New Generalized ANN Based Hybrid Broadband Response Spectra Generator using Physics Based Simulations

Varun Sharma
IIT Mandi: Indian Institute of Technology Mandi
J. Dhanya
IIT Mandi: Indian Institute of Technology Mandi
Maheshreddy Gade (maheshreddy@iitmandi.ac.in)
Indian Institute of Technology Mandi
S. Jayalakshmi
King Abdullah University of Science and Technology

Research Article

Keywords: Physics Based Simulations, PSHA, ANN, hybrid broadband 36 ground motion, spectral ordinate matching.

Posted Date: June 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1672857/v1

New Generalized ANN Based Hybrid Broadband Response Spectra Generator using Physics Based Simulations

Varun Sharma[1], J. Dhanya[2], Maheshreddy Gade[1*] and S. Jayalakshmi[1]

[1,2,3*]School of Engineering, IIT Mandi, Kamand, Mandi, 175005, Himachal pradesh, India.
[4]Division of Physical Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

*Corresponding author(s). E-mail(s): maheshreddy@iitmandi.ac.in;
Contributing authors: d21029@students.iitmandi.ac.in; dhanyaj17@gmail.com; jayalakshmi.sivasubramonian@kaust.edu.sa;

Abstract

To evaluate the seismic risk associated with infrastructures, site-specific seismic hazard studies are needed to be performed. Further, for non-linear time history analysis, one requires broadband ground motion. In modern times using physics-based simulations (PBS) for deriving the ground motion for future earthquakes has got due consideration. The PBS helps in decreasing the uncertainties related to the hazard estimation compared to GMPEs. There is a certain threshold to the PBS methods with the limited computational facilities. Hence, certain hybrid methods are required to attain a Broadband spectra for the simulated ground motion. This study uses a new ANN-based model to generate broadband ground motion spectra using the low-frequency spectral acceleration from PBS, source, path, and site parameters as input variables. A detailed parametric study and performance evaluation was made to

identify the optimal input parameters in conjunction with the best-suited ANN architecture. The performance of the ANN model is demonstrated for Iwate ($\boldsymbol{M_w}$ 6.9,2008) earthquake. We found that the predicted values from the developed ANN model are in good agreement with the recorded data. Furthermore, time histories are generated using the spectral ordinate matching technique from the estimated broadband spectra.

Keywords: Physics Based Simulations, PSHA, ANN, hybrid broadband ground motion, spectral ordinate matching.

1 Introduction

Several past earthquakes have caused huge societal and life damage to mankind. The ever-growing population, economic activities,and rising infrastructure demand escalate the seismic risk. In order to minimize the losses and to come up with better engineering solutions seismic hazard analysis is required. Hazard estimation provides a basic framework for building codes, insurance companies and mitigation strategies. In order to ensure public safety against a possible earthquake event, researchers are trying to come up with earthquake forecasting capabilities equivalent to that of weather forecast (Maechling et al., 2007). Although the target is far from accomplished, incorporating physics laws in ground motion simulations is a step forward. Traditionally, characterization of earthquake hazard is done using Probabilistic Seismic Hazard Analysis (PSHA) (McGuire and Arabasz, 1990). The PSHA framework includes several empirical models that are typically subjected to variability in source, site, and path modeling. The PSHA framework includes several empirical models that are typically subjected to variability in source, site, and path modeling. The corresponding variability might result in significant uncertainties in the hazard model, reducing the predictions' accuracy. One of the key contributors to the uncertainty in the seismic hazard framework is Ground Motion Prediction Equations(GMPEs). GMPEs represent the

probability of exceedance of an intensity measure for an earthquake event at a particular site accounting for source, site, and path effects. Although GMPEs are easy to compute, these equations are regression-based and do not completely capture various physical phenomena like the source directivity and basin effect observed during an earthquake (Graves et al., 2011). In addition, even though computationally efficient, GMPEs do not provide earthquake acceleration time histories that are required for complex non-linear problems. Also, the obtained Intensity Measure (IM) from GMPEs is an average of several sites spatially distributed over a large region (ergodicity) rather than multiple ruptures occurring at a single site, which results in high aleatory uncertainty. Further, the ergodic assumption in GMPEs shows an upward bias in hazard estimation (O'Connell et al., 2007). By including inelastic wave propagation effects that are site-specific through physics-based simulations (PBS), the aleatory uncertainty in the hazard estimation can be reduced (Milner et al., 2021). The PBS replaces the GMPE term and generates IMs based upon simulated ground motions, which reduces uncertainty in the PSHA framework and develops more realistic earthquake hazard models. With the advancement in computational infrastructure, researchers have developed physics-based earthquake simulators (Paolucci et al., 2015; Cui et al., 2013; Shaw et al., 2018; Graves et al., 2011). In some of the recent studies, researchers proposed to replace the traditional GMPEs in PSHA with PBS results (Maechling et al., 2007). Platforms like Cybershake (Graves et al., 2011) and Cybershake NZ (Bradley, Tarbali, Lee, Huang, Motha, Bae, Polak, Zhu, Schill, Patterson, and Lagrava, Bradley et al.) have generated PBS for over 415,000 and 11,362 rupture variations, respectively.

With all the benefits of PBS, the computational requirement remains a major challenge. The available infrastructure for PBS involves a frequency

threshold, i.e., 1-1.5Hz of accuracy (Paolucci et al., 2018). Generally, earthquake ground motion is comprised of broadband frequencies(0-10Hz), PBS for which requires a very fine mesh size and large detailing of topological and geological features, which demands a huge computational cost. Recently, researchers are coming up with techniques to reduce these computational requirements (Cui et al., 2013; Shaw et al., 2018; Milner et al., 2021) by including various approximations. However, a standard methodology is yet to be achieved. A suitable alternative would be to develop broadband ground motion using hybrid methodologies combining low-frequency PBS with a suitable high-frequency generator. The effectiveness of the corresponding hybrid approaches is evident as many researchers have successfully implemented them to simulate broadband ground motions (Smerzini and Villani, 2012; Razafindrakoto et al., 2018; Gade and Raghukanth, 2017; Jayalakshmi et al., 2021). With the recent advancements in the field of Artificial Intelligence(AI) and machine learning, researchers are including the AI-based methods in the field of earthquake engineering (Ahmad et al., 2008; Derras et al., 2014; Dhanya and Raghukanth, 2018a; Gade et al., 2021; Mignan and Broccardo, 2020; Dhanya and Raghukanth, 2022). Recently, Paolucci et al. (2018) have proposed an ANN model to predict high-frequency response spectra using PBS-based low-frequency spectra. The ANN model was designed and trained by utilizing the SIMBAD database, which consists of only 500 strong motion records corresponding to 130 earthquake events with M_w range of 5 to 7.3 and epicentral distance (R_{epi}) $<$ 35 km. Further, the efficiency of the proposed methodology was demonstrated by generating Broadband ground motion for the 2012 Po Plain earthquake M_w 6. However, due to the limited range of M_w and R_{epi}, the model is only applicable for specific events. Further, the developed ANN

model uses only Spectral Acceleration (S_a) values for T> T^*(Period corresponding to threshold frequency) as the predictor variables. The application of the corresponding idea in a more general sense demands the inclusion of a more comprehensive dataset alongside source, path, and site parameters as predictor variables. Furthermore, an extensive parametric performance evaluation is needed to identify the best-suited ANN architecture.

The present study aims to develop a prediction model for high-frequency spectra using PBS-based low frequency S_a values, source, path, and site parameters as predictor variables. Further, a methodology is developed to generate Broadband ground motion from the predicted ANN spectra using the low-frequency PBS. The comprehensive dataset available in the NGA-West2 database was utilized in developing the model. Here, moment magnitude (M_w), rupture distance (R_{rup}), shear wave velocity in top 30m (V_{s30}), focal mechanism (FM) and S_a for T $\geq T^*$ are considered as input variables. Additionally, the effectiveness of the developed model is demonstrated by simulating the ground motions for the Iwate (M_w 6.9,2008) earthquake. The model shows fairly accurate results compared with the recorded data for several stations.

2 Database

The present study utilizes spectral acceleration data available in NGA-West 2, which is a comprehensive global database for ground motion time histories Taylor et al. (2007). Here, horizontal rotd50 response spectra available at https://peer.berkeley.edu/research/data-sciences/databases is used. The database constitutes 21540 ground motion records from 600 events as on January 2015. . The spectral acceleration values are for horizontal rotd50 component. The data is filtered (Dhanya and Raghukanth, 2018a) in order to avoid unsuitable records. The filtration led to the elimination of the spurious

records and left with 13552 records from 288 events. The distribution for the sorted data among various ranges of distance and magnitude among various focal mechanism classes is shown in Fig. 1. The range of magnitude (M_w) and rupture distance (R_{rup}) is 3 - 7.9 and 0.05km - 497 km, respectively and V_{S30} ranges from 89 to 2016 m/s. The classification focal mechanism classes are based on rake angle, the ranges of which have been included in the Fig. 1. The filtered database consists of 180 strike-slip, 38 normal and normal oblique, and 70 reverse and reverse oblique events. The focal mechanisms (FM) are assigned with flags 1, 2, and 3 for strike-slip, normal and normal oblique, and reverse and reverse oblique. Additionally, the spectral accelerations values of 91 periods between 0 and 4s are considered for the analysis.

3 ANN Model for Prediction of Short Periods Spectra

ANN models are generally used to develop non-linear relationships between a large set of input and output variables. The inspiration of the methodology is the functioning of the human brain, which can work on several parallel networks of interconnected neurons at an instant. The methodology is simple to use and highly efficient for the generation of predictor models. The architecture of an ANN model consists of a layer of input neurons that is connected with the hidden layers that are further connected with a layer consisting of output neurons (Wang, 2003). Each connection is assigned with certain weights, and the weights are adjusted in such a manner that it matches the desired output. Other important elements of an ANN network are the transfer function and

error function. The general representation of an ANN model is:

$$h_i = \sigma(\sum_{j=1}^{N} W_{ij} x_j + T^{hid}) \tag{1}$$

where h_i is the output parameters, σ is the transfer function , W_{ij} is the weights, x_j is the input parameters and T^{hid} is the threshold associated with hidden neurons.

In this study, a feed-forward neural network with a multi-layer of perceptron (MLP) is developed to predict short period spectral ordinates ($T < T^*$) based upon long period spectral ordinates ($T \geq T^*$) and parameters representing source, path, and site. Here, T^* is the threshold or crossover period. The initial step in generating a neural network was deciding the functional form such that it captures the buried features within the data with minimal variance. For this purpose, several trials were performed using different combinations of input variables and transfer functions (Nayek and Gade, 2022). The corresponding details are included in the sullplementary material. The final form of input and output variables is as follows:

$$log_{10}S_a(T < T^*) = f(log_{10}S_a(T \geq T^*), M_w, log_{10}V_{(S30)}, R_{rup}, log_{10}R_{rup}, FM) \tag{2}$$

where M_w - Moment magnitude, S_a - Spectral Acceleration, R_{rup} - rupture distance, V_{s30} - shear wave velocity in the top 30m of soil and FM - focal mechanism.

The next decision in the development of a neural network is selecting an optimal number of hidden layers and hidden neurons. However, there are no stringent criteria for selecting the number of hidden layers for a neural network. From the past studies, it has been found that in most cases, a single layer of

hidden neurons is sufficient to capture the model characteristics (Wang, 2003). Therefore, this study also employs a single layer of hidden neurons. Several criteria exist for deciding the number of hidden neurons, and researchers have taken the number equal to twice the number of input neurons (Berry and Linoff, 2004). Nevertheless, no improvement has been observed in the model performance when the hidden neurons are greater than the input variables. Therefore, to avoid the problem of over-fitting, we have taken the number of hidden neurons equal to the number of input neurons. Furthermore, an appropriate scaling function is used to bound the data between 0 to 1 or -1 to 1 to maintain uniformity among the variables. Here, we scaled out parameters between range -1 to 1 using the scaling, such that

$$Y = a(Y - Y_{min}) - 1 \qquad where \qquad a = \frac{2}{Y_{max} - Y_{min}} \tag{3}$$

where, Y is the parameter value to be scaled, Y_{max} and Y_{min} the corresponding maximum and minimum values. Further, the procedure includes training, validation, and testing of the ANN model. For this purpose, a random division of data is done into the training, validation, and testing sets in 75, 15, and 15%, respectively. The purpose of the step is to make sure that the model does not show any trend/bias towards a particular input parameter. In the training phase, the error function between the predicted and actual values is minimized using different algorithms. This study utilizes Levenberg-Marquardt (LM) algorithm with the fitnet function available in the Matlab Deep learning toolbox MATLAB (2019). Thus, we finalized a total of 88 cases having varied combinations of input variables and transfer functions in arriving at the final prediction model. In all these cases, the crossover period T^* is taken as 0.75s.

The performance of each of the model architecture is assessed based on the following performance criteria:

1. Pearson correlation coefficient (R)

$$R = \frac{\sum_{i=1}^{N}(O_i - \overline{O_i})(\hat{O}_i - \overline{\hat{O}_i})}{\sqrt{\sum_{i=1}^{N}(\hat{O}_i - \overline{\hat{O}_i})^2 \sum_{i=1}^{N}(O_i - \overline{O_i})^2}} \tag{4}$$

2. Performance Parameter (PP)

$$PP = 1 - \frac{\langle \| O - \hat{O} \|^2 \rangle}{\sigma^2_{\hat{O}}} \tag{5}$$

3. Root Mean Squared Error ($RMSE$)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - \hat{O}_i)^2}{N}} \tag{6}$$

4. Standard deviation of error ($\sigma(\epsilon)$)

$$\sigma(\epsilon) = \sqrt{\frac{\sum_{i=1}^{N}(O_i - \hat{O}_i) - \overline{(O_i - \hat{O}_i)})}{N-1}} \tag{7}$$

Where, $O = log_{10}S_a(T > T^*)$ is the output from the models, N is the number of data-points. The best model will have R and PP values closer to unity and MSR and $\sigma(\epsilon)$ minimum. We observed that the best transfer function varies based on the form of input (Refer Supplement material). Furthermore, the best performance was obtained for the model containing source, path, site characteristics, and long-period spectral accelerations. The corresponding

model performed best with transfer functions as tan-sigmoidal between the input to hidden layers and pure-linear between hidden to output layers. The related functional form of the model is as shown further:

$$log_{10}S_a(T > T^*) = f_2(bias_2 + \sum_{k=1}^{m} W_{k,2} f_1(bias_{k,1} + \sum_{i=1}^{n} W_{ik,1} X_i)) \quad (8)$$

where,

$$f_1(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \qquad and \qquad f_2(x) = x \quad (9)$$

In eqn. 8 the X_i is the input consist of the parameters in eqn. 2, $W_{ik,1}, bias_{k,1}$ are the unknowns between input-hidden layers and $W_{k,2}, bias_2$ are the unknown coefficients between the hidden-output layers.The best performing the network architecture for the crossover period of $T^* = 0.75$s consists of 21 input neurons and 17 output neurons, with 836 unknown weights and biases. The details of the network architecture can be understood from Fig. 3. Since physics-based simulators have different threshold frequencies, ANN models are also developed for additional periods of 0.5 and 1s. The corresponding architecture for crossover periods of 0.5s and 1s consists of 24 input neurons, 14 output neurons, 950 unknown weights and biases, and 18 input neurons, 20 output neurons, 722 unknown weights and biases, respectively. The performance parameters corresponding to each crossover period (T^*) are compiled in Table 1. We obtained comparable performance at all crossover periods. Furthermore, the performance and applicability of these models are investigated in detail in the following section.

4 Performance Analysis of ANN Model

An essential procedure to evaluate the performance of the developed models is through residual analysis. Further, residual analysis is helpful in investigating whether the model has the ability to reproduce major features within the data. Residual is the difference between the actual and the model predicted value, and ideally, the sum and mean of the residual should be zero. In this study, the logarithmic residual between the predicted and the recorded S_a is plotted for three crossover periods (T^*) is presented in Fig. 3. The plot shows that the mean residuals do not contain any trend. Further, it can be concluded that the mean value of residual value oscillates around zero, which depicts the accuracy of the method and proves that there does not exist a bias towards a particular period. Additionally, we observe that the residues are minimum at the crossover period. The corresponding feature of the models will facilitate avoiding any jump that might arise in typical hybrid broadband methodologies (Paolucci et al., 2018). Furthermore, there always exists a debate on ANN usage in natural phenomena. Since there exist several hidden layers in the network, it becomes obligatory to verify whether the developed model follows the physical laws. The variation of the predicted S_a values with M_w , R_{rup} and V_{s30} for different fault mechanisms is shown in Fig. 4 and 5. The corresponding demonstration is made for a crossover period of 0.75s. Here, the long period spectral acceleration values are taken from the mean prediction by Dhanya and Raghukanth (2018b) for the same database. From Fig. 4, It can be observed that the S_a value decreases with an increase in R_{rup}, and the peak period of the spectra is shifting towards long periods with the increase in magnitude. Furthermore, from Fig. 5 it is observed that the spectral acceleration values increase with the decrease in V_{s30}, and the peak period gets shifted to long periods. It is also interesting to note that there are no jumps

at the crossover period. Thus it can be concluded from the performance evaluation that the models are unbiased and able to capture the known physics of ground motion. Another interesting observation is that the model inherently maintains the compatibility at the crossover period. Nevertheless, we would like to state that all models listed in Table 1 performed equally well and can be applied suitably depending on the available computational facility. It would be interesting to observe the application of the ANN model to a real earthquake, a demonstration of that has been presented in the next section.

4.1 Case study of Iwate Earthquake,2008 M_w 6.9

4.1.1 Earthquake Source Model

On 14 June 2008, an earthquake of magnitude 6.9 M_w occurred in the borders of Iwate and Miyagi districts of Japan. The focal mechanism is the reverse fault with W-NW and E-SE as the compression axis. The rupture models from several source-inversion studies are available for the event at http://equake-rc.info/SRCMOD/searchmodels/viewmodel/s2008IWATEx01ASAN/. These rupture models differs in amplitude, location of large slip regions (or asperities), hypocenter location, fault dimensions and temporal parameters. These differences arise due to the assumptions used in each source-inversion. In this study, the rupture models of Cultrera et al. (2013) and Asano and Iwata (2011) have been used . The details of the source parameters for the rupture models are compiled in Table2. The Cultrera et al. (2013) model is based upon a two-staged non-linear kinematic source inversion technique of the fault plane Piatanesi et al. (2007). The Cultrera's model uses Yoffe-type source time function to describe the temporal evolution of slip (Tinti et al., 2005). The Asano and Iwata (2011) model uses multiple time windows in the inversion, the temporal evolution of slip is represented by a

smooth ramp in the slip velocity function, with a constant rise time of 4.5 sec and an average rupture velocity of 2.4km/s (0.8Vs). The corresponding slip and rupture time distributions for the models are shown in Fig. 6.

4.1.2 Low-Frequency Physics Based Simulations

The low-frequency simulations are performed using a discrete-wavenumber finite element method Olson et al. (1984); Spudich and Xu (2002). To consider the slowest shear wave velocity, 2.7km/s a grid resolutions of dz = 150m is used to simulate ground motions up to a maximum frequency of 3Hz. Adopting a 1D Earth structure (Table 3), synthetic waveforms at all recording stations (Joyner-Boore distance Rjb $<$ 150km) are computed. However, it is beyond the scope for the numerical simulations to capture nonlinear site effects due to the resolution of the mesh. Therefore, we correct for the site effects by removing the mean site residuals for each site class at selected natural Periods. The database for the Iwate event is available at NGA West-2,the data consists of spectral ordinates for 150 stations. The low-frequency PBS simulations are performed for all of the recording stations. For representation purpose three stations AKT017, AKT012, and IWTH17 having R_{rup} values 33.8, 58.1 and 73.7 km respectively has been chosen. The low frequency time histories corresponding to both source models for the three stations are shown in Fig. 7. It can be noticed that the arrival time increases with the increase in distance of the station from the fault rupture. Also, for the Cultrera et al. (2013) model the peaks amplitudes are larger, possibly due to the proximity of the stations to largest asperity region of the source.

4.1.3 ANN Model predictions for the Iwate Earthquake

Using the long period PBS as input and the developed ANN model discussed in section 3, the broadband response spectra are developed for all recording

stations of the Iwate Earthquake. Further to check the model performance, the residuals between the simulated and recorded S_a values have been evaluated as shown in Fig. 8. The figure consists of bias for PBS and the developed ANN model (Model-1) indicated in blue and red respectively. It is observed that, among the considered source models, Asano and Iwata (2011) model showed lesser residual and standard deviation compared to Cultrera et al. (2013) model. Such an observation can be attributed to the variability associated with the slip inversion techniques and the data. However, it was noted that the mean bias is very close to zero, pointing toward the efficiency of the developed model. Additionally, it was also noted that the standard deviation in the short-period range is lesser compared to the PBS period range. Now, it will be interesting to check the merits of including path and site characteristics, contrary to the model proposed by Paolucci et al. (2018). Hence, we made a comparison of the same by estimating the bias using the predictions from the ANN model that considered only long-period spectral acceleration as input variables (Model-2, Refer Supplement Material, Sl no: 2 for the corresponding architecture details). The corresponding bias is indicated as green color in Fig. 8. It can be observed that the prediction variability is more for Model-2, and the mean bias also significantly deviated from the zero-line in comparison with that of Model-1. These observations confirm that a better prediction of the short period spectra can be made by including path and site characteristics in the prediction model. Furthermore, a comparison between the recorded and the predicted spectra corresponding to the two ANN models (i.e., Model-1 and Model-2) is made for the three stations (i.e., AKT017, AKT012, and IWTH17) and is shown in Fig. 9. It can be observed from the figure that the predicted spectra from Model-1 are comparable with the recorded spectra. However, in the case of Model-2, the two spectra are not in good agreement with each other, similar to the

observation from the bias plots (Fig. 9). Further, it is interesting to note that Model-1 can generate comparable spectra at short periods, even if PBS spectra are not in line with the recorded values. A trend that is found absent in the case of Model-2. These findings further boost the claim that the inclusion of source and site parameters as predictor variables in the ANN model leads to better prediction of short period spectra. Therefore, it can be concluded that the proposed ANN model assists in developing broadband spectra accurately with lesser computational effort. For non-linear structural analysis, however, we require simulated time histories. The next section represents the generation of broadband ground motion using PBS and ANN predicted spectra.

5 Generation of Broadband Ground Motion

Generation of broadband ground motions using the predicted response spectra is presented in the section. Here, the spectral ordinate matching method, which iteratively modifies the input seed ground motion until it matches the target spectra, is used to generate the ground motions. The transformation is done using either the frequency or wavelet transform methods (Alexander et al., 2014). In this study, the spectral ordinate matching is done in the frequency domain using the code of Ferreira et al. (2020), and PBS is used as input seed ground motion. Due to the deficiency of PBS waveform corresponding to higher frequencies, a broadband time history cannot be achieved directly. Therefore we require a modified seed input that can cater to all frequency ranges. The methodology for the generation of modified seed can be understood using Fig. 10. Here, the initial seed is modified by adding band-pass (1.33 - 15 Hz) filtered white noise. Here, the white noise signal is filtered using the Butterworth band-pass filter in MATLAB. Further, for generating a non-stationary ground motion, the filtered white noise is modified using an envelope function. Here

the envelope function is the realistic exponential function (Takewaki, 2004) given in the following equation:

$$c(t) = exp(-\alpha t) - exp(-\beta t) \tag{10}$$

Where, $\alpha = 0.13$ and $\beta = 0.45$. The added white noise is consistent with the stochastic nature of the earthquake ground motion and provides a complete waveform that can be spectrally matched to the target response spectra. The filtered white noise is then summed up with the PBS signals to generate a modified seed that is appropriate to be used as input for the spectral ordinate matching technique. The generated modified seed consists of intensity corresponding to all frequencies and can be spectrally matched to the ANN generated target spectra as shown in Fig. 11. The recorded and the simulated time histories along with the fourier amplitude spectra is represented in Fig. 12 and 13. It can be inferred from the figure that the generated time histories are in reasonable comparison with the recorded one, having comparable peaks and duration. The Fourier amplitude spectra of the simulated time histories match reasonably well with the recorded values at higher frequencies, even if lower frequency values are not comparable, similar to what was observed before in the response spectra plots. The study for the Iwate earthquake shows that the model can accurately predict response spectra for earthquakes in active shallow crustal regions. Further, we can conclude that the developed methodology can produce reliable results in simulating broadband ground motion for future earthquakes. The spectra obtained from the developed model can be used for physics based seismic hazard studies, and the broadband ground motion can be used for non-linear structural analysis.

6 Summary and Conclusions

This study proposes a new ANN model that can predict a broad-band spectrum using the physics-based low-frequency simulation, source, and site parameters as the predictor variables. The performance study indicates that including all these factors makes the model robust and accurate. The generated model embraces the advantages of the physics-based simulations in simulated ground motion corresponding to a broader spectrum. The statistical study and residual analysis prove the accuracy of the generated model. To show the effectiveness of the methodology Iwate earthquake of 2008, M_w - 6.9 was used as a case study. The results show that the predicted spectra are comparable with the recorded ones. Also, the model can capture the short period S_a values when PBS S_a values are not in line with the recorded values. The model-generated spectra can be used for seismic hazard studies, leading to better hazard mitigation strategies. Further, time histories were obtained using the spectral ordinate matching method from the generated spectra. The simulated time histories are compared with recorded ones for various stations. The results show that generated time histories are in agreement with the recorded ones. Furthermore, broadband ground motion can be generated for scenario events with different source and site parameters. The simulated ground motions is helpful while performing non-linear structural analysis for the earthquake forces. Although the model shows good performance in predicting a real earthquake, a model based upon advanced machine learning techniques will be interesting to explore. Further, NGA-west2 database consist of only shallow crustal earthquakes in active regions. For wider applicability similar model can be developed using other regional/global databases.

7 Statements & Declarations

All the authors involved in the study have contributed equally in the collection of the dataset, generation of the model, simulation for the case study and preparation of the manuscript. For the study no funds or grants were received, neither study corresponds to any financial or non-financial interest. The data and codes generated in this study are available with the corresponding author upon personal request.

References

Ahmad, I., M.H. El Naggar, and A.N. Khan. 2008. Neural network based attenuation of strong motion peaks in Europe. *Journal of Earthquake Engineering 12*(5): 663–680. https://doi.org/10.1080/13632460701758570 .

Alexander, N.A., A.A. Chanerley, A.J. Crewe, and S. Bhattacharya. 2014. Obtaining Spectrum Matching Time Series Using a Reweighted Volterra Series Algorithm (RVSA). *104*(4): 1663–1673. https://doi.org/10.1785/0120130198 .

Asano, K. and T. Iwata. 2011. Characterization of stress drops on asperities estimated from the heterogeneous kinematic slip model for strong motion prediction for inland crustal earthquakes in Japan. *Pure and Applied Geophysics 168*(1-2): 105–116. https://doi.org/10.1007/s00024-010-0116-y .

Berry, M.J. and G.S. Linoff. 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Bradley, B.A., K. Tarbali, R.L. Lee, J. Huang, J. Motha, S.E. Bae, V. Polak, M. Zhu, C. Schill, J. Patterson, and D. Lagrava. Cybershake NZ v19 . 5 :

New Zealand simulation-based probabilistic seismic hazard analysis .

Cui, Y., E. Poyraz, K.B. Olsen, J. Zhou, K. Withers, S. Callaghan, J. Larkin, C. Guest, D. Choi, A. Chourasia, Z. Shi, S.M. Day, P.J. Maechling, and T.H. Jordan. 2013. Physics-based seismic hazard analysis on petascale heterogeneous supercomputers. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*. https://doi.org/10.1145/2503210.2503300 .

Cultrera, G., G. Ameri, A. Saraò, A. Cirella, and A. Emolo. 2013. Ground-motion simulations within ShakeMap methodology: Application to the 2008 Iwate-Miyagi Nairiku (Japan) and 1980 Irpinia (Italy) earthquakes. *Geophysical Journal International 193*(1): 220–237. https://doi.org/10.1093/gji/ggs074 .

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

Dhanya, J. and S.T. Raghukanth. 2018a. Ground Motion Prediction Model Using Artificial Neural Network. *Pure and Applied Geophysics 175*(3): 1035–1064. https://doi.org/10.1007/s00024-017-1751-3 .

Dhanya, J. and S.T. Raghukanth. 2022. Non-linear principal component analysis of response spectra. *Journal of Earthquake Engineering 26*(4): 2148–2167. https://doi.org/10.1080/13632469.2020.1773352. https://arxiv.org/abs/https://doi.org/10.1080/13632469.2020.1773352 .

Dhanya, J. and S.T.G. Raghukanth. 2018b. Ground Motion Prediction Model Using Artificial Neural Network. *Pure and Applied Geophysics 175*(3): 1035–1064 .

Ferreira, F., C. Moutinho, Á. Cunha, and E. Caetano. 2020. An artificial accelerogram generator code written in Matlab. *Engineering Reports 2*(3): 1–17. https://doi.org/10.1002/eng2.12129 .

Gade, M., P.S. Nayek, and J. Dhanya. 2021. A new neural network–based prediction model for Newmark's sliding displacements. *Bulletin of Engineering Geology and the Environment 80*(1): 385–397. https://doi.org/10.1007/s10064-020-01923-7 .

Gade, M. and S.T.G. Raghukanth. 2017. Simulation of strong ground motion for a M W 8.5 hypothetical earthquake in central seismic gap region, Himalaya. *Bulletin of Earthquake Engineering 15*(10): 4039–4065. https://doi.org/10.1007/s10518-017-0146-2 .

Graves, R., T.H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, D. Okaya, P. Small, and K. Vahi. 2011. CyberShake: A Physics-Based Seismic Hazard Model for Southern California. *Pure and Applied Geophysics 168*(3-4): 367–381. https://doi.org/10.1007/s00024-010-0161-6 .

Jayalakshmi, S., J. Dhanya, S. Raghukanth, and P.M. Mai. 2021. Hybrid broadband ground motion simulations in the indo-gangetic basin for great himalayan earthquake scenarios. *Bulletin of Earthquake Engineering*: 1–30 .

Maechling, P., E. Deelman, L. Zhao, R. Graves, G. Mehta, N. Gupta, J. Mehringer, C. Kesselman, S. Callaghan, D. Okaya, H. Francoeur, V. Gupta, Y. Cui, K. Vahi, T. Jordan, and E. Field. 2007. SCEC cybershake workflows-automating probabilistic seismic hazard analysis calculations. *Workflows for e-Science: Scientific Workflows for Grids*: 143–163. https://doi.org/10.1007/978-1-84628-757-2_10 .

MATLAB. 2019. *(R2019a)*. Natick, Massachusetts: The MathWorks Inc.

McGuire, R.K. and W.J. Arabasz. 1990. 12. An Introduction to Probabilistic Seismic Hazard Analysis. *Geotechnical and Environmental Geophysics*: 333–354. https://doi.org/10.1190/1.9781560802785.ch12 .

Mignan, A. and M. Broccardo. 2020. Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. *Seismological Research Letters 91*(4): 2330–2342 .

Milner, K.R., B.E. Shaw, C.A. Goulet, K.B. Richards-Dinger, S. Callaghan, T.H. Jordan, J.H. Dieterich, and E.H. Field. 2021. Toward physics-based nonergodic psha: A prototype fully deterministic seismic hazard model for southern California. *Bulletin of the Seismological Society of America 111*(2): 898–915. https://doi.org/10.1785/0120200216 .

Nayek, P.S. and M. Gade. 2022. Artificial neural network-based fully data-driven models for prediction of newmark sliding displacement of slopes. *Neural Computing and Applications*: 1–13 .

O'Connell, D.R., R. LaForge, and P. Liu. 2007. Probabilistic ground-motion assessment of balanced rocks in the Mojave Desert. *Seismological Research Letters 78*(6): 649–662. https://doi.org/10.1785/gssrl.78.6.649 .

Olson, A.H., J.A. Orcutt, and G.A. Frazier. 1984, may. The discrete wavenumber/finite element method for synthetic seismograms. *Geophysical Journal International 77*(2): 421–460. https://doi.org/10.1111/j.1365-246X.1984.tb01942.x .

Paolucci, R., F. Gatti, M. Infantino, C. Smerzini, A. Güney Özcebe, and M. Stupazzini. 2018. Broadband ground motions from 3D physics-based

numerical simulations using artificial neural networks. *Bulletin of the Seismological Society of America 108*(3A): 1272–1286 .

Paolucci, R., I. Mazzieri, and C. Smerzini. 2015. Anatomy of strong ground motion: Near-source records and three-dimensional physics-based numerical simulations of the Mw 6.0 2012 may 29 po plain earthquake, Italy. *Geophysical Journal International 203*(3): 2001–2020. https://doi.org/10.1093/gji/ggv405 .

Piatanesi, A., A. Cirella, P. Spudich, and M. Cocco. 2007. A global search inversion for earthquake kinematic rupture history: Application to the 2000 western Tottori, Japan earthquake. *Journal of Geophysical Research: Solid Earth 112*(7): 1–14. https://doi.org/10.1029/2006JB004821 .

Razafindrakoto, H.N., B.A. Bradley, and R.W. Graves. 2018. Broadband ground-motion simulation of the 2011 Mw 6.2 Christchurch, New Zealand, earthquake. *Bulletin of the Seismological Society of America 108*(4): 2130–2147. https://doi.org/10.1785/0120170388 .

Shaw, B.E., K.R. Milner, E.H. Field, K. Richards-Dinger, J.J. Gilchrist, J.H. Dieterich, and T.H. Jordan. 2018. A physics-based earthquake simulator replicates seismic hazard statistics across California. *Science Advances 4*(8): 1–9. https://doi.org/10.1126/sciadv.aau0688 .

Smerzini, C. and M. Villani. 2012. Broadband numerical simulations in complex near-field geological configurations: The case of the 2009 Mw 6.3 L'Aquila earthquake. *Bulletin of the Seismological Society of America 102*(6): 2436–2451. https://doi.org/10.1785/0120120002 .

Spudich, P. and L. Xu. 2002. Documentation of software package Compsyn sxv3. 11: programs for earthquake ground motion calculation using complete 1-d green's functions. *International Handbook of Earthquake and Engineering Seismology* (January 2002): 1–74 .

Takewaki, I. 2004. Critical envelope functions for non-stationary random earthquake input. *Computers and Structures 82*(20-21): 1671–1683. https://doi.org/10.1016/j.compstruc.2004.04.004 .

Taylor, I.J., E. Deelman, D. Gannon, and M.S. Shields. 2007. Workflows for e-Science: Scientific Workflows for Grids. *Workflows for e-Science: Scientific Workflows for Grids*: 1–523. https://doi.org/10.1007/978-1-84628-757-2 .

Tinti, E., E. Fukuyama, A. Piatanesi, and M. Cocco. 2005. A kinematic source-time function compatible with earthquake dynamics. *Bulletin of the Seismological Society of America 95*(4): 1211–1223. https://doi.org/10.1785/0120040177 .

Wang, S.C. 2003. *Artificial Neural Network*, pp. 81–100. Boston, MA: Springer US.

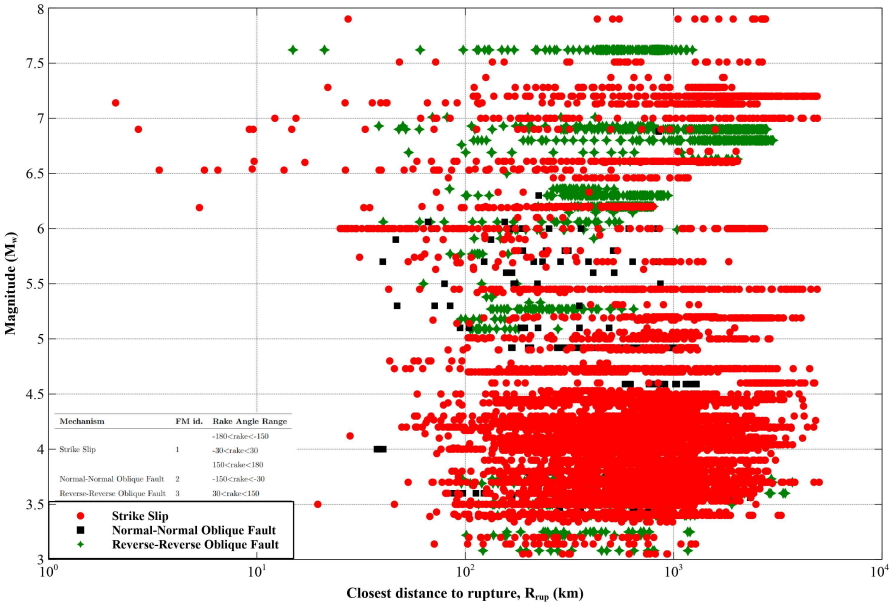

Fig. 1: Distribution of data, with respect to magnitude (Mw), closest distance to fault (R_{rup}) and the earthquake mechanisms, considered for the ANN model development

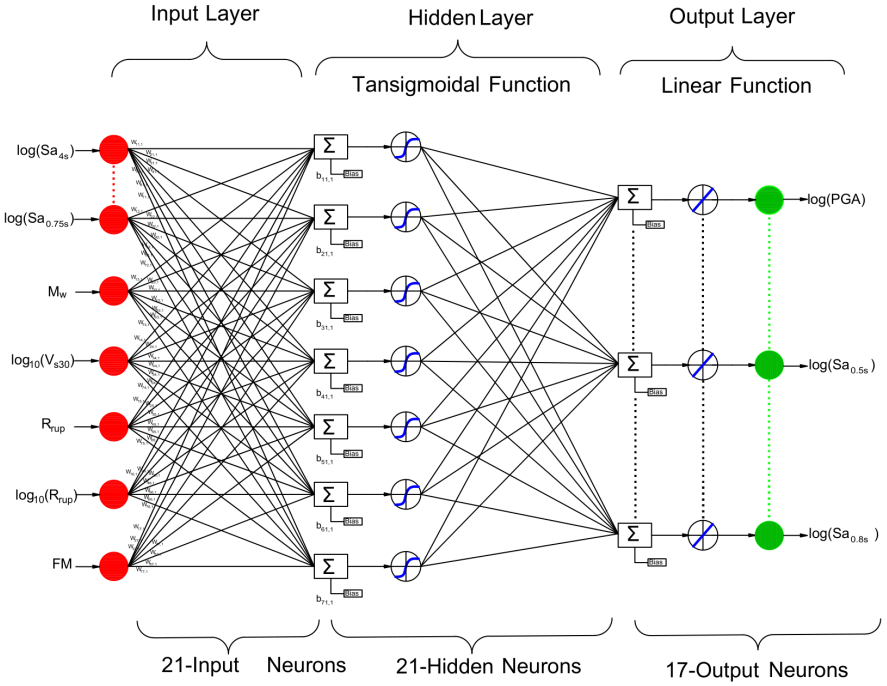

Fig. 2: The architecture of the artificial neural network (ANN) model to predict short period spectra from long period spectra, earthquake source, path and site characteristics. The represented architecture corresponds to that for a cross over period on 0.75s (M_w= moment magnitude, R_{rup} = closest distance to rupture, FM = focal mechanism, V_{S30} = average shear wave velocity of top 30 m, S_a =Spectral Acceleration, T*(cross over period)

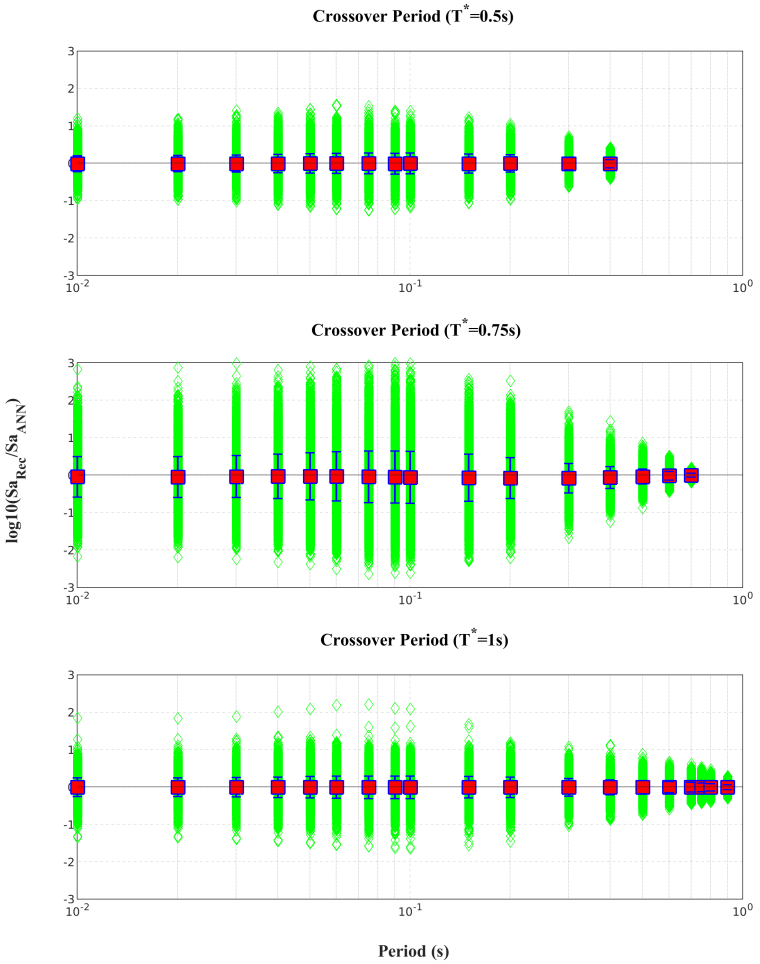

Fig. 3: Residuals along with bias estimation with time period for the ANN models developed from the present study corresponding to different cross over periods (T^*)

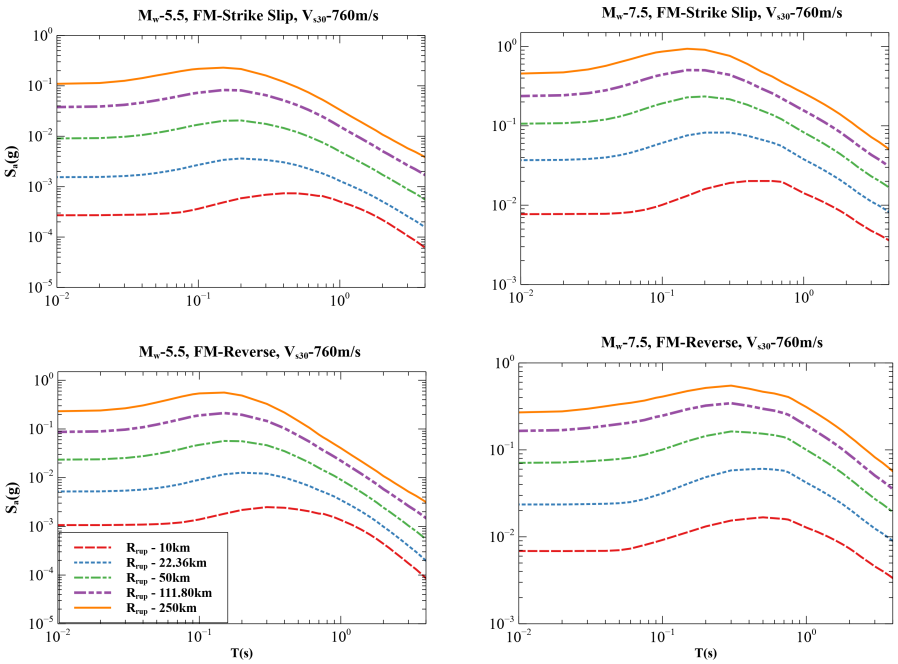

Fig. 4: Variation of Spectral Acceleration (S_a) with respect to Rupture distance (R_{rup}) for different Magnitude (M_w) and Fault Mechanism (FM). Here the model with cross over period T^*=0.75s is considered and low frequency spectral acceleration is taken from the mean value of prediction model developed by Dhanya and Raghukanth (2018)

.

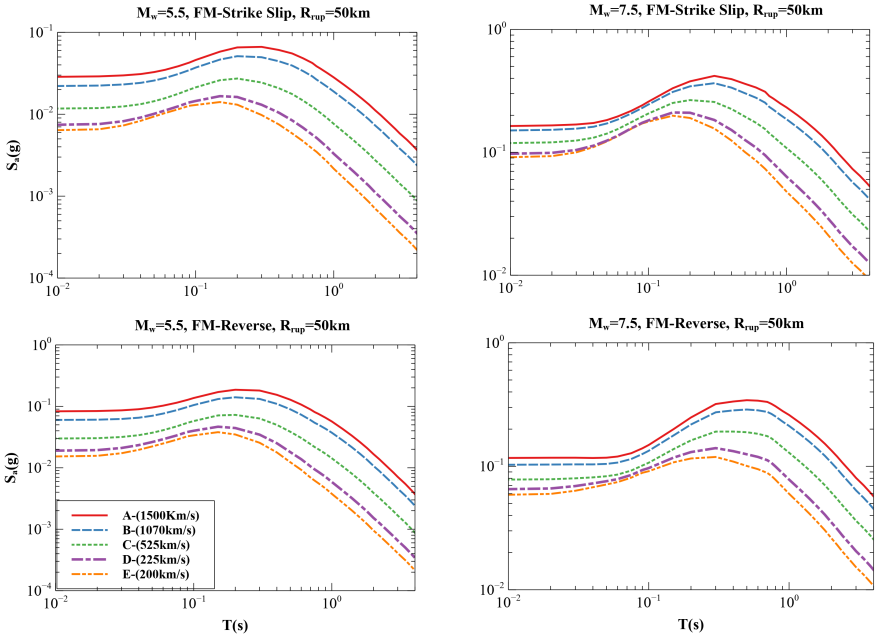

Fig. 5: Variation of Spectral Acceleration (S_a) with respect to Rupture distance (V_{s30}) for different Magnitude (M_w) and Fault Mechanism (FM). Here the model with cross over period T^*=0.75s is considered and low frequency spectral acceleration is taken from the mean value of prediction model developed by Dhanya and Raghukanth (2018)

.

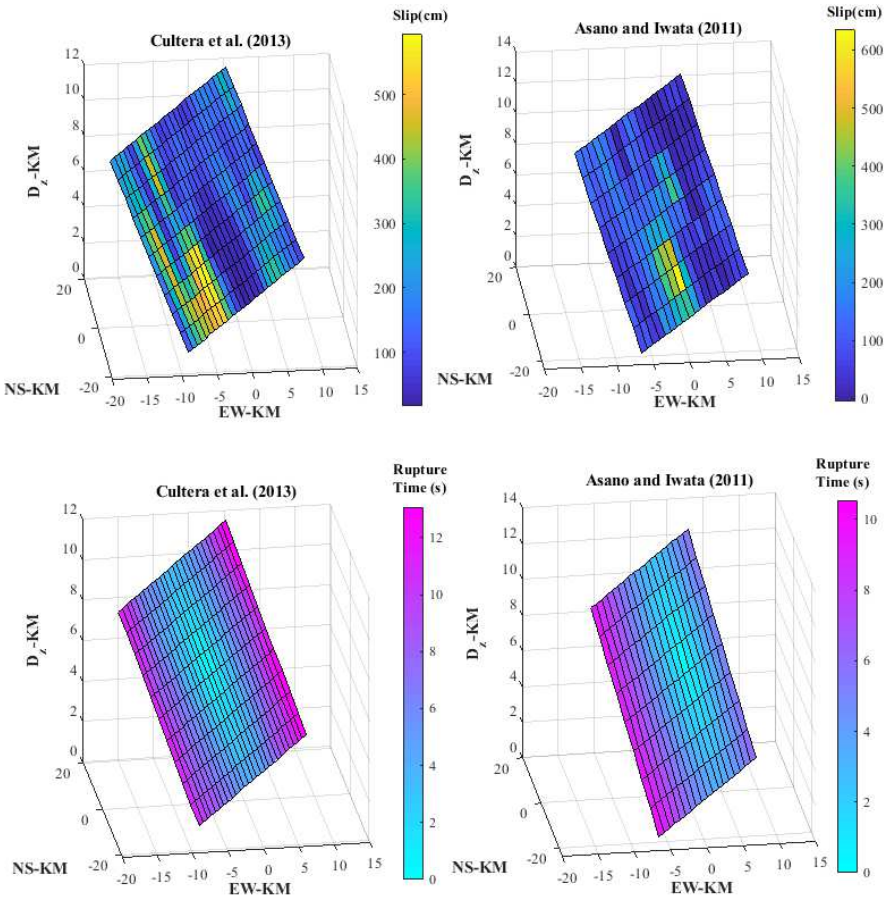

Fig. 6: Slip and rupture time distributions for Iwate Earthquake,2008 ($M_w - 6.9$) (left) Cultrera et al. (2013) model , (right) Asano and Iwata (2011b) model

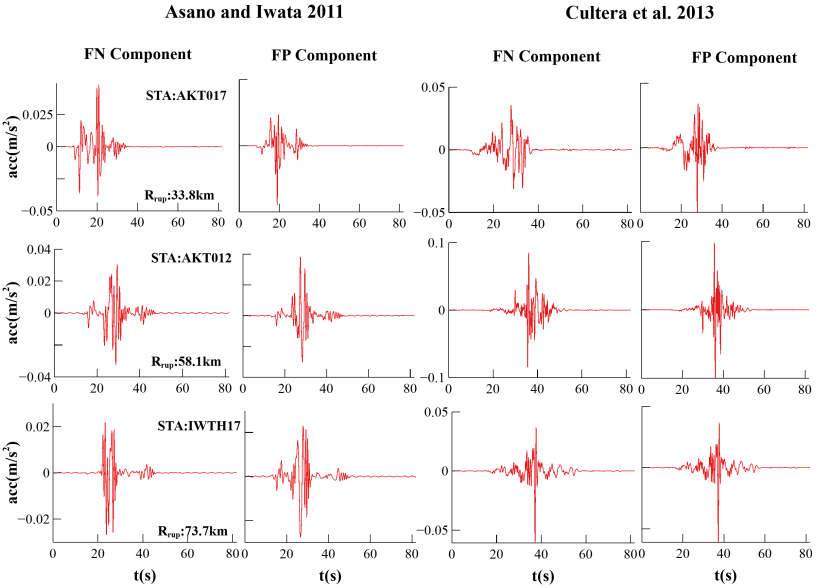

Fig. 7: Physics based input low-frequency ground motion for Iwate Earthquake, 2009 corresponding to different source models (Cultrera et al., 2013; Asano and Iwata, 2011b)

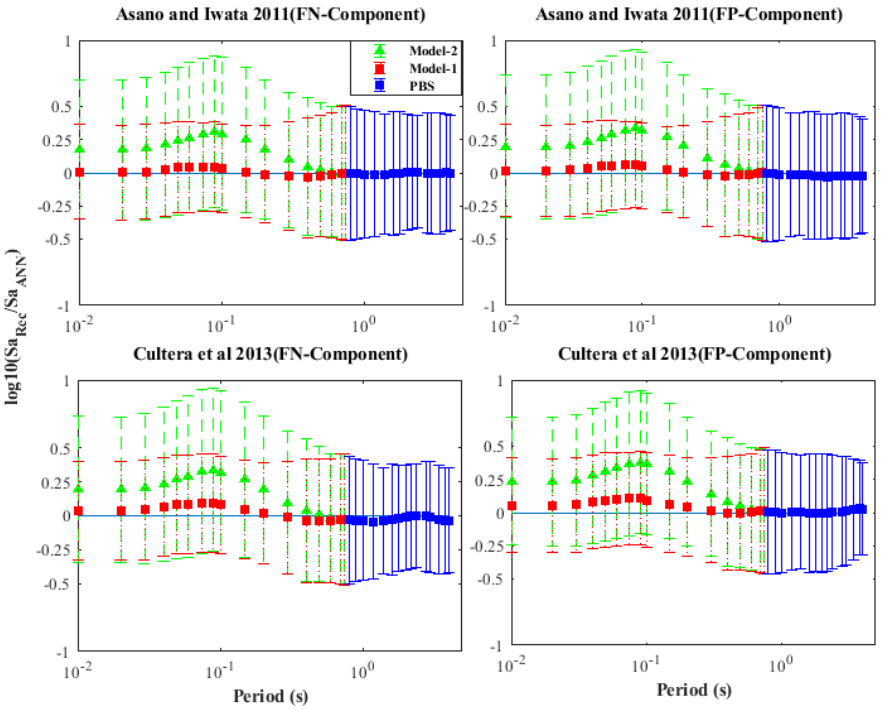

Fig. 8: Bias plots between recorded and simulated spectra for the ANN prediction models corresponding to fault normal and fault parallel components for Iwate Earthquake using the slip fields reported by Asano and Iwata (2011b) and Cultrera et al. (2013) respectively. Here, spectral acceleration corresponds to the PBS is in the higher period part (T^*= 0.75-4s) and from the ANN models with the corresponding LF inputs for high frequency part (T^* <0.75). [Note: input varaible for Model-1 were (M_w, FM, R_{rup}, $log_{10}(R_{rup})$, $log_{10}(V_{s30})$,$log_{10}(Sa \geq T*)$) and Model-2 were ($log_{10}(Sa \geq T*)$)]

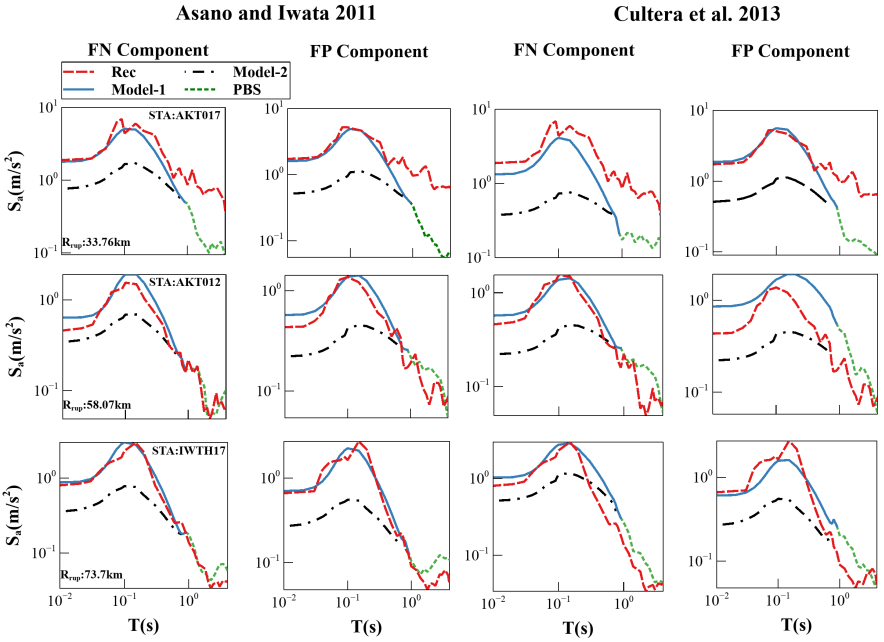

Fig. 9: Comparison plots between recorded, physics based simulation (PBS) and ANN models predicted response spectra at stations AKT017, AKT012 and IWT017 corresponding to Asano and Iwata (2011b) and Cultrera et al. (2013) rupture model

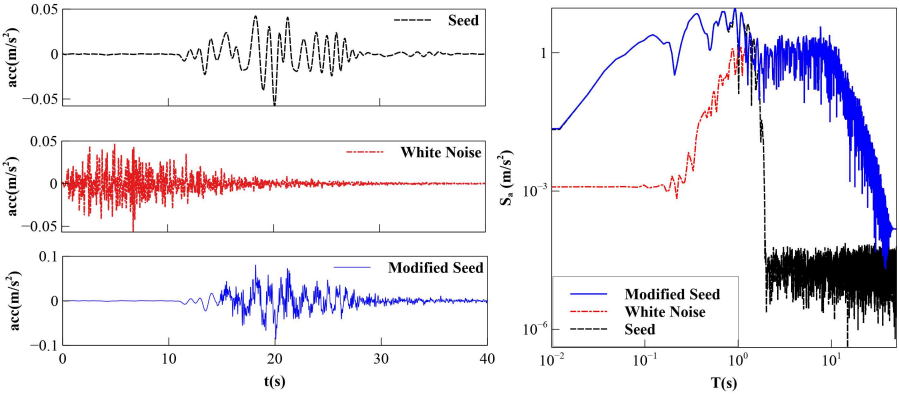

Fig. 10: Procedure for generation of modified seed from seed obtained from Physics Based Simulations (PBS) by addition of filtered white noise

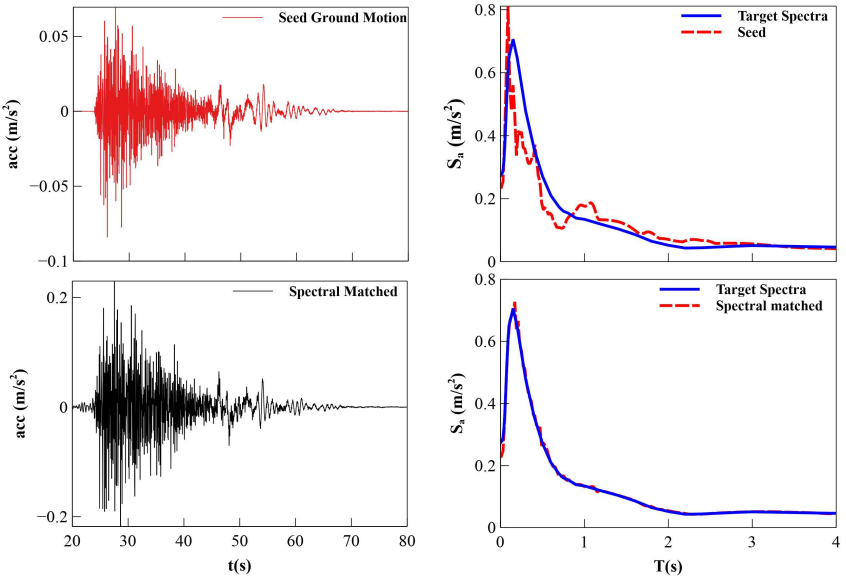

Fig. 11: Spectral ordinate matching representation on the transformation of input seed ground motion signal to output signal that matches with the target spectra

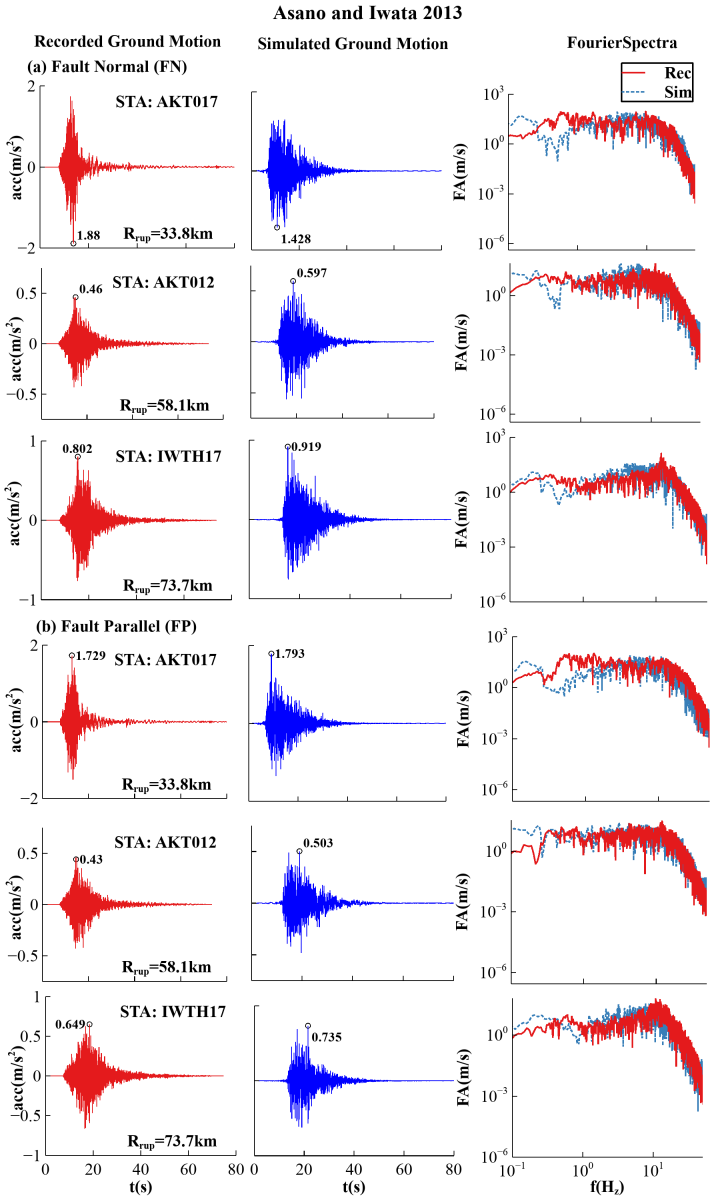

Fig. 12: Time histories and Fourier amplitude plots for the simulated and recorded data corresponding to recording stations AKT017, AKT012 and IWT017 using Asano and Iwata (2011b) source model

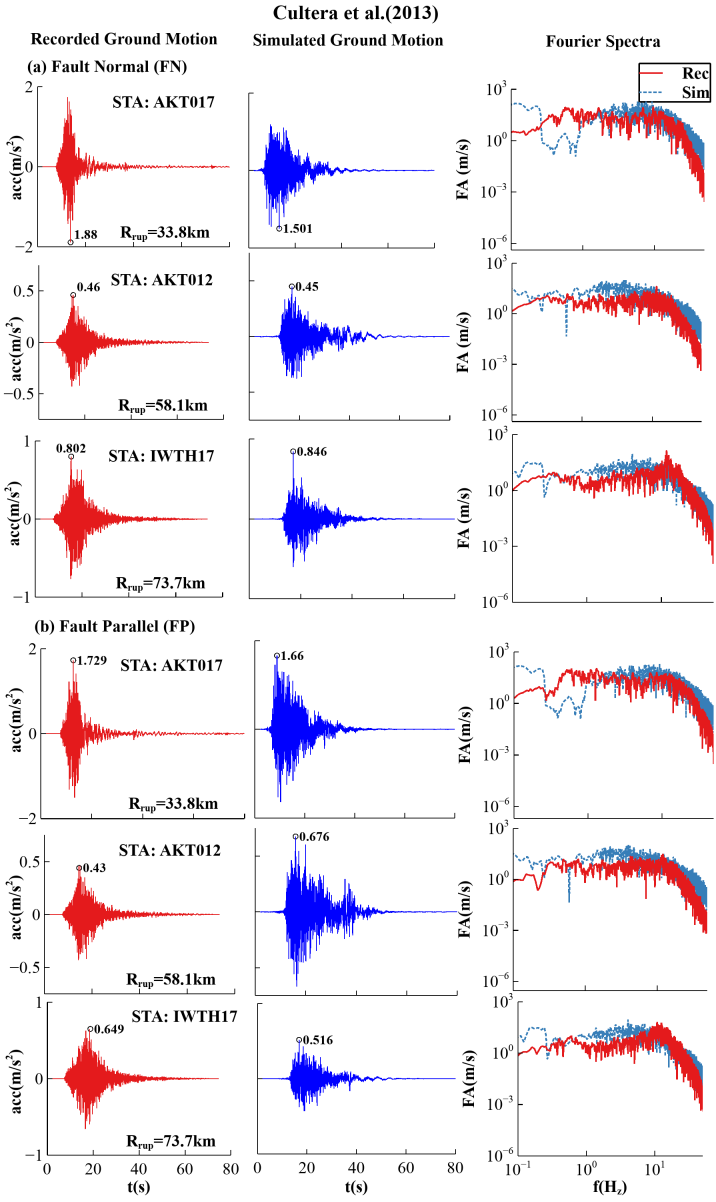

Fig. 13: Time histories and Fourier amplitude plots for the simulated and recorded data corresponding to recording stations AKT017, AKT012 and IWT017 using Cultrera et al. (2013) source model

Table 1: Statistics of the Best Performing ANN models corresponding to various cross-over periods using the with source and site parameters M_w, FM, R_{rup}, $log10(R_{rup})$, $log10(V_{s30})$

No. of inputs	**No. of outputs**	**Crossover Period (T*)**	**Transfer Function**		**Performance Parameters**			
			I-H	**H-O**	R	PP	MSE	σ_ϵ
24	14	0.5	tansig	purelin	0.9756	0.9518	0.0566	0.2379
21	17	0.75	tansig	purelin	0.9766	0.9537	0.0567	0.238
18	20	1	tansig	purelin	0.978	0.9564	0.0561	0.2369

Table 2: Source models detail of Iwate Earthquake,2008 (M_W-6.9) available from SRCMOD database

S.No.	Source Parameter	Asano and Iwata (2011b)	Cultrera et al. (2013)
1	**Fault Dimension**		
	Length (km)	38	42.66
	Width(km)	18	17.38
	Height from Top (m)	0.8	1.34
2	**Hypocentre Location**		
	Longitude (°)	39.030	39.027
	Latitude (°)	140.881	140.878
	Depth(km)	7.77	6.5
3	**Seismic Moment(Nm)**	2.76E+19	3.65E+19
4	**Strike(°)**	209	209
5	**Dip(°)**	51	40
6	**Rake(°)**	108.31	105
7	**Maximum Slip (m)**	5.942	6.36

Table 3: Velocity Profile considered for the physic based simulation of Iwate Earthquake,2008 (M_W-6.9)

Depth (km)	**Vp(Km/s)**	**Vs(km/s)**	**Density (g/cm)**
0	4.8	2.7	1.8
2	5.1	2.9	1.8
4	5.3	3	2.5
14	6.4	3.6	2.8
25	6.8	3.9	2.8
27	7	4	2.8
30	7.2	4.1	2.8

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- electronicSupplement.pdf