
780 *Engineering and Analysis, El Cerrito, California*.

781 Toro, G. R., Abrahamson, N. A., and Schneider, J. F. (1997). Model of strong
782 ground motions from earthquakes in central and eastern north america:
783 best estimates and uncertainties. *Seismological Research Letters*, 68(1):41–
784 57.

785 Walsh, F. R. and Zoback, M. D. (2015). Oklahoma's recent earthquakes and
786 saltwater disposal. *Science advances*, 1(5):e1500195.

787 Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., and Rubinstein, J. L.

(2015). High-rate injection is associated with the increase in us mid-continent seismicity. *Science*, 348(6241):1336–1340.

Yenier, E. and Atkinson, G. M. (2014). Equivalent point-source modeling of moderate-to-large magnitude earthquakes and associated ground-motion saturation effects. *Bulletin of the Seismological Society of America*.

Yenier, E. and Atkinson, G. M. (2015). Regionally adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Application to central and eastern north america. *Bulletin of the Seismological Society of America*, 105(4):1989–2009.

Full Mailing Address for Each Author

Samuel A. Bydlon

Graduate Student

Dept. of Geophysics, Stanford University

Mitchell Building, 397 Panama Mall, Room B55

Stanford, CA 94305, USA

Eric M. Dunham

Associate Professor

Dept. of Geophysics, Stanford University
Mitchell Building, 397 Panama Mall, 3rd floor
Stanford, CA 94305, USA

Abhineet Gupta
Graduate Student
Dept. of Civil and Environmental Engineering, Stanford University,
Blume Center, Bldg. 540, 439 Panama Mall, Room 206
Stanford, CA 94305, USA

List of Figure Captions

Figure 1: Epicenters of earthquakes (blue dots) included in the M_w3-4 ground motion catalog associated with the Oklahoma/Kansas study area. Red dots indicate epicenters of the 2011 M_w 5.6 event near Prague, OK, and the 2016 M_w 5.8 event near Pawnee, OK.

Figure 2: Peak ground velocities (a) and peak spectral accelerations (T=0.2s) (b) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area ground motion dataset. Green dots indicate data M_w 3.0 - 3.5 and black dots indicate data M_w

3.5 - 4.0. Red lines indicate reference Atkinson (2015) M_w 3.5 GMPE (solid indicates mean, dashed +/- 1 standard deviation).

Figure 3: Composite well log data (as received) showing measurements of P- and S-wave speed (top left and top right histograms, respectively) and density (bottom histogram) from more than 20 well logs describing the material structure of the Arbuckle formation used to formulate the material structure representing our target region. At this time we do not have access to individual well logs.

Figure 4: a) Keranen et al. (2014) 1D velocity profile (V_P and V_S) for depths 0 to 15 km. b) 1D velocity profile (V_P and V_S) for depths 0 to 3 km obtained from well logs. At depths >2.7 km, we set the well-log-derived 1D velocity profile equal to the Keranen et al. (2014) 1D velocity profile.

Figure 5: Peak ground accelerations (top) and peak ground velocities (bottom) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred target region 1D material structure (green dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulation produces synthetic ground motion data that agree well with recordings.

Figure 6: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottoms) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred 1D material structure (green dots) and the Keranen et al. (2014) material structure (blue dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulations performed using the preferred 1D material structure better capture the decay rate of ground motion intensities compared to the Keranen et al. (2014) material structure.

Figure 7: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottoms) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred 1D material structure (green dots) and preferred material structure with additional small-scale heterogeneity constrained by composite well logs (blue dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulations performed using the preferred 1D material structure alone better capture the decay rate of ground motion intensities compared to the simulations where small-scale heterogeneity is included, particularly at distances

greater than 10 km. This difference is more pronounced for PSA(T=0.2s) than for PGV.

Figure 8: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area composite ground motion dataset including recorded (gray dots) and simulated (green dots) ground motions.

Figure 9: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area composite ground motion dataset (gray dots). Red line indicates reference Atkinson (2015) M_w 3.5 GMPE. Green line is GMPE for M_w 3.5 events constructed by fitting our composite catalog using Equation 4 (coefficients shown in Table 3).

Figure 10: Plots of residuals (difference between observed and predicted in log units, binned every 10 km) for PGV (left column) and PSA(T=0.2s) (right column). Green squares indicate the mean values of each bin with errors bars indicating +/- 1 standard deviation. The top panel in each column are the residuals between the composite GMPE and the recorded data (gray dots), the middle panel are the residuals between the composite GMPE

and the simulated data (red dots), and the bottom panel are the residuals between the composite GMPE and the combined dataset, where the dots are color coded as either recorded or simulated. The means are near zero or well within 1 standard deviation for all cases, indicating that our simulations are producing ground motion data that has similar amplitude and decay characteristics as recorded data.

Figure A1: Particle velocity (radial, transverse, and vertical components) time series comparisons quantified using the 3-component averaged goodness-of-fit measure described in Appendix A at a station location at 10 km horizontal and 2 km vertical distance from the source between the analytic solution for the LOH.1 verification problem (black) and solutions produced by waveqlab3d (red) for grid spacing of 50 (left) and 25 m (right).

Appendix

Layer-Over-Halfspace Verification Exercise

To establish confidence that waveqlab3d is accurately modeling point moment tensor sources and 3D seismic wave propagation in heterogeneous material structures, we perform the verification exercise LOH.1 designed by Day et al. (2003). LOH.1 is a layer-over-halfspace problem where a point moment

tensor source is set at 2 km depth in a half-space that underlies a 1 km thick low-velocity layer. The grid spacings used for the simulations are 25 and 50 m. Ground motions are generated at 1000 m intervals on the surface at an angle of $\tan^{-1}(4/3) \approx 53.13^{\circ}$ from the tangential direction of the dislocation. We compare our numerically generated ground motions to semi-analytic solutions using the approach of Olsen and Mayhew (2010). Our goodness-of-fit (GOF) measure uses three equally-weighted ground motion intensity metrics: peak ground velocity, peak ground acceleration, and spectral accelerations (5% damping) for periods between 0.1-1 s with 0.01 s spacing and between 1.1-10 s with 0.1 s spacing. For each individual metric, we compute the three-component average (vertical, radial, tangential) of the GOF between the numerical and semi-analytic solution. We then compute a final GOF measure according to Equation 2 of Olsen and Mayhew (2010), with equal weight given to the individual metrics. Figure A.1 shows the comparison between three directional components of ground motion of the numerical and analytic solutions at the farthest station computed (10 km horizontal distance from source on the free surface), along with the three-component averaged GOF measure. We find that waveqlab3d produces solutions with an "excellent" fit (80<GOF<100) relative to the semi-analytic solution for

the cases where grid spacing equals 50 and 25 m. We therefore conclude that waveqlab3d accurately models point moment tensor sources and 3D seismic wave propagation at grid spacing of 50 m and less for frequencies up to 5 Hz (as per the filtering characteristics of the analytic solution defined in the LOH.1 problem documentation (Day et al., 2003)).

Table 1: Preferred 1D Material Structure Used in Simulations

Layer	**Depth (m)**	$\overline{V_P}$	$\overline{V_S}$	$\overline{\rho}$	σ_{V_P}	σ_{V_S}	σ_ρ
1	0-366	2.59	1.33	2.2	10.0	10.0	7.0
2	367-1011	4.42	2.21	2.56	18.1	18.0	12.0
3	1011-1400	3.53	1.76	2.58	18.0	19.6	3.9
4	1400-1593	4.54	2.31	2.66	20.7	21.8	5.8
5	1593-1697	5.82	3.05	2.67	8.6	9.2	2.7
6	1697-1758	3.90	2.19	2.65	18.1	19.1	4.2
7	1758-1823	5.70	3.11	2.69	11.6	10.8	4.2
8	1823-1882	4.05	2.13	2.58	16.5	21.5	9.3
9	1882-2500	6.34	3.44	2.77	8.4	9.2	6.3
10	2500-2700	5.64	3.05	2.68	7.4	8.2	5.6

Average V_P (km/s), V_S (km/s), and density (ρ, expressed in g/cm^3) and normalized standard deviations (σ, expressed in %) used to generate small-scale heterogeneity for the north Oklahoma / southern Kansas target area via the von Karman PSDF. Data obtained from well logs in central Oklahoma. For all depths below 2700 m, average V_P, V_S, and ρ values taken from the Keranen et al. (2014) 1D material structure and normalized standard deviations are taken from Layer 10 of the well log data.

Table 2: Source Specifications of Simulated Earthquakes

Date	**Nearest City**	$\mathbf{M}_w$	**Depth (km)**	**Strike°**	**Dip°**	**Rake°**	$\Delta\sigma$ **(MPa)**
11/07/2015	Medford, OK	3.8	3.9	105	58	-47	32.7
11/11/2015	Medford, OK	3.5	4.0	100	65	-55	23.5
10/17/2015	Anthony, KS	3.4	4.0	280	35	-55	8.6
10/10/2014	Luther, OK	3.2	3.0	285	80	-20	8.4
08/25/2015	Stillwater, OK	3.2	3.0	325	75	-10	9.3
09/16/2015	Pawnee, OK	3.2	3.0	150	80	10	14.0
07/09/2014	Caldwell, KS	3.0	3.4	205	85	65	12.8

Dates, locations, and technical specifications of the simulated earthquakes. Stress drops are computed via Equation 3. Although multiple simulations are performed for each event with varying stress drop, here we report stress drops of the simulations that produced ground motions that best fit the ground motions recorded during the event.

Table 3: GMPE Coefficients for Various Intensity Measures

	PGV	**1 Hz (T=1s) PSA**	**2 Hz (T=0.5s) PSA**	**5 Hz (T=0.2s) PSA**	**PGA**
c_1	-4.388	-0.912	-2.907	-3.868	-4.719
c_2	1.631	0.212	1.296	2.096	2.615
c_3	0.018	0.163	0.051	-0.008	-0.068
c_4	-2.706×10^5	6.218×10^4	-5.122×10^3	-1.281×10^5	2.306
c_5	-0.371	0.063	-0.193	-0.539	-0.652
c_6	-2.706×10^5	6.218×10^4	-5.122×10^3	-1.281×10^5	2.306
c_7	0.160	0.035	0.075	0.156	0.034
c_8	-2.706×10^5	6.218×10^4	-5.122×10^3	-1.281×10^5	2.306
c_9	0.531	0.604	0.530	0.156	0.463
c_{10}	2.706×10^5	-6.218×10^4	5.122×10^3	1.281×10^5	-2.306
c_{11}	0.482	2.238	0.007	0.015	2.272

Coefficients $c_1 - c_{11}$ obtained by fitting the composite recorded/simulated ground motion catalog for the Oklahoma/Kansas target region using Equation 4 for RotD50 horizontal-component ground motion intensity measures PGA (in cm/s^2), PGV (in cm/s), and 5% damped pseudospectral accelerations (in cm/s^2).

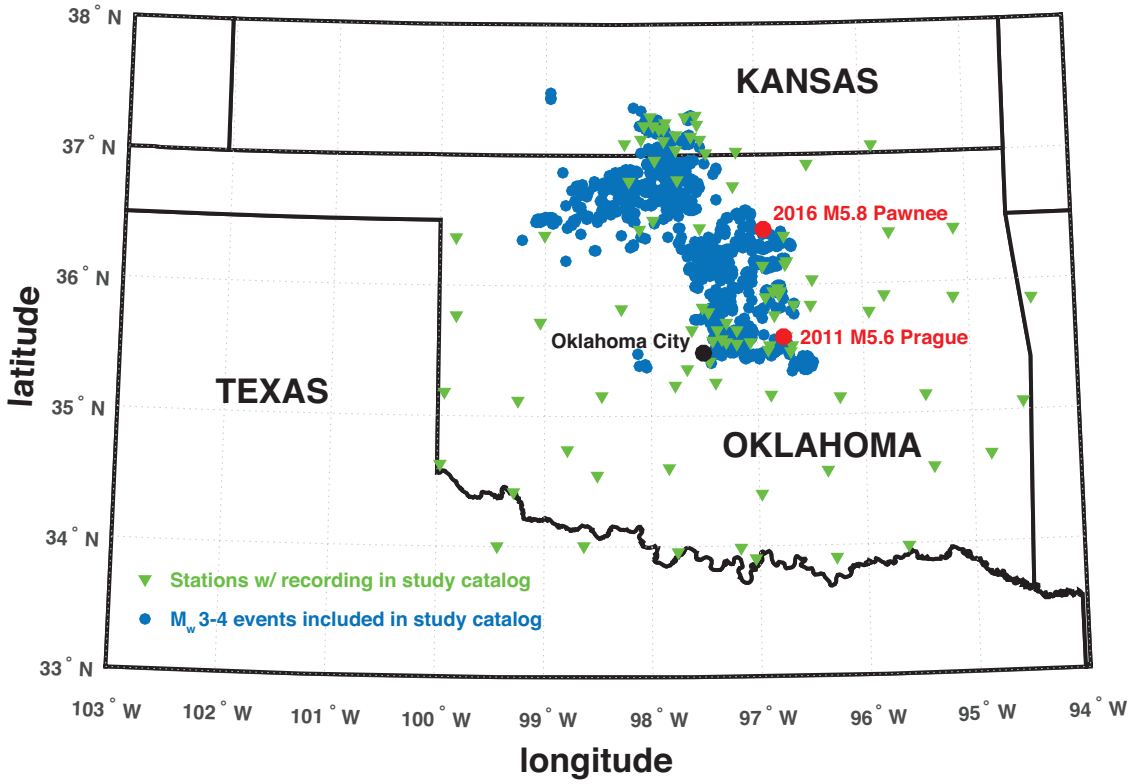

Figure 1: Epicenters of earthquakes (blue dots) included in the M_w3-4 ground motion catalog associated with the Oklahoma/Kansas study area. Red dots indicate epicenters of the 2011 M_w 5.6 event near Prague, OK, and the 2016 M_w 5.8 event near Pawnee, OK.

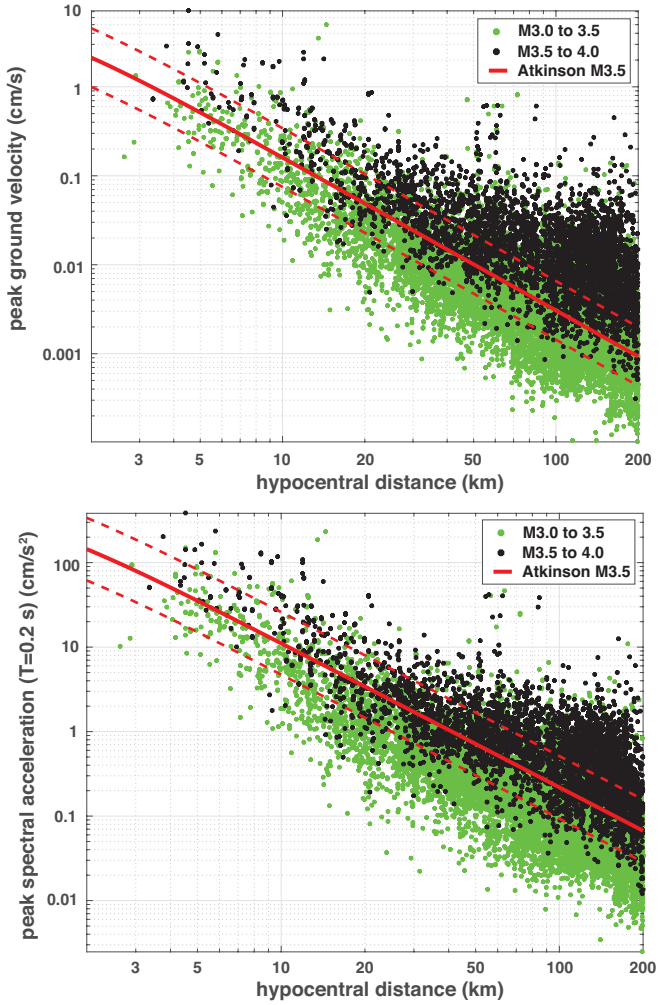

Figure 2: Peak ground velocities (a) and peak spectral accelerations (T=0.2s) (b) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area ground motion dataset. Green dots indicate data M_w 3.0 - 3.5 and black dots indicate data M_w 3.5 - 4.0. Red lines indicate reference Atkinson (2015) M_w 3.5 GMPE (solid indicates mean, dashed +/- 1 standard deviation).

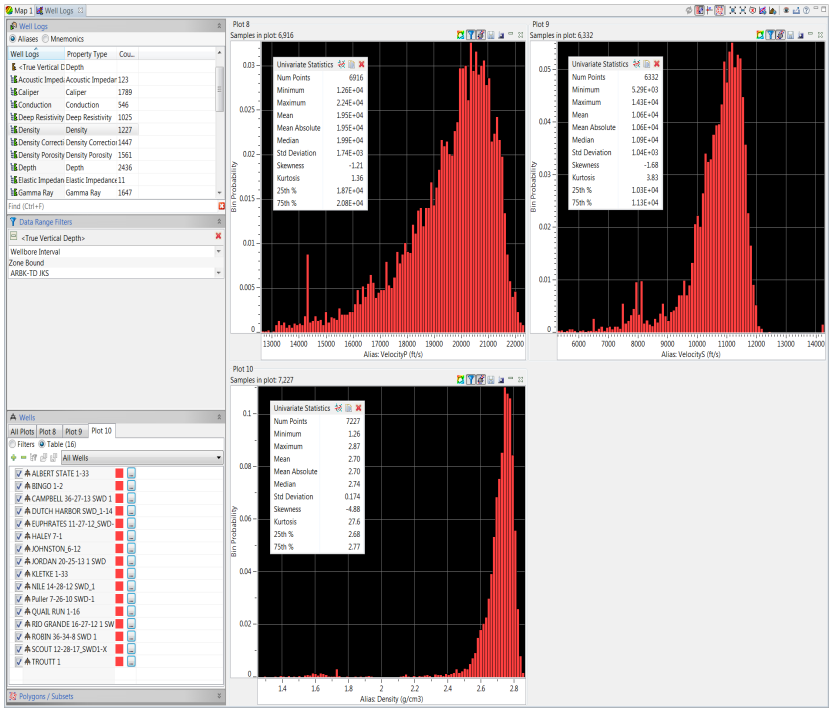

Figure 3: Composite well log data (as received) showing measurements of P and S-wave speed (top left and top right histograms, respectively) and density (bottom histogram) from more than 20 well logs describing the material structure of the Arbuckle formation used to formulate the material structure representing our target region. At this time we do not have access to individual well logs.

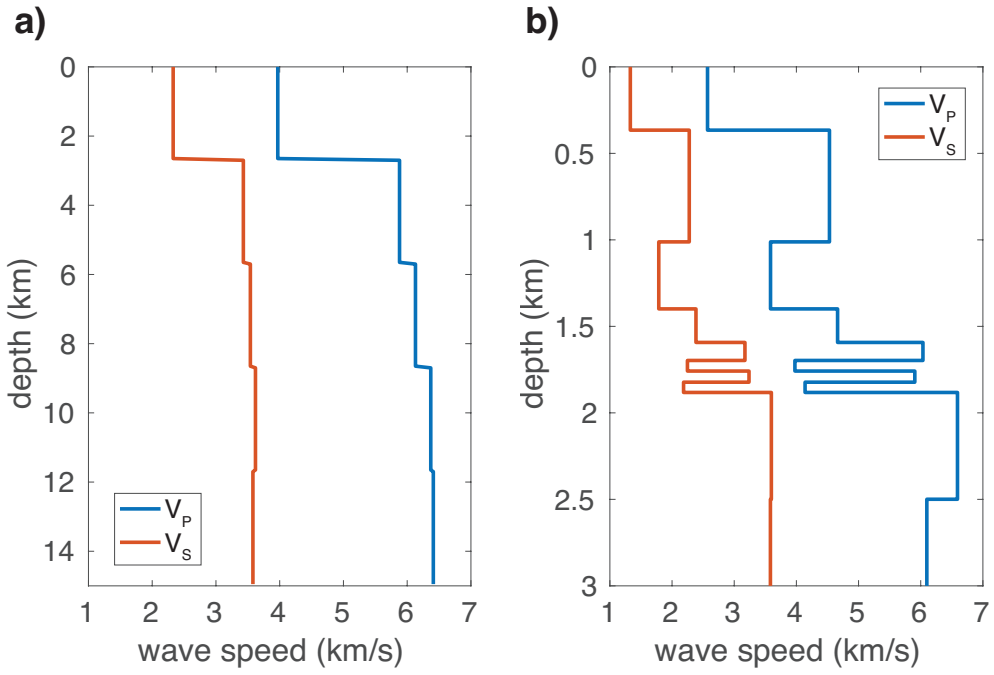

Figure 4: a) Keranen et al. (2014) 1D velocity profile (V_P and V_S) for depths 0 to 15 km. b) 1D velocity profile (V_P and V_S) for depths 0 to 3 km obtained from well logs. At depths >2.7 km, we set the well-log-derived 1D velocity profile equal to the Keranen et al. (2014) 1D velocity profile.

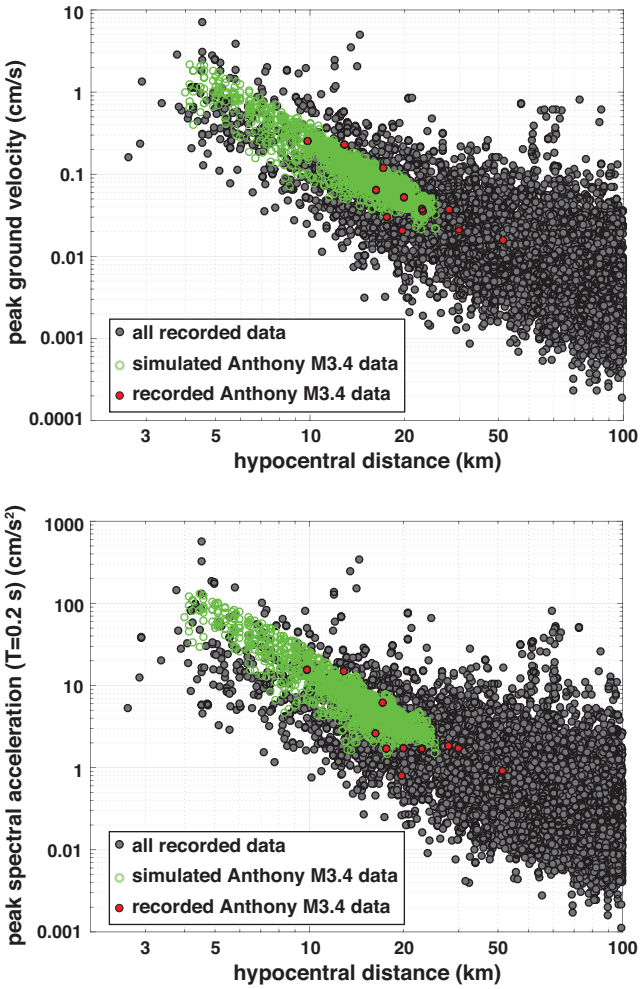

Figure 5: Peak ground accelerations (top) and peak ground velocities (bottom) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred target region 1D material structure (green dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulation produces synthetic ground motion data that agree well with recordings.

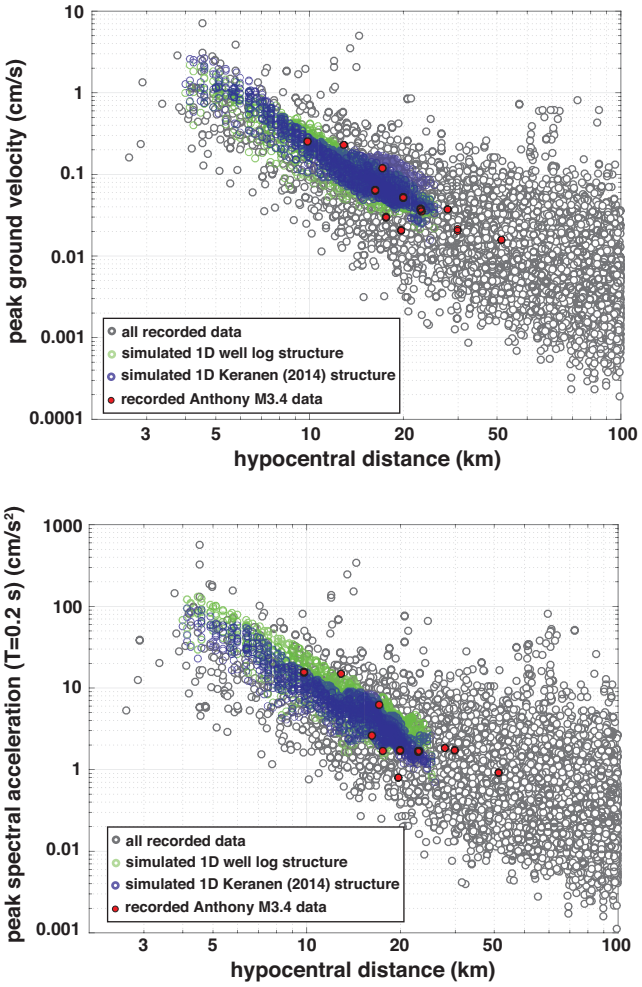

Figure 6: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred 1D material structure (green dots) and the Keranen (2014) material structure (blue dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulations performed using the preferred 1D material structure better capture the decay rate of ground motion intensities compared to the Keranen (2014) material structure.

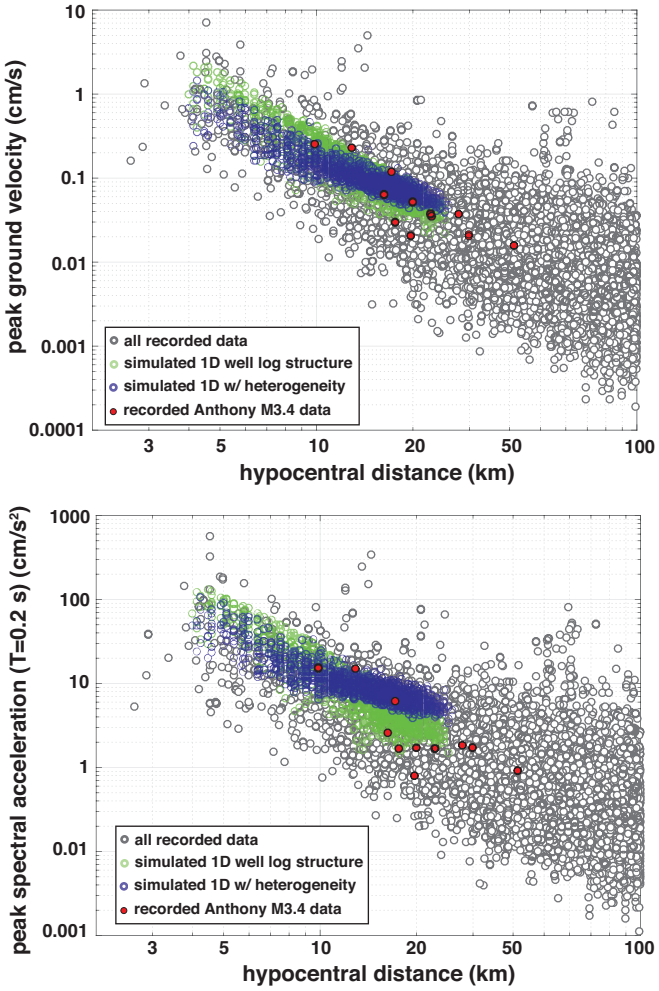

Figure 7: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as a function of hypocentral distance for all recorded ground motions in the target region (gray dots), the synthetic (using f_c= 6.4 Hz) M_w 3.4 Anthony, KS, ground motion data using the preferred 1D material structure (green dots) and preferred material structure with additional small-scale heterogeneity constrained by composite well logs (blue dots), and the recorded M_w 3.4 Anthony, KS, ground motion data (red dots). The simulations performed using the preferred 1D material structure alone better capture the decay rate of ground motion intensities compared to the simulations where small-scale heterogeneity is included, particularly at distances greater than 10 km. This difference is more pronounced for PSA(T=0.2s) than for PGV.

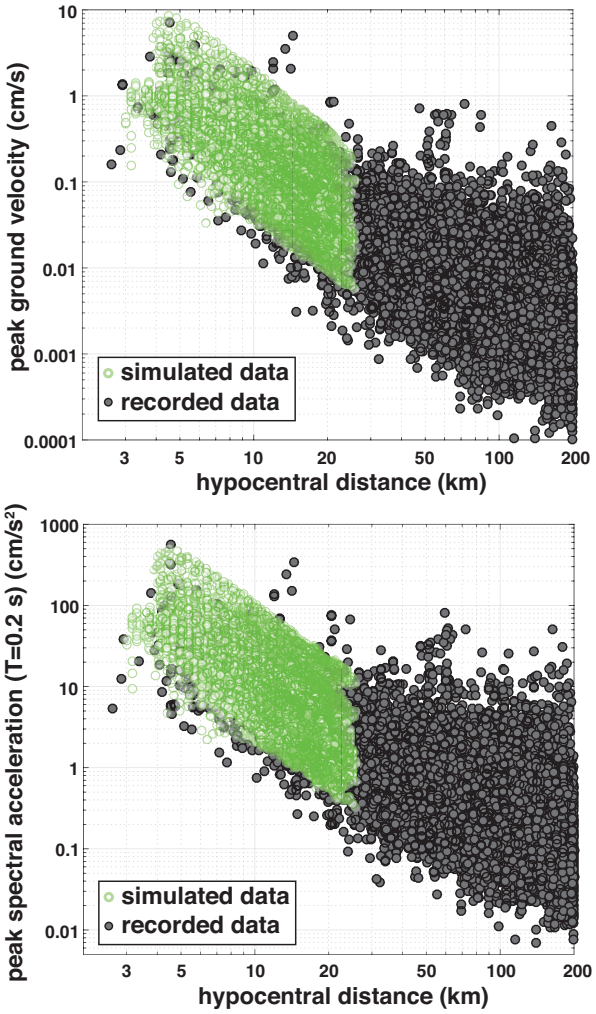

Figure 8: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area composite ground motion dataset including recorded (gray dots) and simulated (green dots) ground motions.

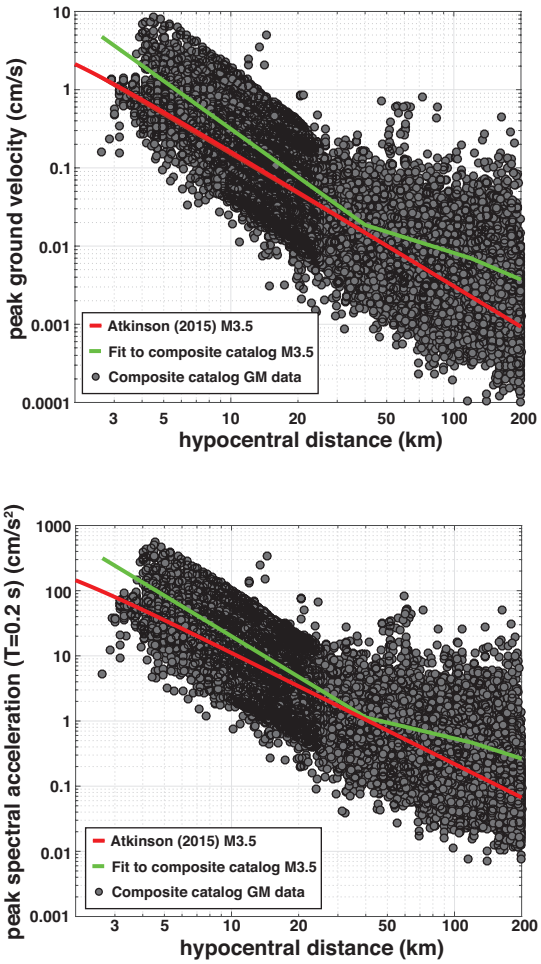

Figure 9: Peak ground velocities (top) and peak spectral accelerations (T=0.2s) (bottom) as functions of hypocentral distance for ground motion data from the M_w 3-4 Oklahoma/Kansas target area composite ground motion dataset (gray dots). Red line indicates reference Atkinson (2015) M_w 3.5 GMPE. Green line is GMPE for M_w 3.5 events constructed by fitting our composite catalog using Equation 4 (coefficients shown in Table 3).

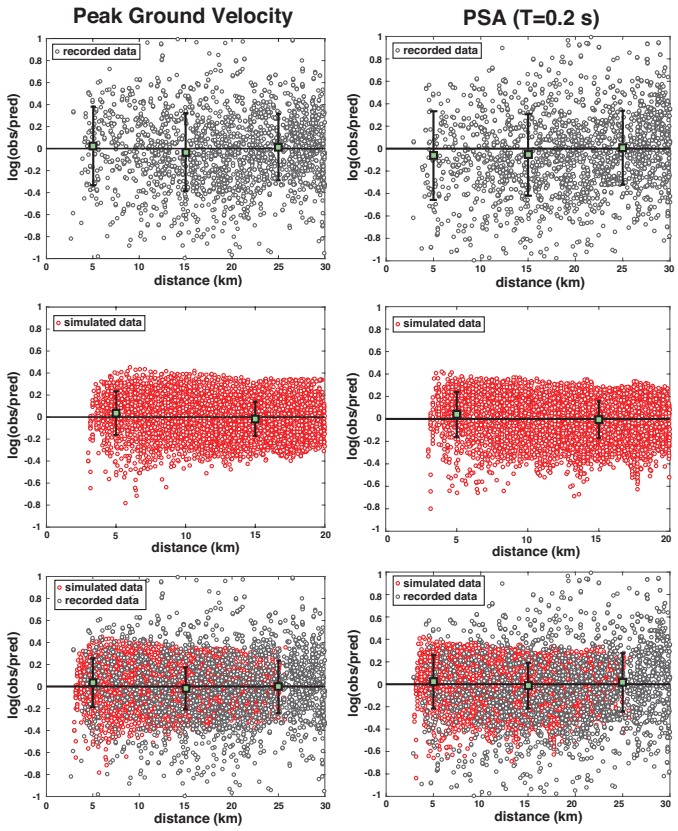

Figure 10: Plots of residuals (difference between observed and predicted in log units, binned every 10 km) for PGV (left column) and PSA(T=0.2s) (right column). Green squares indicate the mean values of each bin with errors bars indicating +/- 1 standard deviation. The top panel in each column are the residuals between the composite GMPE and the recorded data (gray dots), the middle panel are the residuals between the composite GMPE and the simulated data (red dots), and the bottom panel are the residuals between the composite GMPE and the combined dataset, where the dots are color coded as either recorded or simulated. The means are near zero or well within 1 standard deviation for all cases, indicating that our simulations are producing ground motion data that has similar amplitude and decay characteristics as recorded data.

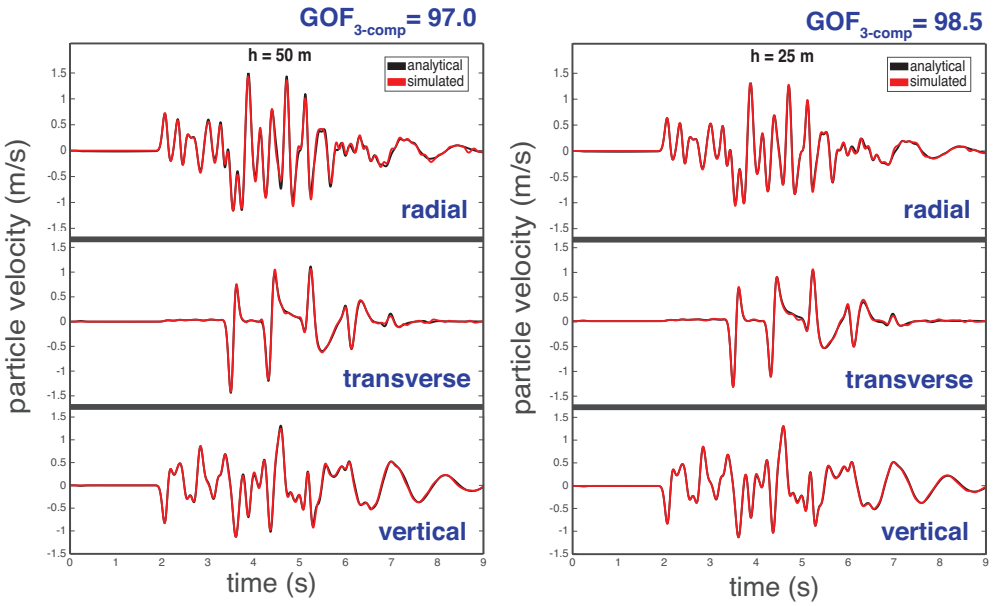

Figure A1: Particle velocity (radial, transverse, and vertical components) time series comparisons quantified using the 3-component averaged goodness-of-fit measure described in Appendix A at a station location at 10 km horizontal and 2 km vertical distance from the source between the analytic solution for the LOH.1 verification problem (black) and solutions produced by waveqlab3d (red) for grid spacing of 50 (left) and 25 m (right).