Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions – Silvestri & Moraci (Eds)

ISBN 978-0-367-14328-2

Deep learning-based site amplification models for Central and Eastern North America

O. Ilhan
University of Illinois at Urbana-Champaign, USA

J.A. Harmon
Hart Crowser, Incorporated, USA

O.A. Numanoglu & Y.M.A. Hashash
University of Illinois at Urbana-Champaign, USA

ABSTRACT: This paper presents deep learning-based site amplification models developed from large-scale simulated site amplification in Central and Eastern North America (CENA). The error evaluation of conventional simulation-based linear and nonlinear response spectrum (RS) and smoothed Fourier amplitude spectrum (FAS) amplification models highlights that fitting whole dataset to predetermined functional forms cannot capture the complex behavior inherent in the simulated amplification in CENA. Deep learning through Artificial Neural Network (ANN) is adopted for a new set of RS and FAS amplification models without the limitations of conventional regression models. This study shows significant improvements over conventional functions by use of ANN-based models: (i) the error in estimation is reduced up to 30% relative to conventional linear and total RS models, (ii) the simulated shallow site response is captured more accurately, and (iii) a continuous model for linear FAS amplification, previously provided as tabulated functions of V_{S30} and soil depth, is produced.

1 INTRODUCTION

Site amplification functions adjustground motions from a reference site condition (usually bedrock) to the ground surface on the basis of site geologic features and reference rock. In the current state of the practice, site amplification is evaluated either empirically where data is available or using supplementary simulations in the case that the data is not sufficient such as CENA, known as a tectonically stable region (Mueller et al. 2015), to develop site amplification terms in ground motion models (GMM's, commonly referred to as amplification functions or attenuation relationships). The site amplification terms in GMM's are predetermined functional forms using site and reference rock parameters for the estimation of ground response in a region of interest and are fit toan empirical ground response dataset by conventional nonlinear or mixed-effects regression techniques. The main objective of this study is to reduce the error of site amplification estimation and to better estimate the simulation-based site amplification behavior at CENA using deep learning techniques such as ANN along with more accurately capturing shallow site amplification.

A parametric study of one-dimensional (1D) site response simulations in Harmon et al. (2019a) was designed to capture uncertainty and variability of site amplification in CENA region mainly attributed to input ground motion, nonlinear dynamic soil properties, and shear wave velocity (V_S). The study uses 1.8 million 1D site response analyses relative to a reference rock condition (V_S = 3000 m/s) of CENA (Hashash et al., 2014). The site amplification database is used to develop modular linear and nonlinear RS and FAS site amplification functions for CENA (Harmon et al., 2019b).

This study presents ANN-based linear and total RS and FAS site amplification models using a subset (90%) of the simulation database of Harmon et al. (2019a) for training these models with application to shallow site RS amplification features of CENA attributed high shallow impedance contrasts. Remaining part (10%) of simulation database is adopted for testing the predictive capability of proposed ANN models. The error evaluation of ANN models is performed by comparison to amplification functions in Harmon et al. (2019b) with lowest error of site amplification estimation amongst all models developed by Harmon et al. (2019b).

2 SIMULATION-BASED SITE AMPLIFICATION AND MODELS FOR CENA

2.1 *Parametric study design for simulation-based site amplification*

The parametric study design for CENA simulations (Harmon et al., 2019a) is composed of (i) thirteen representative profiles from the combination of 10 V_S profiles and 9 geology-dependent soil index and strength properties, (ii) thirty realizations of each of the representative profiles, (iii) three pseudo-random realizations of nonlinear dynamic curves from Darendeli (2001) for a mean and systematically higher and lower G/G_{max} curves at all considered shear strains, (iv) eleven discrete soil horizon depths from 0.0 m (surface rock condition) to 1000.0 m, and (v) six weathered rock zone models for material between the soil and reference rock condition. In total, 70,650 site profiles are developed. The input rock outcrop motions consist of 186 synthetic and recorded rock motions for CENA from NUREG-6729 (McGuire et al., 2001), and 61 motions generated stochastically for CENA reference conditions. The 247 ground motions are uniformly distributed through the 30 V_S realizations of each site profile with three analysis methodologies resulting in1,745,055 site response analyses: 581,685 of each linear, nonlinear, and equivalent linear simulations. Only the linear and nonlinear simulations from that study are used herein.

2.2 *Modular RS and FAS linear and nonlinear site amplification models*

The large-scale simulated site amplification data is used to develop 17 RS site amplification functions, 5 linear and 12 nonlinear (Harmon et al., 2019b). The natural logarithm of totalsite amplification, F_S, is defined as sum of linear and nonlinear components as

$$F_S = ln(amp) = F_{lin} + F_{nl} \tag{1}$$

where F_{lin} is the intensity-independent linear amplification and F_{nl} is the nonlinear site amplification component. In Harmon et al. (2019b) F_{lin} is developed from linear simulations as a function of three parameters: V_{S30} (time-averaged V_S in the top 30 m of a site), and Z_{Soil}(depth to weathered rock) or T_{nat} (site natural period). F_{nl} is modeled using a modified functional form proposed by Chiou and Youngs (2008) and Seyhan and Stewart (2014), which includes V_{S30} and PGA at reference rock and is regressed on the difference between nonlinear and linear simulations. The resulting functional form of eq. (1) can be rewritten as:

$$F_S = F_{lin} + F_{nl} = f(V_{S30}) + f(Z_{Soil})|f(T_{nat}) + f(NL) \tag{2}$$

In this study, the only linear Harmon et al. (2019b) model considered is the L5 model which uses simultaneous V_{S30} and T_{nat} effects and produces the lowest error of site amplification estimation amongst all linear models in Harmon et al. (2019b). In the L5 form, the V_{S30} and T_{nat}effects are, respectively, defined by V_{S30}-scaling (L1) model and f(T_{nat}) functional form including (i) $c_5 R$ term, where R is the Ricker wavelet term to capture the fundamental mode site resonance and (ii) $c_6 T_{nat}$ component to represent the soil damping effects for deep sites. The total amplification model considered in this study are the L5+N2 model, which regresses the linear and nonlinear components separately. N2 model represents the $f(NL)$ term in eq. (2) and adopts V_{S30} and PGA at rock (PGA_r) parameter as driver of nonlinearity. Figure 1 shows the linear L5 and total L5+N2 model estimations of site response for periods of 0.1s along with their residuals defined as natural logarithm of ratio of simulations and model estimations:

The smoothed linear FAS amplification in Harmon et al. (2019b) is provided as tabulated functions of V_{S30} and Z_{Soil}, and the smoothed nonlinear FAS model is analogous to the RS model using the difference between linear and nonlinear analyses due to similar shapes of the nonlinear amplification. Both the linear and nonlinear models for FAS from Harmon et al. (2019b) are adopted in this study.

2.3 *Shallow site amplification at CENA*

The response of shallow sites in CENA, defined as sites with Z_{Soil} less than 30.0 m (Nikolau et al., 2001, Harmon, 2017), have significant amplification due to sharp impedance contrasts between the hard reference rock condition and overlying soil sediment and is mainly controlled by short period portion of site response (Nikolaou et al., 2014). Seismic guidelines such as NEHRP (2015) represent the site factors for short-period range by F_a coefficient, but these factors were developed using strong motion data from sites in Western US (WUS), where sharp stiffness contrast between soil sediment and underlying reference condition is not observed. This may result in underestimation of shallow site response, and lead to unconservative designs.

Figure 2 shows the L5 model estimations of linear site amplification at shallow sites for depths of 5.0 and 30.0 m for periods of 0.1s. L5 model estimations inadequately represent the shallow amplification two ways: (i) the height of first-order peak amplification is underestimated even with the use of c_5R term in L5 form, and (ii) the model cannot capture higher-order peaks for

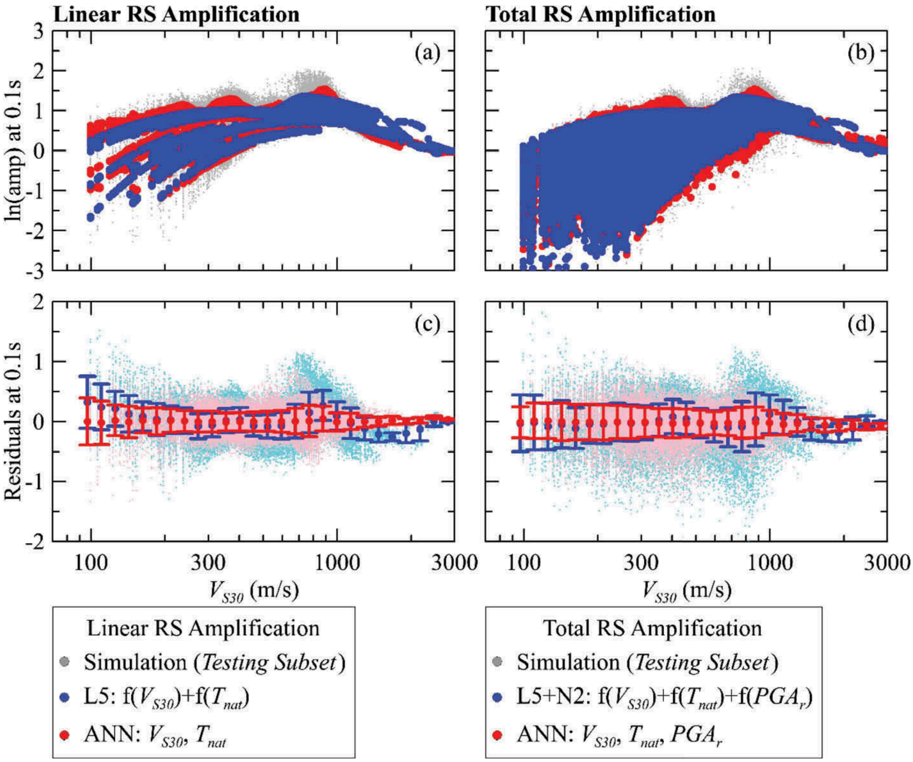

Figure 1. All profile L5 and ANN estimations (a) of linear amplification (testing subset) and corresponding residuals (c), and L5+N2 and ANN estimations (b) of total amplification and corresponding model residuals (d) for 0.1 s (using testing subset, 10% of all simulation data).

short periods (0.1 s). L5+N2 model produces similar estimations for total amplification at shallow sites, but they still cannot capture the full range of the simulation data.

3 DEEP LEARNING-BASED SITE AMPLIFICATION FUNCTIONS FOR CENA

The current state of the practice for the development of attenuation relationsuses mathematical models which relate independent and dependent variables of the ground motion through regression analysis, but various studies show that better estimation of ground motion parameters can be produced by ANN-based site amplification functional forms as compared to classical attenuation relations. Derras et al. (2014) uses asubset of a reference database for seismic ground motion prediction in Europe (RESORCE) and adopts feed-forward type ANN with five input parameters, the moment magnitude (M_w), the Joyner-Boore distance (R_{JB}), the focal mechanism, the hypo central depth and V_{S30}, and one hidden layer of 5 neurons. This model outputs PGA, PGV and 5% damped spectral accelerations (SA) at 62 periods from 0.01 s to 4.0 s. Khosravikia et al. (2018) outlines a framework for ANN-based GMM's using 4,529 ground motion recordings with epicenters in Texas, Oklahoma and Kansas and shows the improvement in model fits by ANN relative to conventional GMM modeling by Hassani and Atkinson (2015). There are other studies (Gullu and Ercelebi, 2007, Alavi and Gandomi, 2011) using ANN to produce GMM's for other regions, and all these models indicate that ANN observably decreases the error of amplification estimation relative to predetermined functions. This work aims to capture the site amplification more precisely as similar to aforementioned studies above but uses much larger site amplification database along with more optimized and novel platform for training ANN models.

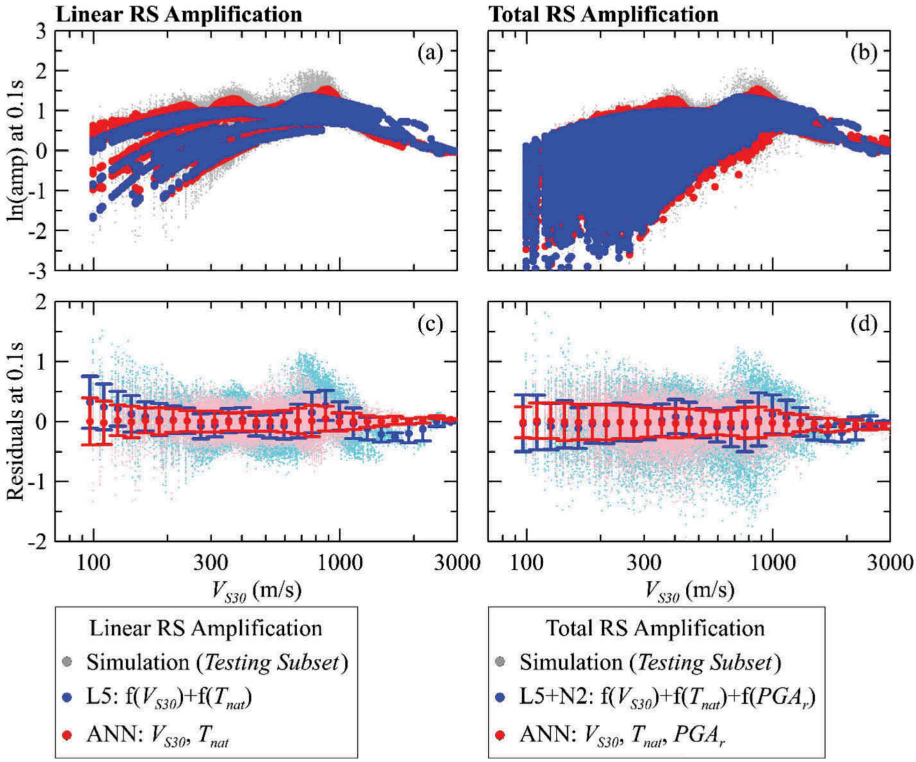

Figure 2. Shallow sites with Z_{Soil} of 5.0 and 30.0 m, L5 and ANN model estimations (a, b) of linear amplification, and L5+N2 and ANN model estimations (c, d) of total amplifications for 0.1s (using testing subset, 10% of all simulation data).

An ANN consists of a collection of inter-connected nodes (or artificial neurons) inspired by the biological neural networks constituting the human brains (McCulloch and Pitts, 1943), and each neuron consists of two parts: the net function and the activation function (Hwang and Hu, 2001). The net function defines the method for the combination of the network inputs inside each neuron as,

$$Y_j = \sum_{\mathrm{j}=1}^{\mathrm{N}} w_{ji} \cdot X_i + b_j \tag{3}$$

where w_{ji} represents the weight connecting i^{th} input to the j^{th} hidden unit, X_i are network inputs, b_j is the bias of the j^{th} unit, and Y_j is the output of the net function. The activation function is used to associate the output of the net function (Y) with the output of the neuron as $a=f(Y)$. Some of the most commonly used activation functions are tangent hyperbolic, linear, and Rectified Linear Unit (ReLU). The net and activation functions are executed in a directed graph structure (or ANN with neurons and external inputs) to perform distributed computing.

The adopted ANN structure of this study (Figure 3a) includes 2 hidden layers, each with 200 nodes and an output layer producing site amplification for selected oscillator periods (22 periods from 0.001 s to 10.0 s for RS and 19frequencies from 0.1 Hz to 33.3 Hz for FAS). The same ANN structure is used for independent modelling of RS and FAS amplification. The activation function is selected as ReLU for hidden layers, which is illustrated in Figure 3b and returns 0 if the input value is negative but gives the value itself if it is positive, and linear for an output layer. The training sample is composed of 90% of all simulation data (492,869 each of linear and nonlinear simulations), and this data is separated into 1232different random batches of 400 simulations to train the samples through 15,000 epochs. The learning rate is selected as 0.00005, and ANN training is performed via Tensor Flow (Abadi et al., 2016). Remaining 10% of simulation data (54,764 each of linear and nonlinear simulations) is used to test the predictive capability of ANN-based models.

Figure 1 shows the ANN estimations for linear and total RS amplification along with the related conventional L5 withV_{S30} and T_{nat} and L5+N2 with V_{S30}, T_{nat} and PGA_r models. The ANN-based models use identical input parameters to corresponding amplification functions. For linear amplification (Figure 1a and c), ANN produces greater height and more accurate location of first-order peak amplification along with the introduction of higher-order peaks at 0.1 s. The ANN estimations of site amplification are more representative of the scatter in the data which can be seen from the reduction of the scatter in the residuals. Similar improvement is obtained in ANN-based modeling of total amplification(Figure 1b and d). Another advantage of using ANN is evidenced by the ability to better model shallow linear and total site amplification (Figure 2). Improvements from the ANN modeling of shallow sites include better capturing of the distinct peak in amplification at V_{S30}? 900 m/s (Figure 2a), the inclusion of higher-order peaks and troughs (Figure 2b) for linear amplification, and variability of total amplification (Figure 2c and d).

Figure 4 shows the performance of two different ANN models for linear and total FAS amplification. The inputs to the linear FAS amplification models are V_{S30} & Z_{Soil} and V_{S30} & T_{nat},

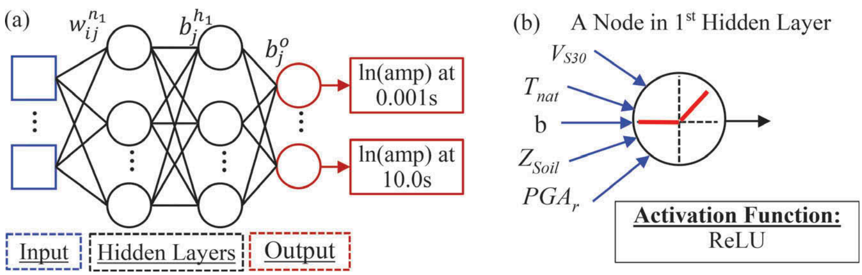

Figure 3. (a) ANN structure for modelling of linear and nonlinear RS and FAS site amplification. (b) Rectified Linear Unit (ReLU) used as activation function for hidden layers.

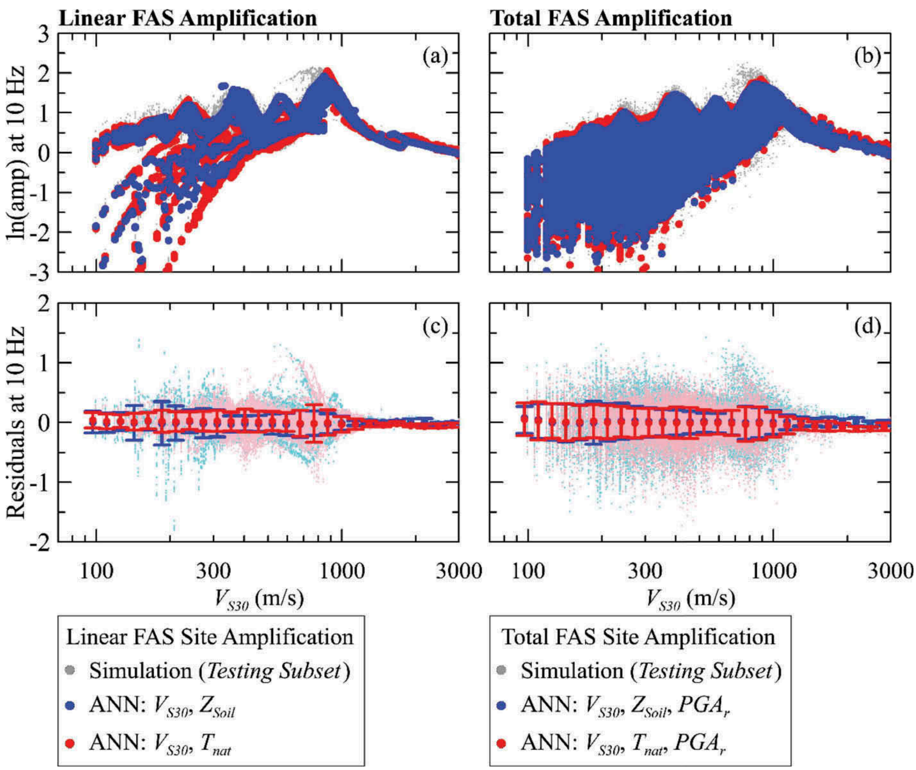

Figure 4. ANN estimations of linear (a) and total (b) FAS site amplification and corresponding linear (c) and total (d) model residuals for 10.0 Hz (using testing subset, 10% of all simulation data).

respectively. The same structure of ANN used for the RS amplification with inclusion of PGA_r is used for nonlinear FAS amplification. ANN models with T_{nat} produce better estimation of site amplification relative to that using Z_{Soil} as observed in the reduction of error in the model residuals (Figure 4c and d)

Error evaluation of ANN-based models along with conventional models is performed by comparison of root-mean-square (RMS) error of L5, L5+N2 and corresponding ANN estimations in Figure 5. The use of ANN's reduces the RMS-error up to approximately 30% both for linear (Figure 5a) and total (Figure 5c) RS amplification estimations relative to conventional models, respectively. For FAS amplification (Figure 5b and d), the effect of using T_{nat} over Z_{soil} as an input to the ANN on reduction in error of estimation is more significant for linear amplification than total amplification.

4 CONCLUSION

Predetermined functional forms using conventional regression techniques can capture simulation-based linear and total site amplification, but lead to significant error of site amplification estimation. This study utilizes ANN method to develop amplification models with considerably less error with application to shallow site amplification at CENA. These models adopt identical input parameters of corresponding simulation-based site amplification models (V_{S30}, T_{nat} and PGA_r) regressed through conventional techniques, and the same structure of ANN is used for modeling of RS and FAS site amplification. The results demonstrate that ANN models can capture the amplitude and location of distinct amplification peaks due to site natural period effects more accurately, introduce higher order amplification peaks, and more

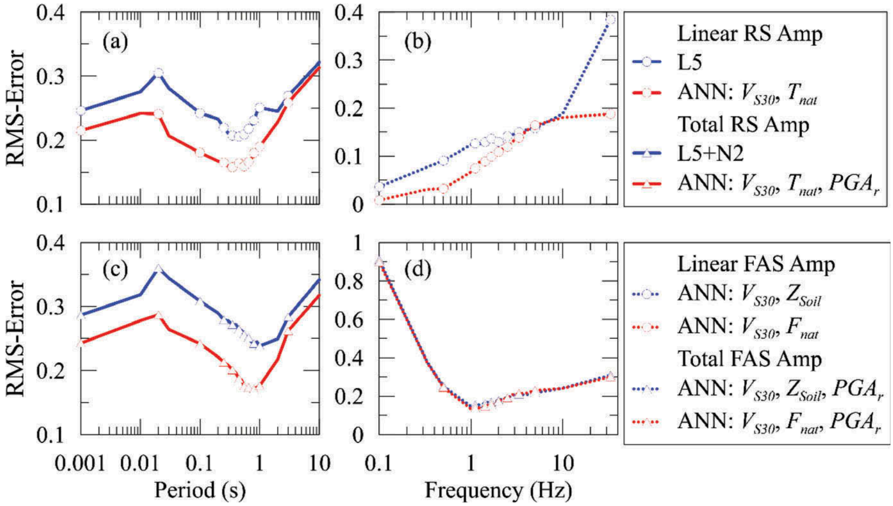

Figure 5. RMS-Error of L5, L5+N2 models and corresponding ANN model estimations of RS (a, c) and FAS (b, d) linear and total site amplification(using training subset, 90% of all simulation data). Same results are obtained for test subset.

properly represent the spread of site amplification data compared to conventional amplification models, which would require additional model terms or more complicated functional forms to estimate these features. In addition, the ANN modeling results in approximately 30% reduction in the standard error estimation of linear and total site amplification, respectively.

Similar ANN-based amplification models were developed for the smoothed FAS. The linear FAS amplification model, previously approximated as a tabulated function of V_{S30} and Z_{Soil}, uses V_{S30} and either of T_{nat} or Z_{Soil} as inputs, and the total FAS amplification is modeled using the same parameters and including PGA_r as driver of nonlinearity. The use of $T_{nat}$with V_{S30}asANN inputs produces lower error of estimation for linear amplification than with Z_{Soil} and V_{S30}, and the effect of T_{nat} in the ANN FAS total amplification model does not show significant reduction in error estimation of linear amplification as it was in RS total amplification.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X, & Kudlur, M. 2016, Tensorflow: a system for large-scale machine learning. In OSDI (Vol.16, pp. 265-283).

Alavi, A. H. & Gandomi, A. H. 2011. Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. Computers & Structures, 89(23-24), 2176-2194.

Boore, D. M. 2005. SMSIM Fortran programs for simulating ground motions from earthquakes, version 2.3. A revision of U.S.Geol.Surv. Open-File Rept. 96-80-A.

Chiou, B. S. J. & Youngs, R. R. 2008. NGA model for average horizontal component of peak ground motion and response spectra, Pacific Engineering Research Center.

Darendeli, M. B. 2001. Development of a new family of normalized modulus reduction and material damping curves,*Ph. D. Thesis*, University of Texas at Austin.

Derras, B., Bard, P. Y., & Cotton, F. 2014. Towards fully data driven ground-motion prediction models for Europe., Bulletin of Earthquake Engineering, 12(1),495-516.

Güllü, H., & Erçelebi, E. 2007. A neural network approach for attenuation relationships: An application using strong ground motion data from Turkey. Engineering Geology, 93(3-4), 65-81.

Hashash, Y. M. A., Kottke, A. R., Stewart, J. P., Campbell, K. W., Kim, B., Moss, C., Nikolaou, S., Rathje, E. M., & Silva, W. J. 2014. Reference Rock Site Condition for Central and Eastern North America. Bulletin of the Seismological Society of America 104(2): 684-701.
Hassani, B. & Atkinson, G. M. 2015. Referenced empirical ground-motion model for eastern North America. Seismological Research Letters, 86(2A), 477-491.
Harmon, J. A. 2017. Nonlinear site amplification functions for central and eastern North America, *Doctoral dissertation*, University of Illinois at Urbana-Champaign.
Harmon, J., Hashash, Y. M. A., Stewart, J. P., Rathje, E. M., Campbell, K. W., Silva, W. J., Xu, B., Musgrove, M., and Ilhan, O. 2019a. Site amplification functions for Central and Eastern North America – Part I: Simulation dataset development, Earthquake Spectra35(2), in press.
Harmon, J., Hashash, Y. M. A., Stewart, J. P., Rathje, E. M., Campbell, K. W., Silva, W. J., and Ilhan, O. 2019b. Site Amplification Functions for Central and Eastern North America - Part II: Model Development and Evaluation, Earthquake Spectra 35 (2), in press.
Hwang, J. N., & Hu, Y. H. 2001. Handbook of neural network signal processing. CRC press.
Khosravikia, F., Zeinali, Y., Nagy, Z., Clayton, P., & Rathje, E. M. 2018. Neural Network-Based Equations for Predicting PGA and PGV in Texas, Oklahoma, and Kansas. arXiv:1806.01052.
McCulloch, W. S., & Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4),115-133.
McGuire, R. K., Silva, W. J., & Costantino, C. J. 2001. Technical Basis for Revision of Regulatory Guidance on Design Ground Motions: *Hazard- and Risk-consistent Ground Motion Spectra Guidelines* (NUREG/CR-6728). R. M. Kenneally.
Mueller, C.S, Boyd, O. S., Petersen, M. D., Moschetti, M. P., Rezaeian, S., & Shumway, A. M. 2015. Seismic Hazard in the Eastern United States. Earthquake Spectra: December 2015, Vol. 31, No. S1, pp. S85-S107
Nikolaou, S., Mylonakis, G., & Edinger, P. 2001. Evaluation of site factors for seismic bridge design in New York City area. Journal of Bridge Engineering, 6(6),564-576.
Nikolaou, S., Pehlivan, M., Richins, J., Lincoln, L., & Deming, P. W. 2014, Seismic Response of Shallow Sites in Eastern US: Implications to the State of Practice, Tenth U.S. National Conference on Earthquake Engineering Frontiers of Earthquake Engineering
Seyhan, E. & Stewart, J. P. 2014. Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. Earthquake Spectra, 30(3),1241-1256.