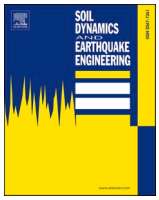
Soil Dynamics and Earthquake Engineering 184 (2024) 108851

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn



Artificial neural network-based ground motion model for next-generation seismic intensity measures

Savvinos Aristeidou, Davit Shahnazaryan, Gerard J. O'Reilly *

Centre for Training and Research on Reduction of Seismic Risk (ROSE Centre), Scuola Universitaria Superiore IUSS Pavia, Palazzo del Broletto, Piazza della Vittoria 15, 27100, Pavia, Italy

ARTICLE INFO

Keywords:
Ground motion model
Intensity measures
Horizontal component definitions
Artificial neural network

ABSTRACT

This paper presents an application of artificial neural networks (ANN) to ground motion modelling. We focused on developing a generalised ground motion model (GGMM) incorporating several seismic intensity measure (IM) types and their inter-IM correlation. These range from classical IMs, such as peak ground acceleration/velocity/displacement, spectral acceleration, and significant duration, to more advanced IMs recently shown to be better descriptors of structural performa nce, such as average spectral acceleration and filtered incremental velocity. Additionally, three different horizontal component definitions were included for the spectral acceleration-based IMs. A total of nine input ground motion causal parameters are required to use the GGMM developed, based on ground motion records from the NGA-West2 database. ANN was used to perform the regression, which differs from the approaches used in many existing ground motion models (GMMs), and gives the possibility to regress all IMs simultaneously in one model. A mixed-effects regression approach was adopted for the regression and the quantification of the inter- and intra-event variability of the GGMM estimation. The correlations between the IMs were also quantified and briefly presented here, which allows for a more refined prediction of seismic shaking and a unified treatment of prediction and IM correlations. This will allow more advanced record selection for non-linear dynamic analyses to be performed, which can consider several facets of ground shaking currently overlooked in many works. We evaluated the performance of the developed GGMM using several metrics and compared it to various existing GMMs developed with either the classical approach or machine learning methods. The results show that the proposed GGMM exhibits very good predictions, especially considering the wide range of IMs tackled. Lastly, this methodology has the flexibility of being able to add more IMs or horizontal component definitions seamlessly.

1. Introduction

Ground motion models (GMMs) are an essential part of seismic hazard analyses and form the basis for seismic risk assessments, shake maps, loss estimation, seismic design of structures, and more, which comprise both fields of earthquake engineering and seismology. GMMs estimate the distribution of expected ground motion intensity and its associated uncertainty, given a set of causal parameters (e.g., magnitude, source-to-site distance, etc.) at a given site location. Many different intensity measures (IMs) can be used to characterise the ground motion shaking intensity at the site of interest. For example, several past studies [1–5] have examined IMs for different structural typologies to identify the benefits and drawbacks of each in various contexts. There is also a growing interest in using cumulative intensity-based IMs (e.g., significant duration) together with peak response amplitude-based IMs (e.g., spectral acceleration), which has sparked the development of a plethora of GMMs to estimate different types of IMs over the years [6]. However, these GMMs (e.g., Ref. [7–10]) predict the IMs independently, with each available GMM being based on a different ground motion database (or at least applying different filtering criteria) and different regression models for the fit. This leads to some degree of heterogeneity, which can be mitigated by developing a generalised ground motion model (GGMM) to estimate different types of IMs collectively [11]. Using independent GMMs to estimate assorted ground motion IMs for the same earthquake scenario can possibly introduce unwanted bias, since they use different regression datasets, which is then propagated into the seismic analysis and risk assessment results. It is important to note that more consistent correlation models can be produced for the set

* Corresponding author.
E-mail address: gerard.oreilly@iusspavia.it (G.J. O'Reilly).

https://doi.org/10.1016/j.soildyn.2024.108851
Received 20 March 2024; Received in revised form 20 June 2024; Accepted 12 July 2024
Available online 18 July 2024

of output IMs since they are estimated using the same GGMM and, therefore, the same filtered ground motion database, which mitigates these aforementioned potential issues.

GMMs can be divided into two types: parametric and non-parametric. Traditionally, parametric models are employed, where fixed functional forms are used to fit a set of coefficients based on empirical data (e.g., Refs. [8,12–15]). These predefined functional forms are derived by observing the ground motion characteristics, like amplification, attenuation and faulting mechanism, leading to increasingly complex functional forms when more effects are considered. Furthermore, these functional forms also differ depending on the IM being estimated, meaning that specific functional forms work better for specific IMs ([16]; section 4.4). Another potential drawback of parametric models is that overly simplistic functional forms may possess limited capability to accurately estimate complex ground motion characteristics. Nonetheless, it should also be stated that parametric GMMs have the advantage of allowing the analyst to maintain some control over the analytical functional forms adopted to ensure that their trends are indeed reflective of the actual physics of the seismological phenomena, as we currently understand them. This is especially helpful where only limited data are available, however with the enriched NGA-West2 dataset researchers can easily opt for machine-learning algorithms.

Meanwhile, researchers have recently been exploring the potential of non-parametric models, where data-driven regression techniques are employed to develop GMMs [11,17]. They have the advantage (which may also be viewed as a disadvantage) of not requiring any predefined analytical equations as input. This stemmed from recent applications of advanced non-parametric models, such as machine learning algorithms, artificial neural networks (ANN), fuzzy logic, etc., in the field of earthquake engineering. There are many alternative machine learning algorithms that were applied in ground motion modelling through the years, such as Support Vector Machine, genetic programming, Gene expression Programming, ensemble decision tree models (e.g., random forest), and sometimes even combination of different machine learning algorithms. For the development of this model, the ANN algorithm was preferred over the aforementioned methods, since the modeller has more control over the model, by tuning the model parameters and structure. The more recent XGBoost algorithm [18] was also tried, but dropped primarily due to the difficulty in adjusting the algorithm to prevent overfitting for this kind of problem, and the difficulty to understand and interpret [19]. The overfitting problem was subtle during the fixed-effects procedure, but it was very amplified in the mixed-effects. Additionally, the eventual ANN model (see "Code availability" section) is much easier and faster to load and use. It is acknowledged, however, that relying purely on data-driven approaches is not a perfect solution, as these models work well only where data are available. In ground motion modelling for engineering use, we are typically interested in strong shaking that can potentially cause damage to engineered structures, which requires data from large-magnitude earthquakes, which are less frequent. Hence, there is a danger that inaccuracies may arise when using these data-driven GMMs to predict intensities that are not well recorded, or even beyond the fitting range. It is here that robust verification is needed, and, in some cases, the physical meaning of the parametric functional forms may be advantageous.

ANNs are generally considered semi-parametric models, because they combine aspects of both parametric and non-parametric models. The parametric aspect is that they have a fixed set of parameters (weights and biases) determined by the network architecture. The non-parametric aspect is that they are highly flexible and capable of approximating complex, non-linear functions. For the case of this study, the ANN is labelled as non-parametric since the parameters were chosen through hyperparameter tuning methods, and they were not fixed. However, it should still be noted that it is still not a non-parametric model in the strictest sense. Derras et al. [20] used ANN to develop a GMM for Europe, in which a local search algorithm named the quasi-Newton back propagation technique was used to calculate the unknown coefficients. In a subsequent study, Derras et al. [21] used the same technique on recordings from the NGA-West2 database to model the variability with respect to site conditions. However, that model has the pitfall of getting trapped into local minima and also does not account for the fault mechanism, which is known to have a notable effect on ground motion modelling.

Therefore, in this study, a novel GGMM was developed using a robust ANN algorithm that can be used to estimate a wide variety of IMs. Nine input parameters were chosen to be included in this model to capture as many ground motion characteristics as possible to predict many IMs from the same model. Among them, the most popular are the moment magnitude, M_w, the rupture distance, R_{rup}, the Joyner-Boore distance, R_{jb}, the site shear wave velocity, $V_{s,30}$, and the style of faulting, *SOF*. The output variables are a mixture of traditional and next-generation IMs, which will be explained in further detail in subsequent sections. This is one of the main contributions of the proposed model: the ability to estimate next-generation IMs more accurately than the scarcely available and simplistic models in the literature. The results and predictions of the proposed GGMM are then presented and compared against the aforementioned recent and well-established GMMs available in the literature. Additionally, the GGMM was trained for three different horizontal component definitions of spectral acceleration, namely, the geometric mean, *RotD50* and *RotD100* definitions [22]. Therefore, the user has the option of choosing their preferred horizontal component definition or estimating the ratios between them for specific seismic hazard conditions.

The following sections describe the ground motion database utilised and the filtering criteria, the predictor and response features used in the ANN model, the model architecture and methodology employed, followed by an evaluation of the model's performance using different performance metrics, visual representation of the model predictions and spectra, while comparing with other relevant GMMs, and a brief description on the imminent correlation models stemmed from this study. The overall structure of this paper is illustrated in Fig. 1.

2. Strong motion database and filtering

To utilise ANN to fit a GGMM, a dataset of ground motion recordings was first required. The NGA-West2 database [23], containing bi-directional ground motion acceleration records with their respective

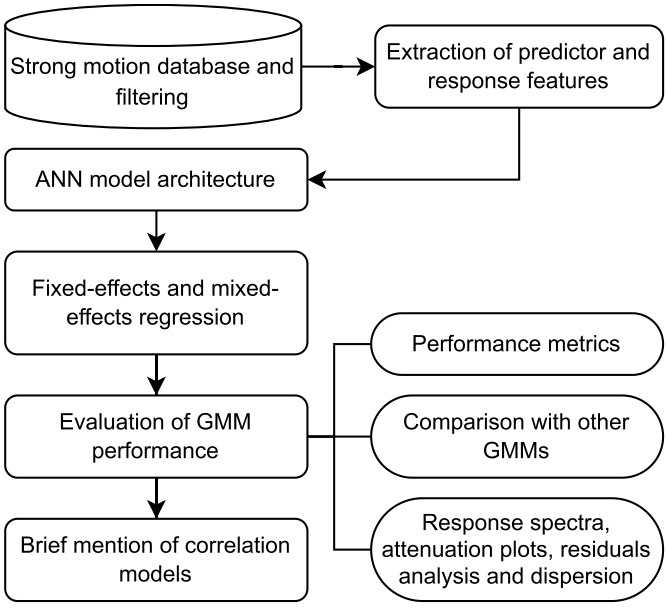


Fig. 1. Flowchart of the paper.

site details, and source information, was adopted. The database was filtered to remove some of the records that may be deemed unsuitable for general use based on the criteria given below, which indirectly form the recommended usage limitations of the GGMM developed herein:

- Only ground motion records from earthquakes with $M_w \geq 4.5$ were utilised. Earthquakes of lower magnitude were omitted as they were assumed to not be strong enough to induce significant non-linear deformations or structural collapse in engineered buildings without a significant amplitude scaling;
- Recordings with R_{rup} greater than 300 km were discarded. It is worth noting that a few researchers have the source-to-site distance limit to be magnitude-dependent [24]. This intends to minimise the potential sampling bias, which can occur at large distances and/or low magnitudes, where ground motions are generally weak, and instruments may only be triggered by stronger-than-average ground motions. However, this bias is considered to be insignificant and therefore, no further actions were taken in this regard;
- Recordings from instruments located on the free field, below the surface, or in the first storey of low-rise structures (fewer than four storeys) were utilised. This was based on the Geomatrix 1st letter code of the NGA-West2 flat-file;
- Events with a hypocentral depth greater than 20 km were discarded;
- Events recorded on sites with higher than 1300 m/s of mean shear wave velocities in the upper 30 m, $V_{s,30}$, were discarded;
- Recordings from all event mechanisms (i.e., strike-slip, normal, reverse, reverse-oblique, and normal-oblique) from active shallow crustal tectonic environments were included;
- Only records whose minimum useable frequency of both components was less than 0.25 Hz were considered;
- Earthquakes with $M_w < 5.5$ and fewer than five recordings were discarded. Earthquakes with $5.5 \leq M_w < 6.5$ and fewer than three recordings were discarded. This was because those earthquakes could be considered to have insufficient number of recordings, and therefore unreliably recorded;
- Recordings were considered only if both horizontal components were available. This was necessary to characterise the different horizontal components of shaking;
- Recordings from aftershocks were excluded since most seismic hazard analyses are performed based on a (Poissonian) recurrence of mainshocks. Therefore, including aftershocks in the regression dataset of the GGMM could introduce an unwanted bias. In this study, a recording is classified as an aftershock if it is defined as a 'Class 2' event with centroid Joyner-Boore distance, $CR_{JB} < 10$ km according to the criteria given in Wooddell and Abrahamson [25], although other classification criteria could have been used.

Based on the above filtering criteria, the final ground motion database included 4135 recordings from 102 earthquakes. The earthquakes were classified into five *SOFs*, including strike-slip (58 earthquakes and 1819 recordings), normal (9 earthquakes and 74 recordings), normal oblique (4 earthquakes and 247 recordings), reverse (21 earthquakes and 1077 recordings) and reverse oblique (10 earthquakes and 918 recordings). Fig. 2 displays the M_w, R_{rup} and $V_{s,30}$ distributions of the filtered database. Additionally, the depth-to-shear wave velocity of 2.5 km/s, $Z_{2.5}$, when missing for some ground motion recordings, was estimated following the prediction equations of Kaklamanos et al. [26].

3. Predictor and response features

Before developing and training the ANN model, it is essential to identify the predictor and response features. The informed selection of predictor features is essential to the robustness and accuracy of the subsequent modelling process. Within the scope of GMMs, past research ([16,27] - Section 4.5 [11]) has highlighted the substantial predictive power of magnitude, *M*, and source-to-site distance, *R*, for most IMs of engineering interest. In addition, several other causal parameters were included to train the ANN models. The full list of predictor features within this study are listed in Table 1, where the associated response features are *PGA*, *PGV*, *PGD* Ds_{595}, Ds_{575}, *Sa(T), FIV3(T),* Sa_{avg2}*(T),* Sa_{avg3}*(T),* which are described in detail below. Users must provide a value for each of the predictor features shown in the first column and can obtain predictions for any of the response features listed in the last column, essentially making it a generalised GMM for the variety of IMs that can be predicted. Also, the minimum and maximum values of each parameter in the filtered database are listed. These values also reflect the recommended usage limits of the model. We note that the GGMM considers multiple depth-related predictor features, which is not necessarily an issue and several models available in the OpenQuake library, for example, follow a similar approach. The inclusion of three different distance metrics may seem peculiar, but it was seen to produce more accurate and reliable predictions, which is in line with the observations of other researchers (e.g., Ref. [11,13]). It is worth noting that users must also be aware that there may be physical limits to the input

Table 1
List of predictor features required when using the GGMM.

Description	Min value	Max value
Moment magnitude, M_w	4.5	7.9
Rupture distance, R_{rup} [km]	0.07	299.59
Hypocentral depth, D_{hyp} [km]	2.3	18.65
Time-averaged shear-wave velocity to 30 m depth, $V_{s,30}$ [m/s]	106.83	1269.78
Style of faulting, *SOF*[a]	0	4
Depth to the 2.5 km/s shear-wave velocity horizon (a.k.a., basin or sediment depth), $Z_{2.5}$ [m]	0	7780
Depth to top of fault rupture, Z_{tor} [km]	0	16.23
Joyner-Boore distance, R_{jb} [km]	0	299.44
Distance measured perpendicular to the fault strike from the surface projection of the up-dip edge of the fault plane, R_x [km]	−297.13	292.39

[a] Encoding for the *SOF* is as follows: 0 for strike-slip, 1 for normal, 2 for reverse, 3 for reverse-oblique, and 4 for normal-oblique.

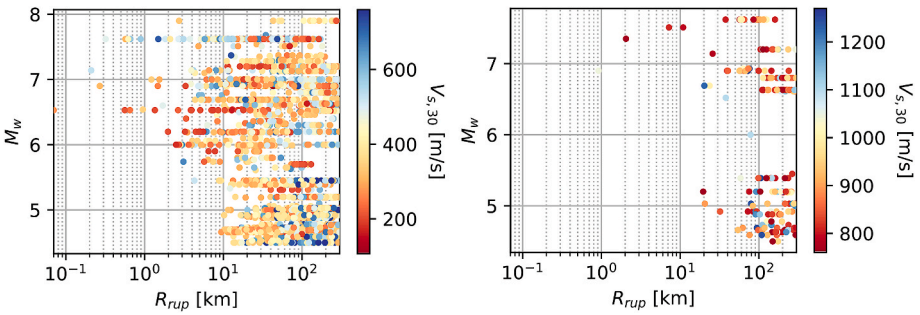


Fig. 2. M_w and R_{rup} distribution of the filtered ground motion database used to fit the GGMM for (left) $V_{s,30} \leq 760$ m/s, and (right) $V_{s,30} > 760$ m/s.

predictor features can take with respect to each other (e.g., $V_{s,30}$ and $Z_{2.5}$); hence, care is required when using such models.

Response features, or IMs, include the peak ground acceleration (PGA), velocity (PGV), displacement (PGD), two definitions of significant duration, *Ds*, given by Equation (1) and (22) definitions of *Sa* at periods ranging from 0.01s to 5.0s, 14 definitions of two different definitions of average spectral acceleration, $Sa_{avg2}(T)$ and $Sa_{avg3}(T)$, at periods ranging from 0.1s to 4.0s given by Equation (3) [28,29], 14 definitions of filtered incremental velocity, *FIV*3, at periods ranging from 0.1s to 4.0s given by Equation (4) [30]. The equations and definitions of some of the output IMs are given in the following.

There are many ways to describe the duration of a strong ground motion [31]; the two most common definitions are bracketed duration and significant duration [8]. The scope here is limited to the significant duration since it is often the preferred definition used in the literature [32]. It is defined as follows:

$$Ds_{xy} = t_y - t_x \tag{1}$$

$$x = \frac{100\%}{I_a}\int_0^{t_x} [a(t)]^2 dt \tag{2}$$

where t_x and t_y are the time stamps on a Husid plot [33] at which x and y per cent of the total Arias intensity, I_a, occurs (as defined in Eq. (16) for x and similarly for y). $a(t)$ stands for the acceleration time history. The most common values of x and y adopted in the literature, included in this study, are $\{x, y\} = \{5\ \%, 75\ \%\}$ and $\{x, y\} = \{5\ \%, 95\ \%\}$, subsequently referred to as Ds_{575} and Ds_{595}, respectively.

Average spectral acceleration, $Sa_{avg}(T)$, has been shown in the literature to be a better overall predictor of structural response, than the classic IM of $Sa(T)$ for the majority of structural typologies [3,4,34]. It can be easily calculated from the geometric mean, which is also the log-average, of a range of $Sa(T)$ values as follows:

$$Sa_{avg}(T) = \left(\prod_{i=1}^{N} Sa(c_i T)\right)^{\frac{1}{N}} \tag{3}$$

where $Sa(c_i T)$ corresponds to the 5%-damped pseudo-acceleration spectral value, c_i is a factor ranging uniformly, $N = 10$ times, from 0.2 to 2.0 and 0.2 to 3.0 for $Sa_{avg2}(T)$ and $Sa_{avg3}(T)$, respectively. Previous research has shown that this spacing scheme is more efficient than a logarithmic one and that the difference between using 10 or 100 periods is negligible, on average [35].

A novel IM, named *FIV*3, proposed by Ref. [30], has shown promising results regarding its efficiency and sufficiency in characterising the collapse performance of buildings. It is briefly summarised mathematically as follows:

$$FIV3 = max\left\{V_{s,max\ 1} + V_{s,max\ 2} + V_{s,max\ 3}, \left|V_{s,min\ 1} + V_{s,min\ 2} + V_{s,min\ 3}\right|\right\} \tag{4}$$

$$V_s(t) = \int_t^{t+\alpha T} \ddot{u}_{gf}(t)dt, \forall t < t_{end} - \alpha T \tag{5}$$

where $V_s(t)$ is a series of incremental velocities (*IV*s) estimated using time segments of αT, $V_{s,max1}$, $V_{s,max2}$, $V_{s,max3}$ are the first, second, and third local largest *IV*s in $V_s(t)$, respectively, and $V_{s,min1}$, $V_{s,min2}$, $V_{s,min3}$ are the first, second, and third local minimum *IV*s in $V_s(t)$, respectively, T is the period of interest, t_{end} is the last instant of time of acceleration time series, and $\ddot{u}_{gf}$ is the filtered acceleration time series using a second-order Butterworth low-pass filter with a cut-off frequency, f_c, equal to βf, where β is a scalar controlling the f_c/f ratio and f is $1/T$. The parameters α of 0.7 and β equal to T are based on Dávalos et al. [14]'s findings.

In addition to the IMs themselves, ground motions are usually recorded in three orthogonal directions in space, so combining these recorded directions into an IM with a specified horizontal component definition is necessary. Several horizontal component definitions have been used in the literature to quantify the intensity of a ground motion on single-degree-of-freedom systems based on the two orthogonal horizontal components, such as arbitrary component, maximum of the two, average, square-root-of-sum-of squares, geometric mean, *GMRotI50*, *RotD50*. Most modern GMMs use the *RotD50* definition [22], as it is considered to be the state-of-the-art horizontal component definition, at least for spectral acceleration IMs, however, also other definitions may be of interest in a seismic risk analysis. In this model, the *RotD50*, *RotD100* and geometric mean were adopted for *Sa* and Sa_{avg}; only the *RotD50* was adopted for *PGA*, *PGV*, and *PGD*; whereas for more advanced IMs (i.e., *FIV3*) and other IMs (i.e., Ds_{575} and Ds_{595}) only the geometric mean definition was adopted.

4. Model architecture

4.1. Fixed-effects with artificial neural network

A *feed-forward* ANN was employed to predict the IMs outlined in Section 3. ANN is a subset of deep learning composed of artificial neurons inter-connecting an input layer, one or more hidden layers, and an output layer [36]. Each neuron performs a simple computation, receiving a signal, applying an activation function, and passing the result through the hidden layers to the output layer, hence the term *feed-forward*. While the neurons in the hidden layers process the information, the neurons in the input layer transmit the input data, and the neurons in the output layer provide the final outputs, or within the scope of this study, the predictions of IMs of interest. Each connection has an associated synaptic weight representing the strength of the connection. Similarly, network neurons are associated with a bias term, which adjusts the point at which the neuron becomes significantly active or inactive based on the total input received, thus influencing the neuron's overall activation behaviour. The synaptic weights of the connections, along with the neurons' biases, represent the neural network's parameters, which are adjusted during the training process to optimise the network's performance. The synaptic weights are used as the multipliers of the outputs of the previous layer, and the bias is a constant added to the outputs before passing through the activation function. The training is typically done through a technique called back-propagation, which uses a gradient descent optimisation [37], where the network tries to minimise the difference between its predictions and the actual target values in the training dataset by adjusting its weights and biases. For a detailed description of neural networks, readers are referred to Haykin

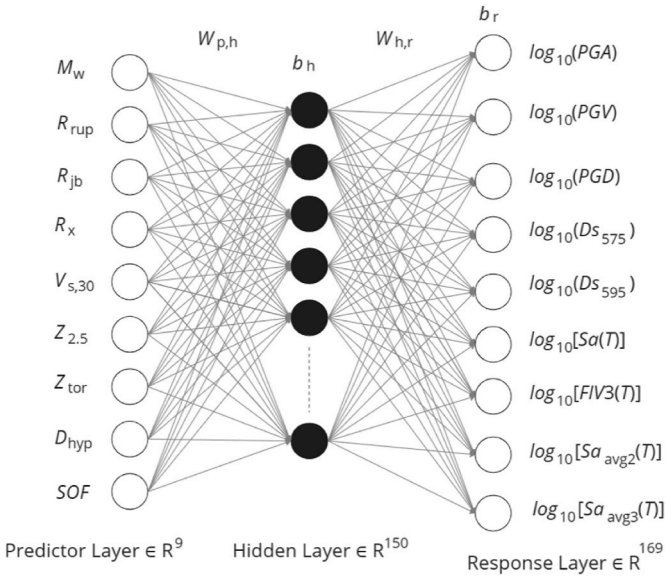


Fig. 3. Architecture of the ANN ground motion model.

[38]. The schematics of the chosen ANN architecture are shown in Fig. 3, and the general expression to predict each IM is as per Equation (6).

$$log_{10}(IM_r) = f_{linear}\left[b_r + \sum_{h=1}^{150} W_{h,r} \bullet f_{tanh}\left(b_h + \sum_{p=1}^{9} W_{p,h} X_p\right)\right] \quad (6)$$

where X_p is the predictor feature p (Table 1), $W_{p,h}$ is the weight of the connection between predictor neuron p and hidden neuron h, b_h is the bias of the hidden neuron h, $W_{h,r}$ is the weight of the connection between hidden neuron h and response neuron r, b_r is the bias of the response neuron r (from 1 to 169), f_{tanh} and f_{linear} are the activation functions of the hidden and response layers, respectively. It should be noted here that the standard deviation term is not shown in equation (6), since the focus here is the fixed-effects regression model itself. The treatment of dispersion is covered in the next subsection.

The ANN training was performed in Python, using the open-source TensorFlow library [39]. The step-by-step implementation, including dataset processing, training of ANNs and generation of predictions, are outlined here. Dataset feature processing and selection of ANN parameters and functions are described as follows:

- Predictor feature engineering: *MinMax* normalisation is adopted to ensure that predictor features are on a similar scale and, hence, have a comparable influence on the model's learning process (e.g., magnitude ranges between $M_w = 4.5$–8, but soil properties can vary between $V_{s,30} = 200$–1200 m/s). This can improve the convergence of the training process and make it less sensitive to the scale of predictor features. Additionally, the initialisation of weights can be more effective, which will facilitate faster convergence and prevent gradients' vanishing or exploding issues. The scaling was done using a range of −3 to 3 instead of 0 and 1, as this range was found to increase the predictive power of this specific model architecture. For what regards *SOF*, one-hot encoding was applied. A lognormal transformation of the predictor features was also tried, but it was seen through manual cross checking with the results of the model that the MinMax normalisation worked better;
- Response (IMs) feature engineering: Similar to predictor features, the response features are scaled to span similar ranges. A log_{10} transformation is applied to the vector of IMs, as it is sometimes adopted in GMMs in the literature, instead of the natural logarithm (e.g., Ref. [17,24,40,41]). The log_{10} transformation limits the response parameter range more than a natural logarithm transformation, and therefore, a more robust fit could be achieved (i.e., better performance metrics);
- Number of hidden layers and neurons: A single hidden layer was employed following a trial-and-error approach, demonstrating that using just one hidden layer was adequate for making predictions. The input layer consisted of 9 neurons, matching the number of predictor features, while the output layer consisted of 169 neurons, corresponding to the number of considered IMs. Concerning the number of neurons in the hidden layer, 150 neurons were chosen, as it produced the model's optimal predictive performance (considering the chosen performance metrics and eventual model dispersion described later). It was observed that using fewer or more neurons led to either underfitting or overfitting, respectively;
- Activation functions: Calculates the output of a neuron. Given the nature of this regression problem, *softmax*, *tanh*, and *linear* activation functions were considered in the input, hidden and output layers, respectively, based on the hyperparameter tuning described later. The *tanh* proved to work better with the range of predictor features, given the *MinMax* normalisation together with passing through the *softmax* activation function. Furthermore, *tanh* in the hidden layer introduces the necessary non-linearity, which enables the network to learn complex patterns from the data. Additionally, *linear* activation function was used for the output layer as it prevents the output values from having an upper or lower limit and can also output negative values to take care of the contradicting effects of some predictor features on different response features;
- Optimisation algorithm and loss function: The loss function employed for optimisation was the mean squared error (MSE) given by Equation (7), and its minimisation was accomplished through the use of the adaptive moment (ADAM) optimisation algorithm [42]. Additionally, the coefficient of determination, R^2, given by Equation (8) was used to determine how well the variation of response features is explained by predictor features in a regression model, where y_i is the ith observed value, $\hat{y}_i$ is the ith predicted value, and $\bar{y}$ is the mean value of n data points;
- Training and testing sets: Before model training, the filtered dataset from Section 2 was randomly split into training and testing sets using an 80:20 ratio.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (8)$$

The next step of the implementation involves training the ANN. To assess the model's performance, a five-fold cross-validation [43] was employed. The training set was randomly partitioned into five equal-sized sets. Five separate ANNs were trained, each using four of the subsets for training and the remaining fifth subset for model prediction validation. The procedure ensures that each subset takes on the role of the validation set for its respective training. The fixed-effect regression metrics for cross-validation were computed as the average of the results from five ANNs. Furthermore, Bayesian optimisation [44] was employed to determine the optimal hyperparameters for the ANN regression model. The objective within the context of this study was to minimise the MSE of the fixed-effects regression by exploring a range of hyperparameters. A summary of the hyperparameters considered is provided below:

- Batch size from 8 to 128: helps balance computational efficiency and model performance. With smaller batch sizes, better model generalisation can be achieved. However, it can be computationally insufficient, as more updates are needed to process the entire dataset. In contrast, larger batch sizes accelerate the training but can hinder model generalisation. Hence, the model is more prone to overfitting;
- Training epochs from 50 to 200: during each epoch, the model passes through all training samples and updates its parameters (weights and biases) based on the loss incurred when making predictions. The updates try to minimise the error and improve the model's performance. While the optimal number of training epochs can improve the model's ability to generalise, with the increasing number of epochs, overfitting may incur. Therefore, early stopping was implemented as a preventive measure against overfitting, which automatically halts training if the model stops improving for 20 consecutive epochs;
- Optimisation algorithm: The following optimisation algorithms were considered: ADAM; root mean square propagation (RMSprop); stochastic gradient descent (SGD); adaptive gradient descent (Adagrad); adaptive learning rate (Adadelta); a variation of ADAM (Adamax); a combination of Nesterov accelerated gradient and Adam (Nadam); follow the regularised leader (Ftrl);
- The learning rate of the optimisation algorithm from $0.5x10^{-3}$ to 0.05: controls the step size during weight updates and influences the convergence speed and stability of the model;
- The activation function of hidden layer: The following activation functions were considered: linear; rectified linear unit (ReLU); leaky ReLU; exponential linear unit (ELU); scaled ELU; *softmax*; hyperbolic tangent (*tanh*).

The approach was utilised to comprehensively evaluate the model's

performance while mitigating the potential risks associated with overfitting (high variance) and underfitting (high bias). The hyperparameters that yielded the best model performance are as follows: *tanh* and *linear* activation functions for the hidden and response layers, respectively; learning rate of $1.04\text{x}10^{-3}$; a batch size of 32; and 100 training epochs.

4.2. Mixed-effects regression

The functional form of the GGMM is given as:

$$\log_{10}IM_i = f_i(\boldsymbol{X},\boldsymbol{\theta}) + \delta b_i \tau_i + \delta w_i \phi_i \quad (9)$$

where $\log_{10}IM_i$ is the logarithm with base 10 of the ith IM; $f_i(\boldsymbol{X},\boldsymbol{\theta}) = \mu_{\log_{10}IM_i|\boldsymbol{X},\theta}$ is the predicted mean output from the ANN model, taking as input a set of causal features (e.g., M_w, R_{rup}, etc.), denoted as $\boldsymbol{X}$; $\boldsymbol{\theta}$ are the 'calibrated coefficients' of the ANN model (i.e., synaptic weights and biases); δb_i and δw_i are the normalised inter- and intra-event (or between- and within-event) residuals of IM_i, respectively; τ_i and ϕ_i are the inter- and intra-event logarithmic standard deviations. Note that herein, the normalised residuals are denoted with lowercase letters (i.e., δb_i and δw_i) and the residuals before normalisation with uppercase letters (i.e., δB_i and δW_i). The main metric to evaluate the model's performance is the total standard deviation, σ. Most recent GMMs have the inter- and intra-event standard deviation models, or just the inter-event standard deviation models, to be magnitude-dependent [45,46]. However, for the sake of simplicity and to not over-complicate this study, the dispersion model was assumed to be magnitude-independent since also a few other studies noticed only minor dependencies on M_w and only for $M_w < 5.5$ [13,47]. To calculate σ, one must first segregate the total residuals between inter- and intra-event residuals, which can be treated as normal variables that ideally should follow a normal distribution with zero mean and standard deviations τ and ϕ, respectively [48]. If the inter- and intra-event residuals are assumed to be mutually independent, then the total standard deviation can be calculated as the square root sum of their variances, given in Equation (10).

$$\sigma = \sqrt{\tau^2 + \phi^2} \quad (10)$$

Taking advantage of this assumption and a better understanding of these two different sources of uncertainty, Abrahamson and Youngs [49] proposed a one-step mixed-effect regression algorithm, using the maximum likelihood approach, to compute the variances τ^2 and ϕ^2. This algorithm is an iterative procedure in which mixed-effects, variances, and model parameters are computed successively. This procedure is now widely applied for the development of GMMs and is hence adopted for the ANN model development here. The adopted algorithm is based on the procedure proposed by Abrahamson and Youngs [49], which is similar to the one used in Derras et al. [20], and can be summarised as follows:

1. Estimate the initial set of ANN model parameters (i.e. [*W*] and {*b*} in Equation (6)), using a fixed-effect training procedure.
2. Estimate τ^2 and ϕ^2 from [*W*] and {*b*}, by maximising the log-likelihood function as given in Abrahamson and Youngs [49], and specifically their Equation (7).
3. Given [*W*], {*b*}, τ^2 and ϕ^2, estimate the random inter-event residuals, δB, as given in Abrahamson and Youngs [49], and specifically their Equation (10).
4. Estimate the new [*W*] and {*b*} using a fixed-effects training procedure for ($\log_{10}IM$ - δB).
5. Repeat steps 2, 3, and 4 until the termination criterion is satisfied. The adopted termination criterion was 0.15 % in terms of the difference between two successive likelihood values.

5. Model performance

5.1. Performance metrics

The performance of the ANN model can be evaluated by comparing the empirical (i.e., recorded) values of IMs with their corresponding model estimations using various metrics. In this study, two of the most common ones (i.e., MSE and R^2) are reported, as described in Equations (7) and (8), while acknowledging that there exist several other metrics in the literature (e.g., ANOVA, LLH, and Akaike information criterion) to test and evaluate a GMM's performance. The resulting average MSE of the (fixed-effects) model, determined through five-fold cross-validation with the optimal parameters, obtained as described in Section 4.1, was found to be 0.080 for the training set and 0.079 for the validation set, showing an overall very low value, with the validation set giving a slightly higher value, as anticipated. Finally, the model corresponding to the optimal parameters, after passing through the mixed-effects regression, was evaluated using the 20 % unseen testing set, and the regression metrics MSE and R^2 associated with each IM are reported in Fig. 4. The eventual average MSE of all IMs were 0.078 for the training set and 0.079 for the test set, again exhibiting high accuracy. It is noteworthy that the average training and validation set MSE after cross-validation with the optimal parameters have very close values to the average training and validation set MSE after passing through the mixed-effects regression, respectively, further validating the model's accuracy. From Fig. 4, it can be seen that the R^2 of the testing set is at times slightly higher or equal to that of the training set, which can be expected but the overall comparison is nonetheless encouraging. At the same time, both values are not excessively low, indicating that while the model has high predictive power, it also avoids overfitting. In the case of MSE, the testing set presents both lower and higher errors than the training set, depending on the IM. For IMs with relatively high MSE, it is the testing set that is higher than the training set. The inverse happens in the case of IMs with low MSE, but to a lesser extent. Reasons for these latter case may include situations where the training data were harder to predict, or had inherently more dispersion that the testing data. Generally, the MSE values are considered to be low, which is another indicator of the model's predictability, and allows the ANN model's performance to be deemed sufficient for general application.

5.2. Comparison with other GMMs

5.2.1. Traditional IMs

To evaluate the quality of GGMM's estimations visually with respect to the available data and other comparable GMMs available in the literature, a few graphical representations were plotted and are described below. To do this, different combinations of input causal parameters were used to illustrate relative trends with respect to other causal parameters. Strike-slip rupture style was assumed, and the average of the observed data was taken for the rest of the features needed for this study and the compared models, except for R_{jb} whose value was indirectly calculated from R_{rup}. Fig. 5 shows the magnitude amplification of *Sa*(0.01s), *Sa*(0.1s), *Sa*(0.5s), and *Sa*(1.0s) for two different rupture distance bins and are compared with the GMMs of Campbell and Bozorgnia [13] (CB14), Dhanya and Raghukanth [17] (DR18), and Fayaz et al. [11] (FXZ21). The horizontal component definitions for *Sa* of these models are *RotD50* for CB14 and FXZ21 and unspecified for DR18. The *RotD50* definition of the proposed model is used for the comparisons. These models are well-established models based on either a classical approach to GMM fitting (i.e., CB14) or machine learning-based approaches (i.e., DR18 and FXZ21). These GMMs were selected to provide a relative comparison but of course, many more models could have been chosen here. The comparison was limited to three to avoid overcrowding in the plots that would inhibit the visualisation.

It can be seen that the estimated values of the proposed GGMM are

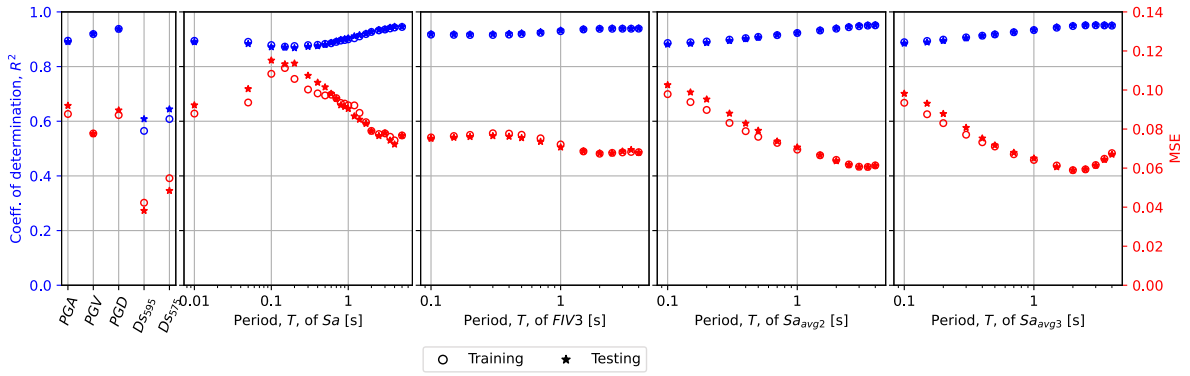


Fig. 4. Training and testing MSE and R^2 values of the ANN model after the mixed effects regression.

generally close to the cloud mean, with a minor deviation in short distances (i.e., 0 km $\leq R_{rup} \leq$ 50 km) in high magnitudes, which is mainly because of the data sparsity. Also, minor deviation from the empirical data is observed in longer distances (i.e., 50 km $\leq R_{rup} \leq$ 100 km) for moderate magnitudes (i.e., $6 \leq M_w \leq 7$), again due to data sparsity. The differences between the compared models and the proposed model are more pronounced in the *Sa* values of longer periods. Generally, all of the compared models give values close to each other and to the proposed model, while also keeping the same trends, giving confidence to the proposed model. For the same IM, the distance attenuation for two different magnitude bins is shown in Fig. 6, along with the comparison of the three previously mentioned models. The FXZ21 model is limited to $R_{rup} \leq 100$ km (and $R_x \geq 0$ km) and is hence not plotted beyond that limit. The estimations of this study capture well the trends of the cloud mean, with the CB14 model also being very close. The other two models are somewhat deviating from the observed mean in some sections of the plot.

It is noted that the mean hypocentral depth of the observed data was used here, which for the two bins of Fig. 5 were 9.80 and 9.83 km, respectively, and for the two bins of the Fig. 6 were 10.10 and 8.13 km, respectively.

In Fig. 7, the distance attenuation of Ds_{595} is illustrated. While amplitude-based quantities decrease with source-to-site distance, the significant duration of ground motion increases due to different waves (e.g., P-, S-, and surface waves) travelling at different velocities and the effect of scattering of those. The estimations of the proposed model are compared with Afshari and Stewart [8] (AS16) and FXZ21 models. For Ds_{595}, both of these models and the proposed model estimate the geometric mean from the two as-recorded horizontal components. While the trends between the proposed model and AS16 are similar, the AS16 model predicts somewhat lower values of significant duration than the proposed model. This difference is thought to be because of the different database filtering criteria of the AS16 model to exclude recordings with unreasonably large durations polluted by high-frequency noise, even though the same database with the proposed model was used (i.e., NGA-West2). The FXZ21 model on the other hand tends to estimate slightly higher values of duration compared to the proposed model.

Overall, the GGMM developed here seems to have good predictability against the observed data when compared to the other models currently available for these traditional IMs of spectral acceleration and significant duration. Similar comparisons were done for the other traditional IMs listed in Table 1 (i.e., *PGA*, *PGD*, *PGV*, Ds_{575} and spectral acceleration at other periods), but were not included here due to space limitations.

5.2.2. Next-generation IMs

The previous section focused on the comparison of traditional IMs, for which several models already exist in the literature. It showed the GGMM proposed here to be of similar quality and bolsters confidence in its general use. In this section, we examine some next-generation IMs that have emerged in recent years that tend not to have a plethora of GMMs available. Hence, while these IMs have been shown in several studies to be quite efficient and sufficient when performing seismic vulnerability and risk analysis, the lack of robust GMMs to predict them is extremely problematic as it means the hazard component of risk is lacking. This study directly addresses this need.

Available GMMs estimating more advanced and complex IMs tend to be more simplistic and can sometimes have limited ranges of application. However, with the proposed model the trends of those next-generation IMs are captured more accurately and for a wide range of ground motion causal parameters. For instance, Figs. 8 and 9 present the distance attenuation and magnitude amplification of *FIV*3(1.0*s*) for two different magnitude bins and it is compared with the recently developed model of Dávalos et al. [14] (DHM20), which to date is the only other GMM available for this IM. The DHM20 model estimates the arbitrary horizontal component definition, whereas the proposed model estimates the geometric mean definition. The DHM20 model is limited to $V_{s,30}$ values between 180 m/s and 360 m/s and takes just two input parameters (i.e., rupture distance and magnitude). Here, the strike-slip mechanism was examined and $V_{s,30}$ was between 180 and 360 m/s for the sake of equal comparison. For the other input parameters, the average was taken, except R_{jb} whose value was calculated from R_{rup}. It can be seen that the proposed model does well in capturing the trends of the cloud mean, while the DHM20 model generally predicts higher values at near and far distances from the source, especially in lower magnitudes.

Another IM that can be classified as next generation is the average spectral acceleration, whose attenuation with distance is illustrated in Fig. 10 for the Sa_{avg3} definition of the IM described in Equation (3) and at a period of 1s. The estimations of the proposed model are compared directly with the model of Dávalos and Miranda [15] (DM21), and with the 'indirect method' to compute the Sa_{avg} values. Regarding the latter method, Kohrangi et al. [50] report the equations that can be used to calculate the mean and standard deviation of Sa_{avg}, denoted $\mu_{\ln Sa_{avg}|rup}$ and $\sigma_{\ln Sa_{avg}|rup}$, respectively, and they are formulated as follows:

$$\mu_{\ln Sa_{avg}|rup} = \left(\frac{1}{N}\right) \bullet \sum_{i=1}^{N} \mu_{\ln Sa(T_i)|rup} \tag{11}$$

$$\sigma_{\ln Sa_{avg}|rup} = \sqrt{\left(\frac{1}{N}\right)^2 \bullet \sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{\ln Sa(T_i),\ln Sa(T_j)} \bullet \sigma_{\ln Sa(T_i)|rup} \bullet \sigma_{\ln Sa(T_j)|rup}} \tag{12}$$

where N refers to the number of $Sa(T)$ values being averaged. $\mu_{\ln Sa(T_i)|rup}$ and $\sigma_{\ln Sa(T_i)|rup}$ are the logarithmic mean and standard deviation of Sa at the ith period in the selected range for a given rupture scenario as obtained from a standard GMM. $\rho_{\ln Sa(T_i),\ln Sa(T_j)}$ is the correlation

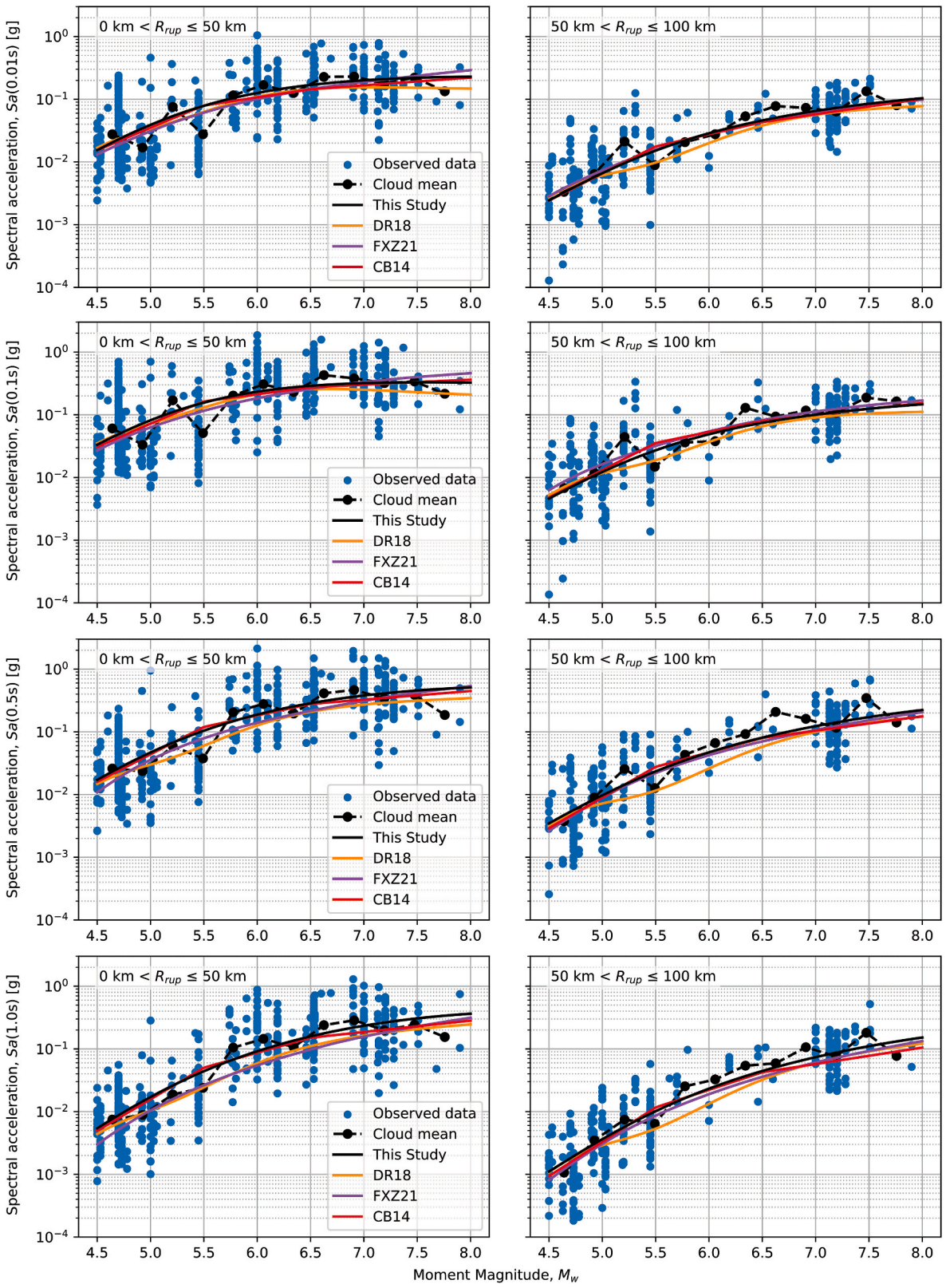


Fig. 5. Magnitude amplification plots of *Sa*(0.01s), *Sa*(0.1s), *Sa*(0.5s), and *Sa*(1.0s) for two different rupture distance bins.

coefficient between ln $Sa(T_i)$ and ln $Sa(T_j)$. Herein, the CB14 model is used as the GMM and the Baker and Jayaram [51] model is used as the correlation model to compute the mean and variance of logarithmic Sa_{avg} values with the 'indirect method' and compare with the output of the proposed model.

It should be noted that the DM21 model estimates the arbitrary horizontal component definition of Sa_{avg}, whereas the *RotD50* definition is chosen for the proposed model. Additionally, since the DM21 model is limited to $V_{s,30}$ values between 180 and 360 m/s, only recordings from sites abiding to those limits were used for the comparisons. It can be

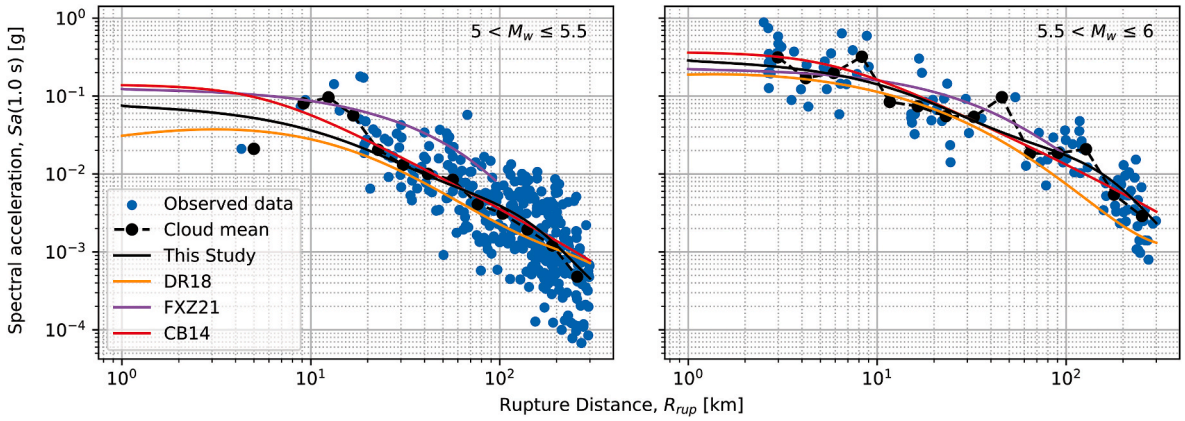


Fig. 6. Distance attenuation plots of *Sa*(1.0s) for two different magnitude bins.

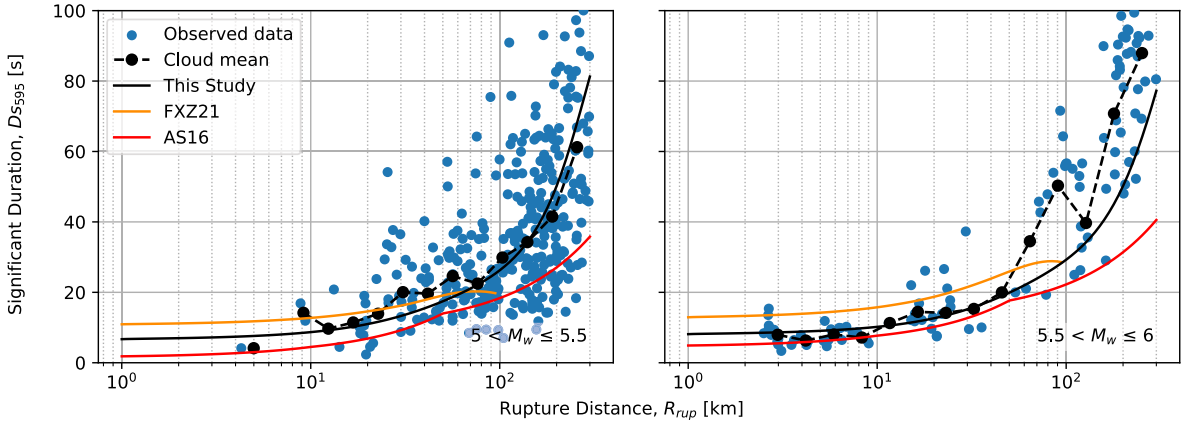


Fig. 7. Distance attenuation plots of Ds_{595} for two different magnitude bins.

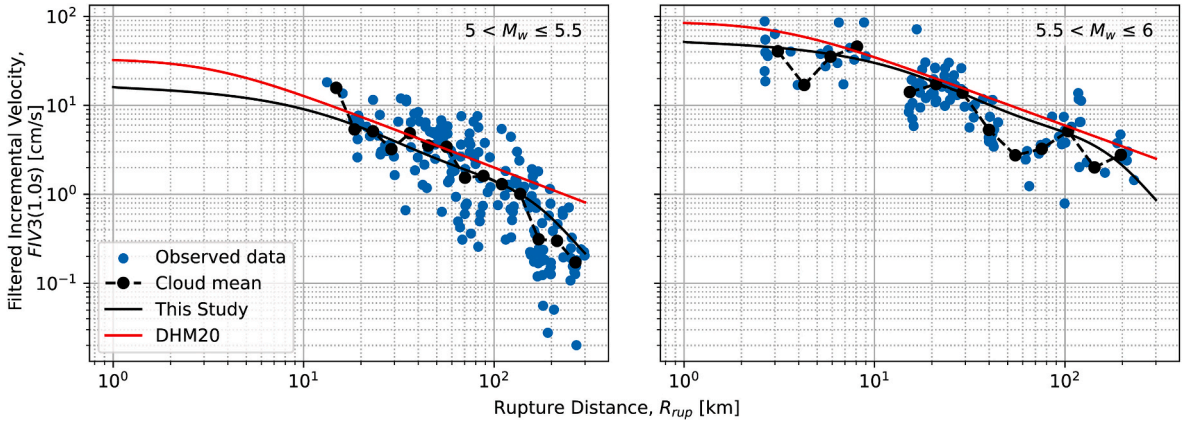


Fig. 8. Distance attenuation plot of *FIV*3(1.0s) for two different magnitude bins.

observed from Fig. 10 that the indirect method does well in capturing the trends of the observed data, since both the GMM and the correlation model come from similar databases (NGA-West2 and NGA, respectively) and have similar filtering criteria. Meanwhile, the direct estimations of DM21 tend to deviate from the binned cloud mean of the data for short and long distances. However, it is important to mention that DM21 suggest that the model should be used between 0 km $\leq R_{jb} \leq$ 150 km, which could explain some of these minor discrepancies. It is worth noting that the indirect method may also be used with the proposed GGMM, where the individual values of $\mu_{\ln Sa(T_i)|rup}$ and $\sigma_{\ln Sa(T_i)|rup}$ are estimated and the correlations are estimated with the ANN-based correlation model presented in Section 6 and in Ref. [52].

Fig. 11 presents the estimations of the proposed GGMM of Sa_{avg3} as a function of magnitude for two different rupture distance bins, along with the empirical data and the comparison with DM21 and the indirect method. It can be observed that the estimations of the proposed model are very close to the ones of the indirect method. The model of DM21 is also close to the proposed model for low, average, and high magnitudes, but slightly deviates in-between. Figs. 12 and 13 present the same information as Figs. 10 and 11, but for Sa_{avg2}. The DM21 is not included in the comparison, because of the different period range used in their calculation of Sa_{avg}, hence only the indirect method is included. It can be

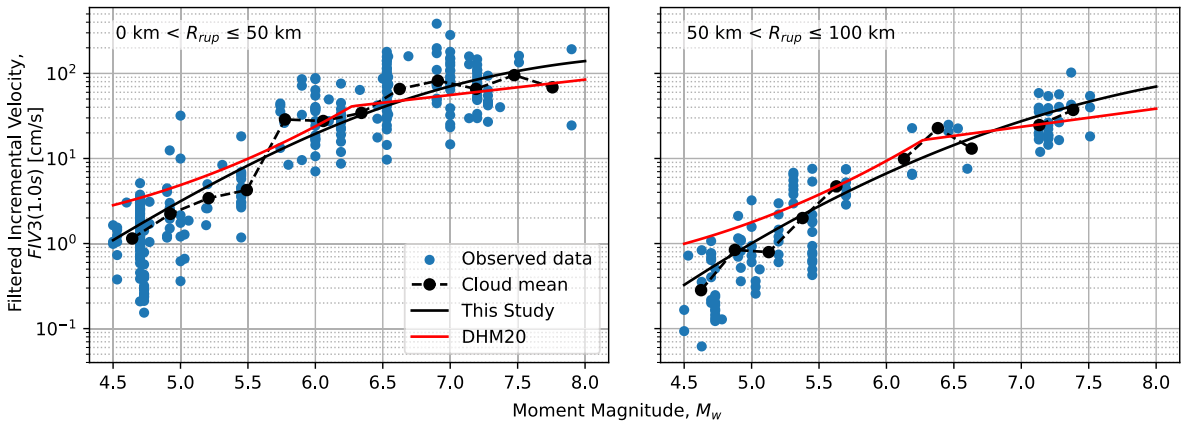


Fig. 9. Magnitude amplification plots of *FIV3*(1.0s) for two different rupture distance bins.

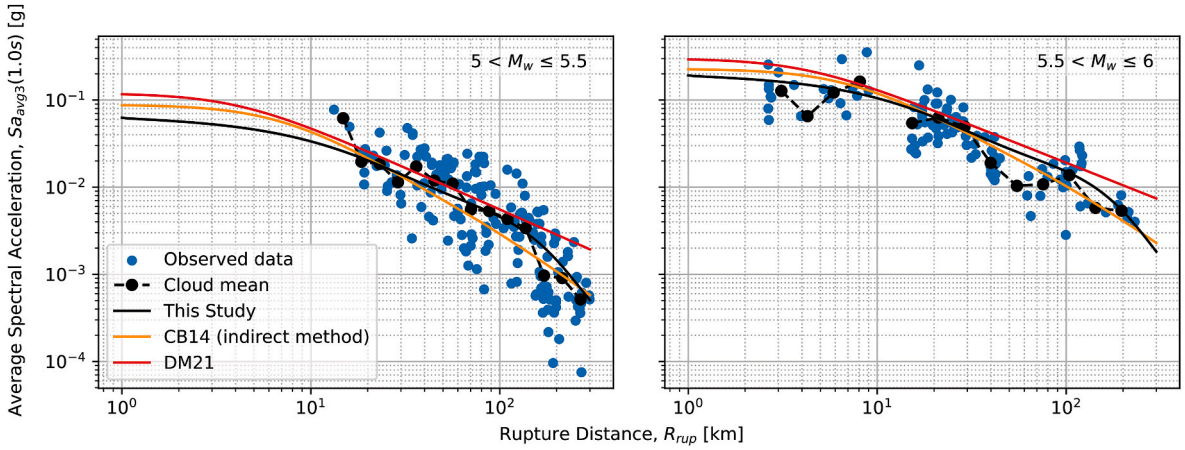


Fig. 10. Distance attenuation plot of Sa_{avg3}(1.0*s*) for two different magnitude bins.

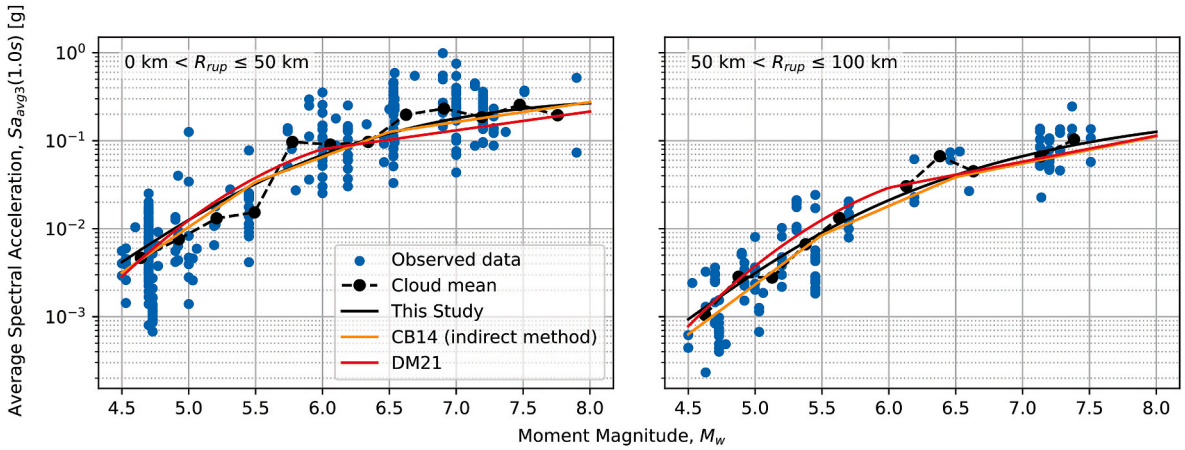


Fig. 11. Magnitude amplification plots of Sa_{avg3}(1.0*s*) for two different rupture distance bins.

seen that the proposed model gives very similar estimations with the calculated values with the indirect method, and both do well in capturing the cloud mean. The only minor exception is the places where data are scarce.

5.3. Response spectra

While the previous sections evaluated the comparisons of the GGMM to available data and other available GMMs, this section provides some brief illustrations of the response spectra that may be obtained for these next-generation IMs investigated and how they are impacted by the main causal parameters. The response spectra, as estimated by the proposed model for *FIV3* and Sa_{avg3}, are presented in Fig. 14 for different ground motion scenarios. On the left panels of the plots, the distance is kept constant while the magnitude is varied, and on the right panels, the magnitude is kept constant while the distance is varied. It can be seen that the *FIV*3 is generally monotonically increasing with period and plateaus after a period of around 0.7s, suggesting that for medium-to

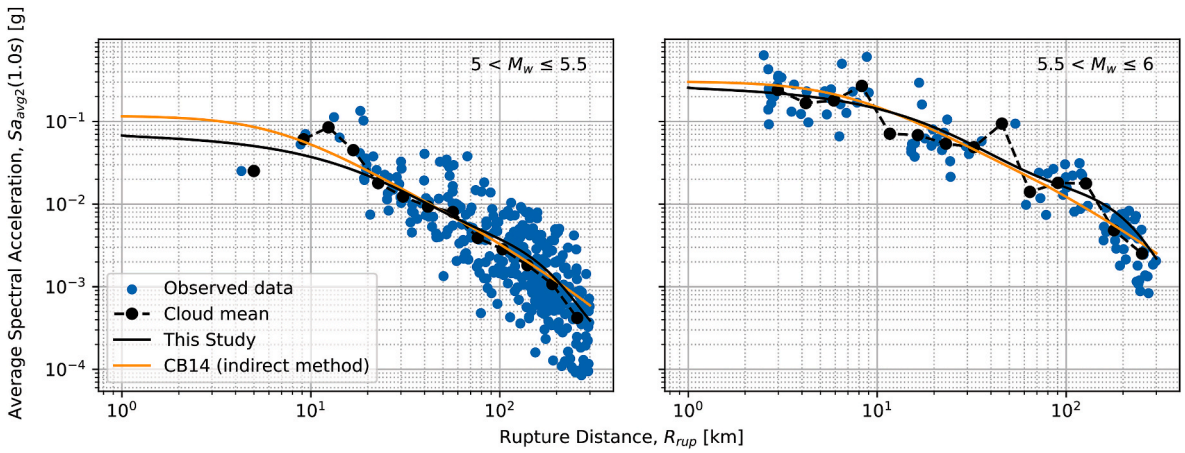


Fig. 12. Distance attenuation plot of $Sa_{\mathrm{avg2}}(1.0s)$ for two different magnitude bins.

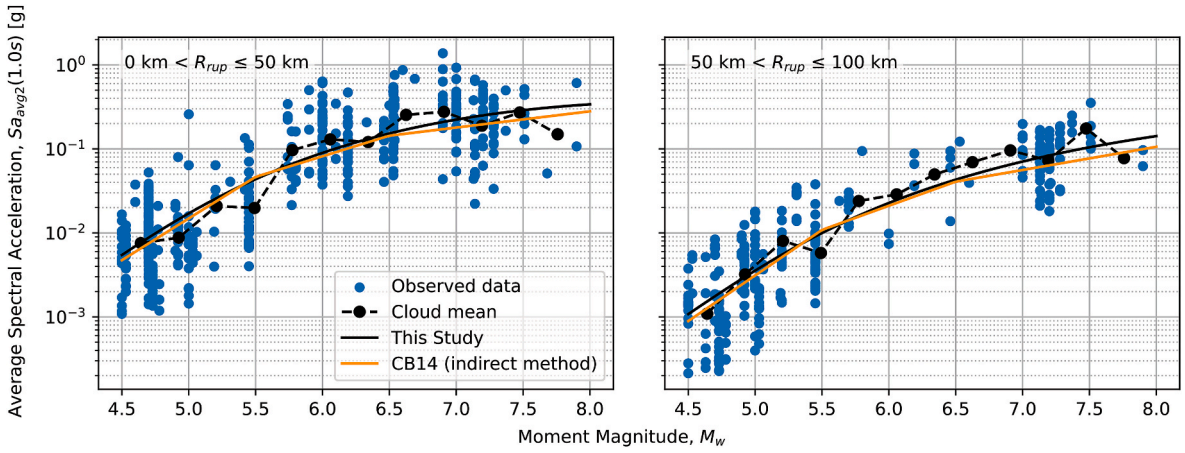


Fig. 13. Magnitude amplification plots of $Sa_{\mathrm{avg2}}(1.0s)$ for two different rupture distance bins.

long-period structures, the *FIV3* IM is almost period-independent. There is an exception for low magnitudes (i.e., 4.5), where there is a mild decrease for long periods. It can also be observed how the effect of the earthquake magnitude is more important than the effect of the source-to-site distance. Similar observations can be made for Sa_{avg3}, but with the difference that it decreases in value with increasing period, which is an expected observation given the similarity in definition to the well-known *Sa* IM.

5.4. Residuals analysis

To ensure that the proposed model did not contain any potential bias with respect to any input parameter, an analysis of the inter-, intra-event, and total residuals was performed. Here, the residuals are defined as *log*10(observations) - *log*10(predictions) and are plotted in Fig. 15 for the inter-, intra-event, and total residuals against three ground motion causal parameters (i.e., M_{w}, R_{rup} and $V_{\mathrm{s,30}}$), respectively, for three IMs (i.e., *Sa*(1.0s), *FIV*3(1.0s), Ds_{595}). It can be observed that there is no notable bias in the binned mean of residuals. Also, there is no significant change in standard deviations versus the GM causal parameters, further corroborating the homoscedasticity assumption for the dispersion model of the proposed GGMM. This assumption suggests that the variability of predicted IM is consistent across different levels of the predictor variables, such as M_{w}, R_{rup} or $V_{\mathrm{s,30}}$. All the observations in Fig. 15 were checked to be valid for all the combinations of IMs, ground motion causal parameters and residual types.

5.5. Model dispersion

The residuals between the values estimated from the proposed GGMM and those observed from the recorded ground motions were used to calculate the inter- and intra-event logarithmic standard deviations. The final standard deviations of all the IMs included in this study are presented in Fig. 16, along with their respective counterparts given from other GMMs available in the literature for relative comparison. All standard deviations were transformed into natural logarithm (i.e., ln) units to have an equal basis for comparison since the fitted GGMM was in terms of log base 10 and other models have used natural logarithm. It can be seen that the total standard deviation of the GGMM is the lowest for most IMs compared to other GMMs available in the literature. This is the case, especially for long-period IMs, although a slight difference was observed at shorter periods, where the intra-event term is slightly higher than CB14, for example, meaning the overall uncertainty is a little higher. Similar trends were observed in the recent GMM of Sedaghati and Pezeshk [53], developed with machine-learning methods. This was investigated, and no specific reason was found to be causing it. It could be argued that previous models used much larger datasets, which could influence the overall variability, but it must be recalled that strict filtering criteria were applied here, as outlined in Section 3, which are expected to decrease the aleatory variability and increase the within-model epistemic uncertainty. In any case, it is noted here that while the mean predictions shown in Fig. 5 may be satisfactory at a shorter period, the uncertainty shown in Fig. 16 is slightly higher using the proposed GGMM. If the analyst is focusing solely on this short period

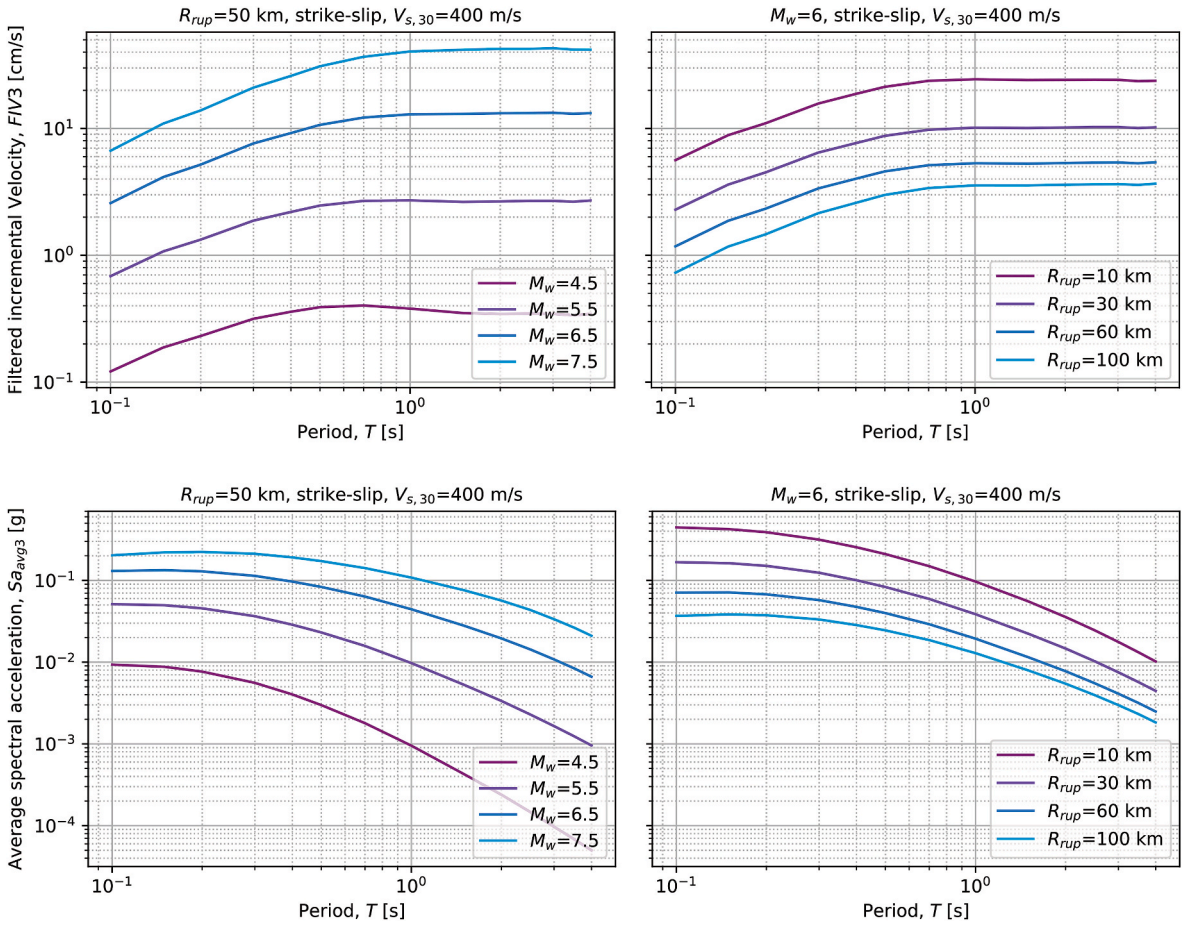


Fig. 14. Model median estimates of response spectra of *FIV*3 (top) and Sa_{avg3} (bottom).

range, other existing models may be more suitable, but from a broader and more generalised perspective, the proposed GGMM tends to have lower uncertainty across several IMs.

Additionally, the proposed model maintains a low inter-event standard deviation and is almost constant throughout all IMs. The high difference between intra- and inter-event standard deviations in this model is likely because of the better characterisation of source effects in comparison to path and site effects. Another reason could be the generally strict filtering criteria applied on the initial strong motion dataset, allowing only recordings from earthquake magnitudes above 4.5, while the source-to-site distance limit was loose, allowing distances from 0 to 300 km.

The comparison of the dispersion of IMs with different horizontal component definitions is illustrated in Fig. 17. It can be seen that the *RotD100* definition presents the highest logarithmic standard deviation from the three, followed by the *RotD50* and then the geometric mean definition. This difference is more pronounced in *Sa* IMs, rather than Sa_{avg2} and Sa_{avg3}, which is an expected result due to the inherent averaging effect of Sa_{avg}. Nonetheless, all horizontal component definitions present very similar standard deviations. The standard deviation of different horizontal component definitions of *Sa* was also studied by Beyer and Bommer [54], but they did not include the *RotD50* definition, as it was not available at the time. The closest definition to *RotD50* would be the *GMRotD50*, for which they found about the same or slightly lower dispersion, than the geometric mean definition. Regarding the *RotD100* definition, denoted as '*MaxD*' in Beyer and Bommer [54], they found its dispersion to be slightly higher than the geometric mean definition, which is aligned with what has been observed here.

6. Correlation modelling

As previously stated, this GGMM, which includes several IMs, finds good utility in creating consistent (i.e., from the same database and GMM) correlation models, which are needed for ground motion record selection and identification of ground motion field used in regional analysis. The general procedure to get those correlation models is outlined in the following. From Equation (9), the total normalised residual, δ_{i}, and total standard deviation, σ_{i}, can be expressed as the sum of inter- and intra-event residuals as:

$$\delta_i\sigma_i = \delta b_i\tau_i + \delta w_i\phi_i \tag{13}$$

This means, by implication, that $\log_{10}\mathrm{IM_i}$ and δ_{i} exhibit a linear relationship. Therefore, the correlation between two IMs is equal to the correlation between the normalised residuals, which in mathematical expression translates to:

$$\rho_{\log_{10}IM_i|\boldsymbol{X},\boldsymbol{\theta},\log_{10}IM_j|\boldsymbol{X},\boldsymbol{\theta}} = \rho_{\delta_i,\delta_j} \tag{14}$$

Then, the correlations of residuals between different IMs can be estimated using the Pearson product-moment correlation coefficient formula:

$$\rho_{x,y} = \frac{\sum_n[(x-\bar{x})(y-\bar{y})]}{\sqrt{\sum_n\left[(x-\bar{x})^2\right]\sum_n\left[(y-\bar{y})^2\right]}} \tag{15}$$

where x and y are generic variables, corresponding to δb_{i} and δb_{j} for inter-event correlation for IMs i and j, and to δw_{i} and δw_{j} for intra-event correlation in this application; $\bar{x}$ and $\bar{y}$ are the sample means and Σ_n is

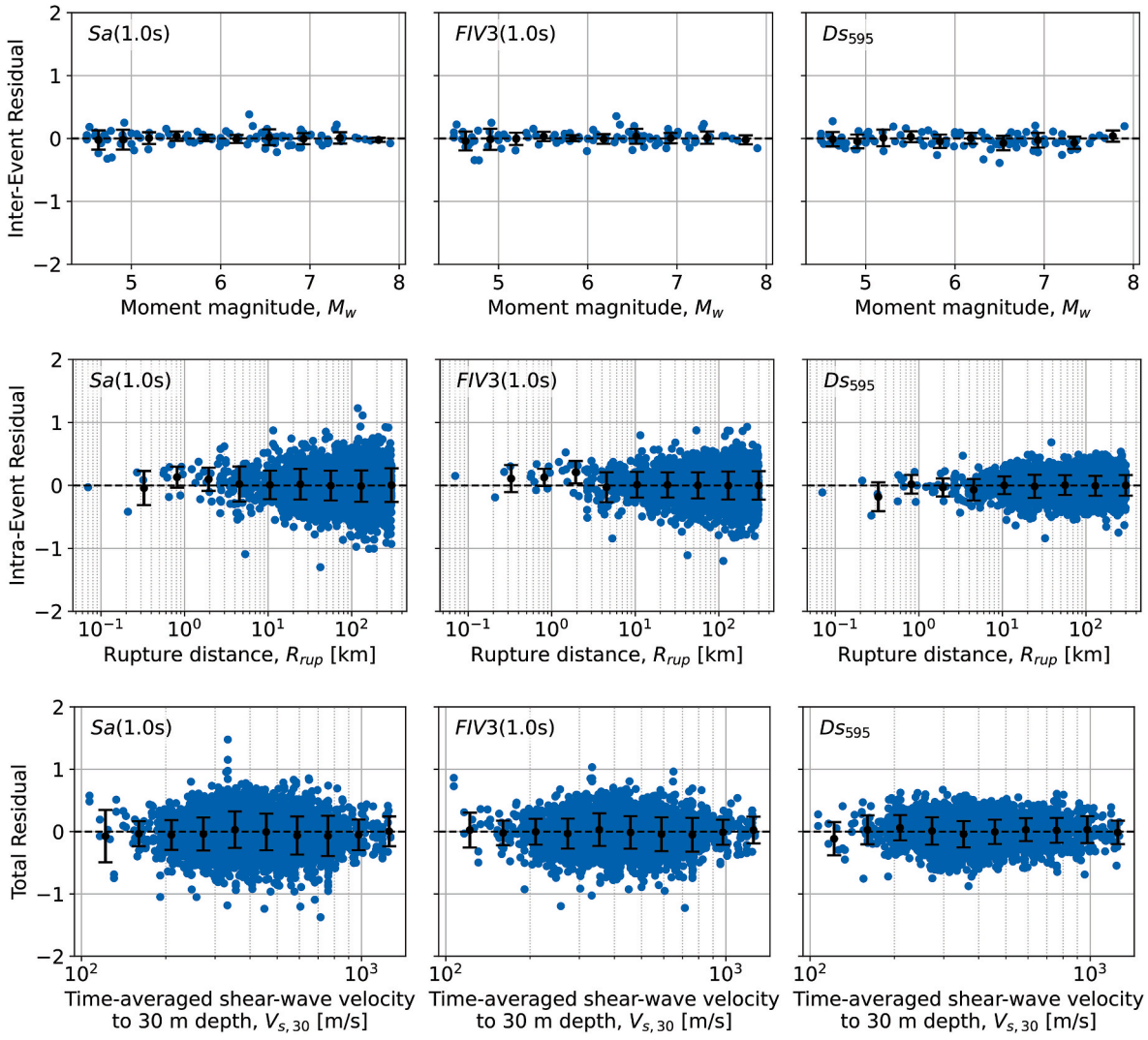


Fig. 15. Inter-, intra-event, and total residuals versus M_w, R_{rup}, and $V_{s,30}$, respectively, for three different IMs. Black dots and error bars represent the binned mean and ± one standard deviation, respectively.

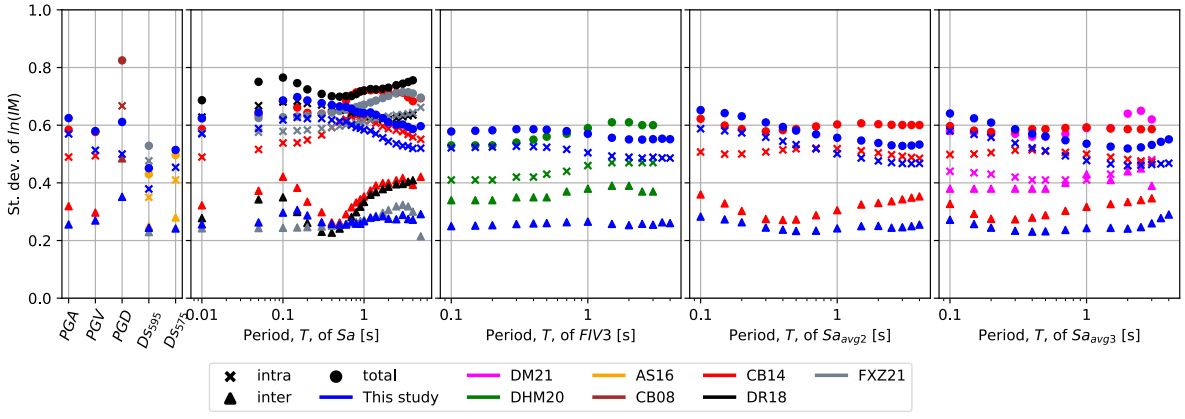


Fig. 16. Inter-, intra-event and total standard deviations of the proposed model for all IMs, compared with models from the literature.

the summation of all n ground motion records. From the definition of correlation coefficient, the correlation between total residuals can be estimated from the inter- and intra-event correlations as follows:

$$\rho_{\delta_i,\delta_j} = \frac{\rho_{\delta b_i,\delta b_j}\tau_i\tau_j + \rho_{\delta w_i,\delta w_j}\phi_i\phi_j}{\sigma_i\sigma_j} \tag{16}$$

Correlation models between *PGA*, *PGV*, *Sa*, *FIV*3, Ds_{595}, Ds_{575}, Sa_{avg2} and Sa_{avg3} were developed using this GGMM and are derived, presented,

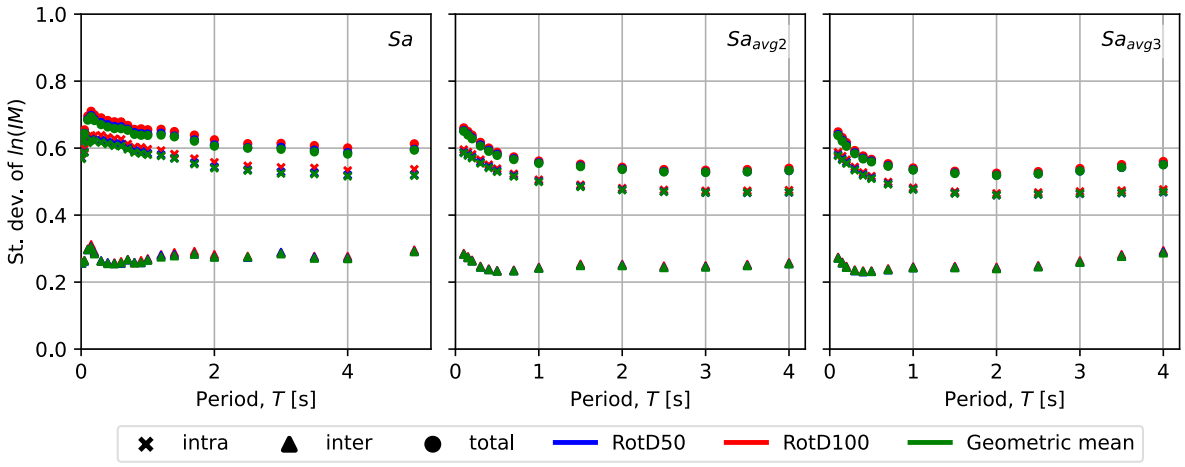


Fig. 17. Comparison of inter-, intra-event and total standard deviations of IMs with different horizontal component definitions.

and discussed in detail in Aristeidou et al. [52]. Interested readers interested in these correlations and their modelling are referred to that work for further details and discussion. As an example, the correlation coefficients between *Sa* and *FIV*3 and between *Ds* and Sa_{avg3} are illustrated in Figs. 18 and 19, respectively.

7. Summary and conclusions

This study proposed a generalised ground motion model (GGMM) to estimate different types of amplitude and cumulative-based intensity measures (IMs) for active shallow crustal earthquakes. These IMs include the peak ground acceleration, *PGA*; peak ground velocity, *PGV*; peak ground displacement, *PGD*; spectral acceleration *Sa*(*T*); two definitions of significant duration, Ds_{575} and Ds_{595}; filtered incremental velocity, *FIV*3(*T*); and two definitions of average spectral acceleration, $Sa_{avg2}(T)$ and $Sa_{avg3}(T)$, meaning a total of 169 output IMs were presented here. Testing and training data were collected from a stringently filtered subset of records from the NGA-West2 strong motion database.

Thorough validation exercises and comparisons with other ground motion models (GMMs) were carried out to demonstrate the suitability of the GGMM. It shows how this framework can effectively capture the complex relationships and interactions between different IMs, which is one of the advantages of this GGMM, as it estimates various IMs in a single model. This helps develop more consistent correlation models between the estimated IMs since they come from the same database and GMM. Meanwhile, more IMs of interest can be seamlessly added to the model's outputs with only minor modifications (e.g., increase of hidden layer neurons). Another advantage observed was that it minimises the dispersion of residuals (aleatory uncertainty) while keeping the two fitting performance metrics (i.e., R^2 and *MSE*) at an optimal level. The

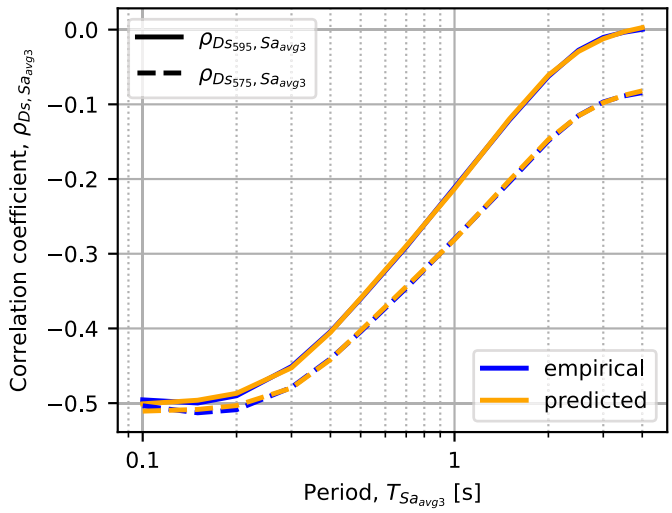


Fig. 19. Empirical and corresponding predicted correlation coefficients between *Ds* and Sa_{avg3}.

logarithmic total standard deviations were low, especially in long-period IMs.

A few limitations of the approaches and methods adopted for the development of this model are that these types of models work well only where data are available, the analyst has less control over the analytical functional forms, which ideally reflect the actual physics of the seismological phenomena, and maybe the homoscedasticity assumption for

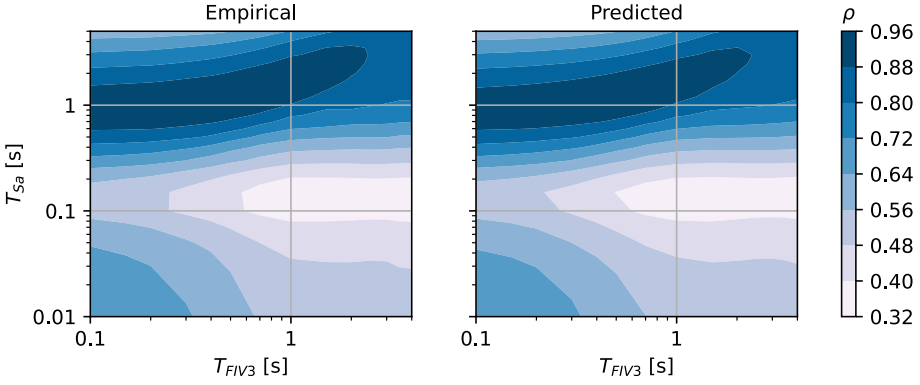


Fig. 18. Empirical and corresponding predicted correlation coefficients between *Sa* and *FIV*3.

treatment of dispersion. The reason for not opting for heteroscedastic dispersion model, is there are already a lot of output IMs from the same model, so adding heteroscedasticity would overcomplicate the model, for small potential benefit.

Overall, the results and comparisons suggest that the proposed model performs excellently in estimating a variety of traditional and next-generation IMs, without any bias against the input features. The proposed model can be used in seismic hazard analysis to derive site-specific uniform hazard spectra, hazard curves for the IM of choice, scenario-based shake maps, and more. This model represents a very useful addition in many situations where more advanced intensity measures, such as filtered incremental velocity or average spectral acceleration, are required. Recent research has highlighted the potential of these intensity measures for a better characterisation of structural response (i.e., sufficiency, efficiency etc.) but to date, there has been little work done to provide suitable models to quantify their hazard (i.e., hazard computability) [3]. This paper has directly aimed at filling that gap. Furthermore, the use of this single GGMM has allowed consistent correlation models to be quantified by Aristeidou et al. [52] which paves the way for more refined record selections following the generalised conditional intensity measure approach [55].

Following the finalisation of the model, the Authors have implemented this GMM in a local version of the well-known PSHA tool OpenQuake. It has been tested and observed to seamlessly integrate with other GMMs available in the literature. The code available online for this GMM takes inputs the nine seismological parameters and outputs the mean and sigma of a specific IM, as with any other GMM. Therefore, even if the model is non-parametric, it can be used like any other classical GMM and easily adapted to any PSHA procedure/code.

Funding

The work presented in this paper has been developed within the framework of the projects "Dipartimenti di Eccellenza", funded by the Italian Ministry of Education, University and Research at IUSS Pavia.

Availability of data and material

Not applicable.

Code availability

The relevant files and functions that can be used to get estimations from the ground motion model presented here are available on GitHub at: https://github.com/Savvinos-Aristeidou/ANN-GGMM.git.

CRediT authorship contribution statement

Savvinos Aristeidou: Writing – original draft, Validation, Software, Methodology, Formal analysis. **Davit Shahnazaryan:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Data curation. **Gerard J. O'Reilly:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Jalayer F, Beck JL, Zareian F. Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory. J Eng Mech 2012;138(3):307–16. https://doi.org/10.1061/(asce)em.1943-7889.0000327.
[2] Kohrangi M, Vamvatsikos D, Bazzurro P. Implications of intensity measure selection for seismic loss assessment of 3-D buildings. Earthq Spectra 2016;32(4):2167–89. https://doi.org/10.1193/112215EQS177M.
[3] O'Reilly GJ. Limitations of Sa(T 1) as an intensity measure when assessing non-ductile infilled RC frame structures. Bull Earthq Eng 2021;19(6):2389–417. https://doi.org/10.1007/s10518-021-01071-7.
[4] O'Reilly GJ. Seismic intensity measures for risk assessment of bridges. Bull Earthq Eng 2021;19(9):3671–99. https://doi.org/10.1007/s10518-021-01114-z.
[5] Otárola K, et al. Impact of ground-motion duration on nonlinear structural performance: Part I: spectrally equivalent records and inelastic single-degree-of-freedom systems. Earthq Spectra 2023;39(2):829–59. https://doi.org/10.1177/87552930231155502.
[6] Douglas J. Ground motion prediction equations 1964–2021. http://www.gmpe.org.uk/gmpereport2014.html. [Accessed 4 December 2023].
[7] Bradley BA. Empirical equations for the prediction of displacement spectrum intensity and its correlation with other intensity measures. Soil Dynam Earthq Eng 2011;31(8):1182–91. https://doi.org/10.1016/j.soildyn.2011.04.007.
[8] Afshari K, Stewart JP. Physically parameterized prediction equations for significant duration in active crustal regions. Earthq Spectra 2016;32(4):2057–81. https://doi.org/10.1193/063015EQS106M.
[9] Campbell KW, Bozorgnia Y. Ground motion models for the horizontal components of Arias intensity (AI) and cumulative absolute velocity (CAV) using the NGA-west2 database. Earthq Spectra 2019;35(3):1289–310. https://doi.org/10.1193/090818EQS212M.
[10] Zafarani H, Soghrat MR. Ground motion models for non-spectral intensity measures based on the Iranian database. J Earthq Eng 2023;27(13):3786–806. https://doi.org/10.1080/13632469.2022.2150334.
[11] Fayaz J, Xiang Y, Zareian F. Generalized ground motion prediction model using hybrid recurrent neural network. Earthq Eng Struct Dynam 2021;50(6):1539–61. https://doi.org/10.1002/eqe.3410.
[12] Campbell KW, Bozorgnia Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. Earthq Spectra 2008;24(1):139–71. https://doi.org/10.1193/1.2857546.
[13] Campbell KW, Bozorgnia Y. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. Earthq Spectra 2014;30(3):1087–115. https://doi.org/10.1193/062913EQS175M.
[14] Dávalos H, Heresi P, Miranda E. A ground motion prediction equation for filtered incremental velocity, FIV3. Soil Dynam Earthq Eng 2020;139:106346. https://doi.org/10.1016/j.soildyn.2020.106346.
[15] Dávalos H, Miranda E. A ground motion prediction model for average spectral acceleration. J Earthq Eng 2021;25(2):319–42. https://doi.org/10.1080/13632469.2018.1518278.
[16] Baker J, Bradley B, Stafford P. Seismic hazard and risk analysis. first ed. Cambridge University Press; 2021. https://doi.org/10.1017/9781108425056.
[17] Dhanya J, Raghukanth STG. Ground motion prediction model using artificial neural network. Pure Appl Geophys 2018;175(3):1035–64. https://doi.org/10.1007/s00024-017-1751-3.
[18] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: ACM; 2016. p. 785–94. https://doi.org/10.1145/2939672.2939785.
[19] Bakouregui AS, et al. Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC columns. Eng Struct 2021;245:112836. https://doi.org/10.1016/j.engstruct.2021.112836.
[20] Derras B, Bard PY, Cotton F. Towards fully data driven ground-motion prediction models for Europe. Bull Earthq Eng 2014;12(1):495–516. https://doi.org/10.1007/s10518-013-9481-0.
[21] Derras B, Bard PY, Cotton F. Site-condition proxies, ground motion variability, and data-driven GMPEs: insights from the NGA-West2 and RESORCE data sets. Earthq Spectra 2016;32(4):2027–56. https://doi.org/10.1193/060215EQS082M.
[22] Boore DM. Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. Bull Seismol Soc Am 2010;100(4):1830–5. https://doi.org/10.1785/0120090400.
[23] Ancheta T, et al. 'PEER NGA-west2 database. Tech Rep PEER 2013/03 2013.
[24] Zafarani H, Soghrat MR. An empirical spectral ground-motion model for Iran using truncated Iranian strong-motion database enriched by near-field records. J Earthq Eng 2024;28(4):922–45. https://doi.org/10.1080/13632469.2023.2226223.
[25] Wooddell KE, Abrahamson NA. Classification of main shocks and aftershocks in the NGA-West2 database. Earthq Spectra 2014;30(3):1257–67. https://doi.org/10.1193/071913EQS208M.
[26] Kaklamanos J, Baise LG, Boore DM. Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. Earthq Spectra 2011;27(4):1219–35. https://doi.org/10.1193/1.3650372.
[27] Cornell C Allin. Engineering seismic risk analysis. Bull Seismol Soc Am 1968;58(5):1583–606.
[28] Vamvatsikos D, Cornell CA. Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape

information. Earthq Eng Struct Dynam 2005;34(13):1573–600. https://doi.org/10.1002/eqe.496.
[29] Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure for collapse risk assessment. Earthq Eng Struct Dynam 2015;44(12):2057–73. https://doi.org/10.1002/eqe.2575.
[30] Dávalos H, Miranda E. Filtered incremental velocity: a novel approach in intensity measures for seismic collapse estimation. Earthq Eng Struct Dynam 2019;48(12):1384–405. https://doi.org/10.1002/eqe.3205.
[31] Bommer JJ, Martínez-Pereira A. Tha effective duration of earthquake strong motion. J Earthq Eng 1999;3(2):127–72. https://doi.org/10.1080/13632469909350343.
[32] Chandramohan R, Baker JW, Deierlein GG. Quantifying the influence of ground motion duration on structural collapse capacity using spectrally equivalent records. Earthq Spectra 2016;32(2):927–50. https://doi.org/10.1193/122813EQS298MR2.
[33] Husid R. Características de terremotos. Análisis general. Revista IDIEM 1969;8(1):21–42.
[34] Kazantzi AK, Vamvatsikos D. Intensity measure selection for vulnerability studies of building classes. Earthq Eng Struct Dynam 2015;44(15):2677–94. https://doi.org/10.1002/eqe.2603.
[35] Eads L, Miranda E. Seismic collapse risk assessment of buildings: effects of intensity measure selection and computational approach, Report No. 184. Stanford University; 2013.
[36] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5(4):115–33. https://doi.org/10.1007/BF02478259.
[37] Kiefer J, Wolfowitz J. Stochastic estimation of the maximum of a regression function. Ann Math Stat 1952;23(3):462–6. https://doi.org/10.1214/aoms/1177729392.
[38] Haykin SS. Neural networks and learning machines. New Jersey: Pearson; 2009.
[39] Abadi M, et al. TensorFlow: large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/; 2015.
[40] Bindi D, et al. Ground motion prediction equations derived from the Italian strong motion database. Bull Earthq Eng 2011;9(6):1899–920. https://doi.org/10.1007/s10518-011-9313-z.
[41] Kalakonas P, Silva V. Earthquake scenarios for building portfolios using artificial neural networks: part I—ground motion modelling. Bull Earthq Eng 2022;(123456789). https://doi.org/10.1007/s10518-022-01598-3 [Preprint].
[42] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: 3rd international conference on learning representations, ICLR 2015; 2014. p. 1–15. http://arxiv.org/abs/1412.6980.
[43] Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc 1984;79(387):575–83. https://doi.org/10.1080/01621459.1984.10478083.
[44] Močkus J. On bayesian methods for seeking the extremum 1975:400–4. https://doi.org/10.1007/3-540-07165-2_55.
[45] Kotha SR, et al. A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. Bull Earthq Eng 2020;18(9):4091–125. https://doi.org/10.1007/s10518-020-00869-1.
[46] Boore DM, et al. A ground-motion prediction model for shallow crustal earthquakes in Greece. Bull Seismol Soc Am 2021;111(2):857–74. https://doi.org/10.1785/0120200270.
[47] Boore DM, et al. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. Earthq Spectra 2014;30(3):1057–85. https://doi.org/10.1193/070113EQS184M.
[48] Atik LA, et al. The variability of ground-motion prediction models and its components. Seismol Res Lett 2010;81(5):794–801. https://doi.org/10.1785/gssrl.81.5.794.
[49] Abrahamson NA, Youngs RR. A stable algorithm for regression analyses using the random effects model. Bull Seismol Soc Am 1992;82(1):505–10. https://doi.org/10.1785/BSSA0820010505.
[50] Kohrangi M, et al. Conditional spectrum-based ground motion record selection using average spectral acceleration. Earthq Eng Struct Dynam 2017;46(10):1667–85. https://doi.org/10.1002/eqe.2876.
[51] Baker JW, Jayaram N. Correlation of spectral acceleration values from NGA ground motion models. Earthq Spectra 2008;24(1):299–317. https://doi.org/10.1193/1.2857544.
[52] Aristeidou S, Shahnazaryan D, O'Reilly GJ. Correlation models for next-generation amplitude and cumulative intensity measures using artificial neural networks. Earthq Spectra 2024:1–24.
[53] Sedaghati F, Pezeshk S. Machine learning–based ground motion models for shallow crustal earthquakes in active tectonic regions. Earthq Spectra 2023;39(4):2406–35. https://doi.org/10.1177/87552930231191759.
[54] Beyer K, Bommer JJ. Relationships between median values and between aleatory variabilities for different definitions of the horizontal component of motion. Bull Seismol Soc Am 2006;96(4 A):1512–22. https://doi.org/10.1785/0120050210.
[55] Bradley BA. A generalized conditional intensity measure approach and holistic ground-motion selection. Earthq Eng Struct Dynam 2010;39(12):1321–42. https://doi.org/10.1002/eqe.995.