SITE AMPLIFICATION PREDICTION MODEL OF SHALLOW BEDROCK SITES BASED ON MACHINE LEARNING MODELS

Yong-Gook Lee[a], Sang-Jin Kim[b], Zeinep Achmet[c], Oh-Sung Kwon[d], Duhee Park[a]*, and Luigi Di Sarno[e]

ABSTRACT

Prediction of the site amplification is of primary importance for a site-specific seismic hazard assessment. A large suite of both empirical and simulation-based site amplification models has been proposed. Because they are conditioned on a few simplified site proxies including time-averaged shear wave velocity up to a depth of 30 m (V_{S30}) and site period (T_G), they only provide approximate estimates of the site amplification. In this study, site amplification prediction models are developed using two machine learning algorithms, which are random forest (RF) and deep neural network (DNN). A comprehensive database of site response analysis outputs obtained from simulations performed on shallow bedrock profiles is used. Instead of simplified site proxies and ground motion intensity measures, matrix data which include the response spectrum of the input ground motion and shear wave velocity profile. Both machine learning based models provide exceptional prediction accuracies of both the linear and nonlinear amplifications compared with the regression-based model, producing accurate predictions of both binned mean and standard deviation of the site amplification. Among two machine learning techniques, DNN-based model is revealed to produce better predictions.

Keywords: Machine learning, random forest, deep neural network, site amplification, site response analysis.

[a] Department of Civil and Environmental Engineering, Hanyang University, Seoul, Korea
[*] Corresponding author. Email: dpark@hanyang.ac.kr (D. Park).
[b] Infrastructure Division, Infrastructure Engineering Group, Infrastructure Geotechnical Engineering Team, Hyundai Engineering and Construction Group, Seoul, Korea.
[c] Department of Civil Engineering, National Technical University of Athens, Zografou, Greece.
[d] Department of Civil & Mineral Engineering, University of Toronto, Toronto, Canada.
e Department of Civil Engineering and Industrial Design, University of Liverpool, Liverpool, UK

1. INTRODUCTION

The vertically propagating shear waves are generally amplified as they radiate upwards from the bedrock through soil layers, which have relatively lower stiffness and density. This phenomenon is referred to as the seismic site amplification. The prediction of site-specific seismic amplification is critical for estimation of the design ground motion and seismic design of various types of structures and facilities.

Regression-based site amplification models that are linked to ground motion models (GMMs) have been developed from both recorded ground motions and numerical simulation outputs. The models are based on site proxies, which include the time-averaged shear wave velocity of top 30 m (V_{S30}), depth at which shear wave velocity (V_S) reaches 1 km/s or greater (Z_1), and natural site period (T_G). They are also conditioned on motion proxies including peak ground acceleration (PGA) and spectral acceleration (SA) at selected periods. Although widely used because of their ease of use, the regression-based site amplification models inevitably contain large levels of uncertainty.

A number of studies proposed to use machine learning (ML) algorithms instead of regression equations to develop site amplification models [1-7]. Kamatchi et al. [3] developed an artificial neural network (ANN)-based methodology to predict site-specific acceleration response using the outputs from one-dimensional (1D) equivalent linear (EQL) site response analyses performed for a selected site in Delhi, India. The input parameters of the ANN model were the moment magnitude of the earthquake (M_w), V_S profile, depth of the soil stratum, damping ratio, and vibration period of the single degree-of-freedom (SDOF) oscillator. Derras et al. [4] also developed an ANN model to predict ground motions using the Reference database for Seismic ground-motion prediction in Europe (RESORCE). The input data set consists of 1,088 recordings from 320 earthquakes. The five parameters that are considered include M_w, Joyner-Boore distance (R_{JB}), V_{S30}, the fault mechanism, and the hypo-central depth. It was revealed that M_w, R_{JB}, and V_{S30} are the most important parameters. The results are compared to the outputs from conventional GMMs. The SAs of the ANN model were reported to be closer to those of the recordings compared with the conventional GMMs. Ilhan et al. [5] developed an ANN-based site amplification model using the calculated site amplification factors for Central and Eastern North America. The ANN model was trained using the 1,745,055 simulation results. Four parameters were selected as the input features: V_{S30}, T_G, depth to weathered rock (Z_{Soil}), and PGA. The ANN models were reported to reduce the root-mean-square (RMS) error by approximately 30% compared with the regression-based models. Roten and Olsen [6] trained a convolutional neural network (CNN) and a multilayer perceptron (MLP) to predict surface-to-borehole amplification functions. The models were trained and tested by using a total of 13,120 events from 662 vertical arrays of the Kiban-Kyoshin network (KiK-net). The spectral frequency, V_S, and a compressional wave velocity (V_P) profile were used as input features of the CNN and MLP model. The results showed that the mean squared logarithmic error (MSLE) of the CNN model is significantly reduced compared with the theoretical amplification and MLP models. Zhu et al. [7] developed the amplification model using a random forest (RF) algorithm with topographic and site proxies. The between-site variability is reduced by up to 38% throughout the whole frequency range. Overall, the ML-based models have been reported to provide more accurate predictions compared with the regression-based models.

Proxy-based models provide quick estimates of the site amplification, but they cannot accurately predict the nonlinear wave propagation through heterogeneous soil layers. Simple models are most appropriate as substitutes for regression-based models, many of which are compatible with GMMs. The local site amplification for a seismic design is most often characterized using the 1D site response analysis, where earthquake induced elastic wave transmission is idealized as a 1D propagation problem modeling only the vertically propagating horizontal shear waves [8-11]. The required inputs for a site response analysis include a *Vs* profile, nonlinear soil curves, and an input acceleration time history.

Whereas a number of proxy-conditioned ML-based models have been developed [1-7], they do not fully utilize their full capability including the potential to use vector or matrix data for which the ordering is crucial. Matrix data in a site response analysis include the response spectrum of the input motion and the *Vs* profile. Such a model, if successfully trained, can potentially replace the numerical site response analysis to predict the seismic amplification.

One may question the need for a rigorous ML-based model considering the relatively low computational cost of performing a 1D site response analysis. However, an effectively trained ML-based model trained with matrix data has a number of advantages compared with a numerical model. An important advantage is that such a ML-based model requires only the input motion response spectrum, instead of the acceleration time history used in a nonlinear site response analysis. It is particularly useful because the uniform hazard or conditional mean spectra calculated from a probabilistic seismic hazard analysis can be directly used without having to develop spectrally matched ground motions.

A rigorous ML-based model potentially has a series of applications. One application is utilization in a probabilistic assessment, which involve the use of a large number of *Vs* profiles and nonlinear curves, along with a series of input motions. Although the computational cost of performing multiple 1D analyses is trivial, developing a significant number of input files and extracting the outputs is extremely time consuming. It can also be utilized for regional assessments including seismic microzonation, which requires a paramount number of analyses. Because of difficulties in performing analyses for a large area, GMMs and associated site amplification factors have been utilized [12]. The technique to fully utilize the broadband response spectrum and *Vs* profile can be applied to develop empirical site amplification models. Good examples are the comprehensive downhole array data including KiK-net and Kyoshin network (K-NET), for which the *Vs* profiles and motion time histories are available. Until now, only proxy-based models have been developed [13]. Additionally, it can be used for extensive probing to better understand the mechanism of site amplification.

The primary challenge of developing such a model is training with a large volume of input data from outputs involving a comprehensive range of site structures and broadband input motions. In this study, ML-based models were developed for site amplification prediction of shallow bedrock sites composed of non-cohesive soils, representative of inland sites of Korea. The results of 1D site response analyses performed for shallow bedrock sites in Korea were utilized to develop and train the ML-based models. Two ML algorithms were used, which are the RF and deep neural network (DNN). The performance of the RF and DNN models are evaluated through comparisons with the site amplifications calculated from numerical analyses. It is also compared with a regression-based site amplification model (hereafter denoted as a regression-based model) that is developed using the identical training dataset.

The focus of this study is to evaluate the potential for an ML-based model to replace the numerical analysis to predict wave propagation for this specific site condition. It should be noted that the proposed ML-based models, which were solely developed from numerical simulation outputs, are not intended to be routinely used for developing site-specific ground motions. For such a purpose, the numerical model needs to be validated against a comprehensive set of recordings including downhole arrays for the site condition of interest. It is also worth noting that the 1D site response analysis model often fails to predict the site amplification due to various physical processes including basin effects and wave scattering, as reported in a number of studies [7, 14, 15]. Although 1D site response analysis does not always provide an accurate prediction of the site amplification, it is worth noting that it is still most often utilized in design to develop site-specific ground motions.

2. TRAINING DATASET

Vertically propagating and horizontally polarized shear waves dominate earthquake ground motion wave field. Therefore, in most cases 1D site response analysis is performed to assess the effect of soil conditions on ground shaking. Site response analysis requires the definition of input ground motions and information on dynamic soil properties such as the V_S profile, the modulus reduction and damping curves. After site response analysis, surface acceleration time series, surface acceleration response spectra, and spectral amplification factors are provided. The most common methods to perform this analysis are the frequency domain EQL [16] and time domain nonlinear analyses [17].

The training dataset used in this study was taken from Aaqib et al. [18]. In their study, a series of linear and nonlinear site response analyses was performed in order to derive a regression-based model for shallow bedrock sites in Korea with a depth to bedrock less than 30 m. The suite of ground motions, the site profiles and the analysis methods used to perform the site response analyses are summarized in the following.

2.1 Input ground motions

A total of 51 recorded ground motions were used in the site response analyses. The recorded rock outcrop ground motions were selected from the NGA-West2 database (https://ngawest2.berkeley.edu/) and U.S. Nuclear Regulatory Guide (NUREG-6729) [19]. Selected recordings from the 2016 Gyeongju and 2017 Pohang earthquakes which occurred in Korea were also used. The M_w and rupture distance (R_{rup}) ranges from 5 to 7.5 and 0 to 100 km, respectively. The significant duration of the motion ($D_{5\text{-}95}$) ranges from 1.89 to 42.78 s. The peak ground acceleration of the recorded rock outcrop ground motions (PGA_{rock}) ranges from 0.01 to 0.50 g. The acceleration response spectrum of all ground motions used in the analyses are shown in

Figure 1. The information on M_w, R_{rup}, and $D_{5\text{-}95}$ of the motions are displayed in Figure 2.

2.2 Site profiles

Forty V_S profiles in inland Korea were used as baseline profiles. The V_{S30} of the baseline profiles range from 227 to 703 m/s. The T_G ranges from 0.08 to 0.48 s. The depth to bedrock (H) ranges from 7 to 29 m. The baseline V_S profiles are shown in Figure 3(a). Due to an insufficient number of measured profiles, additional randomized site characterizations generated from the baseline profiles were used. Twenty randomized V_S profiles were generated for each baseline profile using the procedure of Toro [20]. To avoid the reversals and unrealistic velocity realization in V_S profiles, the distribution of V_S was perfectly correlated and bounded within $\pm 2\sigma \ln V_S$. Examples of realizations of a baseline profile with V_{S30} = 398 m/s and H = 25 m are shown in Figure 3(b). Detailed information including the distribution of V_{S30} for all velocity profiles can be found in Aaqib et al. [18].

2.3 Site response analysis

The site response analyses were performed using 1D site response analysis program DEEPSOIL v7 [17]. The shear modulus reduction and damping curves proposed by Darendeli [21], which is widely used in practice to simulate the nonlinear behavior of soils, were used. The overconsolidation ratio (OCR) was set to 1.0 and the plasticity index (PI) was assumed as zero. The number of cycles of loading and the excitation frequency were defined as 10 and 1.0 Hz, respectively. The horizontal at-rest earth pressure factor (K_0) was calculated using the empirical equation of Jaky [22].

Because the shear modulus reduction curve is reported to provide poor estimate of the nonlinear soil response at strains exceeding 0.5% , shear strength correction has been applied to overcome this shortcoming [23]. The General Quadratic/Hyperbolic (GQ/H) constitutive model of

Groholski et al. [24] implemented in DEEPSOIL v7 [17] was used to apply the strength correction, which was reported to be important even for shallow deposit regions of moderate to low seismicity [25]. To define the shear strength for the GQ/H constitutive model, the Mohr-Coulomb model was used. The cohesion of the Mohr-Coulomb model was calculated as a function of V_S as recommended in Hashash et al. [17]. It is a common practice to estimate the friction angle (ϕ') from standard penetration test (SPT) blow count (*N*) measurement. Because of the unavailability of SPT *N* measurements, they were estimated from the empirical *N* versus V_S correlation developed specifically for Korea by Sun et al. [26]. The friction angle for each layer was then calculated using the empirical *N* versus ϕ' correlation proposed by Wolff [27]. The GQ/H model was fitted to the baseline Darendeli [21] curves using the non-Masing fitting tool described as the modulus reduction and damping curve-fitting procedure (MRDF) [23] implemented in DEEPSOIL v7 [17]. The procedure has been reported to produce curves that fit well with any target curve.

The number of analyses performed for linear and nonlinear analyses was 42,840 (85,680 in total) using 840 V_S profiles and 51 motions. In this study, the results are used to train the ML-based models, as presented in the following sections.

2.4 Reference regression-based site amplification model

Based on the site response analysis, a regression-based site amplification model was developed by Aaqib et al. [18], which is referred to as the AEA21 model hereafter. The amplification model consists of two additive components as follows:

$$Amp\ (f) = F_{lin}(f) + F_{nl}(f) \tag{1}$$

where f is the frequency, $Amp\ (f)$ represents the total site amplification in natural logarithmic units, $F_{lin}\ (f)$ represent linear amplification dependent on V_{S30} and T_G, and $F_{nl}\ (f)$ is the nonlinear component of site amplification dependent on the intensity of the motion. In the following, the symbol, *f*, is omitted in the amplification equations for simplicity.

The linear amplification component, F_{lin}, represents the ratio of the 5% damped surface acceleration response spectrum calculated from a linear analysis to the input ground motion spectrum. Aaqib et al. [18] defines F_{lin} using the following functions, which consists of a flat region at slow V_{S30} followed by a linear region with negative slope:

$$\left(F_{lin}\right)_{V_{S30}} = \begin{cases} c_1 \ln\left(\dfrac{V_L}{V_c}\right) & V_{S30} < V_L \\ c_1 \ln\left(\dfrac{V_{S30}}{V_c}\right) & V_L < V_{S30} < V_c \\ 0 & V_c < V_{S30} \end{cases} \tag{2}$$

The model has three regions with two transitional velocities V_L and V_c. V_L represents the lower end of the flat-region at slow V_{S30}, whereas V_c is the limiting velocity above which there is no amplification. c_1, V_L and V_c are period-dependent parameters. c_1 is the slope parameter which represents the V_{S30} scaling. It was also reported that the prediction of the linear model increases if the T_G is accounted for. The following function developed to capture the T_G effect was proposed:

$$\left(F_{lin}\right)_{T_G} = c_2 R + c_3 T_G \tag{3}$$

Where c_2 and c_3 are period-dependent regression coefficients. R is defined as:

$$R = \frac{2}{\sqrt{3\alpha}\pi^{1/4}}\left[1 - \frac{\beta^2}{\alpha^2}\right]\exp\left[-\frac{\beta^2}{2\alpha^2}\right] \tag{4}$$

where α is a regression coefficient and β is defined as:

$$\beta = \ln\left[\frac{T}{T_G}\right] - \ln(\Upsilon) \tag{5}$$

where T is the spectral period under consideration and Υ is the coefficient for T_G-dependent model. The c_2R is a Ricker wavelet that captures the effect of site resonance of the fundamental mode whereas c_3T_G term captures the amplification at soft sites.

The nonlinear component is used with the linear site amplification to decrease amplification for strong excitations. The nonlinear term is defined as zero for $PGA_{rock} \leq 0.1$ and becomes negative at higher intensities. The regression equation for the nonlinear component is defined as follows:

$$F_{nl} = f_1 + f_2 \ln\left[\frac{PGA_{rock} + f_3}{f_3}\right] \tag{6}$$

where f_1, f_2, and f_3 are coefficients of the model. It should be noted that F_{nl} decreases with an increase in the intensity of the rock motion. The coefficients of the model were shown to be dependent on V_{S30}. The coefficients of the linear amplification and nonlinear model parameters are provided in Aaqib et al. [18].

3. PROPOSED ML- BASED SITE AMPLIFICATION MODEL

As presented in the previous section, ML algorithms that were used to train the site amplification models were DNN and RF. The RF-based model was developed using Scikit-learn package [28], whereas Tensorflow [29] was used for the DNN-based model. The basic principles of an RF as well as a DNN, the differences between ANN and DNN, and the proposed ML-based models are presented in the following section.

After training, the ML-based model must be evaluated in order to assess its generalization performance. The ML-based model which overfits will yield favorable results for the training dataset, but produce poor results for the test dataset. A well-trained model which does not overfit can reproduce results with acceptable accuracy for unseen input features. For accurate results, these input features should have the same value ranges as those used for the training.

This fact can be considered as a shortcoming, but an ML can always be improved and the predictions can be extended for further data ranges by training the network further with new additional datasets.

3.1 Random forest

An ensemble technique is often used to aggregate better predictions by a group of predictors rather than the best prediction by a predictor. An RF is an ensemble of decision trees that consist of a root node, child (or split) nodes, and leaf nodes (i.e., these have no child nodes). Each node has a criterion with respect to one of the input features. From the root node to the leaf node, it is determined whether it is true or false based on the criteria.

The RF algorithm is widely used because of its simplicity and powerful performance in both classification and regression. The RF algorithm [30] was developed by combining a bootstrap aggregating [31] and a random selection of features [32]. Bootstrap aggregating (bagging) is generally preferred to train the RF model. A regression using the RF is performed by controlling hyperparameters: the number of estimators, the maximum number of features, bootstrap, etc.

The bootstrap is a random sampling technique that splits training data into various random subsets. The use of bootstrap produces a higher level of diversity in the subsets and less correlation between the predictors, eventually leading to a lower variance of the RF model. The bagging is performed by randomly sampling the train data with replacement. After bootstrap sampling, the decision trees are trained using corresponding random subsets. The prediction of an individual subset is also calculated by averaging the outputs in the largest number of samples. Finally, the prediction is calculated by averaging predictions of whole subsets.

3.2 Deep neural networks

The ANN is a computing technique designed to simulate the human brain's method of problem-solving. The similarity between the ANNs and the human brain is that both acquire the skills in processing data and finding solutions through training [33]. ANNs consist of simple computing units referred to as "artificial neurons", and each unit is connected to other units via weight connectors. These units calculate the weighted sum of incoming inputs and determine the output using an activation function [34].

The process of calibrating the values of weights and biases of the network is called training of the neural network to perform the desired function correctly [35]. In case of supervised learning, the data is presented in a form of couples (input, desired output) and then the learning algorithm will adapt the weights and biases depending on the error between the predicted output and the real output [35]. This error is calculated using the "loss function" which is defined as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left\| y_i - f\left(x_i\right)\right\|^2 \qquad (7)$$

where N is the number of input-output data sets, f denotes the output estimated by the ANN model, and $\|\cdot\|$ represents a metric that computes the distance between the real and the estimated value of the output. The optimal ANN for a given training set is obtained by minimizing the loss function using an optimization algorithm. The DNN is an ANN with a relatively large number of hidden layers, each providing a different interpretation to the data it feeds on. Hence, critical features of input data can be identified, and hidden patterns of highly complex problems can be found, obtaining superior predictions compared to ANNs. DNNs also tend to perform better with large datasets.

3.3 Input features for the machine learning based model

In order to train ML-based models efficiently, parameters that describe the input ground motions and soil properties must be selected. The natural log of 5% damped spectral acceleration, ln S_a (T), of bedrock motion in the period range from 0.01 to 10 s (total 113 steps) was selected as input feature. With respect to the soil properties, the array of the layer thickness (t) and Vs of the entire soil column was used as input. The maximum number of layers of the soil profiles was set to 29, resulting in a 58×1 vector for each profile. It should be noted that because the orderings of the SA and layer data are crucial in vertical propagation of the seismic waves, this information is retained in the machine learning training. For the RF-based model, along with the aforementioned two sets of input features, M, R_{rup}, V_{S30}, average V_S of the soil profile ($V_{S,soil}$), and T_G were also used as inputs.

3.4 Detailed architecture of the machine learning models

Figure 5 illustrates the architecture of the proposed DNN-based model. For the first group of input features, which is ln S_a (T) of bedrock ground motion, a 113×1 vector, four fully connected hidden layers were created. They have 128, 256, 128, 64 hidden units, respectively. The second group of input features is the site and soil properties, which is a 58×1 vector for each profile. Five hidden layers are used to process this group of features with 64, 128, 256, 128, 64 units, respectively. After processing each group of features individually, a concatenation layer was created to merge the processed information. Four hidden layers were used after the concatenation layer with 1024, 512, 256, 128 hidden units, respectively. The output layer gives the prediction of ln S_a (T) of the surface, which is 113×1 vector. In all hidden layers, except for the output layer, the rectified linear unit (ReLU, [36]) activation function was applied. Batch normalization (BN) is applied after all activation functions, ReLU, to reduce internal covariate shift and achieve a stable distribution throughout training [37]. The linear activation function was applied at the output layer, which gives the prediction of ln S_a (T) of the surface.

For the RF, the feature extraction should be preceded before the training. The input features for the RF-based model consist of 176 variables, which are 113 ln S_a (T) of bedrock ground motion, M_w, R_{rup}, V_{S30}, $V_{S,soil}$, T_G, and 58 soil profile V_S properties. To find the best estimator, GridSearchCV from Scikit-Learn package [28] was used by varying the number of estimators (512, 1,024, 2,048, 4,096) and the maximum number of features (64, 128, 176). The best estimator turned out to be a combination of 2,048 estimators and 176 features. Figure 4 illustrates the schematic of the RF model. The training data are randomly sampled with replacement and divided into N subsets. The N decision trees are generated and assigned to each subset. In the first decision tree, for example, the subset data are divided by the value of V_{S30} at the root node. The data, greater than x_1, are divided again with respect to the value of M_w.

3.5 Training data pre-processing

The dataset was processed before being employed as the input and output of the proposed ML-based models in order to achieve improved accuracy in a relatively small training time. To resolve the skewness, the natural logarithm was applied to the input bedrock motion S_a (T) and the output surface motion S_a (T). The values of the period, T, are not used in the training because the S_a (T) is defined at specific period values.

As described previously, the soil and site properties are a 58×1 vector, considering a maximum of 29 layers in a profile and 2 properties for each layer. However, not all the profiles had this number of layers. As a result, the SQLite table where the soil and site properties were stored had a lot of null values. Having these null values during training affects the accuracy and the training time of the ML models. In order to resolve this issue, all null values were replaced with zeros, while the site and soil property values were used unchanged.

3.6 Training and evaluation

To check whether the trained model overfits, the whole dataset was split into training and test sets. Eighty percent of the dataset, 34,272 data points were used for training. The remaining 20%, 8,568 data points were used to evaluate the performance of the trained ML models. To ensure similar data ranges for the training and test data sets, the site profiles were first grouped into 15 bins based on V_{S30}. The training and test data were selected within each bin. Additionally, 5-fold cross-validation was performed to avoid overfitting. It was shown that the differences between the results of predictions are marginal. Among, five folds, the dataset which produces the lowest residuals compared with the computational outputs was utilized. Mean squared error (MSE) as well as mean absolute error (MAE) were calculated for both training and test sets. Two ML-based models were trained on a Windows-based operating system with 64 GB GPU NVIDIA RTX A6000 and 32GB RAM. The computational times needed for training were approximately 5 hours and an hour for the RF- and DNN-based models, respectively.

The RF-based model was trained with 2,048 estimators and 176 features. With bagging, the training dataset was sampled several times but the whole dataset was not used for training. Assuming the data set has n samples, the probability that one data is not sampled is calculated as follows:

$$\lim_{n \to \infty}\left(1-\frac{1}{n}\right)^{n} = e^{-1} = 0.368 \tag{8}$$

It means that 36.8% of the training data are not sampled, which is called out-of-bag (oob). Thus, the number of samples at the root node is 21,659, which is the remaining 76.2% of the training set. The RF was evaluated using oob scores without sampling an additional validation set.

Before starting the training of the DNN, the layer weights were initialized using Glorot uniform initializer [38] and the biases were set to zero. The Adam optimizer [39, 40] was selected as the optimization algorithm to reduce the MSE, which was used as the loss function. The selection of the hyperparameters was determined by varying the batch size, the learning rate, and the epoch number. The batch size of 512, the learning rate of 0.005, and the epoch of 2,000 showed the lowest MSE, while the training was early stopped before 2,000 epochs.

For both the linear and nonlinear analysis cases, the amplification ratios of the calculated surface response spectra to those of the input motion were trained. The linear amplification, F_{lin}, was trained by the ML-based linear model. The total amplification, *Amp*, was trained by the ML-based nonlinear model. Using equation (1), the nonlinear amplification, F_{nl}, was calculated by subtracting F_{lin} from *Amp*.

4. PREDICTION ACCURACY – RESULTS

The results of MAE and MSE of the training and test sets for both ML-based models are presented in

Table 1. MSE and MAE for both sets are calculated for the natural logarithm of the output acceleration response spectrum. It is clearly shown that the trained network does not overfit, as MSE and MAE of the test set are almost identical with those of the training set.

Figure 6 and Figure 7 plot the calculated and predicted amplifications in log normal unit for linear and nonlinear components, respectively. The amplifications of the AEA21 model, the equations of which are listed in Section 2.4, are also compared with the calculated amplifications. The $\ln(F_{lin})$ is plotted against V_{S30}, whereas $\ln(F_{nl})$ is plotted against PGA_{rock}. Significant scatter of $\ln(F_{lin})$ plotted against V_{S30} is observed. The trends also vary greatly with the spectral periods. The AEA21 model captures the median output favorably, but it fails to predict the variability of the outputs. The level of scatter of the ML-based models fits well with that produced by the site response analyses. In case of $\ln(F_{nl})$, it is shown to decrease with an increase in the intensity of the ground motion. Again, the AEA21 model provides reasonable estimates of the median response, except for spectral periods of 2.0 s. A positive nonlinear component was intentionally constrained for the AEA21 model, and therefore it displays zero values. The AEA21 model is unsuccessful in capturing high level of dispersion. On the contrary, the ML-based models agreeably estimate the nonlinear amplification component.

Figure 8 and Figure 9 illustrate the comparison between the calculated and predicted amplifications by three models with respect to the selected spectral periods. The ML-based models predicted the calculated amplifications, whereas The AEA21 model deviates from the calculated amplifications. Although the AEA21 model can capture the median of $\ln(F_{lin})$, most of $\ln(F_{lin})$ higher than 1.0 were not considered. However, the ML-based models can predict high levels of $\ln(F_{lin})$ throughout the spectral periods. The nonlinear components of the AEA21 model converge to unity at T=5.0 s. The ML-based models predicted well with the nonlinear components at this period, which results in overcoming the range of amplifications.

The calculated MSE values of both linear and nonlinear amplifications are listed in Table 2. The MSE values of the AEA21 model are significantly higher than those of the ML-based models. For the linear amplifications at shorter periods, the MSE values of the RF-based model are slightly higher than those of the DNN-based model. At higher periods, on the contrary, the ML-based models show similar predictions. For the nonlinear amplifications, the MSE values of the RF-based model are slightly higher than those of the DNN-based model.

For the detailed comparisons of calculated and predicted amplifications, the residuals of $\ln(F_{lin})$ and $\ln(F_{nl})$, both calculated as $\ln(F)$ (calculated) – $\ln(F)$ (predicted), are shown in Figure 10 and Figure 11, respectively. Also shown are the binned means of the AEA21 and ML-based models, as well as $\pm 1\sigma$ of the residuals. For the residuals of the $\ln(F_{lin})$, Figure 11, both the AEA21 and DNN-based models produce binned means that are centered around zero residuals. However, the calculated σs show significant differences. The maximum values of σ are 0.332 at T = 2.0 s and 0.141 at T = 0.1 s for the residuals of the AEA21 model and the DNN-based model, respectively. The ML-based model yields greatly lower uncertainty compared with that of the AEA21 model, displaying a pronounced superiority in the prediction performance. For the residuals of $\ln(F_{nl})$, Figure 11, the AEA21 model results in binned means that deviate from zero residuals, whereas the ML-based models are again successful in yielding values well centered around them. The residual σs for the linear and nonlinear components are presented in Figure 12 and Figure 13. For the AEA21 model, the residual σ for $\ln(F_{lin})$ is larger compared with that for $\ln(F_{nl})$. However, the residual σs for both the $\ln(F_{lin})$ and $\ln(F_{nl})$ are extremely low for the ML-based models. By using the ML-based models, the uncertainty caused by the prediction model can be reduced.

The extensive comparisons highlight that the performance of the ML-based models is exceptional. Among the three prediction models, the remarkable performance of the ML-based models to predict the site amplification demonstrates that it can be potentially used as an alternative to the 1D numerical model for the prediction of site amplification for shallow bedrock sites. Although both ML-based models show agreeable fits with the numerical outputs, it is also worth noting that the DNN-based model provides marginally more favorable predictions of the site response, producing lower scatter and standard deviations for both linear and nonlinear amplifications. Considering up to five times lower computational cost for training, it is therefore recommended to use DNN for training the seismic site response.

It should be noted that the favorable predictions of the ML-based models were achieved without using shear modulus reduction and damping curves, as well as shear strength data, as inputs. It is possible that shallow sites subjected to low to moderate intensity motions produced low strain levels, thereby limiting the effect of nonlinearity. Additionally, the use of only cohensionless soil layers representative of inland profiles of Korea may have reduced the variability produced by the soil type. The effect of soil nonlinearity for various soil types and a wide range of site profiles should be comprehensively explored in a future study. However, it is demonstrated that the ML-based models have a capacity to learn the complex nonlinear soil response observed in vertical propagation of shear waves.

The comparison of the regression and ML-based models may be viewed as unfair because of the critical differences in the input features. Whereas the regression model only uses scalar proxies for both the site profile and input motion, the ML-based model utilizes matrix data. It should be noted that the performances would have been not as much different if identical input features were used. The primary purpose of this comparison, as presented in the introduction of this paper, is not to highlight the enhanced performance of the ML algorithms, but to demonstrate the importance of using matrix data in the training. In a classical regression procedure, the application of vector or matrix data is not possible.

5. SUMMARY AND CONCLUSION

In this study, two ML-based site amplification models were developed using a comprehensive database of both linear and nonlinear site response analysis outputs obtained from 1D simulations performed on shallow bedrock profiles. ML algorithms used were RF and DNN, both widely used in practice for training. In performing the site response analyses, the measured and randomized V_S profiles were used. The shear strength adjustment was applied in performing the nonlinear analyses. Whereas scalar data which include site and motion proxies have been used in previous studies, matrix data were used for training, including the response spectrum of the input ground motion and full V_S profile.

For the ML-based models, linear and total amplifications were trained separately. The nonlinear amplification was calculated by subtracting the linear amplification from the total amplification. Linear and nonlinear site amplifications are then compared separately with the simulated outputs. The residuals of the calculated and predicted amplifications are also determined. The performances of the ML-based models are also compared with a regression-based site amplification model which is conditioned on the T_G and V_{S30}, as well as the ground motion intensity.

It is demonstrated that both the linear and nonlinear amplifications predicted with the ML-based models produce exceptional fits with the numerically calculated results and significantly outperform the regression-based model. The regression-based model is successful in providing acceptable binned mean results. However, it cannot capture the pronounced variations in the surface responses.

The DNN-based model produces lower MSEs than those predicted by the RF-based model across despite using fewer input features. Although the RF-based model is an explainable model which can provide information on the relative importance of input features, the model should be trained repeatedly with new input features. However, the DNN-based model can perform the feature extraction and the training simultaneously via its network. Moreover, the computational cost is significantly lower compared with that of the RF-based model. The comparisons highlight that the ML-based models have the potential to replace the numerical model for use in the prediction of the site amplification for shallow bedrock sites if trained with sufficient data that covers a wide range of profiles and motions.

The comparison of the ML-based models with the regression-based model was not intended to demonstrate the higher performance of the former algorithm to yield predictions given an identical set of input features. Rather, it is designed to show the importance of the ability to utilize vector or matrix data in the training. By using the period-dependent SA and depth-dependent V_S array in training, prediction accuracy is dramatically improved.

It should also be highlighted that they were solely developed from numerical simulation outputs, without validation against earthquake recordings. Therefore, the models should be constrained to measurements including downhole arrays for possible application in practice. It should also be noted that the predictions were made for relatively simple site structures, which are shallow bedrock sites with no stress reversals composed of non-cohesive soils. The predictions are shown to be favorable for this However, extensive training should be succeeded to develop a ML-based site response model for a broad range of site profiles.

6. ACKNOWLEDGEMENTS

This work was supported by the Horizon 2020 Programme of the European Commission under the MSCA-RISE-2015-691213-EXCHANGE-Risk grant (Experimental and Computational Hybrid Assessment of NG Pipelines Exposed to Seismic Hazard, www.exchange-risk.eu) and the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 22CTAP-C164148-02).

7. REFERENCES

[1] E. Bojórquez, J. Bojórquez, S.E. Ruiz, A. Reyes-Salazar, Prediction of Inelastic Response Spectra Using Artificial Neural Networks, Math Probl Eng, 2012 (2012) 1-15.

[2] J. Bojórquez, S. Ruiz, E. Bojórquez, Estimation of Inelastic Response Spectra Using Artificial Neural Networks, in: 15 WCEE, Lisbon, 2012, pp. 30343-30452.

[3] P. Kamatchi, J. Rajasankar, G. Ramana, A. Nagpal, A neural network based methodology to predict site-specific spectral acceleration values, Earthquake Engineering and Engineering Vibration, 9 (2010) 459-472.

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

[5] O. Ilhan, J.A. Harmon, O.A. Numanoglu, Y.M. Hashash, Deep learning-based site amplification models for Central and Eastern North America, in: Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions, CRC Press, 2019, pp. 2980-2987.

[6] D. Roten, K.B. Olsen, Estimation of site amplification from geotechnical array data using neural networks, Bulletin of the Seismological Society of America, 111 (2021) 1784-1794.

[7] C. Zhu, F. Cotton, H. Kawase, K. Nakano, How well can we predict earthquake site response so far? Machine learning vs physics-based modeling, Earthquake Spectra, 0 (2022) 87552930221116399.

[8] I.M. Idriss, H.B. Seed, Seismic response of horizontal soil layers, Journal of the Soil Mechanics and Foundations Division, 94 (1968) 1003-1031.

[9] I. Idriss, Response of soft soil sites during earthquakes, in: Proc. HB Seed Memorial Symp., 1990, pp. 273-289.

[10] S.L. Kramer, Geotechnical earthquake engineering, Pearson Education India, 1996.

[11] Y.M. Hashash, D. Park, Non-linear one-dimensional seismic ground motion propagation in the Mississippi embayment, Engineering Geology, 62 (2001) 185-206.

[12] H. Seo, J. Kim, B. Kim, Machine-Learning-Based Surface Ground-Motion Prediction Models for South Korea with Low-to-Moderate Seismicity, Bulletin of the Seismological Society of America, 112 (2022) 1549-1564.

[13] S. Kim, Y. Hwang, H. Seo, B. Kim, Ground motion amplification models for Japan using machine learning techniques, Soil Dynamics and Earthquake Engineering, 132 (2020) 106095.

[14] C. Zhu, F. Cotton, H. Kawase, A. Haendel, M. Pilz, K. Nakano, How well can we predict earthquake site response so far? Site-specific approaches, Earthquake Spectra, 38 (2022) 1047-1075.

[15] E.M. Thompson, L.G. Baise, Y. Tanaka, R.E. Kayen, A taxonomy of site response complexity, Soil Dynamics and Earthquake Engineering, 41 (2012) 32-43.

[16] P.B. Schnabel, SHAKE: A computer program for earthquake response analysis of horizontally layered sites, EERC Report 72-12, University of California, Berkeley., (1972).

[17] Y. Hashash, M. Musgrove, J. Harmon, O. Ilhan, D. Groholski, C. Phillips, D. Park, DEEPSOIL 7.0, user manual, University of Illinois at Urbana-Champaign, (2017).

[18] M. Aaqib, D. Park, M.B. Adeel, Y.M.A. Hashash, O. Ilhan, Simulation-based site amplification model for shallow bedrock sites in Korea, Earthquake Spectra, 37 (2021) 1900-1930.

[19] R.K. McGuire, W.J. Silva, C.J. Costantino, Technical basis for revision of regulatory guidance on design ground motions: Hazard-and risk-consistent ground motion spectra guidelines, Division of Engineering Technology, Office of Nuclear Regulatory Research, 2001.

[20] G. Toro, Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies, Technical Rep, 779574 (1995).
[21] M. Darendeli, Development of a new family of normalized modulus reduction and material damping curves, in: Austin: University of Texas, Austin: University of Texas, 2001.
[22] J. Jaky, The coefficient of earth pressure at rest, J. of the Society of Hungarian Architects and Engineers, (1944) 355-358.
[23] C. Phillips, Y.M. Hashash, Damping formulation for nonlinear 1D site response analyses, Soil Dynamics and Earthquake Engineering, 29 (2009) 1143-1158.
[24] D.R. Groholski, Y.M. Hashash, B. Kim, M. Musgrove, J. Harmon, J.P. Stewart, Simplified model for small-strain nonlinearity and strength in 1D seismic site response analysis, Journal of Geotechnical and Geoenvironmental Engineering, 142 (2016) 04016042.
[25] M. Aaqib, S. Sadiq, D. Park, Y.M.A. Hashash, M. Pehlivan, Importance of Implied Strength Correction for 1D Site Response at Shallow Sites at a Moderate to Low Seismicity Region, in: Geotechnical Earthquake Engineering and Soil Dynamics V, American Society of Civil Engineers, Austin, Texas, 2018, pp. 445-453.
[26] C.-G. Sun, C.-S. Cho, M. Son, J.S. Shin, Correlations between shear wave velocity and in-situ penetration test results for Korean soil deposits, Pure and Applied Geophysics, 170 (2013) 271-281.
[27] T.F. Wolff, Pile capacity prediction using parameter functions, in: Predicted and observed axial behavior of piles: results of a pile prediction symposium, ASCE, 1989, pp. 96-106.
[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, the Journal of machine Learning research, 12 (2011) 2825-2830.
[29] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467, (2016).
[30] L. Breiman, Random forests, Machine learning, 45 (2001) 5-32.
[31] L. Breiman, Bagging predictors, Machine learning, 24 (1996) 123-140.
[32] T.K. Ho, Random decision forests, in: Proceedings of 3rd international conference on document analysis and recognition, IEEE, 1995, pp. 278-282.
[33] S. Haykin, Neural networks and learning machines, 3rd ed., Pearson Education India, 2010.
[34] A. Zayegh, N. Bassam, Neural network principles and applications, London: Pearson, 2018.
[35] S.W. Smith, The scientist and engineer's guide to digital signal processing, 2nd ed., California Technical Publishing, San Diego, California, 1999.
[36] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: 27th International Conference on Machine Learning, Haifa, Israel, 2010.
[37] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, PMLR, 2015, pp. 448-456.
[38] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249-256.
[39] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 2015.

[40] S.J. Reddi, S. Kale, S. Kumar, On the convergence of adam and beyond, in: 6th International Conference on Learning Representations, Vancouver Convention Center, Vancouver, BC, Canada, 2018.

Table 1 Comparison of the MSE and the MAE for the deterministic DNN between the train and the test dataset.

Analysis	Dataset	MSE	MAE
Linear	Training	0.000404	0.014312
	Test	0.000568	0.016076
Nonlinear	Training	0.000644	0.017714
	Test	0.001210	0.022163

Table 2. MSE of three models for linear and nonlinear amplifications

Linear				**Nonlinear**				
***T* (s)**	**AEA21**	**RF**	**DNN**	V_{S30}	***T* (s)**	**AEA21**	**RF**	**DNN**
0.01	0.0667	0.0026	0.0007	**200~250**	**0.01**	0.0955	0.0058	0.0016
0.1	0.0535	0.0020	0.0007		**0.2**	0.1006	0.0059	0.0022
0.2	0.0395	0.0014	0.0005		**0.5**	0.1527	0.0045	0.0021
0.5	0.0147	0.0005	0.0004	**300~350**	**0.01**	0.1024	0.0058	0.0018
1.0	0.0160	0.0002	0.0003		**0.2**	0.1058	0.0045	0.0018
2.0	0.0160	0.0001	0.0002		**0.5**	0.0876	0.0021	0.0012

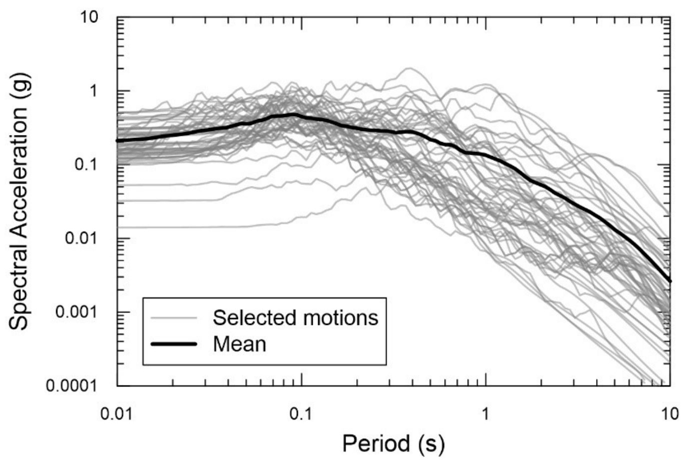

Figure 1. Acceleration response spectra of all ground motions used in the present study.

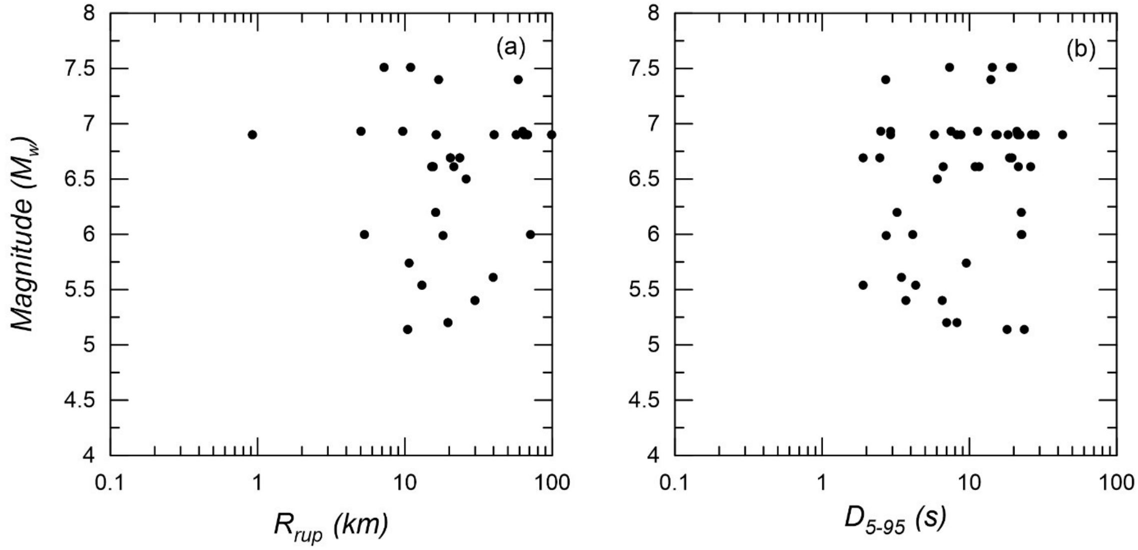

Figure 2. Distribution of moment magnitude (M_w) of input ground motions with respect to (a) rupture distance (R_{rup}) and (b) significant duration ($D_{5\text{-}95}$).

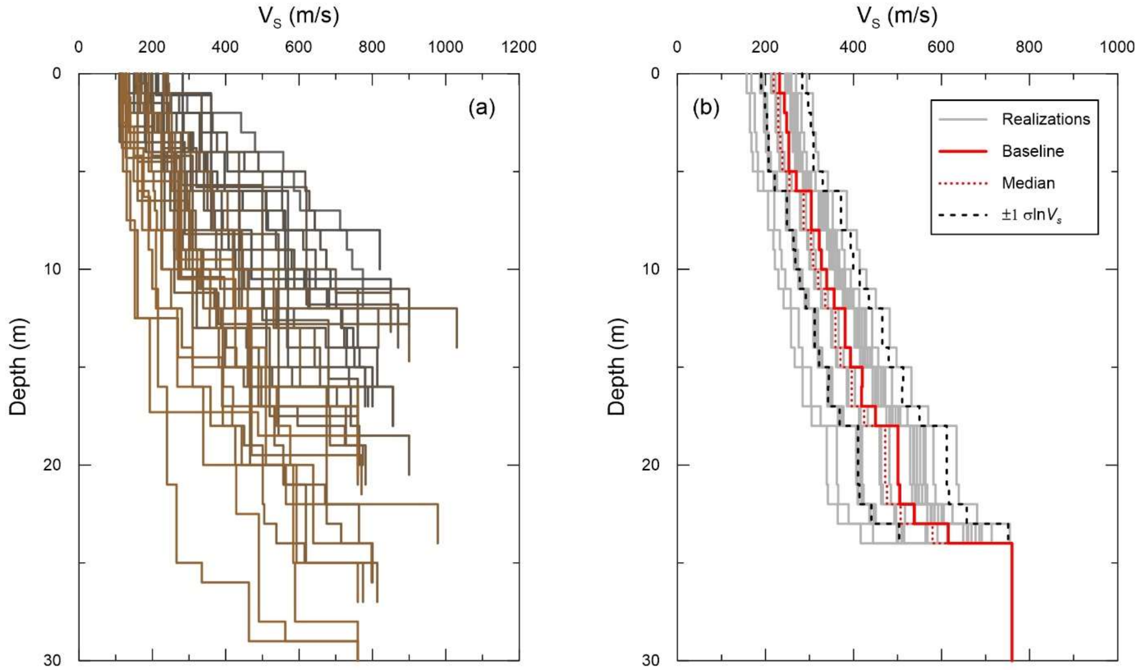

Figure 3. Shear wave velocity profiles: (a) selected 51 baseline profiles, (b) randomized realizations for one selected baseline profile with V_{S30} = 398 m/s.

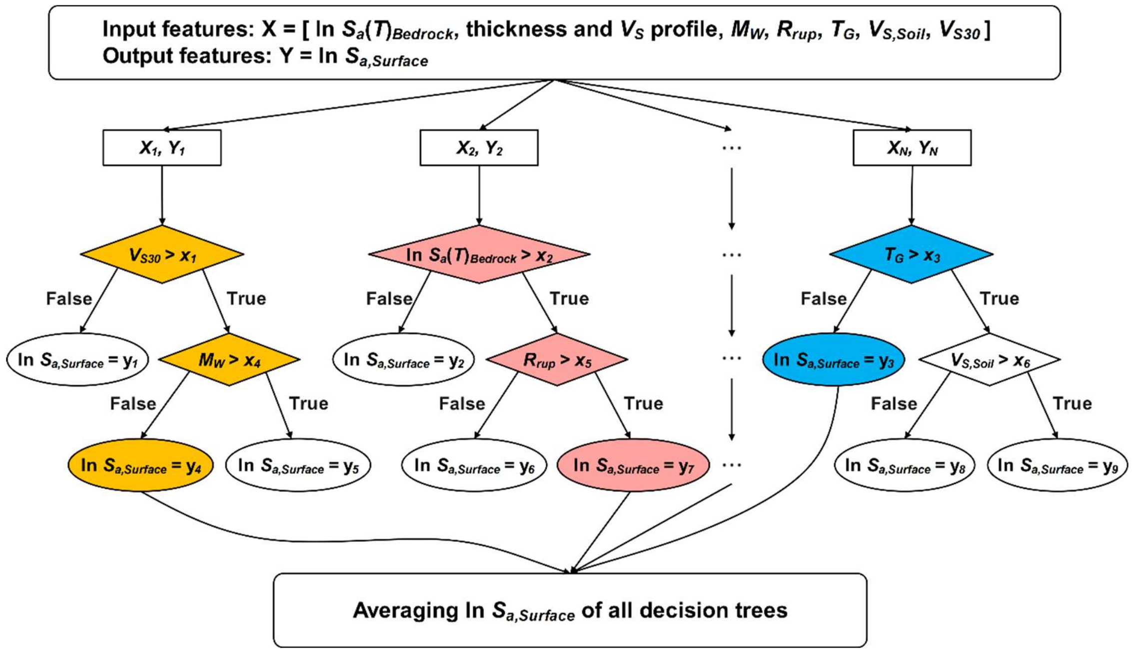

Figure 4. Schematic of the RF algorithm with N decision trees. The decision trees are shown up to depth of 2 with input features. The subset data are sampled N times with replacement. The prediction is calculated by aggregating and averaging ln $S_{a,Surface}$ from all decision trees.

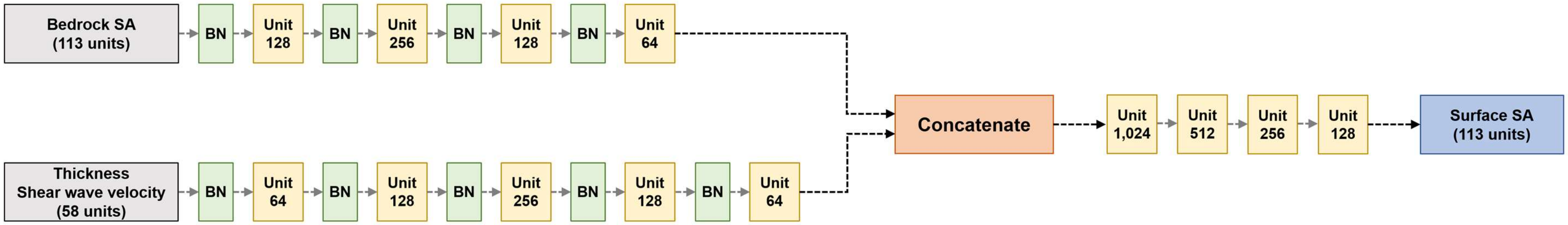

Figure 5. Architecture of the proposed DNN model. BN in the figure represents batch normalization.

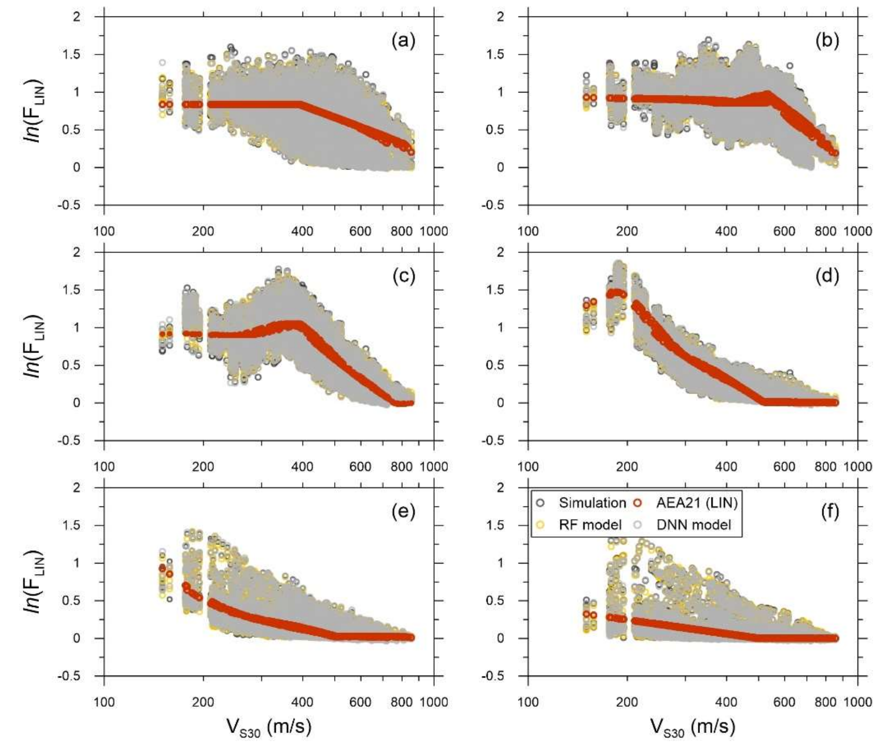

Figure 6. Comparison of the linear amplification components predicted from the AEA21, RF-based and DNN-based models against V_{S30} for selected spectral periods: (a) 0.01 s, (b) 0.1 s, (c) 0.2 s, (d) 0.5 s, (e) 1.0 s, and (f) 2.0 s.

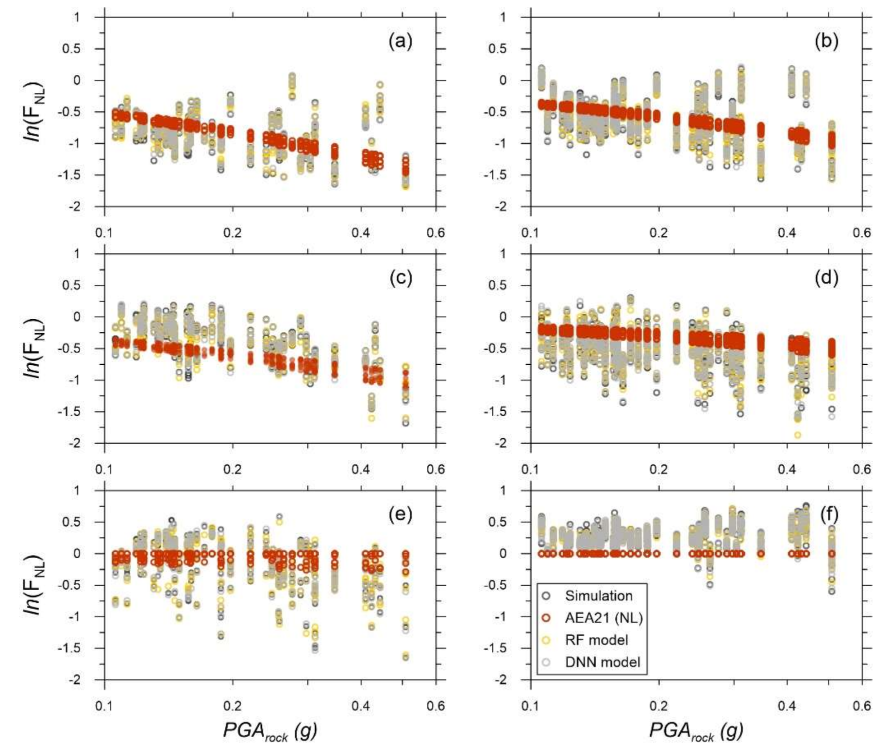

Figure 7. Comparison of the nonlinear amplification components predicted from AEA21, RF-based, and DNN-based models against PGA_{rock} at 0.01 s for (a) 200 < V_{S30} < 250 m/s and (b) 300 < V_{S30} < 350 m/s, at 0.2 s for (c) 200 < V_{S30} < 250 m/s and (d) 300 < V_{S30} < 350 m/s and at 0.5 s for (e) 200 < V_{S30} < 250 m/s and (f) 300 < V_{S30} < 350 m/s.

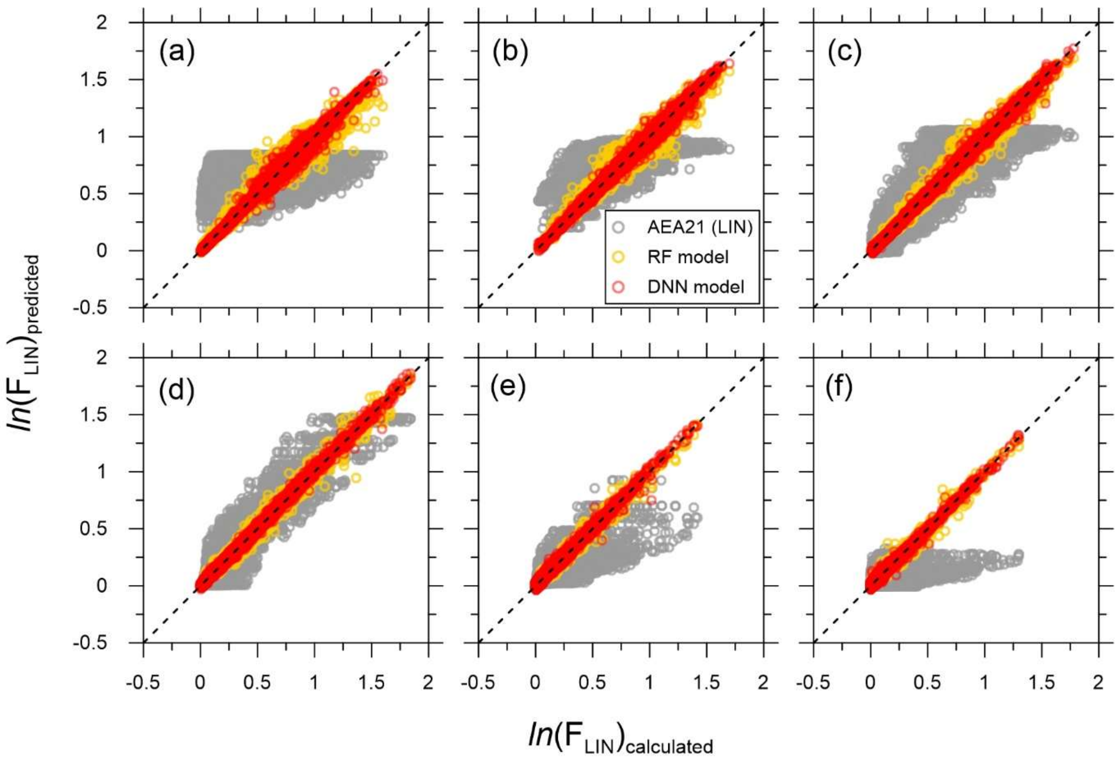

Figure 8. Comparison of the linear amplification components between calculated amplification and predicted amplifications by AEA21, RF-based and DNN-based models for selected spectral periods: (a) 0.01 s, (b) 0.1 s, (c) 0.2 s, (d) 0.5 s, (e) 1.0 s, and (f) 2.0 s.

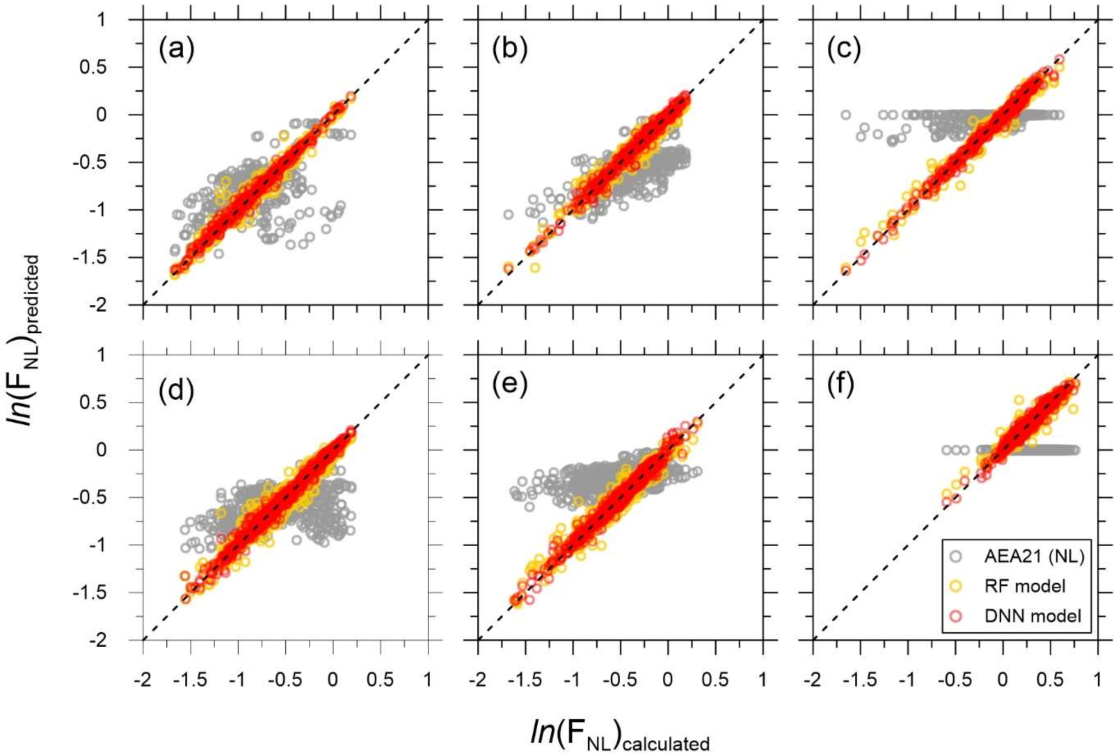

Figure 9. Comparison of the nonlinear amplification components between calculated amplification and predicted amplification by AEA21, RF-based, and DNN-based models at 0.01 s for (a) 200 < VS30 < 250 m/s and (b) 300 < VS30 < 350 m/s, at 0.2 s for (c) 200 < VS30 < 250 m/s and (d) 300 < VS30 < 350 m/s and at 0.5 s for (e) 200 < VS30 < 250 m/s and (f) 300 < VS30 < 350 m/s.

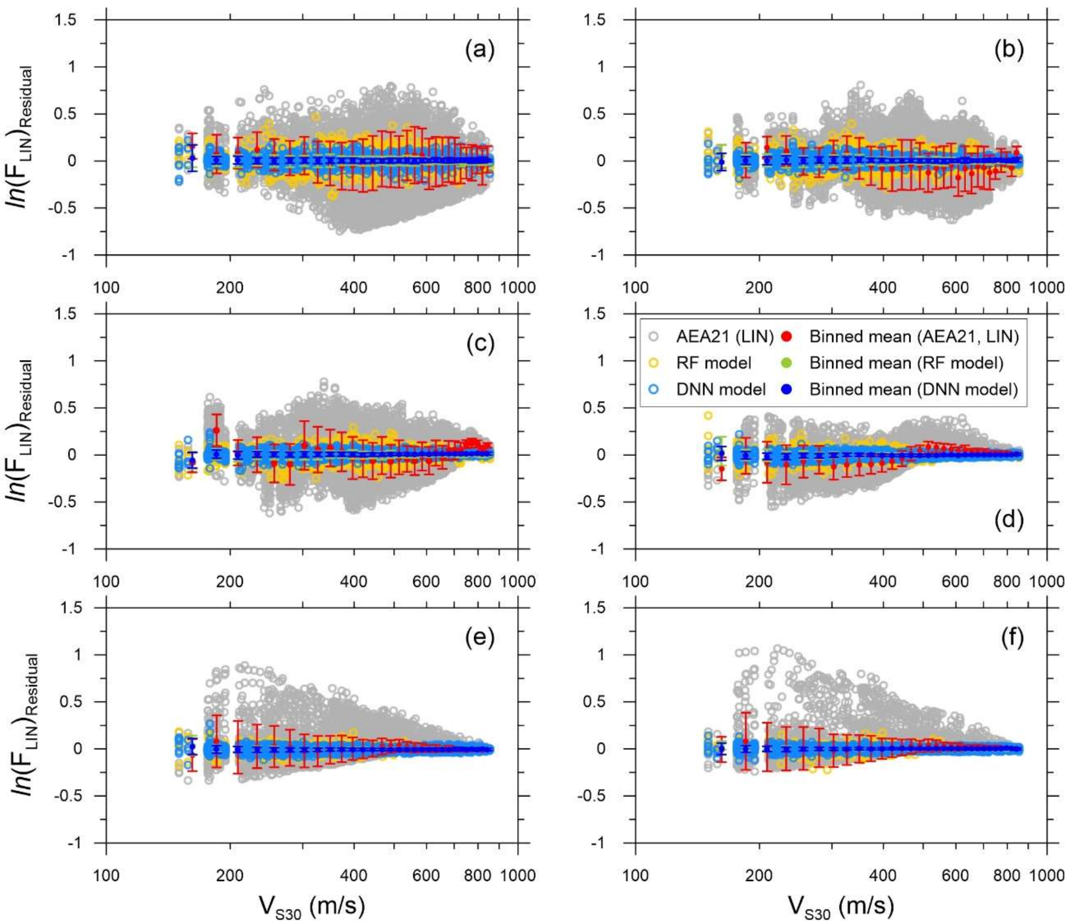

Figure 10. Comparison of the residuals of linear amplification component predicted from AEA21, RF-based and DNN-based models against V_{s30} for selected spectral periods: (a) 0.01 s, (b) 0.1 s, (c) 0.2 s, (d) 0.5 s, (e) 1.0 s, and (f) 2.0s.

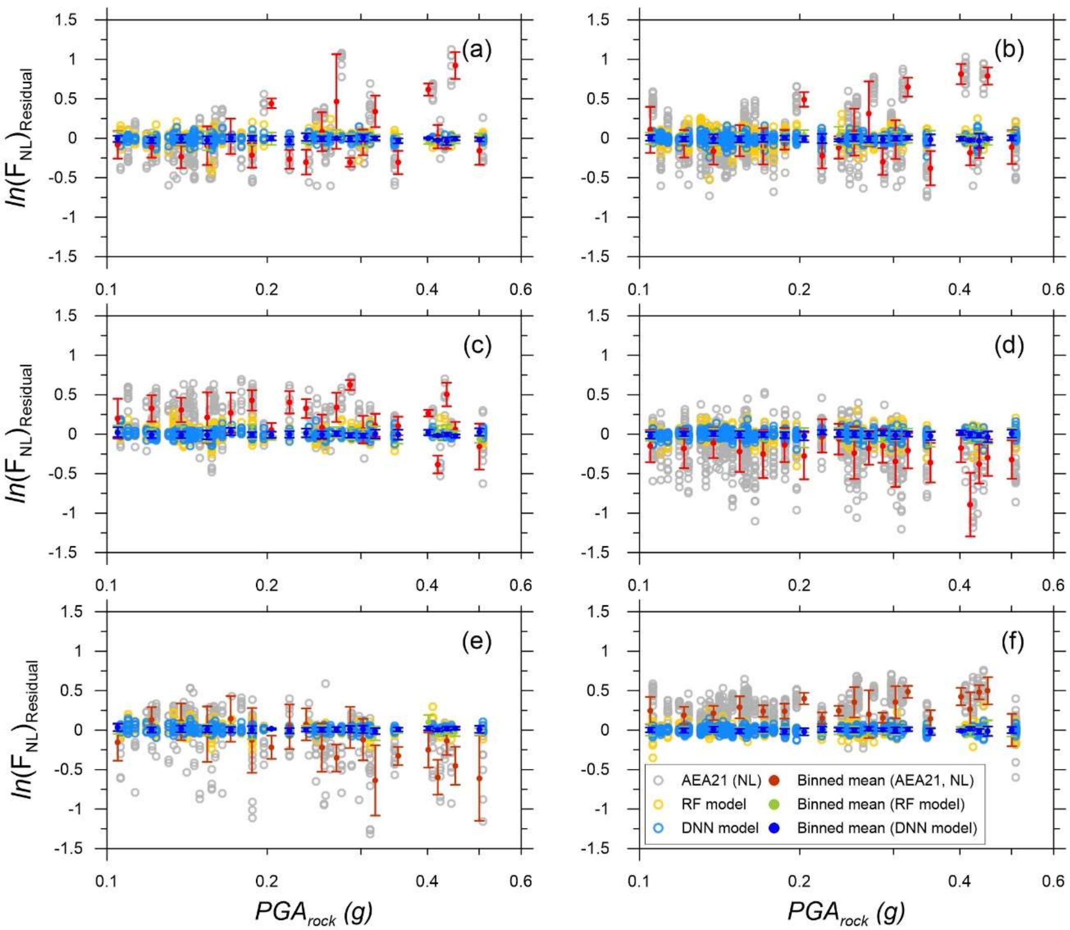

Figure 11. Comparison of the residuals of nonlinear amplification component predicted from AEA21, RF-based and DNN-based models against PGArock at 0.01 s for (a) 200 < VS30 < 250 m/s and (b) 300 < VS30 < 350 m/s, at 0.2 s for (c) 200 < VS30 < 250 m/s and (d) 300 < VS30 < 350 m/s and at 0.5 s for (e) 200 < VS30 < 250 m/s and (f) 300 < VS30 < 350 m/s.

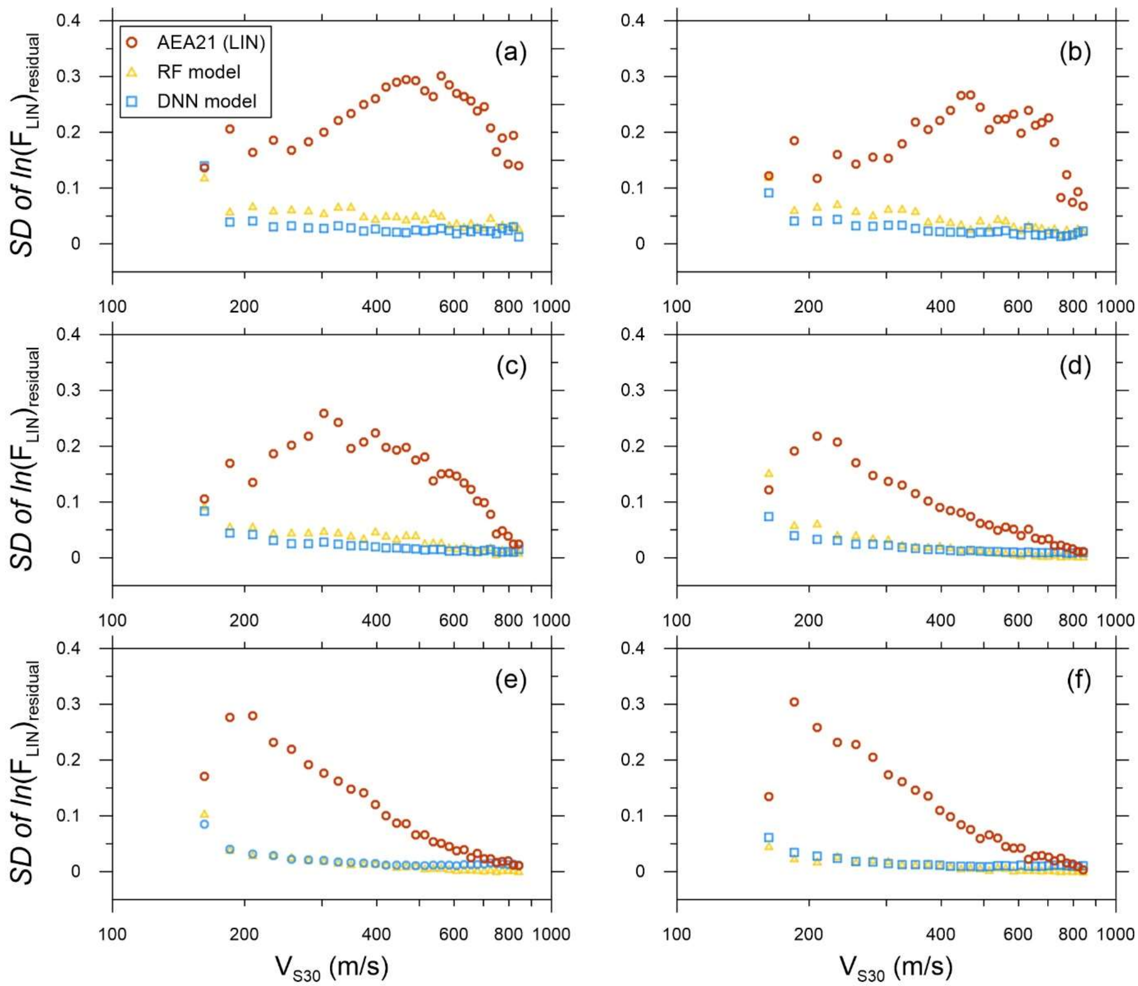

Figure 12. Comparison of standard deviation (SD) of the residuals of linear amplification component predicted from AEA21, RF-based and DNN-based models against Vs30 for selected spectral periods: (a) 0.01 s, (b) 0.1 s, (c) 0.2 s, (d) 0.5 s, (e) 1.0 s, and (f) 2.0s.

587

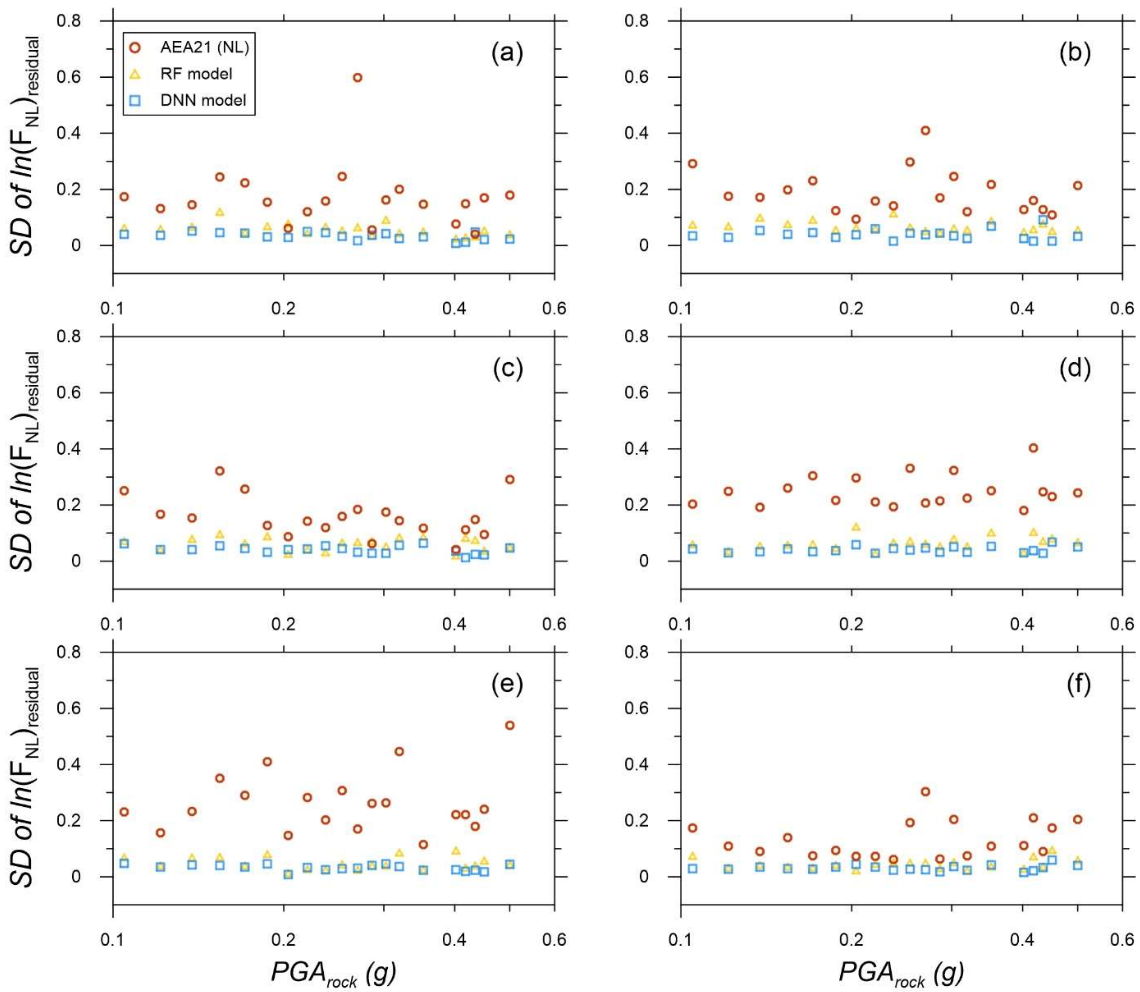

Figure 13. Comparison of standard deviation (SD) the residuals of nonlinear amplification component predicted from AEA21, RF-based and DNN-based models against PGArock at 0.01 s for (a) 200 < VS30 < 250 m/s and (b) 300 < VS30 < 350 m/s, at 0.2 s for (c) 200 < VS30 < 250 m/s and (d) 300 < VS30 < 350 m/s and at 0.5 s for (e) 200 < VS30 < 250 m/s and (f) 300 < VS30 < 350 m/s.