

Artificial Neural Network-Based Framework for Developing Ground-Motion Models for Natural and Induced Earthquakes in Oklahoma, Kansas, and Texas

by Farid Khosravikia, Patricia Clayton, and Zoltan Nagy

ABSTRACT

This article puts forward an artificial neural network (ANN) framework to develop ground-motion models (GMMs) for natural and induced earthquakes in Oklahoma, Kansas, and Texas. The developed GMMs are mathematical equations that predict peak ground acceleration, peak ground velocity, and spectral accelerations at different frequencies given earthquake magnitude, hypocentral distance, and site condition. The motivation of this research stems from the recent increase in the seismicity rate of this particular region, which is mainly believed to be the result of the human activities related to petroleum production and wastewater disposal. Literature has shown that such events generally have shallow depths, leading to large-amplitude shaking, especially at short hypocentral distances. Thus, there is a pressing need to develop site-specific GMMs for this region. This study proposes an ANN-based framework to develop GMMs using a selected database of 4528 ground motions, including 376 seismic events with magnitudes of 3 to 5.8, recorded over the 4- to 500-km hypocentral distance range in these three states since 2005. The results show that the proposed GMMs lead to accurate estimations and have generalization capability for ground motions with a range of seismic characteristics similar to those considered in the database. The sensitivity of the equations to predictive parameters is also presented. Finally, the attenuation of ground motions in this particular region is compared with those in other areas of North America.

Electronic Supplement: Text and figures describing the selection of the hidden layer size of the artificial neural network (ANN) models, as well as sensitivity of ANN models to modeling assumptions

INTRODUCTION

This article aims to develop ground-motion models (GMMs), which are often referred to as attenuation models, for recent

seismic events recorded in Texas, Oklahoma, and Kansas. GMMs are mathematical equations that estimate the intensity measures of ground motions as a function of earthquake magnitude, source-to-site distance, and site conditions. Since 2008, there has been a significant increase in the seismicity rate of this particular region of central and eastern North America (CENA), which is mainly believed to be as a result of human activities such as petroleum development or wastewater disposal (Hough, 2014; Frohlich *et al.*, 2016; Hornbach *et al.*, 2016; Petersen *et al.*, 2016). The November 2011 earthquake with moment magnitude M_w 5.7 in Prague, Oklahoma, the May 2012 earthquake with M_w 4.8 near Timpson, Texas, as well as the September 2016 earthquake with M_w 5.8 in Pawnee, Oklahoma, are three examples of the recent earthquakes in this particular region, which were reported to cause damage to nearby infrastructure (Ellsworth, 2013; Frohlich *et al.*, 2014; Barbour *et al.*, 2017).

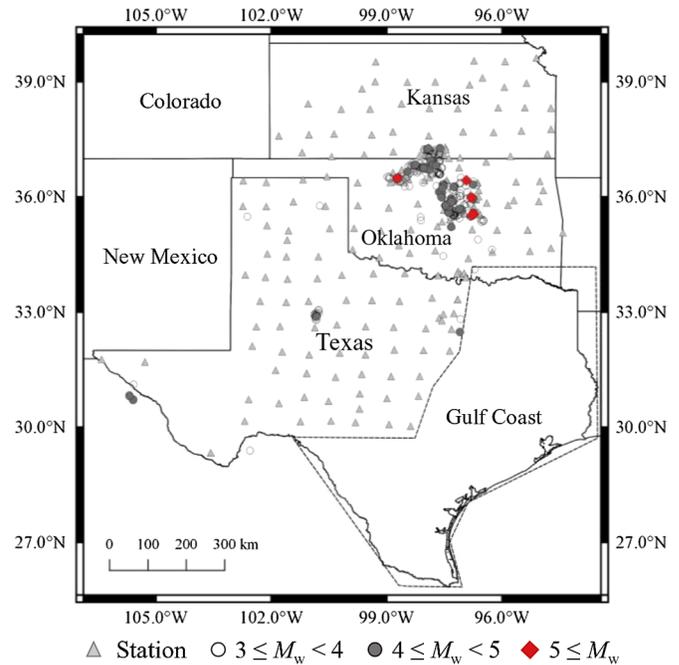
Bommer *et al.* (2016) demonstrated that because induced earthquakes generally occur at shallow depths, they tend to have large ground-motion amplitudes, especially at short hypocentral distances. Such characteristics necessitate further investigations of GMMs for this region, which are keys in investigating and mitigating the hazard proposed by natural and induced earthquakes in that area. In this regard, Atkinson (2015) assumed that the amplitudes of the ground motions from induced events are similar to those of tectonic earthquakes with similar magnitudes and hypocentral distances (R_{hypo}). Small to moderate events (M 3–6) at short hypocentral distances ($R_{\text{hypo}} < 40$ km) from the Next Generation Attenuation-West2 (NGA-West2) database (Ancheta *et al.*, 2014) were used to develop GMMs applicable for induced seismicity. Although the models resulted in good estimation of the intensity measures of the ground motions recorded from induced events, Atkinson (2015) mentioned the need for site-specific GMMs for such events. Later, Gupta *et al.* (2017) assessed ground-motion amplitudes and attenuation for small

to moderate induced and tectonic earthquakes in CENA. Atkinson and Assatourians (2017) described the conditions with which natural-earthquake models can be applied for induced seismicity application. More recently, Khosravikia, Zeinali, *et al.* (2018) showed that the CENA GMMs developed based largely on natural tectonic events cannot be properly used to estimate the intensity measures of the earthquakes in this particular region. Farhadi *et al.* (2018) evaluated the applicability of different GMMs for induced seismicity in CENA, including many GMMs developed as part of the NGA-West2 project (Bozorgnia *et al.*, 2014) and those that had already been developed for CENA. They pointed out that the performance of the models depends on the intensity measure, and no single model performs the best over all of the intensity measures. Novakovic *et al.* (2018) also recommended developing region-specific GMMs for induced seismicity, and they proposed empirically calibrated GMMs for Oklahoma.

The present study proposes new site-specific GMMs for Texas, Oklahoma, and Kansas considering a selected database of 4528 ground motions, including 376 seismic events with magnitudes of 3 to 5.8, recorded over the 4- to 500-km hypocentral distance range in these three states since 2005. The developed GMMs aim to predict peak ground acceleration (PGA), peak ground velocity (PGV), as well as 5% damped elastic pseudospectral accelerations at different periods, denoted by $PSA(T)$, in which T is the considered period. To do so, this study, unlike conventional empirical methods that used regression analysis, proposes a framework in which an artificial neural network (ANN) is used as statistical method.

In conventional empirical methods, the GMMs are built using a regression analysis with predefined linear equations to correlate the ground-motion intensity measures to the predictive parameters. The assumed equations are generally based on physical concepts, which can increase the accuracy of prediction of the statistical method when limited data are available. However, such equations often limit the ability of the method to efficiently simulate complex and unknown behaviors of the ground-motion intensity measures. Some researchers (Güllü and Erçelebi, 2007; Alavi and Gandomi, 2011) have suggested that linear regression analysis has major drawbacks related to the idealization of complex processes, approximation, and averaging of widely varying prototype conditions. However, ANN has the capability of adaptively learning from experience and extracting various discriminators in pattern recognition without predefined functions (Perlovsky, 2001). Therefore, it has the potential to provide more robust predictive models when extensive datasets are used. In the literature, ANN has also been shown to be a promising method in predicting ground-motion characteristics (Kerh and Ting, 2005; Güllü and Erçelebi, 2007; Ahmad *et al.*, 2008; Alavi and Gandomi, 2011; Derras *et al.*, 2012, 2014, 2016).

It is commonly asserted that ANN models behave like “black-box” systems and are not able to show the underlying principles of the prediction. Likewise, it is often believed that ANN models cannot be used by others to predict the outputs unless they retrain their own ANN model, which would



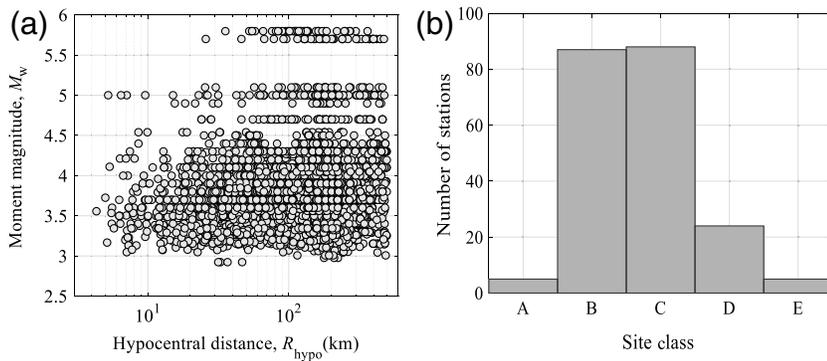
▲ **Figure 1.** Geographic distribution of the events and stations considered in this study. The color version of this figure is available only in the electronic edition.

impede the dissemination of new knowledge generated with ANN models. This issue has been addressed here by converting the ANN models into simple mathematical equations that anyone can easily use to predict ground-motion intensity measures given the input parameters without retraining any ANN model.

The present study evaluates the generalization capability of the proposed GMMs using different criteria presented in the literature. In addition, the effects of the predictive parameters on predicting the intensity measures are also assessed through a sensitivity analysis. Finally, the attenuation of ground motions in the study region is compared with those developed for small to moderate earthquakes of western America, CENA, and those developed to be applicable to induced seismicity.

GROUND-MOTION DATABASE

The database considered in this study (see [Data and Resources](#)) consists of 4528 ground-motion recordings that correspond to 376 different natural and induced seismic events recorded at 209 different seismic stations in Texas, Oklahoma, and Kansas since 2005. Figure 1 shows the location of the considered events and the seismic stations that recorded the considered ground motions. As seen in the figure, all the considered seismic events have magnitudes larger than 3, and they mainly occurred in Oklahoma, especially those with larger magnitudes, that is, $M_w > 5$. The seismic stations are located in all of these three states, resulting in adequate ground-motion recordings in all the three states. As shown in the figure, the seismic stations



▲ **Figure 2.** Characteristics of the ground-motion database, consisting of 4528 ground-motion records from 376 seismic events: (a) magnitude–distance distribution of the database and (b) station number per American Society of Civil Engineers (ASCE) 7-16 site classification.

located on the Gulf Coast Plain are not considered in this study because the Gulf Coast Plain consists of significantly deeper sediments, resulting in different attenuation behavior than the surrounding region (Electric Power Research Institute [EPRI], 2004).

The developed GMMs estimate PGA and PGV as well as PSA values at 20 different periods varying between 0.05 and 2.5 s given earthquake moment magnitude M_w , the hypocentral distance R_{hypo} , and the averaged shear-wave velocity over the top 30 m of soil V_{S30} . The GMMs are based on the geometric mean of the horizontal-component ground-motion amplitudes, which is consistent with previous studies (Yenier and Atkinson, 2015). For the selected ground motions, M_w is either derived from Incorporated Research Institutions for Seismology (IRIS) database (see Data and Resources), or calculated using 1-Hz PSA amplitudes of the vertical component of the ground-motion records (Atkinson and Mahani, 2013; Atkinson *et al.*, 2014). The V_{S30} values for most of the considered ground-motion recording stations are taken from Zalachoris *et al.* (2017), which were determined using the P -wave seismogram method. For a few seismic station locations where P -wave seismogram V_{S30} estimates were not available, the V_{S30} estimates provided by Parker *et al.* (2017) are used, which were determined over the CENA using a hybrid slope-geology proxy method. Finally, the hypocentral distance R_{hypo} , which is defined from the epicentral location and focal depth, is used to allow GMMs to correctly reflect the source-to-site distance attributes of such small to moderate shallow depth earthquakes.

Figure 2a shows the magnitude–distance distribution of the considered ground motions. As seen in the figure, M_w varies between 3.0 and 5.8, representing small to moderate earthquakes. R_{hypo} varies between 4 and 500 km, and approximately 856 records, or 18.9%, have R_{hypo} less than 50 km. Figure 2b shows the number of the stations according to American Society of Civil Engineers (ASCE) 7-16 (ASCE, 2016) site classification. For example, 92 of 209 seismic stations are located on site classes A and B ($760 \text{ m/s} < V_{S30}$), representing rock site conditions according to ASCE 7-16 site

classification. It is observed that V_{S30} varies between 122 and 1706 m/s, representing very soft soil to hard rock, respectively. Furthermore, the hypocentral depth d_f of these seismic events is generally less than 10 km, indicating shallow depth earthquakes. The ranges and statistics of the observed predictive parameters as well as intensity measures for the ground motions considered in this study are shown in Table S1 (available in the electronic supplement to this article). In the following section, the proposed ANN-based ground-motion models for this particular region of CENA are presented and discussed.

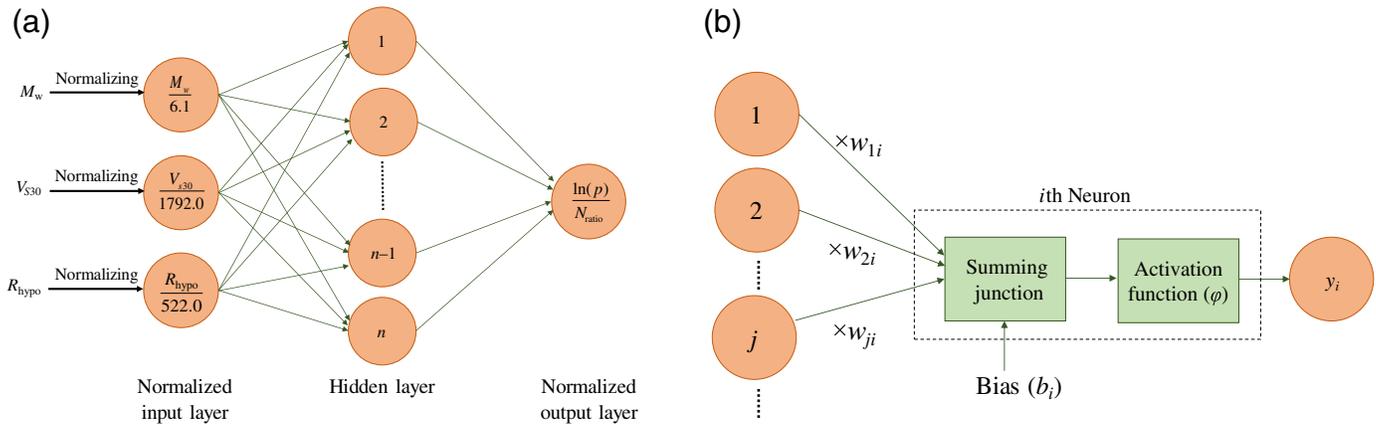
PROPOSED ANN-BASED GROUND-MOTION MODELS

This study uses multilayer perceptron network, which is a kind of ANN with a feed-forward architecture (Cybenko, 1989), to derive site-specific GMMs for Texas, Oklahoma, and Kansas. Figure 3a presents the schematic view of the ANN framework proposed in this study. As seen in the figure, the network here is organized in three different layers as input, hidden, and output layers. The input layer consists of the normalized values of M_w , V_{S30} , and R_{hypo} , and the output layer comprises the normalized output parameter in natural log space. Normalized values of any variable (input or output) are computed by dividing the value by its corresponding normalization ratio N_{ratio} , presented in Table S1. The normalization process prevents saturation of the activation functions within the neural network, thereby resulting in a better estimation of the results. The sensitivity of the ANN models to the normalization process is available in the electronic supplement. It is here assumed that all the proposed ANN models comprise one hidden layer with the same number of neurons. By training different ANN models considering different hidden layer sizes and evaluating their performance (details are available in the electronic supplement), it is concluded that four neurons is the optimal hidden layer size for all outputs. The strength of the connection of neurons in each layer to the neurons of other layers is determined by connection weights.

The general structure of a neuron is shown schematically in Figure 3b. As seen, each neuron receives the outputs of the neurons in the previous layer as inputs. In addition to the outputs of the previous layer, a bias parameter is introduced as an input to each neuron, which acts in the same fashion as the intercept in a regression model. Within each neuron, the summation of the weighted inputs and the bias parameter passes through the activation function to compute the output of the neuron y_i as follows:

$$y_i = \varphi \left(\sum w_{ji} \times x_j + b_i \right), \quad (1)$$

in which x_j is the value of input neuron j ; w_{ji} is the connection weight of the j th neuron from the input layer and the



▲ **Figure 3.** Schematic representation of (a) the artificial neural network (ANN) model and (b) the *i*th neuron of the hidden layer. The color version of this figure is available only in the electronic edition.

considered *i*th neuron; b_i is the bias defined for *i*th neuron; and ϕ is the activation function for the neurons. In the present study, a log-sigmoid function of $\phi(x) = 1/(1 + e^{-x})$ and a linear function of $\phi(x) = x$ are considered as the activation functions of the neurons in the hidden and output layers, respectively. The sensitivity of the ANN models to the selection of the activation functions is available in the ㉔ electronic supplement. The output y_i , in turn, will be used as the input for the neurons in the following layer, if present.

The connection weights as well as bias terms for each neuron are adjusted through the training process of the network. To do so, the Levenberg–Marquardt backpropagation algorithm (Marquardt, 1963), which is a standard nonlinear least-squares optimization algorithm, is implemented. In the training process, the ground-motion database is divided into three different subsets as follows: (1) training subset, which is used to adjust the weights and bias values on the ANN models; (2) validation subset, which is used to minimize overfitting of the models by checking their generalization capability on data they did not train on (overfitting is a modeling error that occurs when the model is too closely fit to a particular set of data and may therefore fail to reliably predict future observations); and (3) testing subset, which is used for testing the final algorithm to confirm the actual predictive power of the models for future data.

In the literature, most of the ANN-based studies randomly divide the ground-motion records into the above-mentioned subsets. However, the ground-motion records from individual earthquake are correlated because of higher-order source effects not accounted for in the model. Thus, there is a moderate risk of overfitting because the ground-motion records in the validation and testing subsets may be correlated with the ones in the training subset, which correspond to the same earthquake. To address this issue, in the present study, instead of randomly dividing the ground-motion records, the 376 seismic events are randomly divided into training, validation, and testing subsets, which approximately consist of 60%, 20%, and 20% of the whole dataset, respectively. Then, for each subset, the

ground-motion records associated with seismic events in that subset are considered for training process.

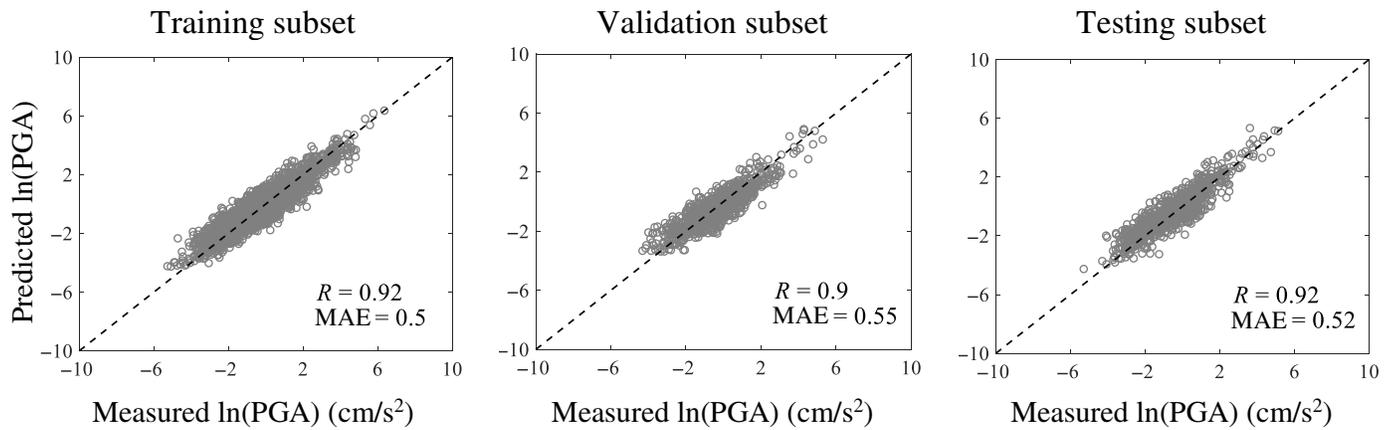
After training the ANN model shown in Figure 3a, it is now turned into mathematical formulation as follows:

$$\frac{\ln(p)}{N_{\text{ratio}}} = b + \sum_{i=1}^4 v_i \times \left[\frac{1}{1 + \exp\left[-\left(w_{1i} \times \frac{M_w}{6.1} + w_{2i} \times \frac{V_{S30}}{1792.0} + w_{3i} \times \frac{R_{JB}}{522.0} + b_i\right)\right]} \right], \quad (2)$$

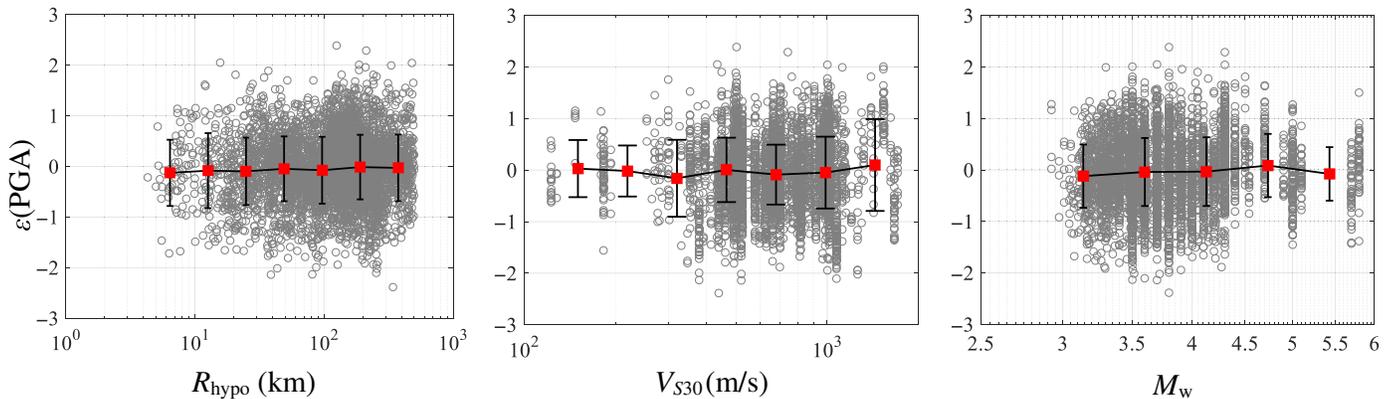
in which p is the predicted value of intensity measure; b is the bias value of output neuron; v_i denotes the connection weights between the *i*th neuron from the hidden layer and output neuron; w_{1i} , w_{2i} , and w_{3i} are the connection weights entering *i*th neuron of the hidden layer from the input layer; and b_i denotes the bias of the hidden layer neurons. The values of w_{1i} , w_{2i} , w_{3i} , and b_i for different outputs are listed in ㉔ Tables S2 and S3, and the values of v_i and b for different outputs are listed in ㉔ Table S4. Using equation (2) and the coefficients presented in the tables, one can easily estimate the ground-motion intensity measures in the study region given the input parameters of the ground motions without retraining the proposed ANN models. The developed GMMs are also implemented into ready-to-use Microsoft Excel Spreadsheet available in ㉔ ANN_Calculation. The reliability and generalization capability of the proposed GMMs are investigated in the following section of the article.

PERFORMANCE ANALYSIS AND MODEL VALIDITY

This section addresses the performance and model validity of the developed ANN models for different intensity measures. To evaluate the performance of the ANN models in predicting the intensity measures, the correlation between the predicted and target values for PGA is shown in Figure 4. Dashed lines in



▲ **Figure 4.** Measured versus predicted values for peak ground acceleration (PGA) for the training, validation, and testing subsets. The values of correlation coefficient R and mean absolute error (MAE) of each subset are also shown in the figure.



▲ **Figure 5.** Residuals of the proposed ground-motion model (GMM) for PGA in respect to input parameters. Squares depict the mean residual and its standard deviation in logarithmically spaced distance bins. The color version of this figure is available only in the electronic edition.

each plot indicate the ideal case in which the predicted values are exactly the same as the target values. As seen in the figure, the data are close to the ideal line, indicating a strong correlation between the estimated and measured values. In addition, Figure 5 plots the distribution of the residuals, ε , defined as the natural log of the ratio of the measured ground-motion parameter to its predicted value, for PGA in relation to the input parameters. The ideal value for this parameter is zero, indicating that the GMM precisely predicts the measured parameter. Residual values above and below zero, respectively, demonstrate that the GMM underestimates and overestimates the measured parameters. As seen in Figure 5, the mean of the residuals for all plots is approximately zero, indicating that the proposed ANN model for PGA on average accurately predicts the intensity measures for the considered ground motions. The same trends are observed for all other intensity measures.

To evaluate external validity of the models, which is defined as the reliability of the model in predicting future data, different sets of criteria available in the literature, as shown in Table 1, are checked. Large values of the correlation coefficients

R between the target and predicted values ($R > 0.8$) demonstrate that the model provides strong correlation between the estimates and the target values (Smith, 1986). In fact, large values of R especially for testing subset (i.e., the subset that the model does not see during training) indicate that they can reliably be used to determine principal ground-motion intensity measures for future data (Pan *et al.*, 2009). Moreover, Golbraikh and Tropsha (2002) suggested that at least one of the slopes of the regression of targets against estimates through the origin and that of estimates against targets through the origin for the testing subset (k and k' , respectively) should be between 0.85 and 1.15. They also suggested that at least one of the correlation coefficients of these regression lines (R_o^2 and $R_o'^2$) should be close to 1.0. To ensure that the model has good predictive power, they proposed that at least either R_o^2 or $R_o'^2$ should be close to R^2 . Finally, Roy and Roy (2008) suggested that values of the modified coefficient of correlation R_m^2 larger than 0.5 can be considered as an indicator of good generalization capability of the model.

Table 1
Considered Criteria for Checking the External Validity of the Models

Criteria	Suggested by
$R = \frac{\sum_{i=1}^m (t_i - \bar{t})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^m (t_i - \bar{t})^2 \sum_{i=1}^m (p_i - \bar{p})^2}} \geq 0.8$	Smith (1986)
$0.85 \leq k = \frac{\sum_{i=1}^m (t_i \times p_i)}{\sum_{i=1}^m p_i^2} \leq 1.15$ or $0.85 \leq k' = \frac{\sum_{i=1}^m (t_i \times p_i)}{\sum_{i=1}^m t_i^2} \leq 1.15$	Golbraikh and Tropsha (2002)
$R_o^2 = 1 - \frac{\sum_{i=1}^m (p_i - t_i^o)^2}{\sum_{i=1}^m (p_i - \bar{p})^2}$ or $R_o^2 = 1 - \frac{\sum_{i=1}^m (t_i - p_i^o)^2}{\sum_{i=1}^m (t_i - \bar{t})^2}$ should close to 1	Golbraikh and Tropsha (2002)
$m_{\text{index}} = \left \frac{R^2 - R_o^2}{R^2} \right \leq 0.1$ or $n_{\text{index}} = \left \frac{R^2 - R_o^2}{R^2} \right \leq 0.1$	Golbraikh and Tropsha (2002)
$R_m^2 = R^2 (1 - \sqrt{ R^2 - R_o^2 }) \geq 0.5$	Roy and Roy (2008)

t_i and p_i are the target and predicted output values, respectively; m is the number of samples; \bar{t} and \bar{p} are the average of the target and predicted output values, respectively; and $t_i^o = k \times p_i$ and $p_i^o = k' \times t_i$.

Table 2
Performance Analysis of Artificial Neural Network Models

Parameters	R_{training}	$R_{\text{Validation}}$	R_{testing}	k	k'	R_o^2	R_o^2	m_{index}	n_{index}	R_m^2
PGV	0.923	0.906	0.918	1.020	0.941	0.986	0.921	0.171	0.094	0.523
PGA	0.924	0.907	0.916	0.982	0.843	0.998	0.903	0.189	0.076	0.505
PSA (0.05 s)	0.927	0.902	0.912	0.955	0.867	0.988	0.912	0.188	0.097	0.503
PSA (0.06 s)	0.927	0.917	0.918	0.953	0.868	0.947	0.923	0.123	0.094	0.572
PSA (0.08 s)	0.916	0.917	0.922	1.034	0.832	0.969	0.899	0.140	0.059	0.556
PSA (0.10 s)	0.916	0.912	0.919	1.049	0.805	0.987	0.907	0.169	0.074	0.525
PSA (0.15 s)	0.904	0.905	0.905	1.035	0.789	0.983	0.881	0.200	0.076	0.487
PSA (0.20 s)	0.890	0.906	0.891	1.051	0.742	0.967	0.911	0.217	0.146	0.465
PSA (0.25 s)	0.898	0.899	0.912	0.985	0.825	0.980	0.909	0.178	0.093	0.511
PSA (0.30 s)	0.907	0.882	0.893	0.988	0.773	0.975	0.875	0.222	0.097	0.462
PSA (0.35 s)	0.896	0.914	0.899	0.972	0.801	0.950	0.887	0.176	0.098	0.503
PSA (0.40 s)	0.901	0.902	0.915	1.019	0.834	0.990	0.930	0.182	0.111	0.510
PSA (0.45 s)	0.910	0.917	0.916	0.974	0.875	0.994	0.898	0.185	0.071	0.508
PSA (0.50 s)	0.916	0.887	0.909	1.024	0.880	0.974	0.904	0.180	0.095	0.507
PSA (0.60 s)	0.927	0.902	0.917	1.028	0.863	0.998	0.909	0.186	0.080	0.508
PSA (0.70 s)	0.923	0.941	0.938	0.968	0.970	0.957	0.998	0.087	0.134	0.636
PSA (0.80 s)	0.930	0.897	0.943	1.023	0.924	0.999	0.938	0.123	0.055	0.595
PSA (0.90 s)	0.929	0.945	0.934	0.966	0.988	0.955	0.999	0.096	0.147	0.620
PSA (1.00 s)	0.939	0.912	0.949	1.012	0.949	0.990	0.979	0.099	0.088	0.632
PSA (1.50 s)	0.938	0.950	0.940	0.986	0.990	0.968	0.990	0.095	0.120	0.627
PSA (2.00 s)	0.949	0.928	0.949	1.019	0.967	0.987	0.971	0.097	0.079	0.635
PSA (2.50 s)	0.953	0.933	0.946	1.020	0.969	0.986	0.971	0.102	0.085	0.625

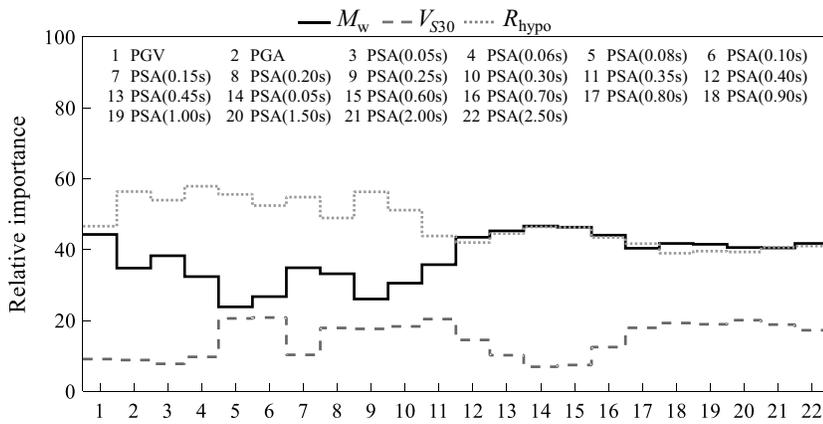
PGA, peak ground acceleration; PGV, peak ground velocity; PSA, pseudospectral acceleration.

Here, the above-mentioned external validation criteria are checked for the proposed ANN models, and the results are shown in Table 2. As seen in the table, the derived models pass all the above-mentioned criteria, which denotes that the developed model can be reliably used to predict the future data. Similar to other statistical methods, the generalization capability of the proposed ANN-based GMMs is limited to the range of the input characteristics considered in this study, and

caution should be exercised when extrapolating beyond the range constrained by the input data.

SENSITIVITY ANALYSIS

Here, the sensitivity of the ANN models to the predictor variables (M_w , V_{S30} , and R_{hypo}) is evaluated. To do so, Garson's algorithm (Garson, 1991) is used to compute the contribution



▲ **Figure 6.** Contribution of the predictive parameters in the developed ANN models.

of each input variable in the ANN output. In this algorithm, the input-hidden and hidden-output weights of the trained ANN models are partitioned, and the absolute values of the weights are taken to calculate the relative importance values. The relative importance values for each ANN model are computed and presented in Figure 6. As shown, for PGA and PSA values at short periods ($T < 0.4$ s), R_{hypo} is the most important parameter, but for PGV and PSA values at long periods ($T > 0.4$ s), R_{hypo} and M_w have similar contributions. Moreover, regardless of which intensity measure is discussed, V_{S30} has the least contribution in predicting the ground-motion intensity measures. It is worth noting that for longer periods ($T > 1$ s), V_{S30} has relatively more contribution ($\sim 20\%$), reflecting amplification effects caused by regional geology.

COMPARISON OF THE GROUND-MOTION MODELS

In this section, the GMMs developed for Texas, Oklahoma, and Kansas are compared with three different sets of GMMs available in the literature as:

1. GMMs developed by Boore *et al.* (2014; hereafter, BSSA14) for western North America as part of the NGA-West2 (Bozorgnia *et al.*, 2014) project with more focus on small to moderate magnitude tectonic earthquakes. It is believed that these events are similar to induced events in terms of key features such as magnitude and distance scaling of ground-motion amplitudes.
2. GMMs developed by Atkinson (2015; hereafter, A15) using small to moderate events (M 3–6) at short hypocentral distances ($R_{\text{hypo}} < 40$) from the NGA-West2 database (Ancheta *et al.*, 2014). The GMMs are developed to be applicable for induced seismicity assuming that the amplitude of motions from induced events is similar to that of tectonic earthquakes with the same magnitudes and hypocentral distances.
3. GMMs developed by Hassani and Atkinson (2015; hereafter, HA15) for CENA. They updated the GMMs

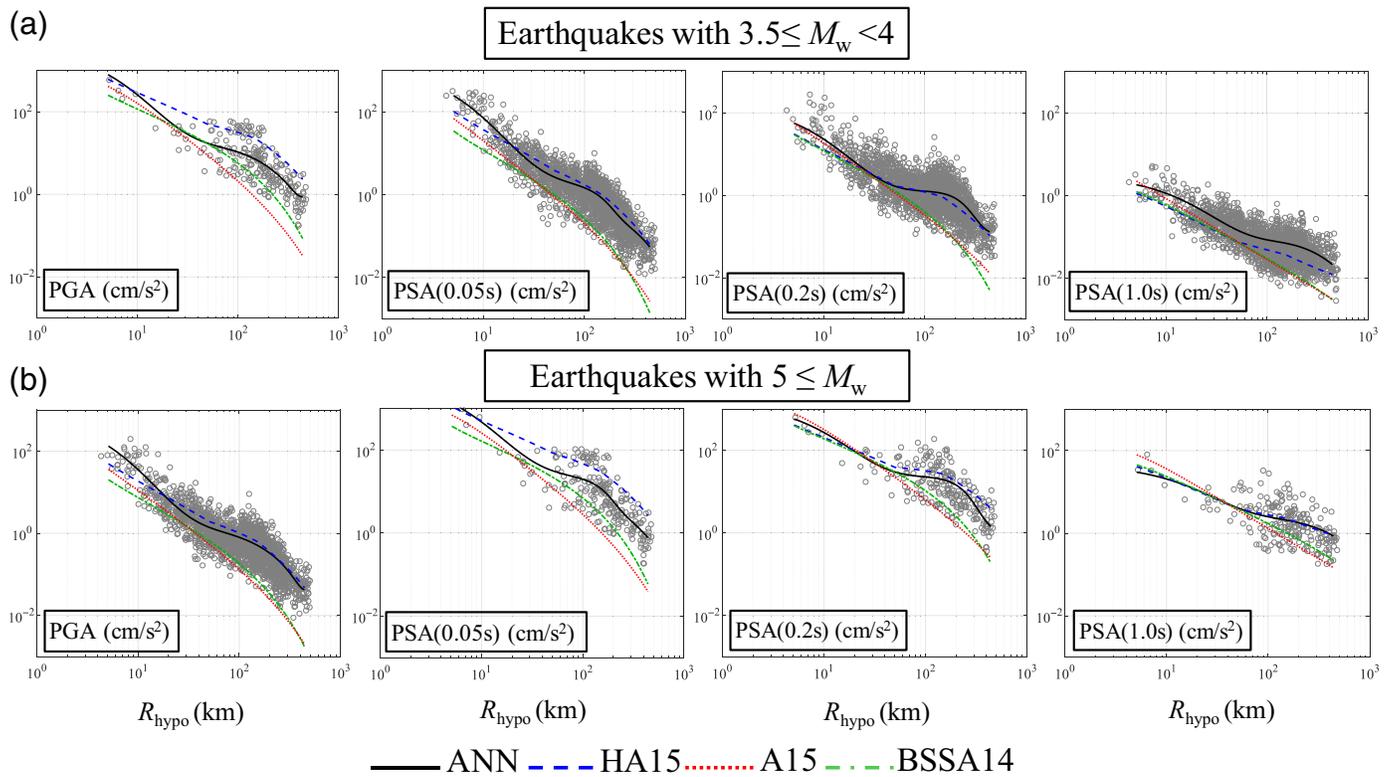
developed by Boore *et al.* (2014) for western North America, applying an adjustment factor to fit the models to the CENA ground-motion database.

Figure 7 demonstrates the intensity–distance relations of the ground motions from the Texas, Oklahoma, and Kansas dataset, as well as the corresponding attenuation curves derived from proposed and above-mentioned GMMs. To plot the attenuation curves from BSSA14 and HA15 in hypocentral distance, it is assumed that $R_{\text{hypo}} \approx \sqrt{R_{\text{JB}}^2 + 5.1^2}$, in which 5.1 km is the average hypocentral depth of the considered data, and R_{JB} is the closest distance to the surface projection of the rupture (Joyner–Boore distance). In fact, similar to pre-

vious studies in the literature (Atkinson, 2015; Hassani and Atkinson, 2015; Khosravikia, Zeinali, *et al.*, 2018), it is assumed that the fault is small enough so that the epicentral distance is approximately equal to the Joyner–Boore distance. As seen in Figure 7, the Texas, Oklahoma, and Kansas motions tend to follow a pronounced trilinear amplitude decay function at regional distances, and ANN models properly capture this behavior. This behavior has been also observed by Khosravikia, Zeinali, *et al.* (2018) and Novakovic *et al.* (2018).

First, the proposed GMMs are compared with those developed for CENA (i.e., HA15). As seen in the figure, for earthquakes with $3.5 \leq M_w < 4.0$, Texas, Oklahoma, and Kansas tend to have higher amplitudes at short distances ($R_{\text{hypo}} < 20$ km) compared with other areas of CENA. This observation is mainly because of the fact that the proposed GMMs are adjusted to the potentially induced seismic events in Texas, Oklahoma, and Kansas, which generally consist of shallower depth earthquakes compared with other areas of CENA (Atkinson and Assatourians, 2017; Khosravikia, Zeinali, *et al.*, 2018). At large distances ($R_{\text{hypo}} > 20$ km), ANN results in similar attenuations as HA15 for PGA and PSA at short periods. However, at longer periods ($T > 0.5$ s), as seen in the plot for PSA(1.0 s), Texas, Oklahoma, and Kansas amplitudes tend to be higher regardless of the distance of the site from the earthquake epicenter, which probably reflects the amplification effects caused by the regional geology of the study region. For earthquakes with larger magnitudes ($M_w > 5$), the proposed amplitudes for PGA and PSA values at shorter periods ($T < 0.5$ s) tend to be lower than those of HA15, reflecting that the application of the CENA GMMs for the study region is likely to overestimate the intensity measures. However, for PSA values at long periods, HA15 and ANN models result in similar attenuations.

Next, the proposed GMMs are compared with those of A15 and BSSA15. As seen in Figure 7, A15 leads to fairly good estimates of the attenuations in Texas, Oklahoma, and Kansas, especially for larger earthquakes ($M_w > 5$) and at shorter hypocentral distances ($R_{\text{hypo}} < 20$ km). However, at larger distances, BSSA14 and A15 predict amplitudes significantly lower



▲ **Figure 7.** Intensity measure to distance relations of the GMMs determined in this study in comparison with [Hassani and Atkinson \(2015;](#) hereafter, HA15) GMMs developed for central and eastern North America, [Atkinson \(2015;](#) hereafter, A15) GMMs developed for small to moderate events at short hypocentral distances with applicability to induced seismicity, and [Boore et al. \(2014;](#) hereafter, BSSA14) developed as part of the Next Generation Attenuation-West2 project. All GMMs are plotted for $V_{S30} = 760$ m/s as well as $M_w = 3.7$ for (a) and $M_w = 5.3$ for (b). The color version of this figure is available only in the electronic edition.

than those observed in the study region, reflecting slower attenuations for Texas, Oklahoma, and Kansas motions at regional distances compared with small to moderate earthquakes in western North America.

CONCLUSION

The recent increase in the seismicity rate of the states of Texas, Oklahoma, and Kansas driven by increased oil and gas production necessitates further investigation of the seismic hazard in CENA. This study presents site-specific GMMs, which are a valuable tool for evaluating and mitigating hazard for that region. The proposed GMMs are mathematical equations that predict PGA, PGV, and 5% damped elastic PSA at different periods given earthquake magnitude, hypocentral distance, and average shear-wave velocity over the top 30 m of soil. This study used a selected database of 4528 ground motions, including 376 seismic events with magnitudes of 3 to 5.8, recorded over the 4- to 500-km hypocentral distance range in these three states since 2005.

This study, unlike many other studies in the literature that used linear regression analysis, outlines a framework in which ANN is used as the statistical method to develop the GMMs. Regression analysis is conducted using predefined linear and

nonlinear equations. Although such equations are based on physical justifications and are helpful in interpreting the derived GMMs, they may limit the predictive power of the GMMs in simulating the complex behavior of ground-motion characteristics. ANN, in contrast, is able to adaptively learn from a dataset without any presumed behaviors. The prediction reliability and generalization capability of the developed ANN models are evaluated against different criteria available in the literature, and the results show that they provide reliable estimates of intensity measures.

It is observed that Texas, Oklahoma, and Kansas motions tend to follow a pronounced trilinear amplitude decay function at regional distances. The GMMs developed for this region are also compared with GMMs developed for small-to-moderate earthquakes of western America, CENA, and those developed to be applicable to induced seismicity. For instance, it is concluded that Texas, Oklahoma, and Kansas amplitudes at long periods or short distances tends to be higher than CENA amplitudes for earthquake with magnitudes of 3.5 to 4. It should be noted that the validity of the proposed ANN models, similar to other data-driven models, is limited to the range of the input characteristics considered in the ground-motion dataset. The use of these models outside the range of variability of the original dataset is not recommended.

DATA AND RESOURCES

The database of ground motions for the present study was retrieved from the Incorporated Research Institutions for Seismology (IRIS) database (<https://www.iris.edu/hq/>, last accessed September 2017) and was processed by Zalachoris and Rathje from The University of Texas at Austin. More information about the database is available at Khosravikia, Potter, *et al.* (2018). The MATLAB computational platform was used to train artificial neural network models (www.mathworks.com/products/matlab, last accessed March 2018). ☒

ACKNOWLEDGMENTS

This work was financially supported by the Texas Department of Transportation (TxDOT) through Grant Number 0-6916, the State of Texas through the TexNet Seismic Monitoring Project, and the Industrial Associates of the Center for Integrated Seismic Research (CISR) at the Bureau of Economic Geology of the University of Texas. The opinions and findings expressed herein are those of the authors and not the sponsors.

REFERENCES

- Ahmad, I., M. H. El Naggari, and A. N. Khan (2008). Neural network based attenuation of strong motion peaks in Europe, *J. Earthq. Eng.* **12**, no. 5, 663–680.
- Alavi, A. H., and A. H. Gandomi (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing, *Comput. Struct.* **89**, nos. 23/24, 2176–2194.
- American Society of Civil Engineers (ASCE) (2016). Minimum design loads and associated criteria for buildings and other structures, *ASCE/SEI 7*, Reston, Virginia.
- Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S.-J. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, *et al.* (2014). NGA-West2 database, *Earthq. Spectra* **30**, no. 3, 989–1005.
- Atkinson, G. M. (2015). Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards, *Bull. Seismol. Soc. Am.* **105**, no. 2A, 981–992.
- Atkinson, G. M., and K. Assatourians (2017). Are ground-motion models derived from natural events applicable to the estimation of expected motions for induced earthquakes? *Seismol. Res. Lett.* **88**, no. 2A, 430–441.
- Atkinson, G. M., and A. B. Mahani (2013). Estimation of moment magnitude from ground motions at regional distances, *Bull. Seismol. Soc. Am.* **103**, no. 1, 107–116.
- Atkinson, G. M., D. W. Greig, and E. Yenier (2014). Estimation of moment magnitude (M) for small events ($M < 4$) on local networks, *Seismol. Res. Lett.* **85**, no. 5, 1116–1124.
- Barbour, A. J., J. H. Norbeck, and J. L. Rubinstein (2017). The effects of varying injection rates in Osage County, Oklahoma, on the 2016 M_w 5.8 Pawnee earthquake, *Seismol. Res. Lett.* **88**, no. 4, 1040–1053.
- Bommer, J. J., B. Dost, B. Edwards, P. J. Stafford, J. van Elk, D. Doornhof, and M. Ntinalexis (2016). Developing an application-specific ground-motion model for induced seismicity, *Bull. Seismol. Soc. Am.* **106**, 158–173.
- Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra* **30**, no. 3, 1057–1085.
- Bozorgnia, Y., N. A. Abrahamson, L. Al Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D.M. Boore, K. W. Campbell, B. S.-J. Chiou, *et al.* (2014). NGA-West2 research project, *Earthq. Spectra* **30**, no. 3, 973–987.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, no. 4, 303–314.
- Derras, B., P. Y. Bard, and F. Cotton (2014). Towards fully data driven ground-motion prediction models for Europe, *Bull. Earthq. Eng.* **12**, no. 1, 495–516.
- Derras, B., P.-Y. Bard, and F. Cotton (2016). Site-condition proxies, ground motion variability, and data-driven GMPEs: Insights from the NGA-West2 and RESORCE data sets, *Earthq. Spectra* **32**, no. 4, 2027–2056.
- Derras, B., P.-Y. Bard, F. Cotton, and A. Bekkouche (2012). Adapting the neural network approach to PGA prediction: An example based on the KiK-net data, *Bull. Seismol. Soc. Am.* **102**, no. 4, 1446–1461.
- Electric Power Research Institute (EPRI) (2004). *CEUS Ground Motion Project Final Report 1009684*, EPRI, Palo Alto, California.
- Ellsworth, W. L. (2013). Injection-induced earthquakes, *Science* **341**, no. 6142, 1225942.
- Farhadi, A., S. Pezeshk, and N. Khoshnevis (2018). Assessing the applicability of ground-motion models for induced seismicity application in central and eastern North America, *Bull. Seismol. Soc. Am.* **108**, no. 4, 2265–2277.
- Frohlich, C., H. Deshon, B. Stump, C. Hayward, M. Hornbach, and J. I. Walter (2016). A historical review of induced earthquakes in Texas, *Seismol. Res. Lett.* **87**, no. 4, 1–17.
- Frohlich, C., W. Ellsworth, W. A. Brown, M. Brunt, J. Luetgert, T. MacDonald, and S. Walter (2014). The 17 May 2012 M 4.8 earthquake near Timpson, East Texas: An event possibly triggered by fluid injection, *J. Geophys. Res.* **119**, no. 1, 581–593.
- Garson, G. D. (1991). Interpreting neural network connection weights, *AI Expert* **6**, 47–51.
- Golbraikh, A., and A. Tropsha (2002). Beware of q^2 !, *J. Mol. Graph. Model.* **20**, no. 4, 269–276.
- Güllü, H., and E. Erçeşlebi (2007). A neural network approach for attenuation relationships: An application using strong ground motion data from Turkey, *Eng. Geol.* **93**, no. 3, 65–81.
- Gupta, A., J. W. Baker, and W. L. Ellsworth (2017). Assessing ground-motion amplitudes and attenuation for small-to-moderate induced and tectonic earthquakes in the central and eastern United States, *Seismol. Res. Lett.* **88**, no. 5, 1379–1389.
- Hassani, B., and G. M. Atkinson (2015). Referenced empirical ground-motion model for eastern North America, *Seismol. Res. Lett.* **86**, 477–491.
- Hornbach, M. J., M. Jones, M. Scales, H. R. DeShon, M. B. Magnani, C. Frohlich, B. Stump, C. Hayward, and M. Layton (2016). Ellenburger wastewater injection and seismicity in North Texas, *Phys. Earth Planet. In.* **261**, 54–68.
- Hough, S. E. (2014). Shaking from injection-induced earthquakes in the central and eastern United States, *Bull. Seismol. Soc. Am.* **104**, no. 5, 2619–2626.
- Kerh, T., and S. B. Ting (2005). Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system, *Eng. Appl. Artif. Intell.* **18**, no. 7, 857–866.
- Khosravikia, F., A. Potter, V. Prakhov, G. Zalachoris, T. Cheng, A. Tiwari, P. Clayton, B. Cox, E. Rathje, E. Williamson, *et al.* (2018). Seismic vulnerability and post-event actions for Texas bridge infrastructure, *FHWA/TX-18/0-6916-1*, Center for Transportation Research (CTR).
- Khosravikia, F., Y. Zeinali, Z. Nagy, P. Clayton, and E. Rathje (2018). Neural network-based equations for predicting PGA and PGV in Texas, Oklahoma, and Kansas, *Fifth Geotech. Earthq. Eng. Soil Dynam. Conference*, Austin, Texas, 10–13 June.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* **11**, no. 2, 431–441.

- Novakovic, M., G. M. Atkinson, and K. Assatourians (2018). Empirically calibrated ground-motion prediction equation for Oklahoma, *Bull. Seismol. Soc. Am.* **108**, no. 5A, 2444–2461.
- Pan, Y., J. Jiang, R. Wang, H. Cao, and Y. Cui (2009). A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine, *J. Hazard. Mater.* **168**, nos. 2/3, 962–969.
- Parker, G. A., J. A. Harmon, J. P. Stewart, Y. M. A. Hashash, A. R. Kottke, E. M. Rathje, W. J. Silva, and K. W. Campbell (2017). Proxy-based V_{S30} estimation in central and eastern North America, *Bull. Seismol. Soc. Am.* **107**, no. 1, 117–131.
- Perlovsky, L. I. (2001). *Neural Networks and Intellect: Using Model-Based Concepts*, Oxford University Press, New York, New York.
- Petersen, M., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016). 2016 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes, *U.S. Geol. Surv. Open-File Rept. 2016-1035*, 52 pp.
- Roy, P. P., and K. Roy (2008). On some aspects of variable selection for partial least squares regression models, *QSAR Comb. Sci.* **27**, no. 3, 302–313.
- Smith, G. N. (1986). *Probability and Statistics in Civil Engineering*, Collins Professional and Technical Books, London, United Kingdom.
- Yenier, E., and G. M. Atkinson (2015). Regionally adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Application to central and eastern North America, *Bull. Seismol. Soc. Am.* **105**, no. 4, 1989–2009.
- Zalachoris, G., E. M. Rathje, and J. G. Paine (2017). V_{S30} characterization of Texas, Oklahoma, and Kansas using the P -wave seismogram method, *Earthq. Spectra* **33**, no. 3, 943–961.

Farid Khosravikia

Patricia Clayton

Zoltan Nagy

Department of Civil Architectural and Environmental

Engineering

The University of Texas at Austin

301E E Dean Keeton Street C1700

Austin, Texas 78712 U.S.A.

farid.khosravikia@utexas.edu

clayton@utexas.edu

nagy@utexas.edu

Published Online 27 December 2018