Geophysical Journal International

Geophys. J. Int. (2024) **236,** 413–429
Advance Access publication 2023 November 01
GJI Seismology

https://doi.org/10.1093/gji/ggad432

ANN-based ground motion model for Turkey using stochastic simulation of earthquakes

Shaghayegh Karimzadeh [1], Amirhossein Mohammadi,[1] Sayed Mohammad Sajad Hussaini [1], Daniel Caicedo,[1] Aysegul Askan [2] and Paulo B. Lourenço[1]

[1] *ISISE, ARISE, Department of Civil Engineering, University of Minho,* 4800-058*, Guimarães, Portugal. E-mail: shaghkn@civil.uminho.pt*
[2] *Departments of Civil Engineering and Earthquake Studies, Middle East Technical University (METU),* 06800*, Ankara, Turkey*

Accepted 2023 October 30. Received 2023 October 17; in original form 2023 June 23

SUMMARY

Turkey is characterized by a high level of seismic activity attributed to its complex tectonic structure. The country has a dense network to record earthquake ground motions; however, to study previous earthquakes and to account for potential future ones, ground motion simulations are required. Ground motion simulation techniques offer an alternative means of generating region-specific time-series data for locations with limited seismic networks or regions with seismic data gaps, facilitating the study of potential catastrophic earthquakes. In this research, a local ground motion model (GMM) for Turkey is developed using region-specific simulated records, thus constructing a homogeneous data set. The simulations employ the stochastic finite-fault approach and utilize validated input-model parameters in distinct regions, namely Afyon, Erzincan, Duzce, Istanbul and Van. To overcome the limitations of linear regression-based models, artificial neural network is used to establish the form of equations and coefficients. The predictive input parameters encompass fault mechanism (FM), focal depth (FD), moment magnitude (M_w), Joyner and Boore distance (R_{JB}) and average shear wave velocity in the top 30 m (V_{s30}). The data set comprises 7359 records with M_w ranging between 5.0 and 7.5 and R_{JB} ranging from 0 to 272 km. The results are presented in terms of spectral ordinates within the period range of 0.03–2.0 s, as well as peak ground acceleration and peak ground velocity. The quantification of the GMM uncertainty is achieved through the analysis of residuals, enabling insights into inter- and intra-event uncertainties. The simulation results and the effectiveness of the model are verified by comparing the predicted values of ground motion parameters with the observed values recorded during previous events in the region. The results demonstrate the efficacy of the proposed model in simulating physical phenomena.

Key words: Machine learning; Computational seismology; Earthquake ground motions.

1 INTRODUCTION

Ground motion models (GMMs) provide estimations of different shaking intensities that characterize strong ground motions from prior knowledge of seismological parameters, such as moment magnitude M_w, fault mechanism (FM), focal depth (FD), average shear wave velocity in the top 30 m (V_{s30}) and various source-to-site distance metrics. Because of their broad applications, GMMs are commonly employed in different fields, such as earthquake engineering and seismology. Moreover, GMMs represent a well-known tool for the prediction of ground shaking intensities, and therefore, their implementation is essential in the context of seismic hazard analyses. A large variety of GMMs have been developed in the past for the prediction of various intensity measures (IMs), including peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), spectral acceleration (SA), or pseudo-spectral acceleration (PSA) at different periods (Boore & Atkinson 2008; Akkar *et al.* 2014; Campbell & Bozorgnia 2014). Nevertheless, two major drawbacks might be pointed out regarding most GMMs currently available. First, most models adopt parametric formulations, which might induce bias in the prediction of IMs (Campbell & Bozorgnia 2012) and second, GMMs rely on the quality of the data set adopted to develop the model, which can be problematic for regions with a moderate to high levels of hazard and lack of recorded accelerograms characteristic of large-magnitude events (Gianniotis *et al.* 2014).

Over the last decade, extensive research on GMMs has been conducted involving parametric formulations and robust mathematical forms. Boore *et al.* (2014) provided prediction equations for computing medians and standard deviations of PGA, PGV and 5 per cent damped PSA for shallow crustal earthquakes using the NGA-West2 database (Ancheta *et al.* 2014). Similarly, GMMs were presented by Bindi *et al.* (2014) valid for Europe and the Middle East, with distances (i.e. Joyner–Boore, R_{JB}; and hypocentral, R_{hypo}) less than 300 km, hypocentral depth up to 35 km and M_w range from 4 to 7.6. Kale *et al.* (2015) proposed a GMM for Turkey and Iran to investigate the potential regional effects on ground motion amplitudes from shallow active crustal earthquakes using a subset of the recently compiled strong-motion database of the Earthquake Model of the Middle East Region project (Şeşetyan *et al.* 2018). Bommer *et al.* (2016) considered the Netherlands seismicity to develop GMMs for spectral ordinates of moderate-to-large-magnitude earthquakes. Likewise, Idini *et al.* (2017) presented a GMM valid for the Chilean subduction zone. The study by Bozorgnia & Campbell 2016a, b) focused on the development of GMMs for vertical components of PGA, PGV and PSA. The model was claimed to be valid for worldwide shallow crustal earthquakes, various types of faulting, M_w from 3.3 to 8.5, and for fault rupture distances ranging from 0 to 300 km. More recently, Bindi *et al.* (2019) introduced a GMM for the prediction of acceleration and displacement spectral ordinates, including country-to-country random effect (Italy, Turkey, Romania, Greece and others). Boore *et al.* (2021) derived a GMM for the horizontal components of PGA, PGV and 5 per cent damped PSA using a wide database of uniformly processed strong-motion data recorded in Greece. Alternatively, the implementation of alternative non-parametric GMMs has increased significantly in the last few years. Wiszniowski (2019) implemented the Fahlman's Cascade Correlation neural network (Fahlman & Lebiere 1990) to generate an improved GMM for the prediction of peak horizontal acceleration as a function of distant metrics in different M_w ranges. Meenakshi *et al.* (2023) adopted artificial neural networks (ANNs) coupled with the genetic algorithm to develop GMMs in the Peninsular India for maximum rotated (RotD50) components of PGA, PGV and 5 per cent damped PSA for periods between 0.01 and 3 s. Sreenath *et al.* (2023) adopted diverse machine learning models to develop a hybrid non-parametric GMM for shallow crustal earthquakes in Europe; the model was developed for a large number of seismic intensities (i.e. PGA; PGV; PGD; cumulative absolute velocity; Arias intensity and significant duration). Based on recorded ground motions in Turkey, Yerlikaya-Özkurt *et al.* (2014) recently derived a GMM for Turkey to predict PGA and PGV using the multivariate adaptive regression splines method.

In the meantime, other researchers have approached the lack of recorded accelerograms characteristic of large-magnitude events by adopting simulation techniques to reproduce synthetic motions. For instance, Ugurhan & Askan (2010) performed stochastic simulation based on the dynamic corner frequency approach proposed by Motazedian & Atkinson (2005) considering the Düzce (Turkey) earthquake that took place on 1999 November 12 (M_w = 7.1). Later, Ozmen *et al.* (2020) studied the same event with an updated simulation approach. Askan *et al.* (2013) investigated the sensitivity to seismic parameters of stochastic simulations using sparse data collected from the 1992 March 13 Erzincan earthquake in eastern Turkey of $M_w = 6.6$. The work of Karimzadeh & Askan (2018) focused on simulations of the historical 1939 Erzincan earthquake in Turkey through the dynamic corner frequency approach using regional seismological information computed from the 1992 earthquake that took place in the same region. Cheloni and Akinci (2020) recently performed stochastic finite-fault simulations to generate high-frequency synthetics motions for the Elazığ earthquake in Turkey of $M_w = 6.8$. A scenario earthquake ground motion data set was also developed for the Gaziantep region in Turkey, which was affected by the recent 2023 Kahramanmaraş events (Arslan Kelam *et al.* 2022). Similarly, stochastic simulation has been employed for simulating the records of past earthquakes, such as: the 1998 July 9 Faial Earthquake (Azores, Portugal) (Karimzadeh & Lourenço 2022); the 2022 February 3 Cay (Turkey) earthquake (Can *et al.* 2021); the 2009 April 6 L'Aquila earthquake (Ugurhan *et al.* 2012) and the 2016 Kumamoto (Japan) earthquake (Zhang *et al.* 2016). Consecutively, other investigations have addressed the validation of synthetic records from a seismological and engineering point of view (Zonno *et al.* 2010; Koboevic *et al.* 2011; Karimzadeh 2019; Karimzadeh *et al.* 2019, 2020; Fayaz *et al.* 2020; Karimzadeh *et al.* 2021a, b). In other studies, large suites of simulated motions have been employed for the development of GMMs valid for the prediction of PGA and spectral ordinates (Campbell 2003; Megawati *et al.* 2005; Withers *et al.* 2020; Raghucharan *et al.* 2021; Sreenath *et al.* 2023).

This paper introduces a novel approach to develop a non-parametric GMM using a database of stochastically simulated records through an ANN implementation in Python. The study focuses on Turkey as the chosen area, driven by its high seismic activity and the scarcity of large-magnitude events in different regions. The choice of Turkey was further motivated by the occurrence of recent catastrophic events of 2023 February 6 in Gaziantep ($M_w = 7.7$) and Elbistan ($M_w = 7.5$). While a non-parametric GMM for estimating spectral ordinates in Turkey has been previously proposed using the XGBoost algorithm (Mohammadi *et al.* 2023), it has limitations in capturing information on large-magnitude events and lacks data set homogeneity. In contrast, the non-parametric GMM presented in this paper relies on a substantial data set of synthetic records generated through the stochastic finite-fault method, encompassing regions such as Afyon, Erzincan, Duzce, Istanbul and Van within Turkey. The developed model aims to predict ground motion IMs such as PGA, PGV as well as various values of PSA in the range 0.03–2.0 s. The effectiveness of the proposed GMM is verified by comparing the predicted values of the ground motion IMs with the observed values recorded during previous events in the selected regions. Additionally, the model's trend is assessed by comparing it with the real data set of Turkey, which includes the most recent events of 2023 February 6. In addition, the developed model is compared against the selected parametric GMM proposed by Kale *et al.* (2015).

2 GROUND MOTION DATA SET

This study utilizes the stochastic finite-fault ground motion simulation approach proposed by Motazedian & Atkinson (2005) to construct the ground motion data set for the GMM. The chosen study area for conducting simulations is selected regions in Turkey characterized by high seismicity and diverse tectonic structures, including Afyon, Erzincan, Duzce, Istanbul and Van. Fig. 1 illustrates the tectonic map of Turkey with their convergence rates Utkucu et al. (2003), highlighting the selected study areas represented by red rectangular boxes. These regions are specifically selected due to the absence of a unified set of recorded ground motions that encompass a wide range of magnitudes, source-to-site distances and site conditions corresponding to past earthquakes. Focusing on these

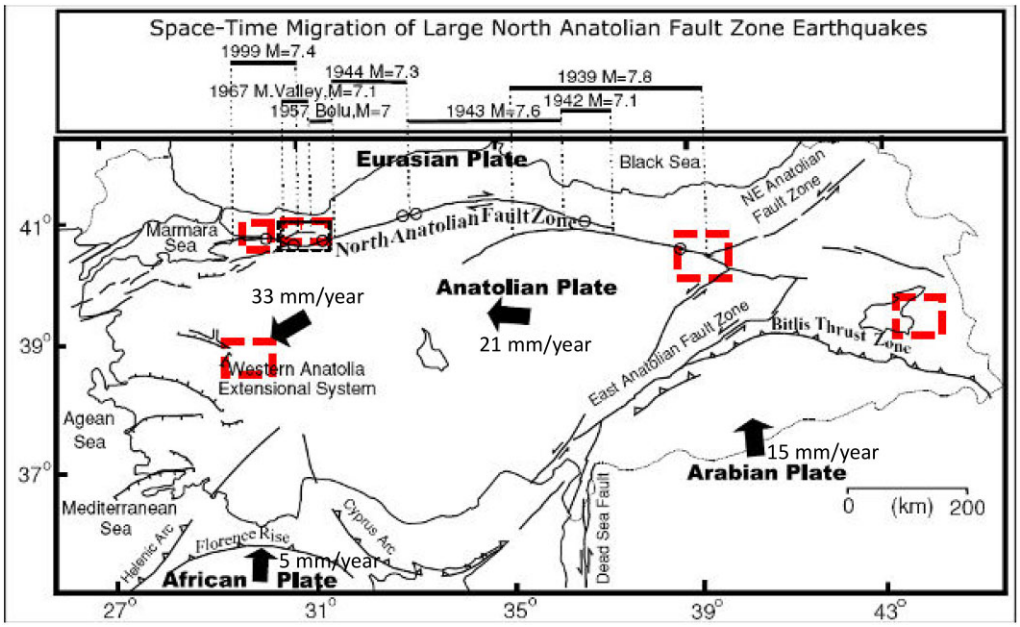

Figure 1. Tectonic map of Turkey displaying epicentres of previous events (modified from the study of Utkucu *et al.* 2003). The red boxes indicate the regions considered in this study.

areas aims to address the need for a comprehensive data set that encompasses a broader spectrum of seismic events and associated ground motions. This section provides details on the methodology employed for ground motion simulation and information regarding the scenario earthquakes and the generated ground motion data set.

2.1. Ground motion simulation method

The generation of synthetic records based on the stochastic point source was originally introduced by Boore (1983), who combined the source spectrum of Aki (1967) and Brune (1970) with the findings of Hanks & McGuire (1981). Initially, the implementation of Boore (1983) relied on the deterministic far-field *S*-wave Fourier amplitude spectrum of acceleration, where random phase angles were incorporated to generate single horizontal acceleration time-series. Subsequently, Beresnev & Atkinson (1997) extended the point-source model of Boore (1983) to include finite-fault effects in simulations. This extension, known as the stochastic finite-fault method, involved discretising the fault plane into smaller subfaults. Each subfault is treated as a stochastic point source, and their contributions are summed in the time domain. However, a limitation of this method is the assumption of a constant corner frequency, which led to the dependence of the total radiated energy on the sizes of the subfaults. To address this limitation, a more recent version of the stochastic method introduced a dynamic corner frequency approach (Motazedian & Atkinson 2005) to model the high-frequency content of the shear wave portion of ground motion records. In this approach, the corner frequency at any given time is defined to be inversely proportional to the area of the subfaults that had ruptured up to that time (Motazedian & Atkinson 2005). Fig. 2 is the schematic distribution of the wave front from a finite-fault source model.

This study employs the stochastic finite-fault method incorporating a dynamic corner frequency concept (Motazedian & Atkinson 2005) to simulate earthquake scenarios in the selected regions in Turkey. In this approach, the fault plane is represented by a collection of smaller subfaults, each of which is considered a stochastic point source (Boore 1983). The acceleration spectrum of each point source (*ij*) is expressed as follows:

$$A_{ij}(f) = \frac{C M_{0_{ij}} H_{ij} (2\pi f)^2}{1 + \left(\frac{f}{f_{c_{ij}}}\right)^2} G(R) e^{-\frac{\pi f R_{ij}}{Q\beta}} S(f) e^{-(\pi \kappa f)}, \quad (1)$$

where

$$C = \frac{R_{\theta\phi} \times \mathrm{FS} \times \mathrm{PRTITN}}{4\pi\rho\beta^3}. \quad (2)$$

In the given equations, various parameters are used to characterize the seismic phenomena: M_0 represents the seismic moment measured in dyne·cm; R_{ij} denotes the distance from the observation point to the subfault indexed as *ij*; β signifies the crustal shear wave velocity, measured in km s^{-1}; Q is the frequency-dependent quality factor; *G(R)* presents the geometric spreading as a function of source-to-site distance (R), *S(f)* represents the soil amplification function; k (kappa) models the linear decay in higher frequencies of the Fourier amplitude spectrum of the *S*-wave portion of the acceleration records, represented in semi-logarithmic space; FS denotes the free surface amplification factor, typically assumed to be 2 and PRTITN is a factor that reflects the partitioning of shear wave energy into two horizontal components. Its assumed value is generally $1/\sqrt{2}\cdot\ \rho$ and represents the crustal density measured in g cm^{-2}; H_{ij} is a scaling factor dependent on frequency, specifically for high frequencies and, finally, $R_{\theta\phi}$-denotes the radiation pattern constant, often considered as 0.55 for shear waves (Atkinson & Boore 1995). It is worth noting that recent investigations (Takemura *et al.* 2016; Kotha *et al.* 2019; Wang *et al.* 2021) underscore the relevance of frequency- and distance-dependent radiation pattern models. Yet, this study is founded on utilizing a consistent radiation pattern coefficient of 0.55, construed as emblematic of an average radiation pattern coefficient.

The term fc_{ij} in eq. (1), which defines the corner frequency of a subfault, is defined as follows:

$$f_{c_{ij}} = N_R(t)^{-\frac{1}{3}} \times 4.9 \times 10^6 \beta \left(\frac{\Delta\sigma}{M_{0-\mathrm{ave}}}\right)^{1/3}, \quad (3)$$

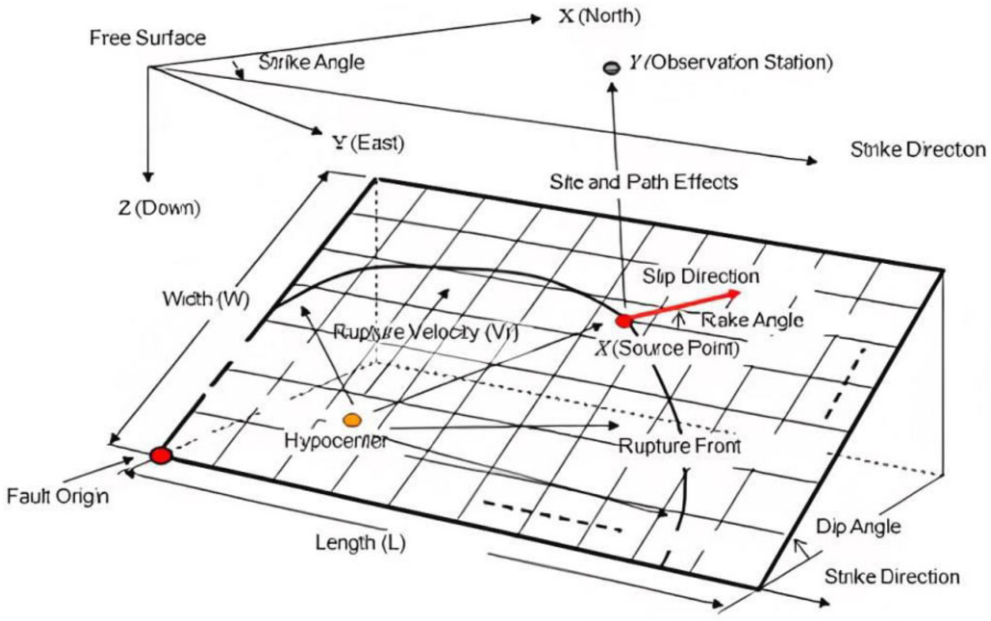

Figure 2. Rectangular finite-fault model illustrating wave propagation (adapted from Hisada 1994).

where $N_R(t)$ represents the total count of subfaults that have experienced rupture by time t, $\Delta\sigma$ denotes the stress drop and $M_{0\text{-ave}}$ indicates the average seismic moment associated with the fault.

The deterministic acceleration spectrum described in eq. (1) is combined with random phases and converted into the time domain for each point source on the fault plane. The individual contributions from each sub-fault are then accumulated in the time domain to produce the overall total acceleration as follows:

$$a(t) = i = \sum_{i=1}^{nl} \sum_{j=1}^{nw} a_{ij}\left(t - \Delta t_{ij} - T_{ij}\right), \tag{4}$$

where, a_{ij} represents the time-series of acceleration specific to the ijth subfault, while $a(t)$ denotes the acceleration of the entire fault. The terms nl and nw represent the number of subfaults considered along the length and width of the rectangular fault plane, respectively. T_{ij} corresponds to the ratio of the subfault radius to the rupture velocity and Δt_{ij} indicates the time delay between each subfault and the observation point.

2.2. Input-model parameters

The ground motion data set in this study encompasses a comprehensive set of time-series derived from simulations conducted in diverse regions of Turkey, including Afyon, Erzincan, Duzce, Istanbul and Van. The simulations cover a wide range of magnitude from 5.0 to 7.5 with an interval of 0.5 in addition to the past events, including the 2002 Afyon ($M_w = 6.6$), 1992 Erzincan ($M_w = 6.9$), 1999 Duzce ($M_w = 7.1$) and 2011 Van ($M_w = 7.1$) earthquakes and a hypothetic scenario ($M_w = 7.4$) in Istanbul. It is noted that the input-model parameters of simulations for Afyon, Erzincan, Duzce and Van were validated against previous seismic events, including the 2002 Afyon ($M_w = 6.6$), 1992 Erzincan ($M_w = 6.6$), 1999 Duzce ($M_w = 7.1$) and 2011 Van ($M_w = 7.1$) earthquakes in Ugurhan & Askan (2010), Aysegul Askan *et al.* (2013), Zengin & Cakti (2014) and Can *et al.* (2021), while, for Istanbul, a hypothetical scenario ($M_w = 7.4$) was validated against GMMs by Uckan *et al.* (2018). In all regions, the simulations involved calibrating the validated parameters specifically for different scenario events. To this end, the dimensions of the fault planes, in terms of length and width, are determined by calibrating them with the earthquake magnitude (M_w) through the empirical equations introduced by Wells & Coppersmith (1994). Similarly, the stress drop is calibrated using the empirical equation proposed by Mohammadioun & Serva (2001). The local site classes are also incorporated in the selected random nodes (Ismet Kanli *et al.* 2006; Ugurhan & Askan 2010; Askan *et al.* 2015; Sahin *et al.* 2016; Akkaya & Özvan 2019). This approach resulted in consideration of three distinct soil types, each characterized by mean V_{s30} values of 255 m s^{-1} (NEHRP soil class D ((US) & (US) 2001)), 310 m s^{-1} (generic soil) and 520 m s^{-1} (NEHRP soil class C ((US) & (US) 2001)), according to the study of Boore & Joyner (1997). By incorporating these soil types, the simulations account for the variability in ground response associated with different soil conditions prevalent within the study regions. Table 1 provides a summary of the input-model parameters utilized for simulations conducted in the selected regions. These parameters encompass source characteristics, path properties and site conditions, offering a comprehensive overview of the key factors considered in the simulations for each respective region. In addition, the table outlines the boundaries for each region in which evenly distributed nodes are chosen for simulations. The respective numbers of stations considered for Afyon, Erzincan, Duzce, Istanbul and Van are 324, 365, 90, 88 and 430.

It is important to note that the simulations conducted in this study have been rigorously verified and validated in earthquake engineering practice. The authors have previously employed these simulations in various studies encompassing different applications (Askan *et al.* 2015; Karimzadeh *et al.* 2017a, b, 2019, 2020, 2021; Karimzadeh & Askan 2018, 2021; Ozmen *et al.* 2020; Can *et al.* 2021; Kelam *et al.* 2022). This extensive practical application and validation serve to enhance the reliability and credibility of the simulation methodology used.

Table 1. Source-, path- and site-related input-model parameters.

Parameter	Value				
Region/boundary	Afyon/29°–31° E 37.4°–38.5° N	Duzce/30°–32° E40°–41° N	Erzincan/39°–40° E 39°–40° N	Istanbul/28.8°–29° E 40.9°–41.1° N	Van/42.3°–44° E 37.5°–40° N
Fault mechanism	Normal	Strike-Slip	Strike-Slip	Strike-Slip	Thrust
Strike (°)	271	264	125	81.5	241
Dip (°)	43	64	90	90	51
Rupture velocity ($m\,s^{-1}$)	3700	3700	3500	3500	3700
Rupture velocity/crustal shear wave velocity ($m\,s^{-1}$)	0.9	0.8	0.8	0.8	0.8
Crustal density ($kg\,m^{-3}$)	3000	2800	2800	2800	2800
Slip weight	Random	Random	Random	Random	Random
Pulsing percentage (per cent)	50	30	50	50	50
Windowing function	Saragoni–Hart	Saragoni–Hart	Saragoni–Hart	Saragoni–Hart	Saragoni–Hart
Geometric spreading	R^{-1} if $R \leq 30$ km $R^{-0.5}$ otherwise	R^{-1} if $R \leq 30$ km $R^{-0.4}$ if $30 < R \leq 60$ km $R^{-0.6}$ if $60 < R \leq 90$ km $R^{-0.8}$ if $30 < R \leq 100$ km otherwise $R^{-0.5}$	$R^{-1.1}$ if $R \leq 25$ km $R^{-0.5}$ otherwise	R^{-1} if $R \leq 30$ km $R^{-0.4}$ if $30 < R \leq 60$ km $R^{-0.6}$ if $60 < R \leq 90$ km $R^{-0.8}$ if $30 < R \leq 100$ km otherwise $R^{-0.5}$	R^{-1} if $R \leq 30$ km $R^{-0.4}$ if $30 < R \leq 60$ km $R^{-0.6}$ if $60 < R \leq 90$ km $R^{-0.8}$ if $30 < R \leq 100$ km otherwise $R^{-0.5}$
Duration model	$T_0 + 0.05R$	$T_0 + 0.05R$	$T_0 + 0.05R$	$T_0 + 0.05R$	$T_0 + 0.1R$
Quality factor	$\mathrm{ggad432}_i eqn013.gif 80 f^{0.5}$	$88 f^{0.9}$	$\mathrm{ggad432}_i eqn015.gif 22 f^{0.68}$	$\mathrm{ggad432}_i eqn016.gif 80 f^{0.45}$	$88 f^{0.9}$
Site amplification	NEHRP soil class D	NEHRP soil classes C and D	Generic soil, NEHRP soil classes C and D	Generic soil, NEHRP soil class D	Generic soil, NEHRP soil class D

2.3. Simulation results

The results of simulations led to a total of 7358 acceleration time-series in Afyon, Duzce, Erzincan, Istanbul and Van. In summary, the data set contains scenarios with M_w ranging from 5.0 to 7.5 and R_{JB} values up to 272 km. For Duzce simulations, the M_w values of the scenario events are 5.0, 5.5, 6.0, 6.5, 7.0, 7.1 and 7.5. For Erzincan, M_w values include 5.0, 5.5, 6.0, 6.5, 6.6, 7.0 and 7.5. For Istanbul, M_w values are 5.0, 5.5, 6.0, 6.5, 7.0 and 7.4, while for Afyon, the M_w range includes 5.0, 5.5, 6.0, 6.5, 6.6 and 7.0. Finally, for the scenario events in Van, the M_w values are 5.0, 5.5, 6.0, 6.5, 7.0 and 7.1. For all regions, three distinct soil types characterized by V_{s30} values of 255, 520 and 310 m s^{-1} are considered. These measures of shear wave velocity correspond to soil types C, D and generic soil, respectively (Boore & Joyner 1997). As reported previously, the number of stations corresponds to 90 for Duzce, 365 for Erzincan, 88 for Istanbul, 324 for Afyon and 430 for Van, leading to a total of 1297 stations. The distribution of FMs within the stations corresponds to 40.7 per cent for the strike-slip mechanism, 35.1 per cent for the thrust and 24.2 per cent for the normal FM (see Fig. 3a). In addition, the distribution of simulations with respect to regions is shown in Fig. 3(b). It is noted that the availability and validity of input-model parameters impact the variability in the number of simulations conducted in different regions (Ismet Kanli *et al.* 2006; Ugurhan & Askan 2010; Askan *et al.* 2015; Sahin *et al.* 2016; Akkaya & Özvan 2019).

Furthermore, the histograms portrayed in Fig. 4 reveal the seismological features of the performed simulations. First, the presence of high-magnitude motions (i.e. M_w values from 7.0 up to 7.5) should be noted, covering in this way the lack of real recorded high-magnitude events. An even occurrence of simulations in the range $5.0 \leq M_w \leq 7.0$ is also observed. In the case of distance metrics, the number of synthetic records reduces as the values of R_{JB} sequentially increase. In the case of V_{s30}, the predominance of simulations for soil type C is clear, with fewer occurrences for generic soil and soil type D. The majority of simulations were performed for depth values less than or equal to 10 km, while the rest of them (approximately 20 per cent) were performed for depth values in the range 10 km $<$ FD $<$ 20 km. Finally, the distribution of PGA and PGV metrics regarding R_{JB} values, and for each FM is depicted in Fig. 5. Independently of the fault type, the distribution of PGA and PGV shows higher values for lower R_{JB} and higher values of M_w. This behaviour is coherent with the actual distribution of PGA and PGV with respect to distant metrics and M_w values, which further validates the performance of simulations.

Finally, Fig. 6 depicts samples of the simulated time-series for the regions under analysis, selected as representative examples of large-magnitude events with higher PGA values. Baseline correction and Butterworth filtering in the range of 0.1–25 Hz are applied for the postprocessing of the signals. The earthquake time-series for Afyon represents a scenario of M_w 7.0, V_{s30} of 255 m s^{-1} and R_{JB} of 62.49 km with a resulting PGA of 395.55 cm s^{-2}. The same soil conditions, M_w of 7.1 and R_{JB} of 30.40 km, are assumed for Duzce, with an estimated PGA of 540.01 cm s^{-2}. The M_w for Erzincan simulation is for a scenario event of 7.5, with V_{s30} of 520 m s^{-1} and R_{JB} of 4.81 km resulting in PGA of 913.10 cm s^{-2}. For Istanbul, M_w of 7.4, V_{s30} of 255 m s^{-1} and R_{JB} of 8.16 km are taken, respectively, resulting in a PGA of 746.45 cm s^{-2}. Ultimately, a value of 354.89 cm s^{-2} in terms of PGA is computed by taking M_w, V_{s30} and R_{JB} values of 7.1, 310 m s^{-1} and 122.68 km, respectively.

3 GROUND MOTION MODELLING METHODOLOGY

The prevailing approach for predicting ground motion IMs, like PGA, PGV or PSA, is to employ GMMs. These models are typically developed using empirical methods that entail performing statistical regression analysis on extensive data sets of ground motion intensities (Bindi *et al.* 2014, 2019; Boore *et al.* 2014, 2021; Kale *et al.* 2015; Bommer *et al.* 2016; Bozorgnia & Campbell 2016a, b; Idini *et al.* 2017). Given the considerable variability or dispersion observed in the ground motion data for each IM, GMMs typically offer a probability distribution of potential ground motion results rather than a single deterministic value. In this study, the GMM to be developed has the following form:

$$\ln\left(y_{ij}\right) = \begin{pmatrix} \ln(\mathrm{PGA}) \\ \ln(\mathrm{PGV}) \\ \ln\left(\mathrm{PSA}_{0.03s}\right) \\ \vdots \\ \ln\left(\mathrm{PSA}_{2s}\right) \end{pmatrix} = f\left(M_w, R_{JB}, V_{s30}, \mathrm{FD}, \mathrm{FM}\right) + \eta_i + \varepsilon_{ij}, \quad (5)$$

where $\ln(y_{ij})$ is the natural logarithm of the interested IM, herein PGA, PGV and PSA. The inter-event residual component is denoted as η_i and the intra-event residual component is denoted as ε_{ij} both in the natural logarithm scale. Finally, i represents the index of the earthquake event, and j represents the index of the station. The functional form in eq. (5) is modelled using the ANN algorithm.

The two components of residuals in GMMs, namely inter-event and intra-event residuals, are assumed to be independent and follow a normal distribution with a mean of zero. The inter-event residual component has a standard deviation of τ, while the intra-event residual component has a standard deviation of σ. To calculate the total standard deviation for a given GMM, the square root of the sum of squares of the two components of residuals is taken. This can be expressed mathematically as follows:

$$\phi = \sqrt{\sigma^2 + \tau^2}. \quad (6)$$

Finally, the inter-event error for the *ith* earthquake event can be described as follows:

$$\eta_i = \frac{\tau^2 \sum_1^{n_i} \delta_{ij}}{n_i \tau^2 + \sigma^2} \approx \frac{\sum_1^{n_i} \delta_{ij}}{n_i} \left(n_i \tau^2 >> \sigma^2\right). \quad (7)$$

Since the number of records in each event is rather large and n_i^2 is much larger than 1, the approximate equation can accurately measure the inter-event residuals (Kubo *et al.* 2020). Finally, the intra-event residuals can be obtained by subtracting the inter-event residuals and predicted IMs from the observed ones.

ANNs are intricate networks comprised of interconnected neural computing elements. They possess the capability to receive input stimuli and adapt to their environment through learning. The process of utilizing ANN involves two phases: learning and recall. In the learning phase, known data sets are employed to train the network by adjusting the weights between the input and output layers. Subsequently, during the recall phase, the network applies the acquired weights to process new inputs and make predictions. ANNs have emerged as a well-established and widely utilized tool across various domains (Flood & Kartam 1994; Abiodun *et al.* 2018). Neural network paradigms are characterized by various nomenclatures. In the context of network architecture, a single-layer network comprises individual input and output units, while a multilayer network incorporates one or more hidden units situated between the input and output layers. The backpropagation neural network is a well-known example of a multilayer neural network (Adamowski &

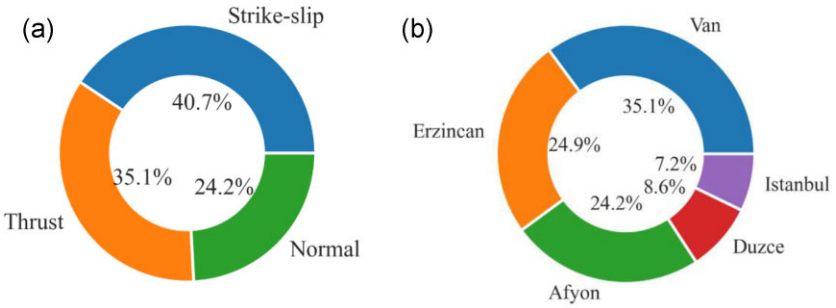

Figure 3. Distribution of simulations regarding (a) FMs and (b) regions.

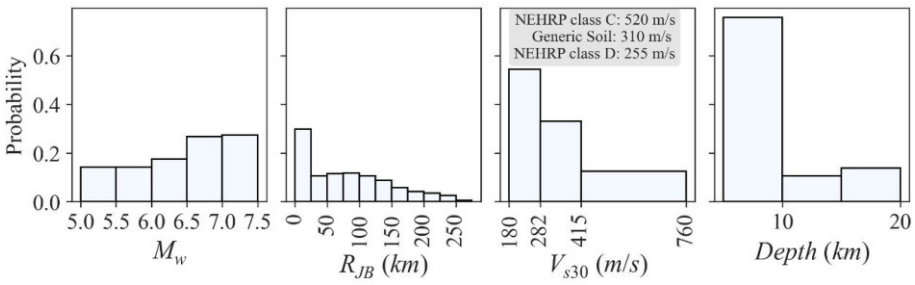

Figure 4. Histograms of seismological features of the simulated Turkish ground motion records.

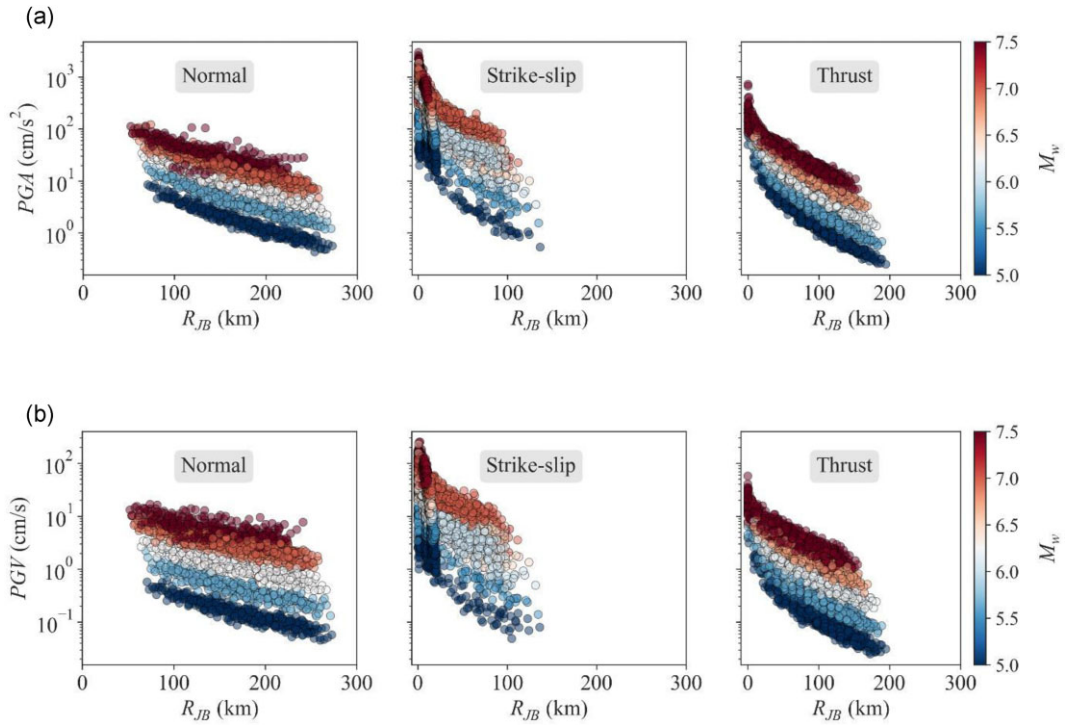

Figure 5. (a) PGA and (b) PGV distributions with respect to the distant metric and faulting mechanism.

Karapataki 2010). It employs a gradient-descent technique to minimize errors during the learning process, thereby facilitating error minimization in the network. ANN attempts to emulate the parallel information processing capability observed in the human brain. The increasing interest in utilizing ANN-based black box models in seismology and earthquake engineering problems can be attributed to their ability to model nonlinear multivariate problems effectively (Möller *et al.* 2009; Dhanya & Raghukanth 2018; Paolucci *et al.* 2018; Khosravikia & Clayton 2021; Kalakonas & Silva 2022a, b; Mohammadi *et al.* 2023). ANNs have demonstrated promising performance in capturing the complex relationships present in such problems, making them valuable tools in earthquake engineering applications.

Here, the Sklearn multilayer perception regressor (Pedregosa *et al.* 2011), a widely recognized and established Python package, is utilized. In conjunction with cross-validation, the random search

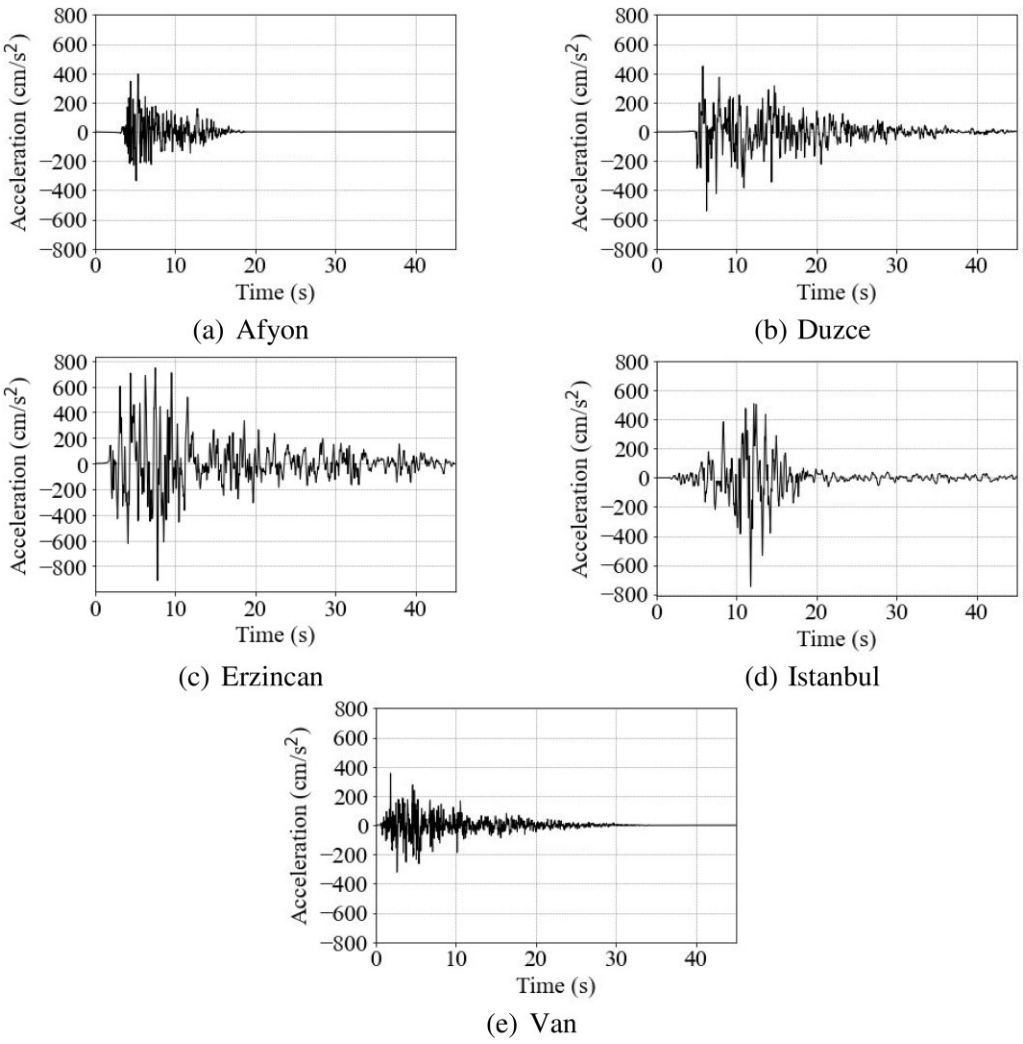

Figure 6. Samples of simulated time-series for the regions under analysis.

technique is employed to determine the optimal configuration for the number of hidden layers, the number of nodes and the activation function within the model. The data set is divided, with 80 per cent of the data allocated for training the model and the remaining 20 per cent used only for model validation. The input parameters utilized in the model include the input-model parameters of this study include M_w, R_{JB}, V_{s30}, FD and FM. The output layer of the model consisted of the prediction of ground motion IMs, namely PGA, PGV and PSA in the period range of 0.03–2.0 s. This study utilizes a one-layer neural network architecture based on the random search optimization technique. The architecture includes an input layer, a hidden layer consisting of 50 neurons, and an output layer. The provided schematic in Fig. 7 illustrates the structural representation of the employed ANN algorithm.

4 RESULTS AND DISCUSSIONS

This section provides an overview of the results obtained from the developed ANN-based GMM. Following that, the developed GMM is subjected to a validation process in which its pattern is compared with all the recorded ground motion data sets from Turkey, encompassing the latest events up to 2023. Specifically, the success of the GMM in estimating the intensity parameters of past real events in Turkey, with a focus on the Afyon, Erzincan, Duzce, Istanbul and Van regions, is evaluated to further assess its performance.

4.1. Performance of ANN-based GMM

In this section, the performance of ANN-based GMM is assessed through a set of statistical metrics, including root-mean-square error (RMSE), coefficient of determination (R^2), Pearson correlation coefficient (r) and mean-absolute-percentage error (MAPE). These metrics provide insight into the model's accuracy, fit, correlation and relative error, allowing for a comprehensive evaluation of its performance. Fig. 8 presents the evaluated metrics for the proposed GMM across all considered IMs, namely, ln(PGA), ln(PGV) and ln(PSA) at periods ranging from 0.03 to 2 s. The values of RMSE, R^2 and r metrics fall within a narrow range, indicating a consistent performance across all IMs without any notable variation for a specific IM. The mean RMSE value is approximately 0.3, slightly increasing toward ln(PSA) at longer periods. In a similar trend, as can be seen in Fig. 8, there is a slight decline in the R^2 and r metrics for ln(PSA) as the periods increase. The mean R^2-value is nearly 0.97, suggesting a robust fit between the model and the data. The mean r-value is

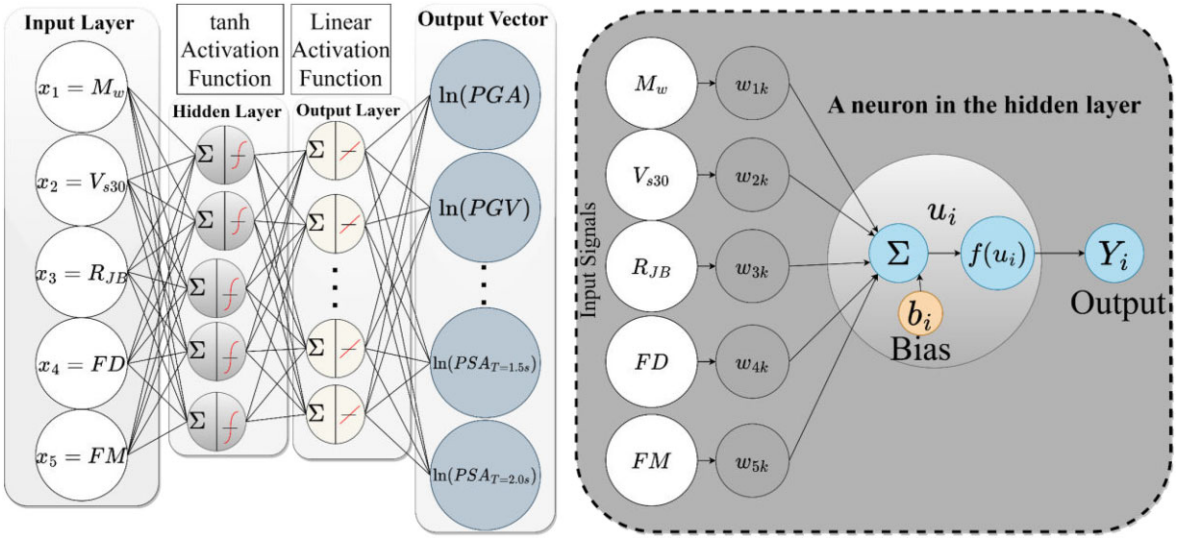

Figure 7. Structure of the ANN model and illustration of artificial neurons of the hidden layer.

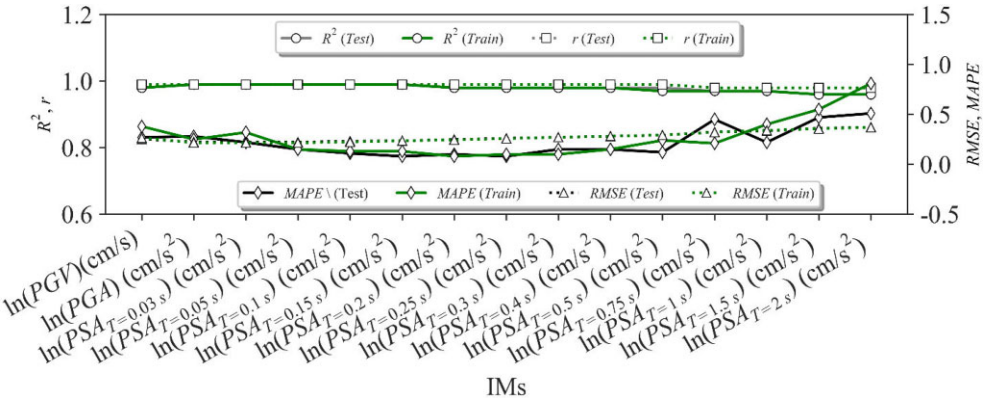

Figure 8. Model performance metrics in terms of RMSE, R^2, r and MAPE for IMs including ln(PGA), ln(PGV) and ln(PSA) at periods ranging from 0.03 to 2 s.

roughly 0.98, indicating a strong linear correlation between the estimated and observed IMs Furthermore, the similarity between the RMSE, R^2 and r values of the train and the test data sets suggests a consistent level of accuracy, fit and correlation across both data sets, respectively. This implies that the model is performing reliably in terms of these metrics regarding the unseen data. On the other hand, the MAPE values exhibit a slightly wide range, indicating varying levels of prediction accuracy across considered IMs. The MAPE values for ln(PSA) within periods of 0.05–0.4 s remain in a tight range with a mean value of roughly 0.1 across the train and test data sets. A similar trend is observed for ln(PGA) with a mean MAPE value of 0.2. However, the MAPE values increase for ln(PGV) and ln(PSA) at periods higher than 0.4 s. Additionally, there is less consistency across the train and test data sets for ln(PGV) and ln(PSA) at long periods. The differences observed between MAPE, and RMSE are an anticipated outcome due to the inherent nature of these metrics, which assess distinct aspects of error. MAPE emphasizes the relative magnitude of errors, whereas RMSE takes into account the overall magnitude of errors. As a result, MAPE tends to be more sensitive to outliers and the scale of the data within the model. However, the agreement between, RMSE, R^2 and r values is higher than that of MAPE, indicating a strong performance of the model.

Subsequently, the model's bias with respect to the input variables, namely, M_w, R_{JB} and V_{s30} is assessed through the analysis of residuals. To this end, the total uncertainty is divided into the inter-event (τ) and intra-event (σ) uncertainties, demonstrating the standard deviation of residuals attributed to the earthquake source and site characteristics, respectively. Fig. 9 shows the distribution of inter-/intra-event and total uncertainties for ln(PGA), ln(PGV) and ln(PSA) at periods of 0.03 s up to 2 s. Overall, the inter-event residual is consistently smaller than the intra-event residual across all IMs. The intra-event residual of ln(PSA) tends to increase at longer periods, leading to higher total uncertainty.

A closer analysis of residuals is conducted by selecting sample IMs, ln(PGA), ln(PGV) and ln(PSA) at periods of 0.2, 0.5, 1.0 and 2.0 s. These IMs are selected to encompass a frequency bandwidth including low , intermediate and high frequencies. Fig. 10 shows the distribution of inter-event residuals with respect to M_w for selected IMs Likewise, Figs 11 and 12 show the distribution of intra-event residuals in relation to R_{JB} and V_{s30} for the same IMs, respectively. The inter-event residuals vary between −0.5 and 0.5, while the intra-event residuals show a broader range between −1.0 and 1.0, consistent with other studies (Akkar *et al.* 2014; Kale *et al.* 2015; Mohammadi *et al.* 2023). The fitted green lines in Figs 10–12 indicate the mean of residuals corresponding to the

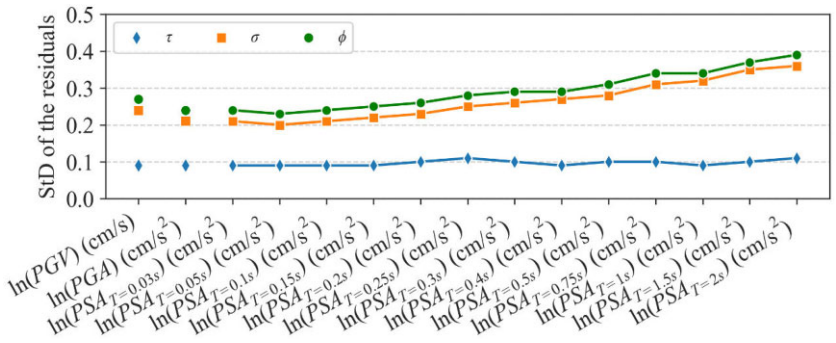

Figure 9. Distribution of the inter-event (τ) and intra-event (σ), and total uncertainties (φ) for IMs including ln(PGA), ln(PGV) and ln(*PSA*) at periods ranging from 0.03 to 2 s.

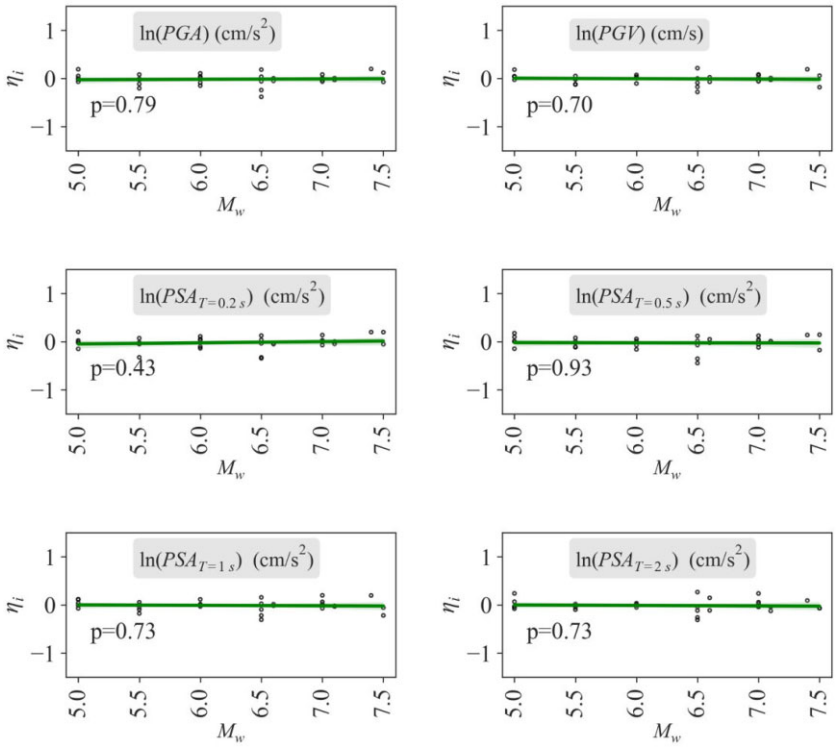

Figure 10. Distribution of the inter-event residual (η) with respect to M_{w} for ln(PGA), ln(PGV) and ln(PSA) at periods of 0.2, 0.5, 1.0 and 2.0 s.

explanatory variables, while the shaded area around these lines represents the 95 per cent confidence interval for the true mean of the residuals. There is no discernible trend in the inter- and intra-event mean residual across all IMs, indicating the unbiasedness of model errors. The tight confidence intervals further support this observation. Additionally, we employ *p*-values at a significance level of 0.05 to examine the null hypothesis regarding the unbiasedness of model errors. *p*-values close to 1.0 imply less bias toward the input parameters. In general, all *p*-values exceed 0.05 by far across all IMs, implying the absence of any trend in the mean residual. Thus, the model does not exhibit systematic bias toward M_{w}, R_{JB} and V_{s30}. However, as can be seen in Fig. 10, the confidence interval of the mean residual is slightly wider as the magnitude of the event increases.

4.2. Validation of the developed ANN-based GMM

To ensure that the proposed GMM captures the characteristics of recorded strong ground motions, the results are assessed across various magnitudes (M_{w}) and distances (R_{JB}) by considering soil class C ($V_{s30} = 520$ m s^{-1}), FM of strike-slip and the mean FD. Fig. 13 illustrates the variation of selected IMs, including PGA, PGV and PSA at periods of 0.2, 0.5, 1 and 2 s with respect to M_{w} for various R_{JB} values of 1, 30 and 70 km. The median of the selected IMs and a range of two standard deviations are considered. Filled and unfilled dots, respectively, represent IMs obtained from real and simulated earthquake events. The results are also compared with a selected parametric GMM developed for Turkey by Kale *et al.* (2015). It is clear that higher M_{w} and lower R_{JB} result in elevated levels of all selected IMs Similarly, Fig. 14 shows the variation of

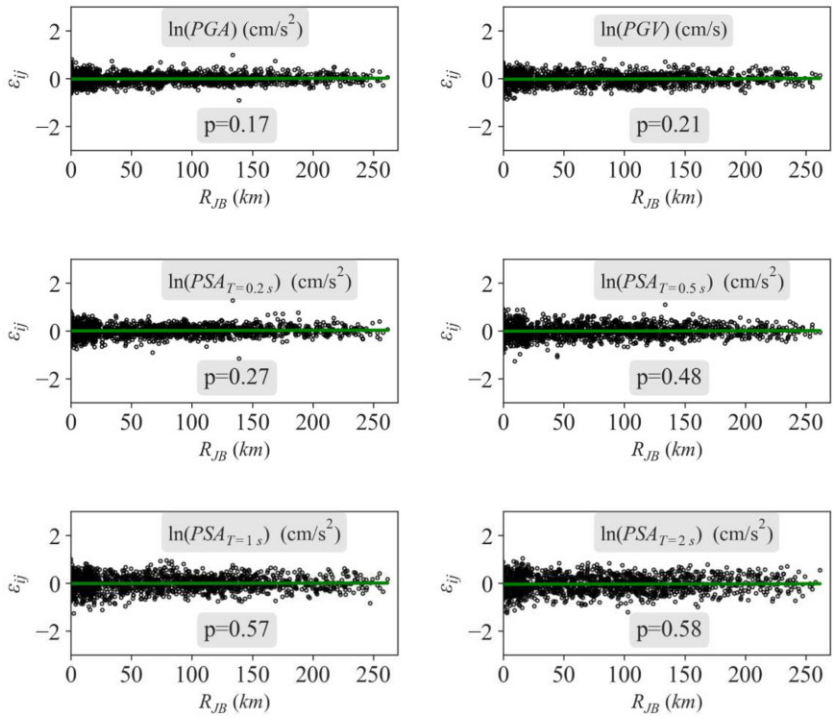

Figure 11. Distribution of the intra-event residual (ε) with respect to R_{JB} for ln(PGA), ln(PGV) and ln(PSA) at periods of 0.2, 0.5, 1.0 and 2.0 s.

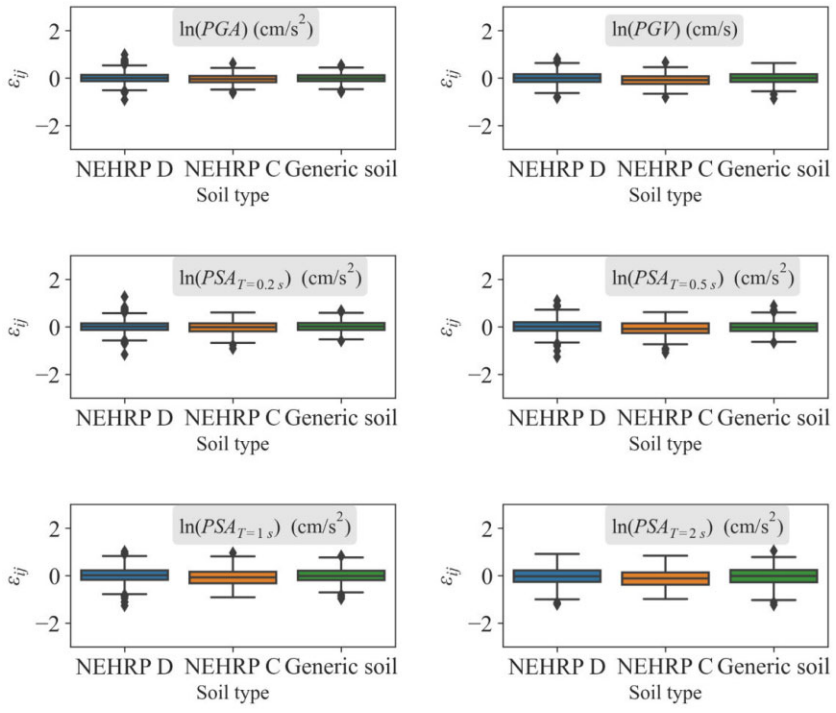

Figure 12. Distribution of the intra-event residual (ε) with respect to V_{s30} for ln(PGA), ln(PGV) and ln(PSA) at periods of 0.2, 0.5, 1.0 and 2.0 s.

the same IMs with respect to R_{JB} for different M_{w} values of 5.5, 6.5 and 7.5. Similarly, the results are compared with the GMM of Kale *et al.* (2015). An increase in R_{JB} leads to a decrease in PGA, PGV and PSA levels at all periods. The proposed GMM effectively captures this distance-dependent attenuation. As shown in Fig. 14, higher M_{w} is associated with higher ground motion amplitudes consistent with the former observation. This is expected as higher M_{w} corresponds to higher energy release during an earthquake. There are multiple real records, including those from the recent Turkey 2023 event, which fall within the considered seismological criteria. As shown in Fig. 14, the proposed GMM effectively captures the sample IMs obtained from these real records, particularly for strong events ($M_{\mathrm{w}} = 7.5$), mostly within two standard deviations.

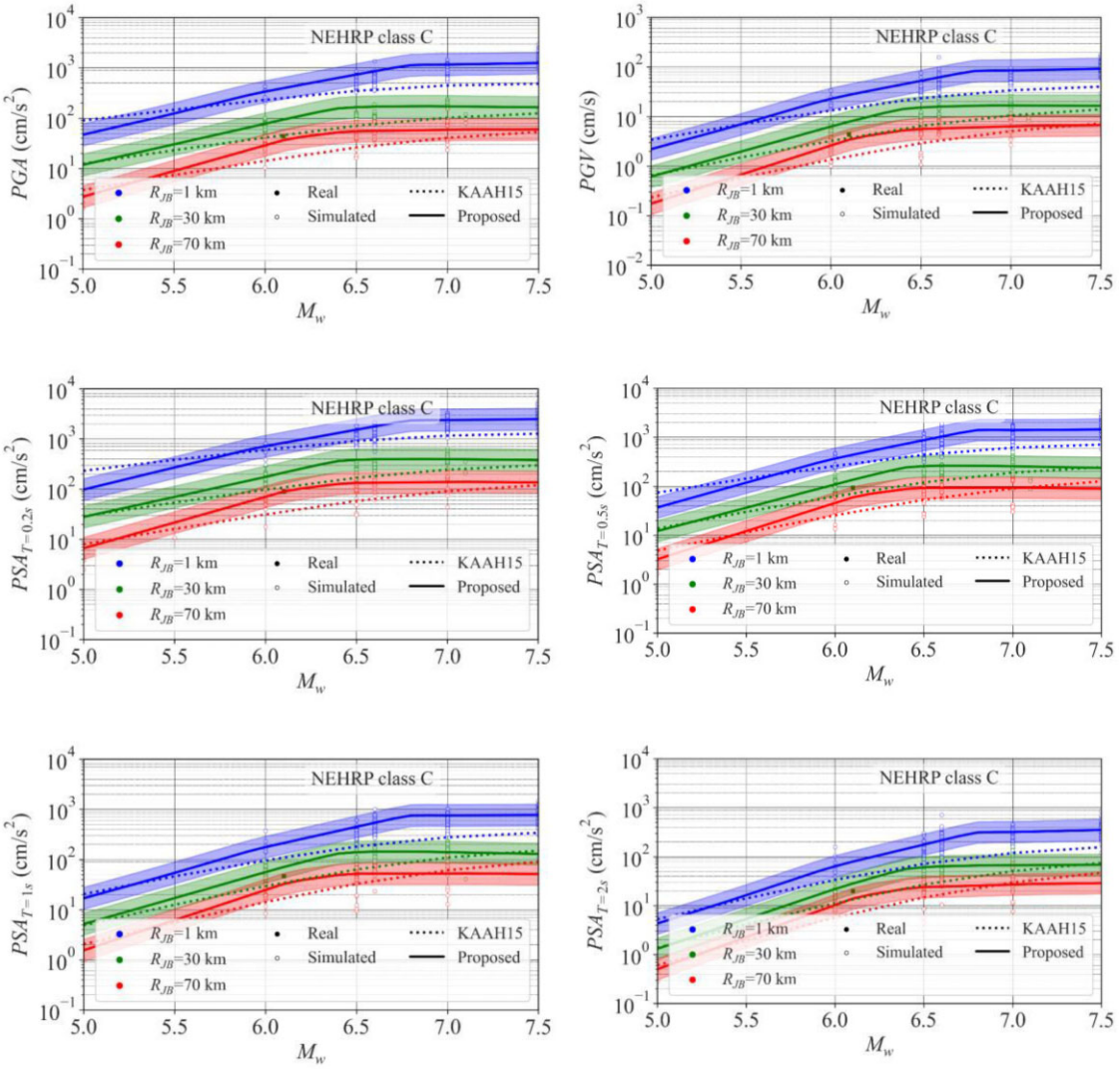

Figure 13. Variation of PGA, PGV and PSA at periods of 0.2, 0.5, 1 and 2 s with respect to M_w for faut mechanism of strike-slip, $V_{s30} = 520$ m s^{-1} and R_{JB} = 1, 30 and 70 km.

Finally, the comparisons show that for large-magnitudes the ANN-based GMM developed by this study performs more effectively than the parametric GMM, especially for the 2023 Kahramanmaraş earthquakes in Turkey. This implies the reliable performance of the proposed GMM regarding unseen data.

The capability of the proposed GMM to capture the geometric and inelastic attenuation is further investigated. To this end, the variation of PSA regarding R_{JB} for soil class C ($V_{s30} = 520$ m s^{-1}), FM of strike-slip and two M_w of 5.0 and 7.5 are illustrated in Fig. 15(a). Results demonstrate that as distance increases, the peak value of PSA decreases and shifts toward longer periods as observed in previous studies (Dhanya & Raghukanth 2018; Mohammadi *et al.* 2023). The seismic energy dissipates as it propagates away from the source due to geometric and inelastic attenuation (Boore 2003), resulting in lower levels of PSA. However, higher frequency contents tend to attenuate faster with distance, leading to a shift of PSA peaks toward longer periods. This shift is affected by M_w and is less for $M_w = 7.5$ compared to 5.0, as illustrated in Fig. 15(a). Similarly, the performance of the proposed GMM in representing the effects of soil is assessed. For this purpose, soil class C, generic soil and soil class D are considered with a representative mean V_{s30} of 520, 310 and 255 m s^{-1}, respectively, according to NEHRP soil classification ((US) & (US) 2001) and Boore & Joyner (1997). Fig. 15(b) depicts the variation of PSA regarding soil classes for R_{JB} =10 km and two M_w of 5.0 and 7.5. The results clearly indicate that when transitioning from stiffer soil (type C) to softer soil (type D), there is a notable increase in PSA level, especially for longer periods. Additionally, the peak of the spectra tends to shift towards longer periods. Results indicate that the magnitude of the earthquake influences the extent of this peak shift, which aligns with the fundamental principles of the earthquake's physics (i.e. the corner frequency is lower for large M_w and thus, large events have enhanced longer periods).

The performance of the developed model is further assessed by analysing its ability to predict the ground motion IMs for real events that occurred in the regions where there are available simulated motions. Specifically, the model's performance is evaluated for the 2002 Afyon ($M_w = 6.6$), 1992 Erzincan ($M_w = 6.6$), 1999 Duzce ($M_w = 7.1$) and 2011 Van ($M_w = 7.1$) earthquakes. Table 2 represents detailed information on these events. Table 3 compares the real ground motion IMs, namely PGA and PGV, with the predicted values obtained from the developed GMM at three selected stations with different seismological information for each event. The

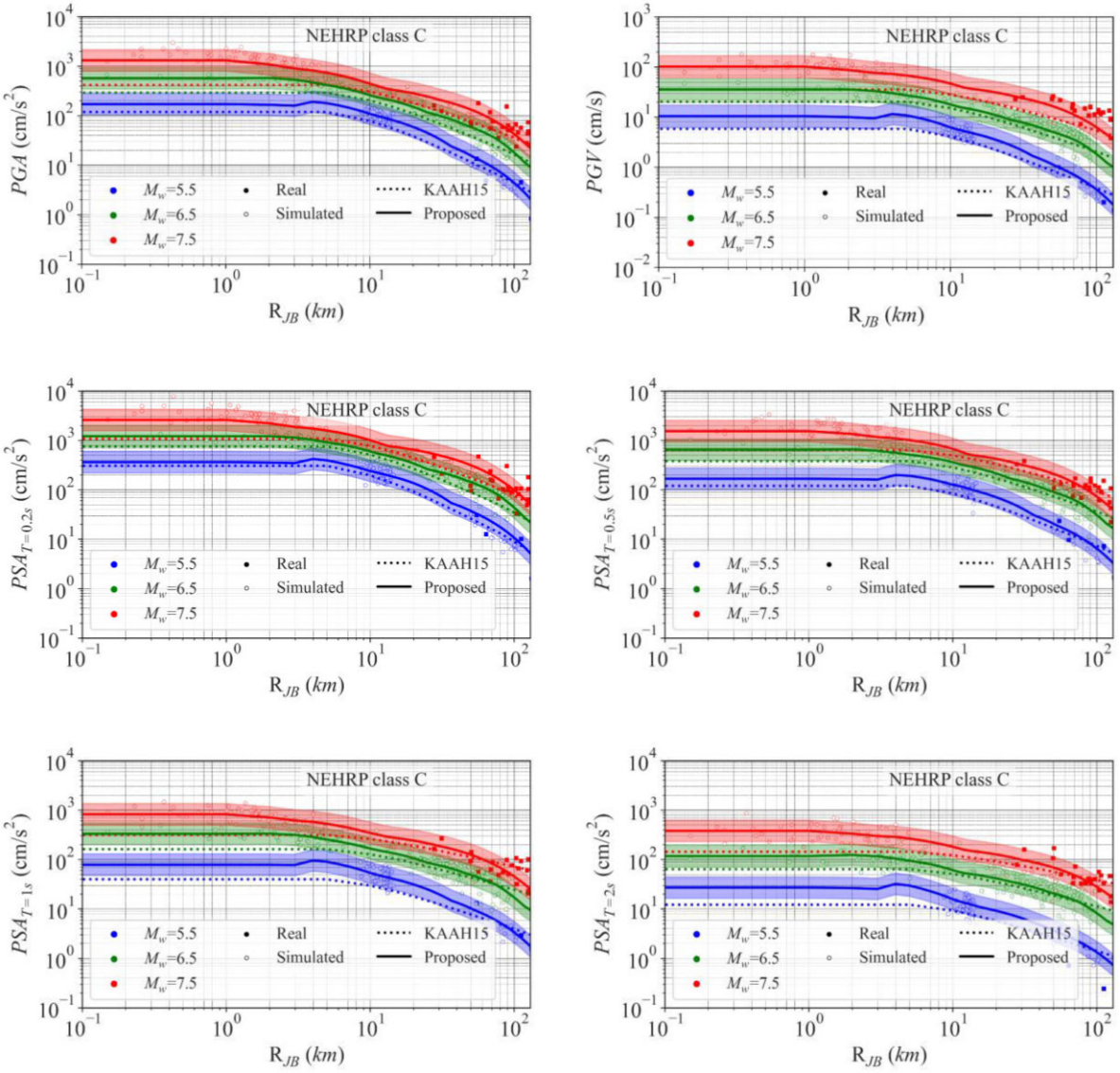

Figure 14. Variation of PGA, PGV and PSA at periods of 0.2, 0.5, 1.0 and 2.0 s with respect to R_{JB} for FM of strike-slip, $V_{s30} = 520\ \mathrm{m\,s^{-1}}$, and $M_{\mathrm{w}} = 5.5$, 6.5 and 7.5.

comparison demonstrates that, for the majority of stations, the estimated values fall within two standard deviations of the actual values for all events, with both underestimation and overestimation of the recorded data. This observation provides additional validation for the proposed model and simulated data set, affirming their suitability in accurately predicting the characteristics of the recorded strong motions.

It is worth noting that the stochastic ground motion simulation method effectively replicates observed ground motions for medium- to high-frequency contents, predominantly due to the influence of seismic wave scattering, as highlighted in Sato *et al.* (2012). However, propagation path effects resulting from deterministic velocity structures become prominent for short-frequency contents. Consequently, the current method's reproducibility diminishes for low-frequency (long-period) content (IMs, including PGV and PSA at 1 and 2 s). While addressing this limitation by integrating deterministic velocity structures through the finite-difference, spectral element, or finite-element methods (Mai *et al.* 2010; Pitarka *et al.* 2022), simulations for long-period ranges would enhance the method's accuracy.

In summary, the developed ANN-based GMM, utilizing homogeneous earthquake data, demonstrates the ability to accurately capture real ground motion attenuation patterns, thus eliminating the need for complex nonlinear regressions with numerous coefficients. However, the traditional approach remains crucial in situations characterized by limited data availability when compared to all machine-learning-based non-parametric GMMs. This limitation is often due to its reliance on established equations rooted in fundamental physical principles. It is noteworthy that the recent earthquakes that occurred in Turkey in 2023 have notably improved the quality and quantity of near-field from large-magnitude events. Therefore, it is imperative that future research endeavours prioritize the incorporation of this high-quality data into machine-learning-based GMMs.

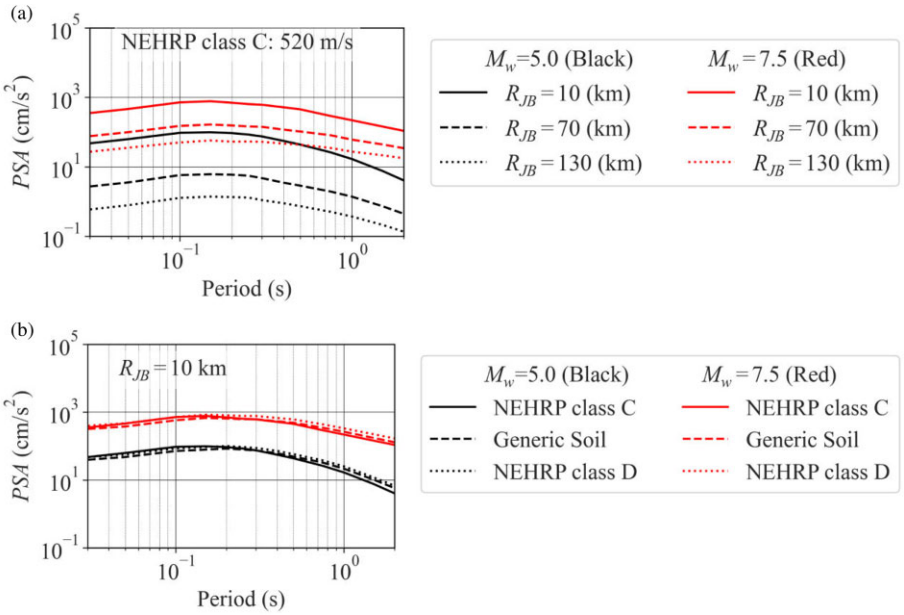

Figure 15. Variation of PSA for different (a) R_{JB} values of 10, 70 and 130 km for soil class C (V_{s30} = 520 m s^{-1}) and (b) soil classes for $R_{\rm JB}$ =10 km based on FM of strike-slip and two $M_{\rm w}$ of 5.0 and 7.5.

Table 2. Detailed information on the past large-magnitude earthquakes within the region

Event	Date	Local time	Epicentre latitude	Epicentre longitude	$M_{\rm w}$	FM	FD
2002 Afyon	3-Feb-2002	7:11	38.6	31.2	6.6	N	9.60
1992 Erzincan	13-Mar-1992	17:18	39.7	39.6	6.6	SS	10.00
1999 Duzce	12-Nov-1999	18:57	40.8	31.2	7.1	SS	11.23
2011 Van	23-Oct-2011	13:41	38.6	43.5	7.1	R	16.00

Table 3. Comparing real PGA and PGV with predicted values from the developed GMM for past earthquakes in Turkey at selected stations

Region	Station	$R_{\rm JB}$(km)	V_{s30} (m s^{-1})	PGA$_{\rm obs}$(cm s^{-2})	PGV$_{\rm obs}$(cm s^{-1})	PGA$_{\rm pred}$ (cm s^{-2})	PGV$_{\rm pred}$ (cm s^{-1})
						$\mu \pm 2\phi$	$\mu \pm 2\phi$
Afyon	301	52	226	108.93	10.42	[39.27–94.28]	[4.02–12.31]
	4302	133	243	23.00	3.37	[14.59–35.17]	[1.87–5.72]
	6401	144	285	6.59	1.38	[14.89–35.90]	[1.95–5.97]
Duzce	1612	102	197	22.03	2.55	[18.48–44.55]	[2.50–7.67]
	8101	0	282	455.98	73.49	[494.66–1192.57]	[40.89–125.31]
	1604	156	459	8.25	1.98	[5.96–14.36]	[0.82–2.50]
Erzincan	2402	5	314	438.34	93.19	[344.56–830.70]	[27.08–83.00]
	2403	62	433	73.35	4.94	[29.45–71.01]	[2.64–8.09]
	2405	58	320	32.30	7.35	[36.47–87.93]	[3.70–11.33]
Van	401	98	295	16.59	4.95	[11.18–26.95]	[1.50–4.60]
	1211	171	463	4.26	1.82	[3.34–8.04]	[0.48–1.47]
	4901	130	315	8.46	2.08	[7.61–18.34]	[1.10–3.38]

5 CONCLUSIONS

In this study, a local GMM for Turkey is developed, utilizing a homogeneous data set created through the simulation of region-specific records. The stochastic finite-fault approach is employed, using the validated input-model parameters for selected regions in Turkey with large past earthquakes, including Afyon, Erzincan, Duzce, Istanbul and Van. To overcome the limitations of traditional linear regression-based models, an ANN is utilized herein to establish the predictive equations and coefficients. The predictive input parameters include FM, FD, $M_{\rm w}$, $R_{\rm JB}$ and V_{s30}. The simulation results include spectral ordinates (PSA) within a specific period range (0.01–2.0 s), PGA and PGV. The uncertainty of the GMM is quantified through the analysis of residuals, providing insights into inter- and intra-event uncertainties. The developed GMM and simulation results are compared with the real data set of Turkey. The following main conclusions are drawn from the analysis conducted in this study:

(i) A homogeneous ground motion data set covering a large range of magnitudes, source-to-site distances and soil classes is developed for different regions in Turkey. The data set consists of 7359 records, covering different FMs including normal, reverse and strike-slip, a range of $M_{\rm w}$ between 5.0 and 7.5, $R_{\rm JB}$ in the range of 0–272 km

and soil classes of C, D and generic soil as proposed in Boore & Joyner (1997). A comparison of the trend of these data sets with the recorded ground motion data set demonstrates the validity of the simulations within the region. Therefore, these data sets are available for engineering use.

(ii) The analysis of residuals yields satisfactory levels of uncertainty across all spectral values. The residuals are further examined to assess inter- and intra-event uncertainties concerning explanatory variables. Specifically, the inter-event residual is examined with respect to magnitude, while the intra-event residual is investigated considering soil and distance information within the data set. The results indicate that the inter-event uncertainty for all spectral values is smaller than the intra-event uncertainty. Additionally, it is observed that neither the inter-event nor the intra-event residuals exhibit significant bias with respect to the input variables. These findings indicate a consistent performance of the GMM with respect to input parameters.

(iii) The proposed GMM effectively captures the physical characteristics observed in real earthquakes regarding the magnitude, source-to-site distance and soil type. In particular, the GMM features distance-dependent attenuation, geometric and anelastic attenuation, and soil amplification effects.

(iv) The capability of the proposed GMM in estimating the ground motion amplitudes even for large-magnitude events, including the most recent ones in Turkey in 2023, is confirmed by comparing results with the unseen data from Turkish real record data sets.

(v) The model's effectiveness is further verified by comparing the predicted ground motion parameters with observed values recorded during previous events in the region. Overall, the research validates the suitability of the proposed model and simulated data set in accurately simulating seismic phenomena in Turkey. The utilization of an ANN-based GMM offers a notable advantage in comprehensively capturing the intricate and nonlinear attributes inherent in ground motion data sets, in contrast to the parametric GMMs. This advancement holds particular significance, as the ANN-based model alleviates the stringent limitations imposed by conventional GMMs in terms of prescribed functional forms and the determination of unknown coefficients. Furthermore, the existing parametric GMM exhibit a constraint in adequately representing intensity levels, especially for events with large magnitudes. This deficiency is addressed through the proposed model introduced in this research.

Finally, the validity of the proposed model is restricted to the examined regions and other areas with similar tectonic characteristics. It is essential to emphasize that the model's applicability is solely limited to the FMs, FDs, magnitudes, distances and soil conditions considered in this study. Additionally, using a stochastic finite-fault simulation technique enhances the model's ability to reproduce medium- to high-frequency ground motion records accurately. Yet, incorporating more accurate frequency- and distance-dependent radiation pattern models can augment the fidelity and precision that characterize the outcomes of ground motion simulation endeavours, particularly in enhancing low-frequency content representation. In addition, deterministic wave propagation studies using numerical models such as finite-element, finite-difference and spectral element are imperative for a more precise simulation of the low-frequency contents. Thus, to improve the accuracy and overcome limitations of the proposed model, especially in different regions, future research should focus on implementing the suggested ANN-based GMM in other tectonic zones while also addressing the need to enhance the representation of low-frequency content. These improvements can be achieved by constructing region-specific simulated data sets and employing hybrid ground motion simulation approaches. Moreover, the continuous enhancement in data quality and quantity, as evident from events such as the 2023 February 6 Kahramanmaraş earthquakes in Turkey, highlights the pressing necessity for future research to prioritize the initial integration of this superior data into synthetic data sets and, in the following stages, into machine-learning-based GMMs. These efforts will facilitate advancements in accuracy and expand the applicability of the proposed model across various geological settings.

ACKNOWLEDGMENTS

This work was partly financed by FCT/MCTES through National funds (PIDDAC) under the R&D Unit Institute for Sustainability and Innovation in Structural Engineering (ISISE), under reference UIDB/04029/2020, and under the Associate Laboratory Advanced Production and Intelligent Systems ARISE under reference LA/P/0112/2020. This study has been partly funded by the STAND4HERITAGE project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 833123), as an advanced grant. This work is financed by national funds through FCT—Foundation for Science and Technology, under grant agreement 2020.08876.BD attributed to the second author. This work is financed by national funds through FCT—Foundation for Science and Technology, under grant agreement UI/BD/153379/2022 attributed to the third author. *Shaghayegh Karimzadeh*: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualisation, Writing—original draft, Writing—review & editing. *Amirhossein Mohammadi*: Formal analysis, Investigation, Methodology, Resources, Visualisation, Writing—original draft, Writing—review & editing. *Sayed Mohammad Sajad Hussaini*: Formal analysis, Investigation, Writing—original draft, Writing—review & editing. *Daniel Caicedo*: Formal analysis, Investigation, Writing—original draft, Writing—review & editing. *Aysegul Askan*: Data curation, Resources, Writing—review & editing. *Paulo B. Lourenço:* Funding acquisition, Resources, Supervision, Writing—review & editing.

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. & Arshad, H., 2018. State-of-the-art in artificial neural network applications: a survey. *Heliyon,* **4,** e00938, Elsevier. https://doi.org/10.1016/j.heliyon.2018.e00938.

Adamowski, J. & Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *J. Hydrol. Eng.,* **15,** 729–743, American Society of Civil Engineers.

Aki, K., 1967. Scaling law of seismic spectrum. *J. geophys. Res.,* **72,** 1217–1231.

Akkar, S., Sandıkkaya, M.A. & Bommer, J.J., 2014. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull. Earthq. Eng.,* **12,** 359–387.

Akkaya, İ. & Özvan, A., 2019. Site characterization in the Van settlement (Eastern Turkey) using surface waves and HVSR microtremor methods. *J. appl. Geophys.,* **160,** 157–170.

Ancheta, T.D. *et al.*, 2014. NGA-West2 database. *Earthq. Spectra,* **30,** 989–1005.

Arslan Kelam, A. *et al.*, 2022. An evaluation of seismic hazard and potential damage in Gaziantep, Turkey using site specific models for sources, velocity structure and building stock. *Soil Dyn. Earthq. Eng.,* **154**. 107129. https://doi.org/10.1016/j.soildyn.2021.107129.

Askan, A., Karimzadeh, S., Asten, M., Kiliç, N., Şişman, F.N. & Erkmen, C., 2015. Assessment of seismic hazard in the erzincan (Turkey) region: construction of local velocity models and evaluation of potential ground motions. *Turkish J. Earth Sci.,* **24,** 529–565.

Askan, A., Sisman, F.N. & Ugurhan, B., 2013. Stochastic strong ground motion simulations in sparsely-monitored regions: a validation and sensitivity study on the 13 March 1992 Erzincan (Turkey) earthquake. *Soil Dyn. Earthq. Eng.,* **55,** 170–181.

Atkinson, G.M. & Boore, D.M., 1995. Ground-motion relations for eastern North America. *Bull. seism. Soc. Am.,* **85,** 17–30.

Beresnev, I.A. & Atkinson, G.M., 1997. Modeling finite-fault radiation from the ωn spectrum. *Bull. seism. Soc. Am.,* **87,** 67–84.

Bindi, D., Kotha, S.R., Weatherill, G., Lanzano, G., Luzi, L. & Cotton, F., 2019. The pan-European engineering strong motion (ESM) flatfile: consistency check via residual analysis. *Bull. Earthq. Eng.,* **17,** 583–602.

Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Puglia, R. & Augliera, P., 2014. Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset. *Bull. Earthq. Eng.,* **12,** 391–430.

Bommer, J.J., Dost, B., Edwards, B., Stafford, P.J., Elk, J.v., Doornhof, D. & Ntinalexis, M., 2016. Developing an application-specific ground-motion model for induced seismicity. *Bull. seism. Soc. Am.,* **106,** 158–173.

Boore, D.M., 1983. Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bull. seism. Soc. Am.,* **73**. https://doi.org/10.1785/BSSA07306A1865.

Boore, D.M., 2003. Simulation of ground motion using the stochastic method. *Pure appl. Geophys.,* **160,** 635–676, Springer.

Boore, D.M. & Atkinson, G.M., 2008. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthq. Spectra,* **24,** 99–138.

Boore, D.M., *et al.*, 2021. A ground-motion prediction model for shallow crustal earthquakes in Greece. *Bull. seism. Soc. Am.,* **111,** 857–874.

Boore, D.M. & Joyner, W.B., 1997. Site amplifications for generic rock sites. *Bull. seism. Soc. Am.,* **87,** 327–341.

Boore, D.M., Stewart, J.P., Seyhan, E. & Atkinson, G.M., 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq. Spectra,* **30,** 1057–1085.

Bozorgnia, Y. & Campbell, K.W., 2016a. Ground motion model for the vertical-to-horizontal (V/H) ratios of PGA, PGV, and response spectra. *Earthq. Spectra,* **32,** 951–978.

Bozorgnia, Y. & Campbell, K.W., 2016b. Vertical ground motion model for PGA, PGV, and linear response spectra using the NGA-West2 database. *Earthq. Spectra,* **32,** 979–1004.

BRUNE, J.N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. geophys. Res.,* **75,** 4997–5009.

Campbell, K.W., 2003. Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America. *Bull. seism. Soc. Am.,* **93,** 1012–1033.

Campbell, K.W. & Bozorgnia, Y., 2012. A comparison of ground motion prediction equations for arias intensity and cumulative absolute velocity developed using a consistent database and functional form. *Earthq. Spectra,* **28,** 931–941.

Campbell, K.W. & Bozorgnia, Y., 2014. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq. Spectra,* **30,** 1087–1115.

Can, G., Askan, A. & Karimzadeh, S., 2021. An assessment of the 3 February 2002 Cay (Turkey) earthquake (Mw=6.6): modeling of ground motions and felt intensity distribution. *Soil Dyn. Earthq. Eng.,* **150,** 106832.

Cheloni, D., and & Akinci, A. 2020. Source modelling and strong ground motion simulations for the 24 January 2020, M w 6.8 Elazığ earthquake, Turkey, *Geophysical Journal International,* **223,** 1054–1068.

Dhanya, J. & Raghukanth, S.T.G., 2018. Ground motion prediction model using artificial neural network. *Pure appl. Geophys.,* **175,** 1035–1064.

Fahlman, S.E. & Lebiere, C., 1990. The cascade-correlation learning architecture. edited by Jordan, Michael I., LeCun, Yann, and & Solla, Sara A.. *Advances in Neural Information Processing,* Vol., **2**. mitpress.

Fayaz, J., Dabaghi, M. & Zareian, F., 2020. Utilization of site-based simulated ground motions for hazard-targeted seismic demand estimation: application for ordinary bridges in Southern California. *J. Bridge Eng.,* **25**. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001634.

Flood, I. & Kartam, N., 1994. Neural networks in civil engineering. II: systems and application. *J. Comput. Civil Eng.,* **8,** 149–162, American Society of Civil Engineers.

Gianniotis, N., Kuehn, N. & Scherbaum, F., 2014. Manifold aligned ground motion prediction equations for regional datasets. *Comput. Geosci.,* **69,** 72–77.

Hanks, T.C. & McGuire, R.K., 1981. The character of high-frequency strong ground motion. *Bull. seism. Soc. Am.,* **71,** 2071–2095.

Hisada, Y., 1994. An efficient method for computing Green's functions for a layered half-space with sources and receivers at close depths. *Bull. seism. Soc. Am.,* **84,** 1456–1472.

Idini, B., Rojas, F., Ruiz, S. & Pastén, C., 2017. Ground motion prediction equations for the Chilean subduction zone. *Bull. Earthq. Eng.,* **15,** 1853–1880.

Ismet Kanli, A., Tildy, P., Prónay, Z., Pinar, A. & Hermann, L., 2006. Vs30 mapping and soil classification for seismic site effect evaluation in Dinar region, SW Turkey. *Geophys. J. Int.,* **165,** 223–235.

Kalakonas, P. & Silva, V., 2022. Earthquake scenarios for building portfolios using artificial neural networks: part II—damage and loss assessment. *Bull. Earthq. Eng.,* https://doi.org/10.1007/s10518-022-01599-2.

Kalakonas, P. & Silva, V., 2022. Seismic vulnerability modelling of building portfolios using artificial neural networks. *Earthq. Eng. Struct. Dyn.,* **51,** 310–327.

Kale, Ö., Akkar, S., Ansari, A. & Hamzehloo, H., 2015. A ground-motion predictive model for iran and turkey for horizontal PGA, PGV, and 5% damped response spectrum: investigation of possible regional effects. *Bull. seism. Soc. Am.,* **105,** 963–980.

Karimzadeh, S., 2019. Seismological and engineering demand misfits for evaluating simulated ground motion records. *Appl. Sci. (Switzerland),* **9,** 4497. https://doi.org/10.3390/app9214497.

Karimzadeh, S. & Askan, A., 2018. Modeling of a historical earthquake in Erzincan, Turkey (Ms∼7.8, in 1939) using regional seismological information obtained from a recent event. *Acta Geophys.,* **66,** 293–304.

Karimzadeh, S. & Askan, A., 2021. Collection of microseismic intensity data: a model for Turkey. *Arab. J. Geosci.,* **14**. https://doi.org/10.1007/s12517-021-06812-1.

Karimzadeh, S., Askan, A. & Yakut, A., 2017a. Assessment of simulated ground motions in Earthquake engineering practice: a case study for Duzce (Turkey). *Pure appl. Geophys.,* **174,** 3589–3607.

Karimzadeh, S., Askan, A., Yakut, A. & Ameri, G., 2017b. Assessment of alternative simulation techniques in nonlinear time history analyses of multi-story frame buildings: a case study. *Soil Dyn. Earthq. Eng.,* **98,** 38–53.

Karimzadeh, S., Hussaini, S.M.S., Funari, M.F. & Lourenço, P.B., 2021a. On the effect of different code-based ground motion selection approaches for the estimation of the seismic demand of masonry structures by using real ground motion data set. https://doi.org/10.1002/essoar.10509375.1.

Karimzadeh, S., Kadas, K., Askan, A. & Yakut, A., 2021b. Comparison of real and simulated records using ground motion intensity measures. *Soil Dyn. Earthq. Eng.,* **147,** 106796. https://doi.org/10.1016/j.soildyn.2021.106796.

Karimzadeh, S., Kadasa, K., Askanb, A., Erberikb, M.A. & Yakutb, A., 2020. Derivation of analytical fragility curves using SDOF models of

masonry structures in Erzincan (Turkey). *Earthq. Struct.,* **18**. https://doi.org/10.12989/eas.2020.18.2.249.

Karimzadeh, S. & Lourenço, P.B., 2022. Stochastic ground motion simulation of the 9th of July 1998 Faial Earthquake (Azores, North Atlantic). *Authorea Preprints*, Authorea. https://doi.org/10.1002/essoar.10508875.3.

Karimzadeh, S., Ozsarac, V., Askan, A. & Erberik, M.A., 2019. Use of simulated ground motions for the evaluation of energy response of simple structural systems. *Soil Dyn. Earthq. Eng.,* **123,** 525–542.

Khosravikia, F. & Clayton, P., 2021. Machine learning in ground motion prediction. *Comput. Geosci.,* **148,** 104700, Elsevier. https://doi.org/10.1016/j.cageo.2021.104700.

Koboevic, S., Guilini-Charrette, K., Castonguay, P.X. & Tremblay, R., 2011. Selection and scaling of NBCC 2005 compatible simulated ground motions for nonlinear seismic analysis of low-rise steel building structures. *Can. J. Civ. Eng.,* **38**. https://doi.org/10.1139/l11-094.

Kotha, S.R., Cotton, F. & Bindi, D., 2019. Empirical models of shear-wave radiation pattern derived from large datasets of ground-shaking observations. *Sci. Rep.,* **9,** 981, Nature Publishing Group UK London. https://doi.org/10.1038/s41598-018-37524-4.

Kubo, H., Kunugi, T., Suzuki, W., Suzuki, S. & Aoi, S., 2020. Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Sci. Rep.,* **10**. https://doi.org/10.1038/s41598-020-68630-x.

Mai, P.M., Imperatori, W. & Olsen, K.B., 2010. Hybrid broadband ground-motion simulations: combining long-period deterministic synthetics with high-frequency multiple S-to-S backscattering. *Bull. seism. Soc. Am.,* **100,** 2124–2142.

Meenakshi, Y., Vemula, S., Alne, A. & Raghukanth, S.T.G., 2023. Ground motion model for Peninsular India using an artificial neural network. *Earthq. Spectra,* **39,** 596–633.

Megawati, K., Pan, T.-C. & Koketsu, K., 2005. Response spectral attenuation relationships for Sumatran-subduction earthquakes and the seismic hazard implications to Singapore and Kuala Lumpur. *Soil Dyn. Earthq. Eng.,* **25,** 11–25, Elsevier.

Mohammadi, A., Karimzadeh, S., Banimahd, S.A., Ozsarac, V. & Lourenço, P.B., 2023. The potential of region-specific machine-learning-based ground motion models: application to Turkey. *Soil Dyn. Earthq. Eng.,* **172,** 108008, Elsevier. https://doi.org/10.1016/j.soildyn.2023.108008.

Mohammadioun, B. & Serva, L., 2001. Stress drop, slip type, earthquake magnitude, and seismic hazard. *Bull. seism. Soc. Am.,* **91,** 694–707.

Möller, O., Foschi, R.O., Quiroz, L.M. & Rubinstein, M., 2009. Structural optimization for performance-based design in earthquake engineering: applications of neural networks. *Struct. Saf.,* **31,** 490–499.

Motazedian, D. & Atkinson, G.M., 2005. Stochastic finite-fault modeling based on a dynamic corner frequency. *Bull. seism. Soc. Am.,* **95,** 995–1010.

NEHRP soil classification (US) & (US), 2001. *NEHRP Recommended Provisions (National Earthquake Hazards Reduction Program) for Seismic Regulations for New Buildings and Other Structures,* Building Seismic Safety Council.

Ozmen, E., Karimzadeh, S. & Askan, A., 2020. Broadband ground motion simulation within the City of Düzce (Turkey) and building response simulation. *Pure appl. Geophys.,* **177,** 2353–2373.

Paolucci, R., Gatti, F., Infantino, M., Smerzini, C., Özcebe, A.G. & Stupazzini, M., 2018. Broadband ground motions from 3D physics-based numerical simulations using artificial neural NetworksBroadband ground motions from 3D PBSs using ANNs. *Bull. seism. Soc. Am.,* **108,** 1272–1286, GeoScienceWorld.

Pedregosa, F. *et al.*, 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.,* **12**. 2825–2830.

Pitarka, A., Akinci, A., Gori, P.D. & Buttinelli, M., 2022. Deterministic 3D ground-motion simulations (0–5 Hz) and surface topography effects of the 30 October 2016 Mw 6.5 Norcia, Italy, earthquake. *Bull. seism. Soc. Am.,* **112,** 262–286.

Raghucharan, M.C., Somala, S.N., Erteleva, O. & Rogozhi, E., 2021. Seismic attenuation model for data gap regions using recorded and simulated ground motions. *Nat. Hazards,* **107,** 423–446.

Sahin, A., Sisman, R., Askan, A. & Hori, M., 2016. Development of integrated earthquake simulation system for Istanbul. *Earth Planets Space,* **68**. https://doi.org/10.1186/s40623-016-0497-y.

Sato, H., Fehler, M.C. & Maeda, T., 2012. *Seismic Wave Propagation and Scattering in the Heterogeneous Earth,* Springer Science & Business Media.

Şeşetyan, K. *et al.*, 2018. The 2014 seismic hazard model of the Middle East: overview and results. *Bull. Earthq. Eng.,* **16,** 3535–3566.

Sreenath, V., Podili, B. & Raghukanth, S.T.G., 2023. A hybrid non-parametric ground motion model for shallow crustal earthquakes in Europe. *Earthq. Eng. Struct. Dyn.,* **52,** 2303–2322.

Takemura, S., Kobayashi, M. & Yoshimoto, K., 2016. Prediction of maximum P-and S-wave amplitude distributions incorporating frequency-and distance-dependent characteristics of the observed apparent radiation patterns. *Earth Planets Space,* **68,** 1–9.

Uckan, E., Umut, Ö., Sisman, F.N., Karimzadeh, S. & Askan, A., 2018. Seismic response of base isolated liquid storage tanks to real and simulated near fault pulse type ground motions. *Soil Dyn. Earthq. Eng.,* **112,** 58–68.

Ugurhan, B. & Askan, A., 2010. Stochastic strong ground motion simulation of the 12 november 1999 Düzce (Turkey) earthquake using a dynamic corner frequency approach. *Bull. seism. Soc. Am.,* **100,** 1498–1512.

Ugurhan, B., Askan, A., Akinci, A. & Malagnini, L., 2012. Strong-ground-motion simulation of the 6 April 2009 L'Aquila, Italy, earthquake. *Bull. seism. Soc. Am.,* **102,** 1429–1445.

Utkucu, M., Nalbant, S.S., McCloskey, J., Steacy, S. & Alptekin, Ö., 2003. Slip distribution and stress changes associated with the 1999 November 12, Düzce (Turkey) earthquake (Mw = 7.1). *Geophys. J. Int.,* **153,** 229–241.

Wang, T., Xie, X. & Ji, L., 2021. Stochastic finite-fault simulation of the MS 7.0 Lushan earthquake based on frequency-and distance-dependent radiation patterns. *Bull. seism. Soc. Am.,* **111,** 3387–3402.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. seism. Soc. Am.,* **84,** 974–1002.

Wiszniowski, J., 2019. Estimation of a ground motion model for induced events by Fahlman's Cascade Correlation Neural Network. *Comput. Geosci.,* **131,** 23–31.

Withers, K.B., Moschetti, M.P. & Thompson, E.M., 2020. A machine learning approach to developing ground motion models from simulated ground motions. *Geophys. Res. Lett.,* **47**. https://doi.org/10.1029/2019GL086690.

Yerlikaya-Özkurt, F., Askan, A. & Weber, G.W., 2014. An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method. *Eng. Optim.,* **46,** 1651–1668.

Zengin, E. & Cakti, E., 2014. Ground motion simulations for the 23 October 2011 Van, Eastern Turkey earthquake using stochastic finite fault approach. *Bull. Earthq. Eng.,* **12,** 627–646.

Zhang, L., Chen, G., Wu, Y. & Jiang, H., 2016. Stochastic ground-motion simulations for the 2016 Kumamoto, Japan, earthquake. *Earth Planets Space,* **68**. https://doi.org/10.1186/s40623-016-0565-3.

Zonno, G., Oliveira, C.S., Ferreira, M.A., Musacchio, G., Meroni, F., Mota-de-Sá, F. & Neves, F., 2010. Assessing seismic damage through stochastic simulation of ground shaking: the case of the 1998 Faial earthquake (Azores Islands). *Surv. Geophys.,* **31,** 361–381.

APPENDIX A. A WEB-BASED APPLICATION SOFTWARE

This study uses Streamlit to build a graphical user interface tool that provides easy access to the GMM developed by ANN. The code is available at https://github.com/amirxdbx/GMM_Turkey_simulation. The interface of the tool is available at https://amirxdbx-gmm-turkey-simulation-deploy.streamlit.app/. The user needs to define the characteristics of a scenario earthquake in terms of M_w, R_{JB}, V_{s30} and FM. The outcome of the software is given in terms of IMs, including PGA, PGV and PSA.