This is the pre-peer reviewed version of the following article: Karimzadeh, S., Mohammadi, A., Salahuddin, U., Carvalho, A. and Lourenço, P.B., 2024. Backbone ground motion model through simulated records and XGBoost machine learning algorithm: An application for the Azores plateau (Portugal). Earthquake Engineering & Structural Dynamics, 53(2), pp.668-693, which has been published in final form at https://doi.org/10.1002/eqe.4110. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Backbone Ground Motion Model Through Simulated Records and XGBoost Machine Learning Algorithm: An Application for the Azores Plateau (Portugal)

Shaghayegh Karimzadeh[1*], Amirhossein Mohammadi[2], Usman Salahuddin[3], Alexandra Carvalho[4], Paulo B. Lourenço[5]

[1*] Post-Doctoral Researcher, Department of Civil Engineering, ISISE, ARISE, University of Minho, Campus de Azurém, 4800 - 058, Guimarães, Portugal, shaghkn@civil.uminho.pt, Corresponding Author
[2] PhD candidate, Department of Civil Engineering, ISISE, ARISE, University of Minho, Campus de Azurém, 4800-058, Guimarães, Portugal, amir.xdbx@gmail.com
[3] Post-Graduate Student, Department of Civil Engineering, 62032, Camerino, Italy, University of Camerino, busman.salahuddin@studenti.unicam.it
[4] Assistant Researcher, LNEC, Structural Department 1700-111, Lisbon, Portugal, xanacarvalho@lnec.pt
[5] Professor, Department of Civil Engineering, ISISE, ARISE, University of Minho, Campus de Azurém, 4800 - 058, Guimarães, Portugal, pbl@civil.uminho.pt

Abstract

Azores Islands are seismically active due to the tectonic structure of the region. Since the 15th century, they have been periodically shaken by approximately 33 moderate to strong earthquakes, with the most recent one in 1998 (M_w =6.2). Nonetheless, due to insufficient instrumental seismic data, the region lacks a uniform database of past real records. Ground motion simulation techniques provide alternative region-specific time series of prospective events for locations with limited seismic networks or regions with a seismic gap of catastrophic earthquake events. This research establishes a local ground motion model (GMM) for the Azores plateau (Portugal) by simulating region-specific records for constructing a homogeneous dataset. Simulations are accomplished in bedrock using the stochastic finite-fault approach by employing validated input-model parameters. The simulation outcomes are compared with global empirical models, as well as a recently developed model specifically designed for the Pan-European dataset. A probabilistic numerical technique, namely the Monte-Carlo simulation, is employed to estimate the outcome of uncertainty associated with these parameters. The results of the simulations are post-processed to predict the peak ground motion parameters in addition to spectral ordinates. This study uses XGBoost to circumvent the difficulties inherent to linear regression-based models in establishing the form of equations and coefficients. The input parameters for prediction are moment magnitude, Joyner and Boore distance, and focal depth. The quantification of GMM uncertainty is accomplished by analysing the residuals, which provides insight into both inter- and intra- event uncertainties. The outcomes demonstrate the effectiveness of the suggested model in simulating physical phenomena.

Keywords: Stochastic finite-fault ground motion simulation, Ground motion model (GMM), XGBoost, Azores plateau (Portugal)

1 Introduction

Throughout history, earthquakes have been the leading cause of human casualties from natural hazards, resulting in significant economic losses, especially in regions with high seismic activity. Despite representing only 3% of people affected by natural disasters, they account for 58% of all disaster-related fatalities and 21% of all known economic losses [1]. Global population exposure to an earthquake of moderate to severe intensity has increased by 93% during the last 40 years [2]. Exposure is expected to expand even further with population growth and urbanisation. Evaluation of seismic hazard is the inherent, unavoidable, component of risk mitigation studies, particularly in earthquake-prone zones. Seismic hazard analysis can be accomplished through either deterministic or probabilistic approaches [3]. Ground motion models (GMMs), which estimate seismic intensity measures (IMs) for different scenario events, are essential for seismic hazard analysis. To date, there are over 485 empirical GMMs developed globally for Peak Ground Acceleration (PGA), 316 models for spectral ordinates, and 18 backbone GMMs in the literature [4,5]. These models frequently employ seismological parameters, such as earthquake magnitude, fault mechanism, source-to-site distance, and site conditions, as their primary explanatory variables.

In regions with adequate seismic networks and historical event data, recorded ground motions can be utilised to develop GMMs. If such data is unavailable, ground motion Intensity Measures (IMs) can be estimated through the use of simulated databases generated by various empirical, numerical, and analytical methodologies [6]. The simulated dataset can better represent the regional seismicity of the study areas [7–12]. Douglas [5] lists 87 GMMs developed based on simulated records for the prediction of PGA and elastic response spectral ordinates. Among these models, 42 are derived based on stochastic simulation approaches [13], while the rest are derived based on other ground motion simulation methods [14–16]. In the literature, most GMMs are developed based on linear regression [17–20]. However, capturing the nonlinear behaviour of the existing ground motion databases using linear regression is challenging. To overcome this, Machine Learning (ML) algorithms can be used to capture the nonlinear and complex behaviour of the data. In engineering applications associated with the development of GMMs, artificial neural networks, random forest, fuzzy logic, gradient boosting, and eXtreme Gradient Boosting (XGBoost) are prevailing among ML techniques [21–27]. The study of [5] summarises 39 nonparametric GMMs worldwide [18].

The literature review reveals that the number of region-specific GMMs is limited due to the need for high-quality data for large-magnitude or near-field events. Therefore, challenges remain in establishing a homogenous ground motion dataset that is indicative of the seismological characteristics of a specific region [5]. To this end, this study aims at developing an ML-based backbone GMM for the region of the Azores Plateau (AP) in Portugal by generating a homogenous dataset. In the initial stage, the tectonic activity and the fault zones of the Central and Eastern Azores region, including the islands of Faial, Pico, São Jorge, Graciosa and Terceira, are discussed in detail. In spite of moderate to high seismicity, the region lacks recorded motions. The stochastic finite-fault approach by [28] is employed to overcome this issue, and a homogenous ground motion database of 247,710 records is constructed. Simulations are performed for various scenario events with a magnitude range of 5.0 to 6.8, and a bin size of 0.1. All simulations are performed using region-specific input-model parameters proposed by [29]. In the simulations, the uncertainty regarding the rupture of representative active faults is considered through Monte Carlo Simulation (MCS). The simulated dataset is compared with the empirical models developed by [30] (BA08), [20] (AC10), and [31] (ASB14). Finally, to develop the region-specific GMM in the study area, the XGBoost algorithm is implemented [32]. The performance of the model is evaluated through coefficient of determination (R^2), Pearson corelation coefficient (r), mean absolute percentage error (MAPE), and root-mean-square error (RMSE). The present study demonstrates the capability of the proposed approach in capturing the complex behaviour of earthquake motions.

2 Study Area: Active Tectonics in the Central and Eastern Azores Islands

The Azores region is dominated tectonically by the triple junction between the North American Plate, the Eurasian Plate, and the African Plate known as the Azores Triple Junction (ATJ) (Figure 1). AP has exceptionally shallow bathymetry with a roughly triangular configuration restricted by a 2-kilometre bathymetric line [33]. This may be related to a hotspot above the mantle plume. S and P wave velocities, geochemical fingerprints, gravity, crustal thickness, and uneven topography all indicate the existence of a plume in the region [34]. The Azores archipelago's central and eastern island groups are located near the western section of the Eurasian-African plate boundary [35]. Azores Gibraltar Fracture Zone is the Atlantic part of this plate boundary having three sections with distinctive morphology and seismotectonic regimes [35]. The Azorean part of the fracture zone corresponds to a wide shear zone, which accommodates the differential spreading rates north and south of the ATJ [33,36], and is considered to migrate towards north from east fracture zone [37].

The central and eastern parts of AP are in a diffuse and complex deformation zone that was sheared under a dextral trans-tensile regime, as shown by the region's active tectonics and volcanism. The two primary fault systems, each consisting of two sets of dips in opposing directions, represent the fault pattern. The fault geometry and kinematics indicate a maximum horizontal tensile stress axis in the NE-SW direction, a maximum horizontal compressive stress axis in the NW-SE direction, and a vertical intermediate compressive stress axis. Kinematic measurements suggest a second stress field in the eastern São Miguel and Graciosa islands that may alternate with the first one in time [38,39]. The interplay between volcanic activity, surface faulting, and subaerial geomorphological processes of denudation are reflected in the morphology of the islands. Middle Pleistocene to Holocene era formations are affected by these morphologies and structures.

Tectonically regulated volcanism occurs along faults (fissure volcanic systems) or at the intersection of faults (central volcanoes). Low- to moderate-size earthquakes, the majority of which occur at shallow depths (about

40 km), are indicative of the strong seismicity of this area [40]. The first seismic event recorded on Faial was in 1614, when an earthquake struck Terceira Island. However, the major earthquakes were recorded only in the 20th century, causing significant damage in the island, in 1924, 1926, 1980 and 1998 (Figure 1). The earthquakes of 1 January 1980 with a moment magnitude (M_w) of 6.9 and 9 July 1998 (M_w=6.2) were the last two damaging events to impact the Azores Islands [41], and are examples of large magnitude events that inflicted significant damage. The 1998 event had an epicentre located offshore, about 10 km NE of Faial Island. The maximum recorded intensity was VIII in NE Faial towards the epicentre, and significant local amplification effects were observed [42]. This event was not limited to Faial Island but was also felt in other islands like Pico and São Jorge; the maximum intensity recorded in Pico was VII, while in São Jorge, it reached VI. As a result, nine people died, with more than 150 people injured and over 1 500 houses damaged. The heavier destruction was caused at Riberinha and Espalhafatos, where the maximum intensity was observed, given the presence of many old stone masonry buildings in the area, which are highly vulnerable to seismic events. A 19th century bridge also collapsed during the process [41].

The neotectonics of the five islands in the central and eastern AP, namely Faial, Pico, São Jorge, Terceira, and Graciosa, are described in detail in the subsequent sections

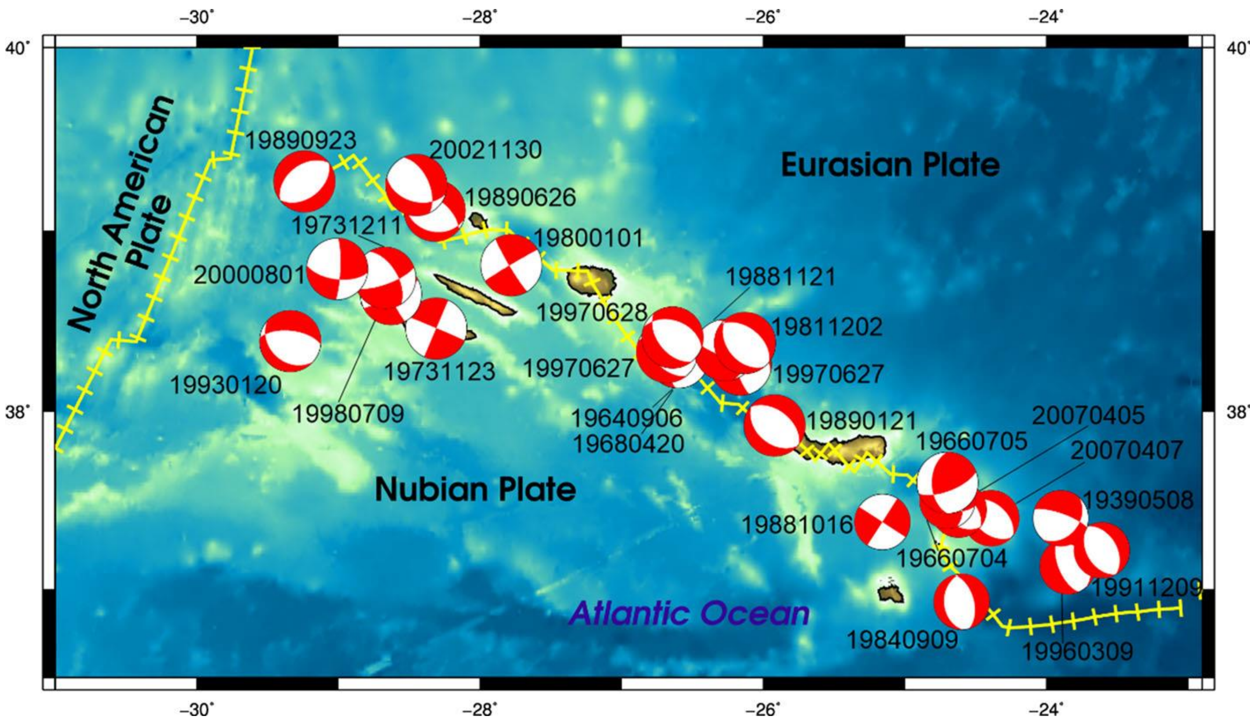


Figure 1 Map of central and eastern Azores with tectonic plates and records [29]

2.1 Faial Island

Faial Island is 21 km long, up to 14 km wide, and rises to an altitude of 1043 m at Cabeço Gordo. Two primary fault mechanisms on the island are WNW-ESE and NNW-SSE trending [43]. A WNW-ESE trending graben structure named the Pedro Miguel Graben, made up of normal dextral faults, distinguishes the eastern portion of Faial Island. The southern section of the structure is defined by the north-dipping Rocha Vermelha, Espalamaca, and Flamengos faults. At the same time, the northern half of the graben is formed by a group of faults dipping to the south (namely, the Ribeirinha, Eastern and Western segments of Lomba Grande faults). The Lomba do Meio and Lomba de Baixo faults have NE dipping south of the Caldeira. The Flamengos fault is connected by the Lomba de Baixo fault (Figure 2). Table 1 summarises information on the active faults in Faial Island. It is evident that the fault mechanism is predominantly normal for all, with the maximum expected moment magnitude (M_w) of 6.6 due to the potential rupture of the Espalamaca fault.

It should be emphasised that a total of 49 stations shown by triangular symbols in Figure 2 have been considered for the simulations that are carried out in the subsequent section.

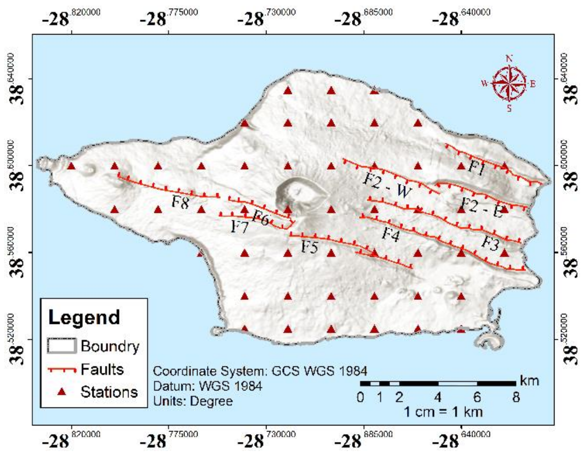


Figure 2 Active tectonics in Faial Island with the stations used for simulations shown by triangular symbols.

Table 1. Information on the active faults of Faial Island [35]

No	Fault Name	Fault Rupture Length (km)	$M_{w\text{-}max}$	Fault Mechanism	Strike (°)	Dip (°)
F1	Ribeirinha	12.5	6.3	Normal	115	75
F2-E	Lomba Grande Eastern segment	12.5	6.3	Normal	115	80
F2-W	Lomba Grande Western segment	12.5	6.3	Normal	115	80
F3	Rocha Vermelha	14.0	6.4	Normal	290	55
F4	Espalamaca	20.3	6.6	Normal	295	70
F5	Flamengos	11.5	6.3	Normal	290	70
F6	Lomba do Meio	4.0	5.2	Normal	295	70
F7	Lomba de Baixo	4.0	5.2	Normal	300	50
F8	Capelo	8.8	5.8	Normal	290	90

2.2 Pico Island

Pico is 46 km long, up to 15.8 km wide with the highest altitude of 2351 m at Montanha do Pico. The tectonic structure of this island is consistent with that of Faial Island. Two major fault systems are present: WNW–ESE tending system with the normal dextral Lagoa do Capitão and Topo fault zones defining the Brejos Graben, which is covered by Pico volcano in the west and NNW–SSE where less frequent conjugate faults are identified by volcanic alignments including Cabeço do Sintrão fault [43]. Figure 3 presents the tectonic map of Pico Island, while Table 2 provides details on the main active faults. Using the correlations of [44], the maximum expected M_w for this island ranges from 6.1 to 6.6 with a normal faulting mechanism.

It is worth noting that the subsequent section of the study uses a total of 119 stations for simulations, which are indicated by triangular symbols in Figure 3.

Table 2. Information on the active faults of Pico Island [35]

No	Fault Name	Fault Rupture Length (km)	$M_{w\text{-}max}$	Fault Mechanism	Strike (°)	Dip (°)
P1	Lagoa do Capitao	8.8	6.2	Normal	120	80-90
P2	Topo	7.5	6.1	Normal	285	70-90
P3	Cabeço do Sintrão	21	6.6	Normal	293	-

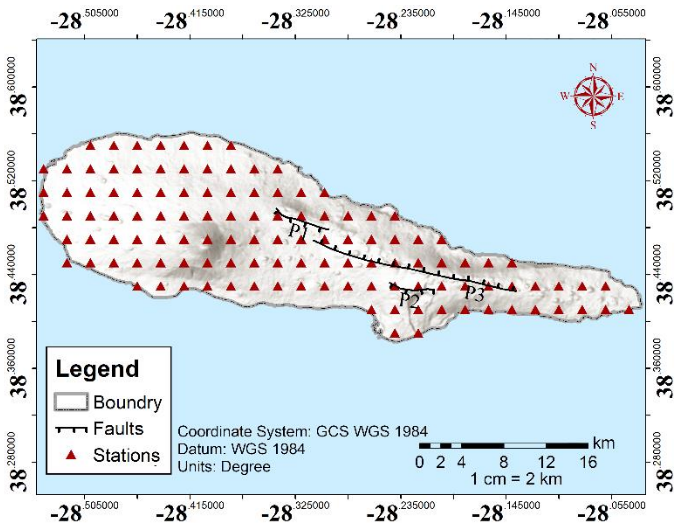


Figure 3 Active tectonics in Pico Island with the stations used for simulations shown by triangular symbols.

2.3 São Jorge Island

The São Jorge Island is 54 km long and 7 km wide, with the highest altitude of 1053 m at the Pico da Esperança. The same fault systems seen in Faial and Pico characterise São Jorge Island tectonically: one normal dextral set of faults trending WNW-ESE, dipping north and south in both directions, and one normal left-lateral series of faults trending NNW-SSE [43]. The younger western half of the island is dominated by the Picos and Pico do Carvão fault zones, two major normal dextral WNW-ESE fault zones. In the eastern region, the Urze-São João fault, which exhibits a continuous scarp that is 10 km long, and the Cume Faja do Belo fault are the most significant WNW-ESE trending tectonic structures [43]. The major NNW-SSE trending faults are the Ribeira Seca fault which separates the western São Jorge from the eastern part, and the Serra de Topo fault. The tectonic map and the information on the active faults of this island are presented in Figure 4 and Table 3, respectively. According to fault dimension, the maximum anticipated M_w due to rupture of the faults, all with a normal fault mechanism, ranges between 6.1 and 6.8 [44].

It should be pointed out that Figure 4 displays a total of 70 stations that are utilised for simulations in the subsequent section of the study, and these stations are indicated by triangular symbols.

Table 3. Information on the active faults of Sao Jorge Island [35]

No	Fault Name	Fault Rupture Length (km)	$M_{w\text{-}max}$	Fault Mechanism	Strike (°)	Dip (°)
SJ1	Picos	33	6.8	Normal	120	90
SJ2	Pico Carvão	12	6.3	Normal	285	75-90
SJ3	Urze-São João	15	6.4	Normal	304	80
SJ4	Cume Faja do Belo	7.4	6.1	Normal	120	70
SJ5	Serra do Topo	7.2	6.1	Normal	140	-
SJ6	Ribeira Seca	7.3	6.1	Normal	160-170	-

2.4 Terceira Island

The elliptically shaped Terceira Island rises to 1 021 m above sea level at Santa Barbara and has a major axis that is 30 km long and trends WNW-ESE. Terceira is affected by three significant earthquakes in 1614, 1841, and 1980. The epicentres of the first two seismic events likely occurred on the island or in the area in the vicinity of offshore, while the third one occurred offshore between the islands of Terceira, Graciosa, and São Jorge [45,46]. Terceira Island's primary tectonic features are NW-SE trending faults. The Lajes Graben, which cuts across Terceira's older NE region, dominates the island's tectonic structure. Three significant normal-dextral NW-SE trending faults are the NE-plunging Fontinhas, and Cruz do Marco faults to the SW along with the SW-dipping Lajes Fault in the NE. The Santa Bárbara Graben, which spans the Santa Bárbara volcano, is the second major structure. Maximum anticipated magnitudes range from M_w of 5.9 to 6.4 according to [44].

It is noteworthy that Figure 5 illustrates a total of 104 stations that have been utilised for simulations in the subsequent section of the study, and these stations are marked by triangular symbols.

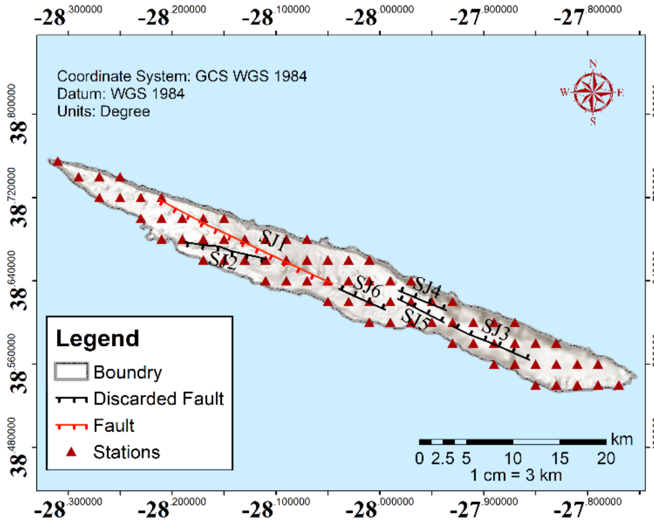


Figure 4 Active tectonics in São Jorge Island with the stations used for simulations shown by triangular symbols.

Table 4. Information on the active faults of Terceira Island [35]

No	Fault Name	Fault Rupture Length (km)	$M_{w\text{-}max}$	Fault Mechanism	Strike (°)	Dip (°)
T1	Lajes	8.2	6.1	Normal	138	70-90
T2	Fontinhas	9.0	6.2	Normal	313	-
T3	Cruz do Marco	5.0	5.9	Normal	310	70
T4	Santa Bárbara	12.9	6.4	Normal	308	70

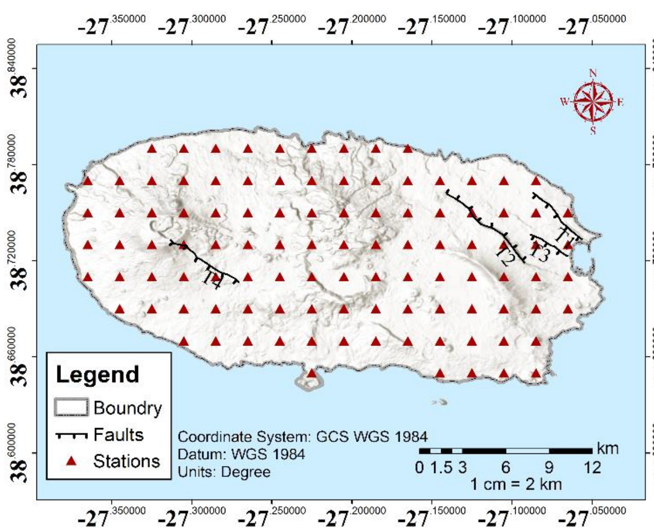


Figure 5 Active tectonics in Terceira Island with the stations used for simulations shown by triangular symbols.

2.5 Graciosa Island

Graciosa is an elliptical-shaped volcanic island with a 12 km length and 7 km width and a maximum elevation of 402 m on the caldera's southern rim. This island has an NW-SE normal tectonic structure. Several faults

trending NW-SE to NNE-SSW, which have grown into significant fault scarps, cut through the old volcanic complexes of Serra das Fontes and Serra Branca [38,47]. Among them, there are the North and South Serra Branca Faults, the South Serra das Fontes Fault along with the Saúde-Hortelã Fault. The East Serra das Fontes fault, which is symbolised by a massive scarp facing southeast, is the main NNE-SSW-trending faulting mechanism. Correlations provide maximum anticipated M_w ranging from 5.7 to 5.9 due to the rupture of all faults [44].
It is worth mentioning that Figure 6 displays a total of 17 stations that have been employed for simulations in the subsequent section of the study, and these stations are identified by triangular symbols.

Table 5. Information on the active faults of Graciosa Island [35]

No	Fault Name	Fault Rupture Length (km)	$M_{w\text{-}max}$	Fault Mechanism	Strike (°)	Dip (°)
G1	Saúde-Hortelã	5.0	5.9	Normal	140	-
G2	South Serra das Fontes	4.6	5.8	Normal	126	-
G3	North Serra Branca	4.8	5.9	Normal	302	-
G4	South Serra Branca	3.2	5.7	Normal	305	-
G5	East Serra das Fontes	4.6	5.8	Normal	340	-

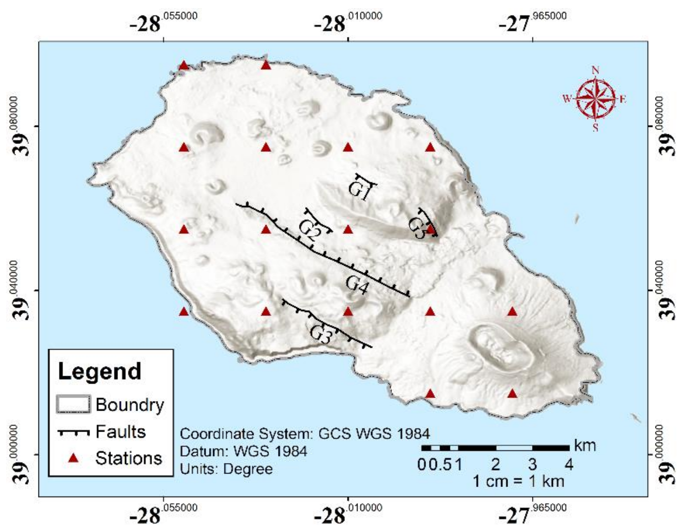


Figure 6 Active Tectonics in Graciosa Island with the stations used for simulations shown by triangular symbols.

3 Region-Specific Simulated Strong Ground Motion Database

The Eastern and Central Azores Islands have similar neotectonics and geology. By referring to Section 2, it is evident that most of the faulting structures responsible for major events can be identified on Faial Island. Yet, the maximum expected M_w in Faial Island is reported as 6.6, while this value is 6.8 due to the potential rupture of the Picos fault in São Jorge Island. In this study, in order to cover all magnitude ranges in GMM, possible fault ruptures on both Faial and São Jorge Islands are modelled as scenarios to represent the potential earthquakes in the AP. A simulated database is formed due to the assumption in the rupture of ten faults, nine in Faial Island as listed in (Table 1) and one in São Jorge Island as the first fault listed in Table 3. These faults are shown in red colour in Figure 2 and Figure 4. Following the ground motion simulation methodology proposed, the scenario events, and input-model parameters with the generated database are discussed in detail next.

3.1 Ground Motion Simulation Methodology

The stochastic methodologies encompass both point-source and finite-fault methods. The point-source approach, initially proposed by [48], includes the following shear wave acceleration spectrum at an observation point:

$$A(f) = CM_o\left[(2\pi f)^2 / \left(1+(f / fc)^2\right)\right]G(R)e^{-\pi fR/(Q(f)\beta)}e^{(-\pi f\kappa_0)}S(f) \tag{1}$$

where M_0 represents the seismic moment in Nm, while f_c denotes the corner frequency in hz, C is the scaling constant representing the radiation pattern for shear waves, amplification on free surface amplification, the division of horizontal components into two, crustal density, and shear wave velocity. The term in the squared parenthesis corresponds to an ω^{-2} source spectrum, as proposed by [49]. $G(R)$ is geometrical spreading representing distance (R) dependent attenuation, and $Q(f)$ is the quality factor representing frequency (f) dependent anelastic attenuation, κ is zero-distance kappa for upper crust attenuation, and $S(f)$ is the frequency dependent soil amplification factor.

The point-source approach was later expanded to introduce the finite-fault method [50–52]. In subsequent developments, Boore [53] enhanced the approach proposed by [52]. Additional modifications and improvements were made, such as scaling high-frequency motions based on the integral of the squared acceleration spectrum rather than the integral of the squared velocity spectrum. Moreover, the truncation of the sub-fault time series was eliminated. In this modified version, the duration of the sub-fault motions is determined by the inverse of the corner frequency associated with each sub-fault.

Using the EXSIM12 platform [54], this study employs the latest version of the stochastic finite-fault ground motion simulation methodology to model acceleration time series of scenario earthquakes [55]. The algorithm proposed by [52], which was developed based on the original FINSIM code by [51], is enhanced in this technique by adding the improvements suggested by [53]. The low-frequency component of the simulations is strengthened by the improved stochastic method. By considering factors including earthquake magnitude, fault geometry, strike, dip, slip distribution, density, and rupture velocity, this method can recognise the fault rupture. To receive the seismic signal in the time domain at any observation site, the source contribution is combined with the attenuation parameters and site effects.

The ruptured fault plane is depicted as a grid of smaller sub-sources in the stochastic finite-fault approach by assuming a point-source for each sub-source with a ω^{-2} source spectrum, as proposed by [49]. Depending on how far a sub-source is from the hypocentre, each sub-source ruptures with an appropriate time delay. The time domain summation of the contributions from the delayed sub-sources is carried out as follows:

$$A(t) = \sum_{i=1}^{N} H_i Y_i\left(t+\Delta t_i + T_i\right) \tag{2}$$

where A(t) represents the total seismic signal at time t, N is the total number of sub-sources, Y_I demonstrates the seismic signal of i^{th} sub-source which is its inverse Fourier transform [50,53], Δt_I is the sum of the fracture initiation and time delay due to the distance of the i^{th} sub-source from the hypocentre, the term T_I relates to a fraction of rise time considered for additional randomisation and finally, the term H_I resembles the normalisation factor of the i^{th} sub-source introduced for the conservation of energy with the following formula:

$$H_i = \frac{M_0}{M_{0i}}\sqrt{\sum_j\left(\frac{f_0^2 f_j^2}{f_0^2 + f_j^2}\right)^2 \Big/ N\sum_j\left(\frac{f_{0i}^2 f_j^2}{f_{0i}^2 + f_j^2}\right)^2} \tag{3}$$

where f_0 is the corner frequency of the main fault plane, f_j is the j^{th} frequency ordinate, M_0 is the total seismic moment, and the terms M_{0i} and f_{ci} are the seismic moment and corner frequency of the respective i^{th} sub-source formulated as follows:

$$M_{0i} = M_0 ´ S_i \Big/ \mathring{a}_{i=1}^{N} S_i \tag{4}$$

$$f_{ci} = 4.9\times10^6\beta_s\left(\frac{\Delta\sigma}{p\times M_0}\right)^{\frac{1}{3}} \text{ where } p = \begin{cases} \dfrac{N_R}{N} & \text{if } N_R < N\times PP \\ PP & \text{if } N_R \geq N\times PP \end{cases} \tag{5}$$

where s_I is the slip of the i^{th} sub-source in Equation (2). In Equation (3), the term N_R represents the total number of sub-sources which are activated when the i^{th} sub-source triggers and $\Delta\sigma$ is the stress drop in bars. The term PP is the pulsing percentage. The algorithm is based on a dynamic corner frequency approach where the corner frequencies of the activated sub-sources descend with rupture progress until reaching a specified level which is *PP*. For the rest sub-sources, the corresponding corner frequency remains constant.

3.2 Input Parameters

In this study, simulations are performed on a total of 23 scenario events with varying magnitudes and ruptured fault planes. The information on the considered scenario events is summarised in Table 6. Karimzadeh and Lourenço [56] simulated the 1998 Faial (M_w =6.2) event and provided region-specific input-model parameters based on simulation validations against observed motions from this event. In this study, the validated input-model parameters of [56] are calibrated for the scenario events. To account for uncertainty in the parameters representing source and attenuation effects, however, they are here assumed to be random variables.
Table 7 gives information on the deterministic input-model parameters, whereas Table 8 lists the probabilistic parameters and their Probability Distribution Functions (PDFs). To this end, the study by [29] is used to implement regional models with their PDFs. Parameters of

Table *8* are utilised to perform 30 MCSs for every event, each with distinct combinations. Finally, simulations are performed in a total of 359 nodes at bedrock, as displayed above in Figure 2 to Figure 6.

Table 6. Information on the scenario events

Scenario	M_w	Strike (°)	Dip (°)	Length (km)	Width (km)	Fault No	Region
1	6.3	115	75	12.5	12.5	F1	Faial
2	6.1	115	75	12.0	12.0	F1	Faial
3	5.5	115	75	7.0	5.5	F1	Faial
4	6.3	115	80	12.5	12.5	F2E	Faial
5	6.2	115	80	12.3	12.3	F2E	Faial
6	6.3	115	80	12.5	12.5	F2W	Faial
7	5.4	115	80	6.0	5.0	F2W	Faial
8	6.4	290	55	14.0	14.0	F3	Faial
9	5.9	290	55	10.0	9.0	F3	Faial
10	5.3	290	55	5.0	5.0	F3	Faial
11	6.6	295	70	20.3	14.3	F4	Faial
12	6.5	295	70	19.0	14.0	F4	Faial
13	6.0	295	70	11.0	10.0	F4	Faial
14	6.3	290	70	12.5	12.5	F5	Faial
15	5.7	290	70	8.5	6.5	F5	Faial
16	5.2	295	70	4.0	4.0	F6	Faial
17	5.0	295	70	4.0	3.0	F6	Faial
18	5.2	300	50	4.0	4.0	F7	Faial
19	5.1	300	50	4.0	3.5	F7	Faial
20	5.8	290	90	8.8	8.7	F8	Faial
21	5.6	290	90	8.0	6.0	F8	Faial
22	6.8	120	90	26.0	16.0	SJ1	Sao Jorge
23	6.7	120	90	33.0	18.0	SJ1	Sao Jorge

Table 7. Deterministic input-model parameters

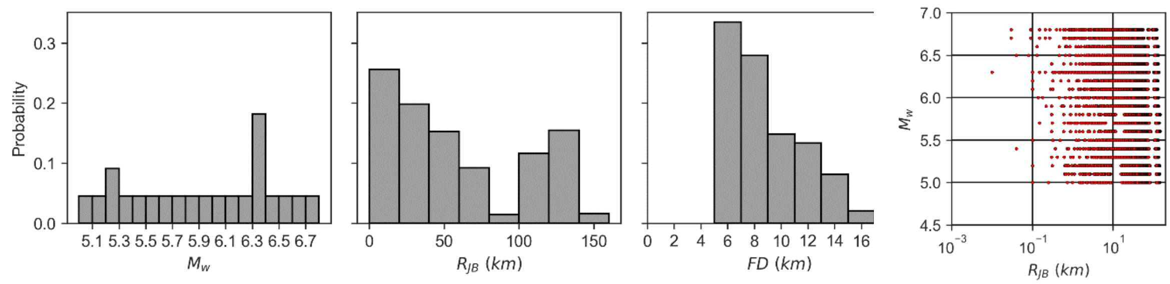
Parameter	Value
Crustal Thickness, D (km)	13
Crustal Density (g/cm^3)	$Depth = 0.0\ km \rightarrow 2.67$ $Depth = 2.5\ km \rightarrow 2.77$ $Depth = 8.0\ km \rightarrow 2.86$ $Depth = 14.0\ km \rightarrow 2.93$ [29]
Shear Wave Velocity (km/s)	$Depth = 0.0\ km \rightarrow 3.1$ $Depth = 2.5\ km \rightarrow 3.7$ $Depth = 8.0\ km \rightarrow 4.2$ $Depth = 14.0\ km \rightarrow 4.6$ [29]
Shear Wave Velocity/Crustal Velocity	0.8
Geometric Spreading	$R^{-1.0}$ $R \le 1.5D\,km$ $R^{0.0}$ $1.5D\,km < R \le 2.5D\,km$ [29] $R^{-0.5}$ $R > 1.5D\,km$
Duration Model (R in km)	T_0+0.1R
Window Type	Saragoni-Hart
Damping	5%
Slip Weight	Random
Iseed	309

Table 8. Probabilistic input-model parameters

Parameter	Value	PDF
Hypocentre Location	Along the length and width	Uniform
Pulsing Percent	30-50	Uniform
Kappa	0.075±0.02 [29]	Uniform
Stress Drop (bars)	110±20 [29]	Lognormal
Quality Factor	$(76\pm11)\,f^{0.69\pm0.09}$ [29]	Lognormal

3.3 Ground Motion Database

Simulations of this study result in 247,710 ground motion records for the entire AP. Figure 7a shows histograms in terms of the seismological features of the Azores ground motion dataset, including M_w, Joyner and Boore distance (R_{JB}), and Focal Depth (FD). The scenario events range in M_w from 5.0 to 6.8, grouped into 0.1 magnitude bins. The probability is highest for a magnitude of 6.3, corresponding to the characteristic earthquake of several fault planes. The R_{JB} ranges between 0 to 150 km representing more near-field data than far-field. Lastly, FD changes between 5.0 and 17.0 km, which is indicative of shallow events. The distribution of M_w versus R_{JB} is presented in Figure 7b. Finally, the normalised 5% damped Pseudo Spectral Acceleration (PSA) of the Azores dataset for different magnitude intervals is shown in Figure 8. This figure displays the normalised PSA by their respective PGA values. In addition, the mean PSA, and a range of probable spectral values by one standard deviation (σ) above and below the mean are presented. In accordance with physics, simulations of earthquakes with greater magnitudes exhibit higher spectral ordinates at higher periods when contrasted with earthquakes of smaller magnitudes. Moreover, the scatter plots serve as evidence of the uniformity of the dataset across all magnitudes and distances, highlighting one of the advantages of the ground motion simulations.



a) b)

Figure 7 a) Histograms of seismological characteristics of the Azores ground motion dataset, and b) Distribution of M_w versus R_{JB} for the Azores ground motion dataset.

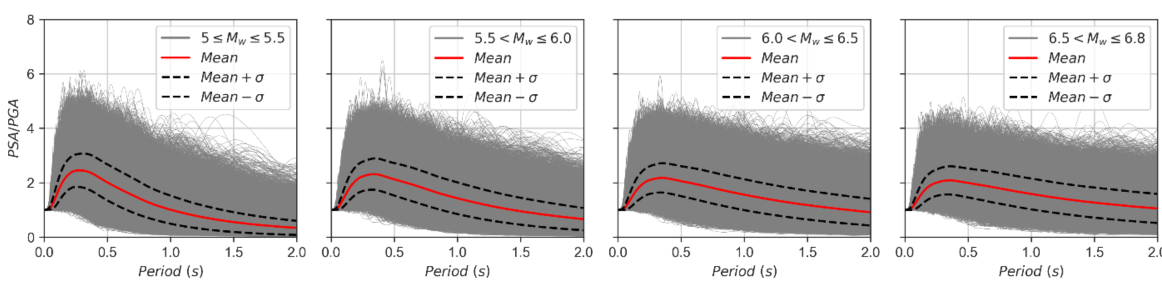


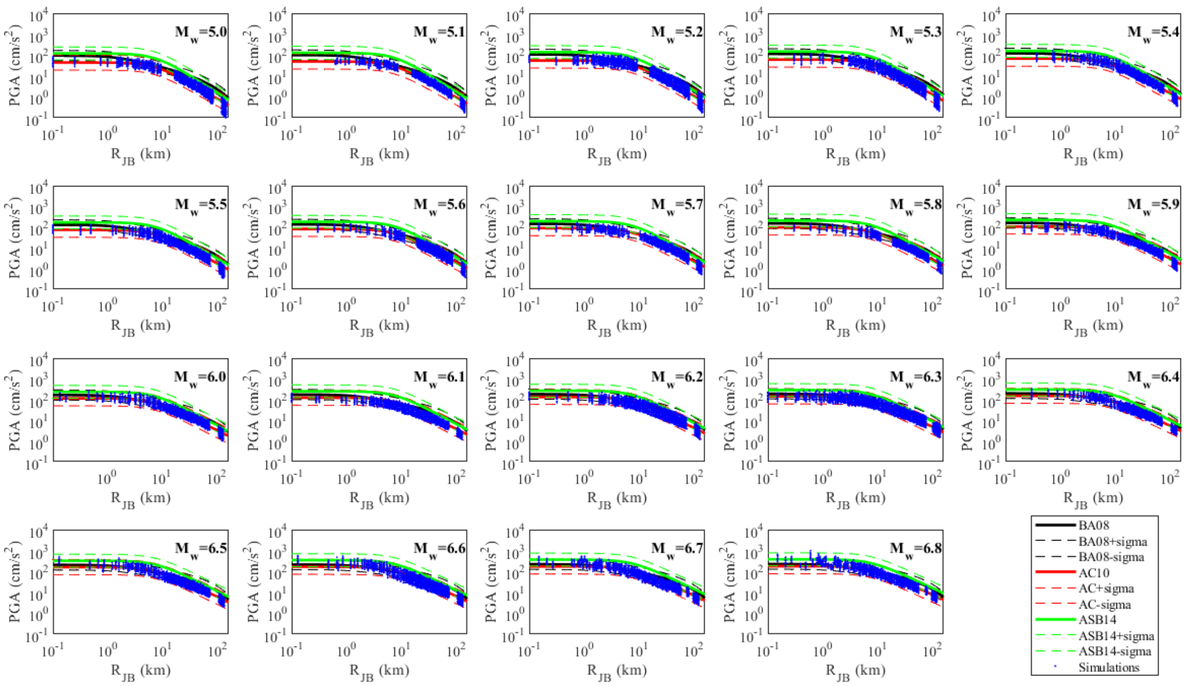
Figure 8 Normalised 5% damped PSA of the Azores dataset.

3.4 Validations of the Simulated Dataset

Simulations are typically validated by comparing their ability to estimate observed records of past earthquakes or, in cases where recorded motions are not available, by comparing their trend with appropriate existing GMMs. As previously stated, the input-model parameters used for simulations of this study are calibrated based on region-specific parameters that have already been validated for the 1998 Faial event with a magnitude of 6.2 [56]. On the other hand, the validation process for simulated ground motions with existing GMMs involves comparing the simulated ground motion outcomes to the expected values generated by established equations. This is done to assess the accuracy and reliability of the simulated data and to determine if it aligns with the expected results generated by the equations. If the simulated data matches the predictions of the equations, it can be considered a validation of the simulation methodology. This process is often used in seismology research and engineering to evaluate the performance of ground motion simulations and to determine their suitability for practical applications.

In this study, the simulated dataset is validated using well-known GMMs Boore and Atkinson [30] (BA08), Akkar and Çağnan [20] (AC10), and Akkar et al. [31] (ASB14). The comparison of the simulated dataset for different magnitudes against the GMMs is plotted in Figure 9. According to the results, the simulations are consistent with the AC10 relationship for PGA regardless of the magnitude. However, the PGV results vary depending on the magnitude of the scenario event. For magnitudes smaller than 5.9, the results align more with BA08, while as the magnitude increases, the trend becomes closer to the attenuation of the ASB14 model. This further emphasises the significance of developing GMMs specific to the region.

(a)



(b)

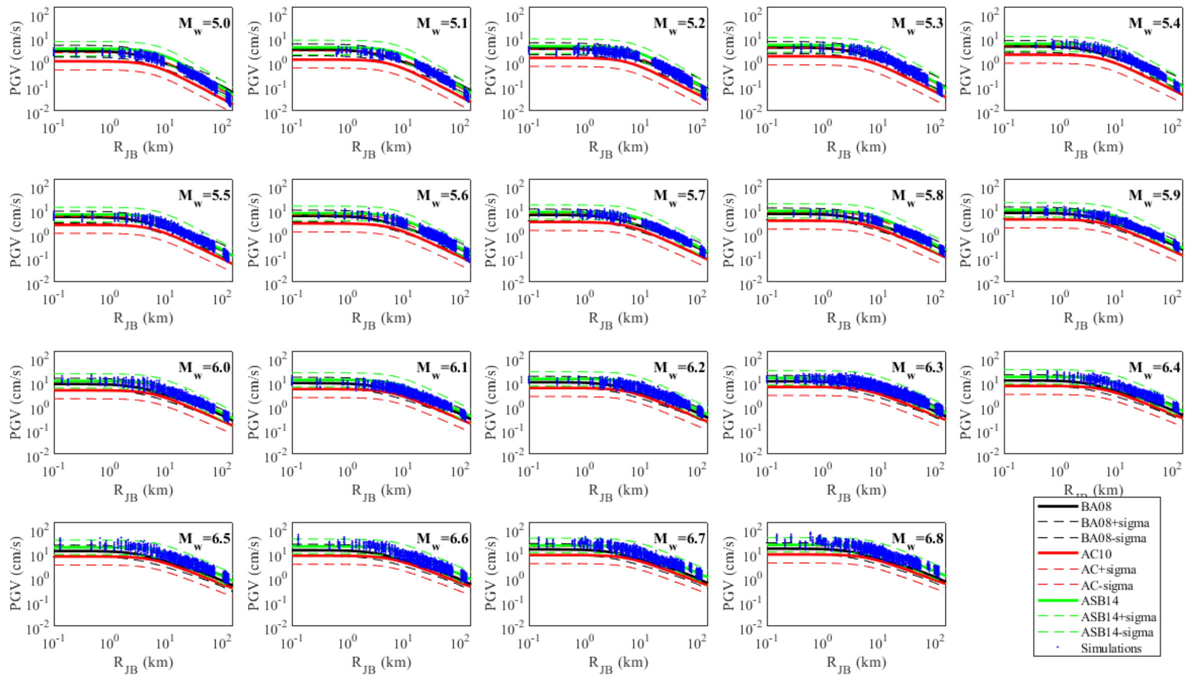


Figure 9 Attenuation of simulations in terms of (a) PGA and (b) PGV against empirical GMMs.

4. Ground Motion Modelling Methodology

The common method for predicting ground motion IMs, such as PGA, PGV or PSA, is to utilise GMMs. These models are typically developed through empirical approaches that involve statistical regression analysis of large datasets of ground motion intensities [57]. Since there is a significant amount of variability or scatter in the observed ground motion data for each IM, GMMs generally provide a probability distribution of possible ground motion outcomes, instead of a single deterministic value:

$$\ln\left(y_{ij}\right)=\begin{pmatrix}\ln\left(\mathrm{PGA}\right)\\ \ln\left(\mathrm{PGV}\right)\\ \ln\left(\mathrm{PSA}_{0.03s}\right)\\ \vdots\\ \ln\left(\mathrm{PSA}_{2s}\right)\end{pmatrix}=f\left(M_w,R_{JB},FD\right)+\eta_i+\varepsilon_{ij} \tag{6}$$

where η_i is the inter-event residual component and ε_{ij} is the intra-event residual component in the natural logarithm scale, i denotes the index of the earthquake event, and j represents the station's index. The functional form in Equation 6 is modelled using XGBoost algorithm. Two components of residuals in GMMs, namely inter-event and intra-event residuals, are assumed to be independent, normally distributed random variables with zero mean and standard deviations of τ and σ, respectively. The inter-event and intra-events residuals are assumed independent; therefore, the total standard deviation for a given GMM is calculated as the square root of the sum of squares of the two types of residuals. This is expressed mathematically as:

$$\phi=\sqrt{\sigma^2+\tau^2} \tag{7}$$

where ϕ is the total standard deviation, σ is the intra-event standard deviation, and τ is the inter-event standard deviation.

The total residual δ_{ij} is obtained by:

$$\delta_{ij}=\eta_i+\varepsilon_{ij}=\ln IM_{ij}^{sim}-\ln IM_{ij}^{m} \tag{8}$$

where $\ln IM_{ij}^{sim}$ is the simulated value (in terms of PGA, PGV or PSA) and $\ln IM_{ij}^{m}$ is the GMM prediction value. The inter-event error for each earthquake event can be described as follows:

$$\eta_i = \frac{\tau^2 \sum_{1}^{n_i} \delta_{ij}}{n_i \tau^2 + \sigma^2} \approx \frac{\sum_{1}^{n_i} \delta_{ij}}{n_i} \quad (n_i \tau^2 >> \sigma^2) \tag{9}$$

As in this study the number of records in each event is large (n_i =359) and $n_i\tau^2$ is much larger than σ^2, the approximate equation can accurately measure the inter-event residuals [58]. Finally, the intra-event residuals can be obtained as follows:

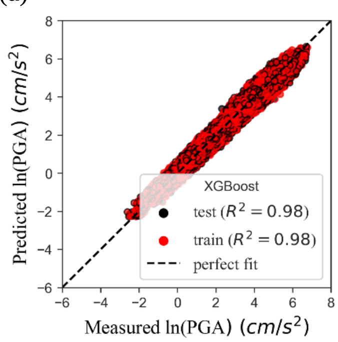
$$\varepsilon_{ij} = \delta_{ij} - \eta_i \tag{10}$$

The GMM of this study is developed using the XGBoost algorithm. XGBoost [32] is a powerful ML algorithm that has become increasingly popular in recent years due to its superior performance in various applications. This is an ensemble-based learning algorithm that combines multiple decision trees to make accurate predictions by minimizing prediction errors. It is based on the gradient boosting framework, which involves iteratively adding new decision trees to the model and optimizing the model's parameters to minimize the loss function. XGBoost's unique features include its ability to handle missing values, its built-in regularization techniques to prevent overfitting, and its capability to handle both regression and classification tasks. Additionally, the algorithm has been shown to be highly scalable, making it suitable for large datasets. Despite its high performance, XGBoost requires careful parameter tuning and validation to achieve optimal results in a specific application. Tuning the model parameters is difficult but important as the accuracy of the predictions done with ML algorithms highly depends on them. Bayesian optimization is a powerful mathematical technique that can be used to efficiently tune hyperparameters of complex models. It is particularly effective in optimizing black-box functions that take a long time to evaluate [59]. This approach has gained popularity in fine-tuning hyperparameters of ML algorithms due to its flexibility in optimizing derivative-free functions [60]. In comparison to generic optimization techniques such as grid and random search, Bayesian optimization is considered to perform better [60,61]. Therefore, it is employed in this study to optimize the hyperparameters of the XGBoost model.

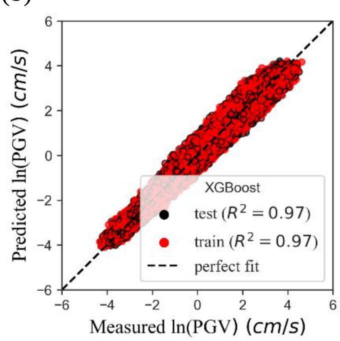
5. Results and Discussion

The section presents the outcomes of the developed GMM, whereby Figure 10 assesses the efficacy of the model by examining the concurrence between the predicted and observed values of the chosen IMs, which include PGA, PGV, and PSA at T=0.3 s and T=1.5 s in the natural logarithmic scale. Plots are generated for the training and testing datasets and juxtaposed against the ideal fit, with a lower degree of variation to the ideal fit indicating superior model performance. The analysis revealed that the model performed well for both datasets, as evidenced by the coefficient of determination exceeding 0.95, implying a high degree of accuracy in the predictions.

(a)



(b)



(c)

(d)

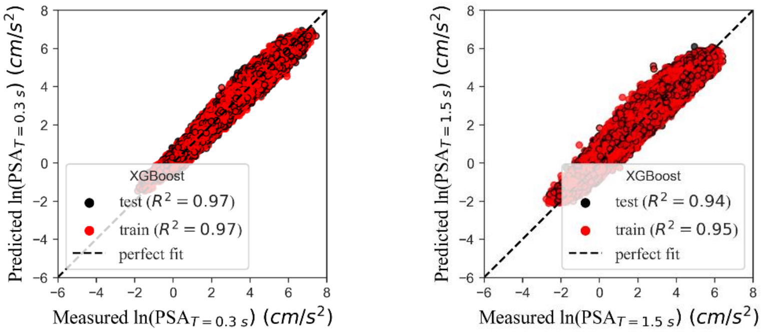


Figure 10 Observed versus predicted values of the developed GMM for the selected IMs including (a) ln(PGA), (b) ln(PGV), (c) ln($PSA_{T=0.3\ s}$), and (d) ln($PSA_{T=1.5\ s}$).

The performance indicators of the developed model, including R^2, r, MAPE, and RMSE are presented in Figure 11. The results indicate that for all IMs, the model's performance is acceptable, with both indicators R^2 and r exceeding 0.90. However, for PGV and larger periods of the PSA, a decrease in these parameters is observed when compared to PGA and PSA with smaller periods (less than 0.3 s). This decrease is further confirmed by the error indices, namely RMSE and MAPE, where an increase is observed by these indicators. These findings suggest that the model may perform better for PGA and spectral ordinates of shorter periods compared to PGV and spectral ordinates of longer periods, which should be taken into consideration when applying the model in practice. This observation is also consistent with the existing empirical GMMs [20,30,31].

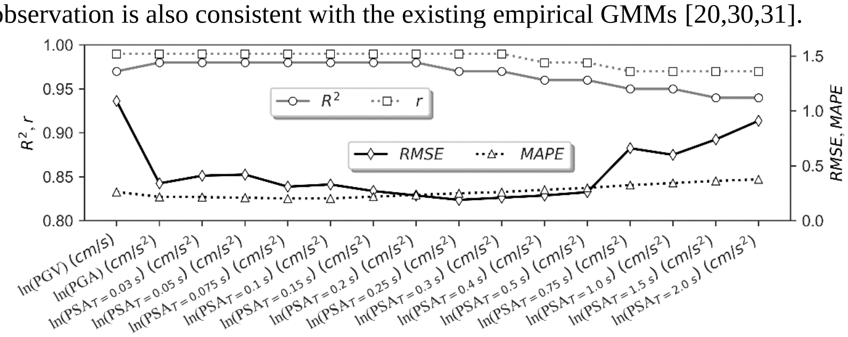
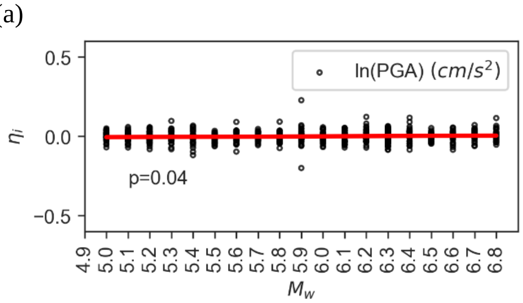


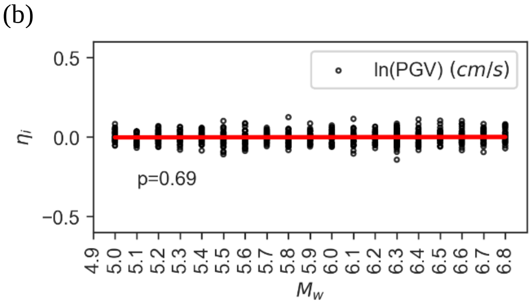
Figure 11 Model performance indicators for different IMs.

The model's potential bias is evaluated by examining the inter-event and intra-event uncertainties in relation to source- and site-related parameters, M_w and R_{JB}, respectively, as shown in Figure 12 and Figure 13. The residuals are found to be unbiased for all M_w and R_{JB} ranges, as evidenced by the absence of any observable patterns in the mean residual across all considered IMs, and the inter-event and intra-event residuals are consistent with previous research [62], ranging from -1.5 to 1.5 and -0.2 to 0.2, respectively. P-values are also calculated and displayed at a significance level of 0.05 to verify the null hypothesis of unbiased estimates. The GMM is deemed independent of explanatory variables because the mean residuals for all IMs fluctuate around zero. Furthermore, unlike GMMs based on real records, the uncertainty of residuals remains constant as magnitude increases or distance decreases, indicating a significant advantage of ground motion simulations over real datasets.

(a)



(b)



(c)

(d)

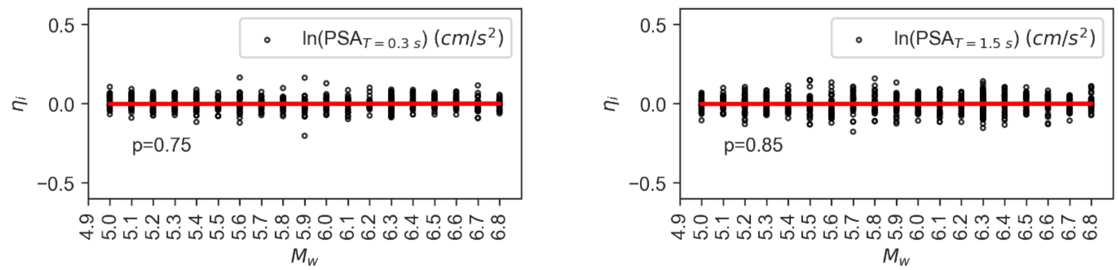


Figure 12 Distribution of the inter-event residuals with respect to M_w for the selected IMs including (a) ln(PGA), (b) ln(PGV), (c) ln(PSA$_{T=0.3\ s}$), and (d) ln(PSA$_{T=1.5\ s}$).

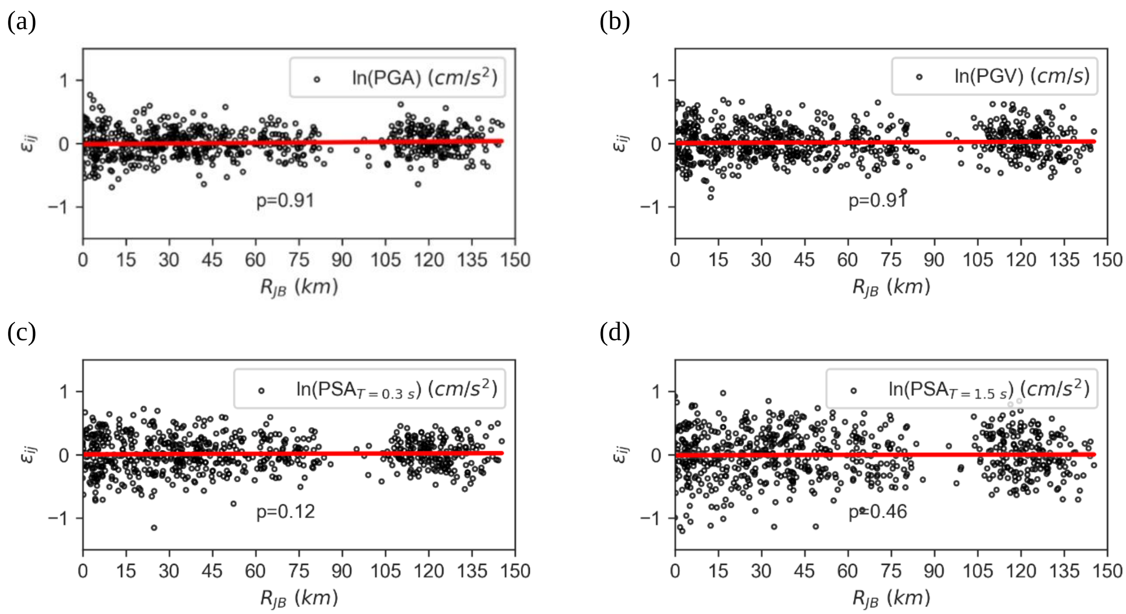


Figure 13 Distribution of the intra-event residuals with respect to R_{JB} for the selected IMs including (a) ln(PGA), (b) ln(PGV), (c) ln(PSA$_{T=0.3\ s}$), and (d) ln(PSA$_{T=1.5\ s}$).

Figure 14 presents the standard deviation of inter-event, intra-event and total residuals for PGA, PGV, and all spectral ordinates. The analysis shows that the inter-event uncertainty for all spectral values is smaller than the intra-event uncertainty across all period ranges, which is consistent with previous literature. However, the smaller range of inter-event uncertainty could be attributed to the use of the same ground motion simulation approach in a single region. The total residual ranges from 0.2 to 0.4, with an increase in value observed with an increase in the spectral period, in line with previous findings. Furthermore, the XGBoost-GMM model exhibited an acceptable uncertainty range and performed well when compared to existing models [62].

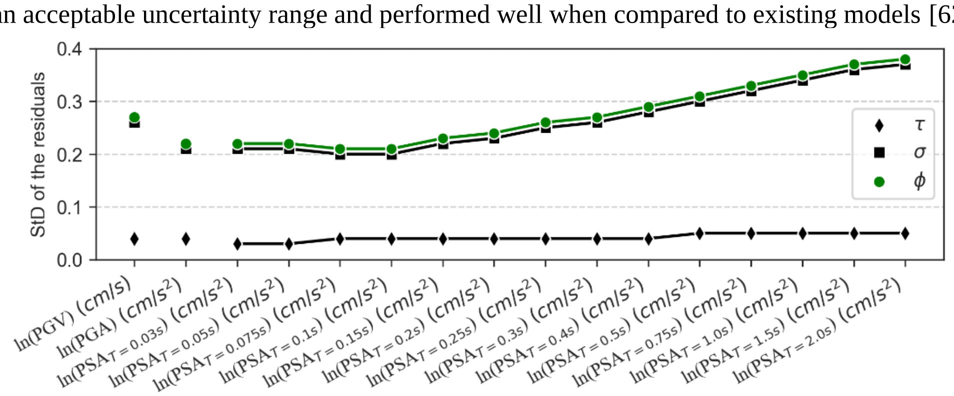
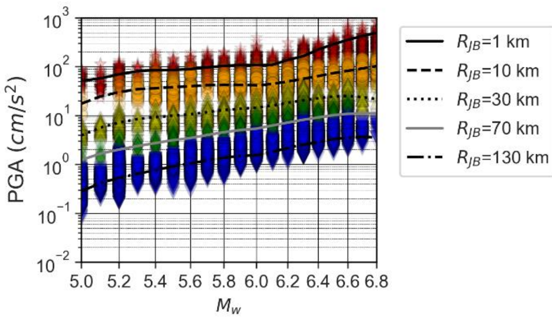


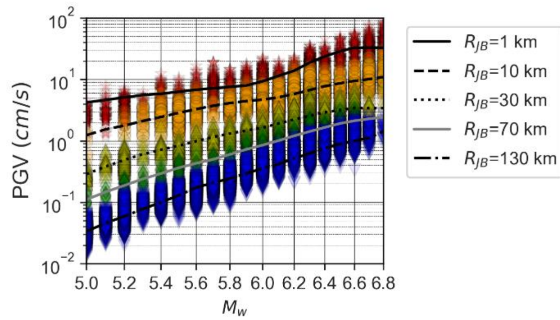
Figure 14 Standard deviation of the inter-event, intra-event, and total residuals for PGA, PGV, and all spectral ordinates.

The proposed GMM is subjected to further evaluation to determine its ability to represent physics-based phenomena regarding the behaviour of real earthquakes. To this end, the results for various magnitude and distance combinations, utilising the FD of 8.0 km, are compared. Figure 15 illustrates the estimated PGA, PGV, and PSA for periods T=0.3 s and T=1.5 s, for a range of magnitudes between 5.0 and 6.8. This evaluation is conducted for five R_{JB} values of 1 km, 10 km, 30 km, 70 km, and 130 km. The results indicate that an increase in magnitude and a decrease in the distance leads to a corresponding rise in the levels of PGA, PGV, and PSA at all period ranges. It can be inferred that the model's ability to accurately capture the trends in earthquake records demonstrates its capability to represent actual earthquake data.

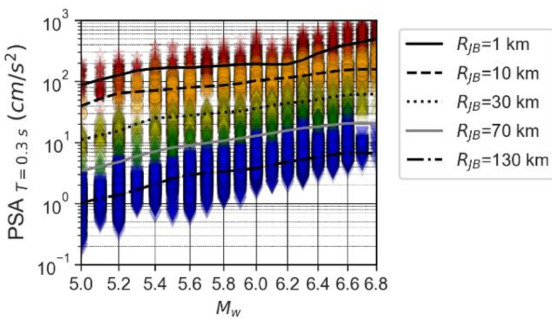
(a)



(b)



(c)



(d)

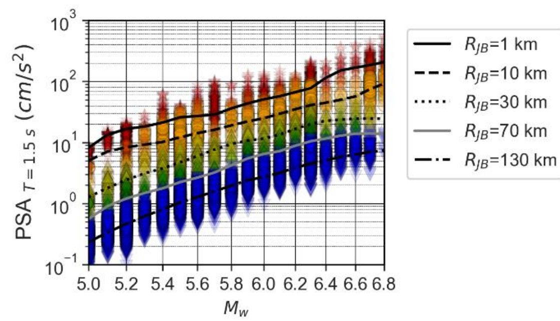
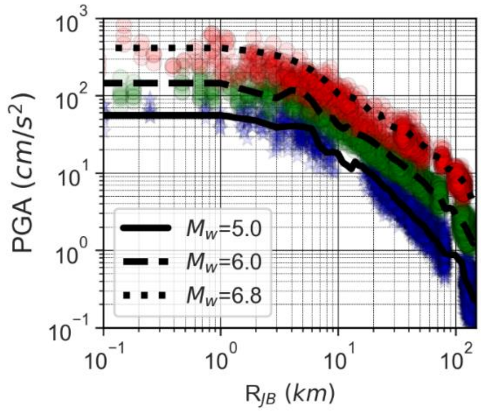
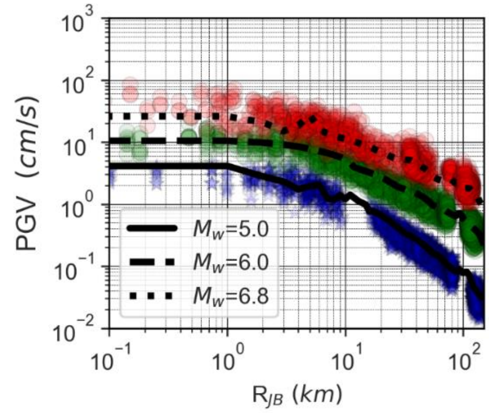


Figure 15 Variation of the selected IMs including (a) ln(PGA), (b) ln(PGV), (c) ln($PSA_{T=0.3\ s}$), and (d) ln($PSA_{T=1.5\ s}$) with respect to M_w using R_{JB} of 1 km, 10 km, 30 km, 70 km, and 130 km and FD of 8.0 km.

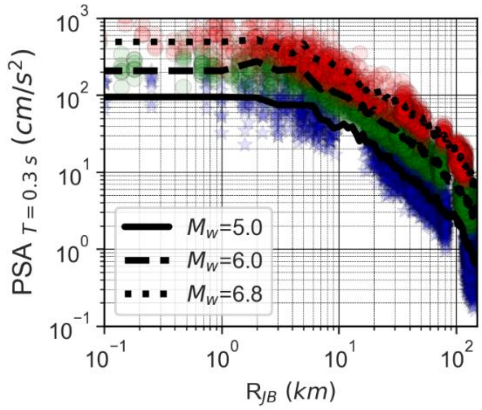
Furthermore, the trend of the GMM is compared against the change in R_{JB} for different magnitudes (5.0, 6.0, and 6.8) using various values of R_{JB} between 0-150 km. The results are plotted in Figure 16. The outcomes show that an increase in R_{JB} leads to a decrease in the PGA, PGV, and PSA levels at all period ranges, indicating that the proposed GMM effectively captures the distance-dependent attenuation. Consistent with the former observation, an increase in magnitude results in an increase in the ground motion amplitudes. It is also evident that for large magnitudes and smaller distances, the performance of the model remains the same, which emphasises the advantage of simulations over the use of real datasets.

(a) (b)





(c)



(d)

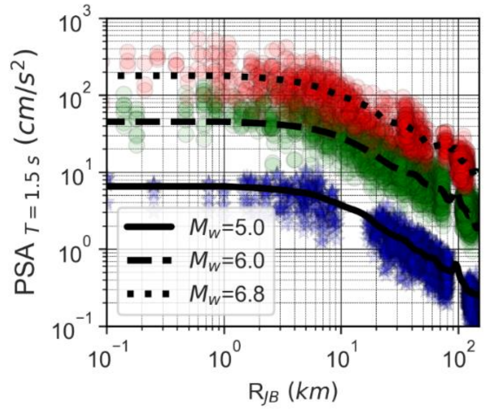


Figure 16 Variation of the selected IMs including (a) ln(PGA), (b) ln(PGV), (c) ln($PSA_{T=0.3\ s}$), and (d) ln($PSA_{T=1.5\ s}$) with respect to R_{JB} using M_w of 5.0, 6.0, and 6.8 and FD of 8.0 km.

Subsequently, the variation of PSA concerning R_{JB} is examined for three distinct moment magnitudes (5.0, 6.0, and 6.8). The results are illustrated in Figure 17. The results reveal that as the distance increases, the peak value of the PSA shifts towards longer periods, which aligns with the physical characteristics of distance-dependent damping of ground motions. Furthermore, in accordance with established earthquake physics, the event magnitude determines the extent to which the peak shifts.

Overall, the findings of the study suggest that the proposed XGBoost-GMM is capable of capturing the behaviour of empirical GMMs with minimal seismological data and without the need for nonlinear regression with multiple coefficients. The model estimates PGA, PGV, and PSA between periods of 0 and 2 s for the Azores Plateau, and its implementation requires fewer computations (as detailed in Appendix A). Furthermore, the proposed XGBoost-GMM has the potential to be applied in future studies for simulations performed on the surface considering local soil effects. Finally, the proposed model represents a promising approach for estimating ground motion parameters for seismic hazard analysis in the Azores Plateau.

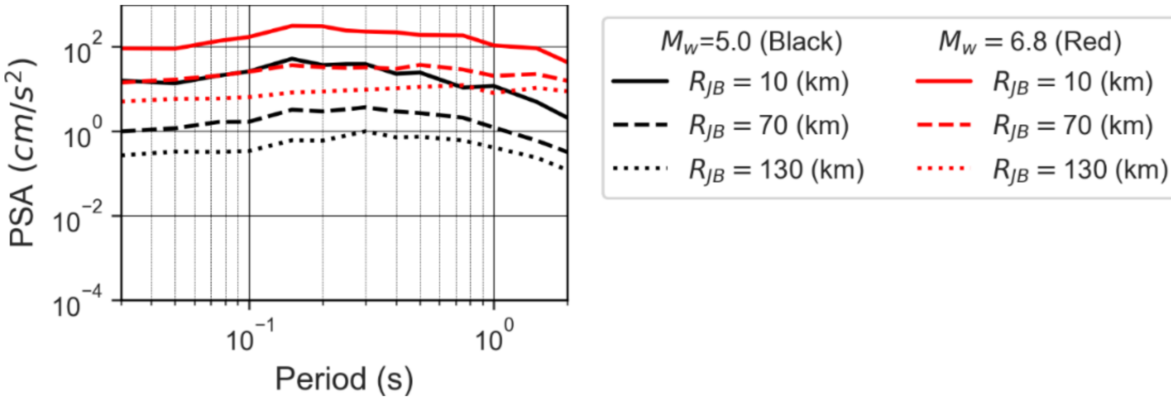


Figure 17 Variation PSA with respect to R_{JB} (including 10 km, 70 km, and 130 km) using M_w of 5.0 and 6.8 and FD of 8.0 km.

6. Conclusions

This paper proposes a machine learning-based backbone ground motion model for the Azores Plateau in Portugal, which is built using a simulated, homogenous dataset for the region. The study first discusses the tectonic activity and fault zones in the Central and Eastern Azores region, where despite the high seismic activity, there is a lack of recorded ground motion data. To address this challenge, the study uses the stochastic finite-fault approach to generate a region-specific ground motion dataset through scenario event simulations, accounting for the uncertainty in the rupture of active faults and path attenuation. To cover all magnitude ranges, this study models possible fault ruptures on both Faial and São Jorge Islands as scenarios to represent potential earthquakes in the Azores Plateau. A simulated database is generated by assuming the rupture of ten faults, nine on Faial Island, and one on São Jorge Island with maximum magnitude boundary of 6.8.

The results of simulations are verified against the well-known empirical ground motion models. The XGBoost algorithm is then used to develop a region-specific GMM, which is known for its accuracy, flexibility, and computational speed in regression problems. Finally, the performance of the developed GMM is compared to the real records of the 1998 Faial earthquake.

A concise summary of the findings of this study is as follows:

- The neotectonics and geology of the Eastern and Central Azores Islands are found to be similar, with most faulting structures responsible for major events identified on Faial Island. However, the potential rupture of the Picos fault in Sao Jorge Island can result in a maximum expected M_w of 6.8, which is higher than that reported for Faial Island (i.e., M_w=6.6).
- The developed simulated dataset effectively mimics the real behaviour of earthquake motions, as evidenced by the increase in amplitudes of spectral ordinates at higher periods with an increase in earthquake magnitude. This observation further validates the accuracy and correctness of the simulations.
- This study successfully validates the simulated dataset using well-known GMMs, including Boore and Atkinson [30] (BA08), Akkar and Çağnan [20] (AC10), and Akkar et al. [31] (ASB14), as the trend of simulated intensity measures lies within one standard deviation of the predicted mean values from these equations. The results indicate that the simulations are consistent with the AC10 relationship for PGA regardless of the magnitude, while PGV results vary depending on the magnitude of the scenario event. This observation highlights the importance of developing region-specific GMMs to ensure accurate ground motion simulations for practical applications.
- The developed XGBoost-GMM demonstrated a strong performance for both the training and testing datasets, with a high level of agreement observed between the predicted and observed values of the selected IMs, such as PGA, PGV, and PSA at T=0.3 s and T=1.5 s. Furthermore, acceptable model performance values are obtained for all IMs, with the coefficient of determination and correlation coefficient values exceeding 0.90. However, a decrease in these performance indicators is observed for PGV and larger periods of the PSA compared to PGA and PSA with smaller periods, as confirmed by error indices. This suggests that the model may perform better (i.e., with less uncertainty) for PGA and spectral ordinates of shorter periods compared to longer periods and PGV, consistent with existing empirical GMMs. These results should be taken into consideration when applying the model in practice.

- The study evaluates the potential bias of the developed GMM by analysing the inter-event and intra-event uncertainties with respect to source and site parameters. The residuals are found to be unbiased across all considered intensity measures and ranges of magnitude and distance, as evidenced by the absence of observable patterns in the mean residual. P-values also confirm the null hypothesis of unbiased estimates. The GMM is deemed independent of explanatory variables, as the mean residuals for all intensity measures fluctuate around zero. Additionally, unlike GMMs based on real records, the residual uncertainty remains constant with the increasing magnitude or decreasing distance, highlighting the advantage of using simulated ground motions over real datasets.
- The developed XGBoost-GMM model exhibited an acceptable uncertainty range and performed well compared to existing models, as shown by the inter-event, intra-event, and total residuals for PGA, PGV, and all spectral ordinates. The analysis revealed that the inter-event uncertainty is smaller than the intra-event uncertainty for all spectral values, consistent with previous literature, but could also be attributed to the use of the same ground motion simulation approach in a single region. The total residual increases with an increase in the spectral period, in line with previous findings.
- Based on the evaluation conducted, the XGBoost-GMM model is found to effectively mimic real earthquake phenomena by demonstrating magnitude-dependent increase and distance-dependent decrease. The analysis reveals that as the distance increases, the peak value of the PSA shifts towards longer periods, which aligns with the physical characteristics of distance-dependent damping of ground motions. The event magnitude also determines the extent to which the peak shifts, which is consistent with established earthquake physics. Furthermore, the performance of the model remains the same for large magnitudes and smaller distances, which highlights the advantage of simulations over real datasets. These findings provide evidence of the model's reliability in effectively simulating and analysing earthquake events.

Overall, the findings of this study have significant implications for seismology research and engineering. The XGBoost GMM is limited in its ability to extrapolate beyond the input range of predictor variables due to the machine learning algorithm's lack of adherence to underlying physical formulations. Therefore, it is advisable to utilize the model that has been developed for the Azores Plateau for shallow seismic events with a magnitude (M_w) range of 5.0 to 6.8 and a distance (R_{JB}) of up to 150 km in the bedrock. Finally, it should be acknowledged that the GMM proposed in this study serves as a backbone model. However, to account for soil effects, future research should be carried out with additional verifications on the seismic data collected from the stations that captured the Faial earthquake of 1998. This would enhance the accuracy and reliability of the GMM for seismic hazard assessment in the future.

Acknowledgement

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 (doi.org/10.54499/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 partially financed by national funds through FCT - Foundation for Science and Technology, under grant agreement 2020.08876.BD attributed to the second author.

References

1. Mizutori M, Guha-Sapir D. Human cost of disasters: an overview of the last 20 years (17 p.). Belgium: Undrr. Retrieved in 2020, January 21 2020.
2. Martino P, Daniele E, Thomas K, Alice S, Aneta F, Christina C. Atlas of the Human Planet 2017: Global Exposure to Natural Hazards 2017.
3. McGuire RK. Deterministic vs. probabilistic earthquake hazards and risks. *Soil Dynamics and Earthquake Engineering* 2001; **21**(5): 377–384.

4. Douglas J. Calibrating the backbone approach for the development of earthquake ground motion models. *Best Practice in Physics-Based Fault Rupture Models for Seismic Hazard Assessment of Nuclear Installations: Issues and Challenges Towards Full Seismic Risk Analysis* 2018.
5. Douglas J. Ground motion prediction equations 1964–2021. *Dept Civil Environ Eng, Univ Strathclyde, Glaskow, UK* 2022: 670.
6. Rezaeian S, Xiaodan S. Stochastic ground motion simulation 2014.
7. Askan A, Karimzadeh S, Asten M, Kilic N, ŞİŞMAN FN, Erkmen C. Assessment of seismic hazard in the Erzincan (Turkey) region: construction of local velocity models and evaluation of potential ground motions. *Turkish Journal of Earth Sciences* 2015; **24**(6): 529–565.
8. Tsioulou A, Taflanidis AA, Galasso C. Modification of stochastic ground motion models for matching target intensity measures. *Earthquake Engineering & Structural Dynamics* 2018; **47**(1): 3–24. DOI: 10.1002/eqe.2933.
9. Vetter CR, Taflanidis AA, Mavroeidis GP. Tuning of stochastic ground motion models for compatibility with ground motion prediction equations. *Earthquake Engineering & Structural Dynamics* 2016; **45**(6): 893–912. DOI: 10.1002/eqe.2690.
10. Askan A, Karimzadeh S, Bilal M. Seismic intensity maps for North Anatolian Fault Zone (Turkey) based on recorded and simulated ground motion data, Book Chapter, Neotectonics and Earthquake Potential of the Eastern Mediterranean Region 2017.
11. Karimzadeh S, Ozsarac V, Askan A, Erberik MA. Use of simulated ground motions for the evaluation of energy response of simple structural systems. *Soil Dynamics and Earthquake Engineering* 2019; **123**: 525–542.
12. Karimzadeh S, Kadas K, Askan A, Erberik MA, Yakut A. Derivation of analytical fragility curves using SDOF models of masonry structures in Erzincan (Turkey). *Earthquakes and Structures* 2020; **18**(2): 249–261.
13. Atkinson GM, Boore DM. Recent trends in ground motion and spectral response relations for North America. *Earthquake Spectra* 1990; **6**(1): 15–35.
14. Campbell KW. 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. *Bulletin of the Seismological Society of America* 2003; **93**(3): 1012–1033.
15. Douglas J. Ground motion estimation equations 1964-2003' Intermediary Report 2006(December).
16. Megawati K, Pan TC, Koketsu K. Response spectral attenuation relationships for Sumatran-subduction earthquakes and the seismic hazard implications to Singapore and Kuala Lumpur. *Soil Dynamics and Earthquake Engineering* 2005; **25**(1): 11–25.
17. Rietbrock A, Strasser F, Edwards B. A stochastic earthquake ground-motion prediction model for the United Kingdom. *Bulletin of the Seismological Society of America* 2013; **103**(1): 57–77.
18. Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra* 2014; **30**(3): 1025–1055. DOI: 10.1193/070913EQS198M.
19. Yaghmaei-Sabegh S, Karimzadeh S, Ebrahimi M, Ozsarac V, Du W. A new region-specific empirical model for prediction of ground motion significant duration in Turkey. *Bulletin of Earthquake Engineering* 2022; **20**(10): 4919–4936. DOI: 10.1007/s10518-022-01417-9.
20. Akkar S, Çağnan Z. A local ground-motion predictive model for Turkey, and its comparison with other regional and global ground-motion models. *Bulletin of the Seismological Society of America* 2010; **100**(6): 2978–2995. DOI: 10.1785/0120090367.
21. Alavi AH, Gandomi AH. Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Computers and Structures* 2011; **89**(23–24): 2176–2194. DOI: 10.1016/j.compstruc.2011.08.019.
22. Khosravikia F, Clayton P. Machine learning in ground motion prediction. *Computers and Geosciences* 2021; **148**. DOI: 10.1016/j.cageo.2021.104700.
23. Raghucharan MC, Somala SN, Erteleva O, Rogozhi E. Seismic attenuation model for data gap regions using recorded and simulated ground motions. *Natural Hazards* 2021; **107**(1): 423–446. DOI: 10.1007/s11069-021-04589-w.

24. Seo H, Kim J, Kim B. Machine-Learning-Based Surface Ground-Motion Prediction Models for South Korea with Low-to-Moderate Seismicity. *Bulletin of the Seismological Society of America* 2022; **112**(3): 1549–1564. DOI: 10.1785/0120210244.
25. Sreenath V, Podili B, Raghukanth STG. A hybrid non-parametric ground motion model for shallow crustal earthquakes in Europe. *Earthquake Engineering & Structural Dynamics* 2023. DOI: 10.1002/eqe.3845.
26. Mohammadi A, Karimzadeh S, Amir S, Ozsarac V, Lourenço PB. The potential of region-specific machine-learning-based ground motion models : Application to Turkey. *Soil Dynamics and Earthquake Engineering* 2023; **172**(January): 108008. DOI: 10.1016/j.soildyn.2023.108008.
27. Yerlikaya-Özkurt F, Askan A, Weber GW. An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method. *Engineering Optimization* 2014; **46**(12): 1651–1668. DOI: 10.1080/0305215X.2013.858141.
28. Assatourians K, Atkinson G. EXSIM12: A Stochastic Finite-Fault Computer Program in FORTRAN 2012.
29. Carvalho A, Reis C, Vales D. Source and high-frequency decay parameters for the Azores region for stochastic finite-fault ground motion simulations. *Bulletin of Earthquake Engineering* 2016; **14**: 1885–1902.
30. Boore DM, Atkinson GM. 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. *Earthquake Spectra* 2008; **24**(1): 99–138.
31. Akkar S, Sandıkkaya MA, Bommer JJ. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bulletin of Earthquake Engineering* 2014; **12**(1): 359–387. DOI: 10.1007/s10518-013-9461-4.
32. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016; **13-17-August-2016**: 785–794. DOI: 10.1145/2939672.2939785.
33. Lourenço N, Miranda JM, Luis JF, Ribeiro A, Madeira J, Needham HD. Morpho-tectonic analysis of the Azores Volcanic Plateau from a new bathymetric compilation of the area. *Marine Geophysical Research* 1998; **20**(3): 141.
34. Silveira G, Stutzmann E, Davaille A, Montagner JP, Mendes-Victor L, Sebai A. Azores hotspot signature in the upper mantle. *Journal of Volcanology and Geothermal Research* 2006; **156**(1–2): 23–34.
35. Madeira J, Brum da Silveira A, Hipólito A, Carmo R. Chapter 3 active tectonics in the central and eastern Azores islands along the Eurasia–Nubia boundary: a review. *Geological Society, London, Memoirs* 2015; **44**(1): 15–32.
36. Vogt PR, Jung WY. The Terceira Rift as hyper-slow, hotspot-dominated oblique spreading axis: A comparison with other slow-spreading plate boundaries. *Earth and Planetary Science Letters* 2004; **218**(1–2): 77–90. DOI: 10.1016/S0012-821X(03)00627-7.
37. Heezen BC, Tharp M, Ewing M. *The floors of the oceans*. vol. 65. Geological Society of America New York; 1959.
38. Hipólito A, Madeira J, Carmo R, Gaspar JL. Neotectonics of Graciosa Island (Azores): A contribution to seismic hazard assessment of a volcanic area in a complex geodynamic setting. *Ann Geophys* 2013; **56**(6): 1–18.
39. Carmo R, Madeira J, Ferreira T, Queiroz G, Hipólito A. Chapter 6 Volcano-tectonic structures of São Miguel Island, Azores. *Geological Society, London, Memoirs* 2015; **44**(1): 65–86.
40. Bezzeghoud M, Adam C, Buforn E, Borges JF, Caldeira B. Seismicity along the Azores-Gibraltar region and global plate kinematics. *Journal of Seismology* 2014; **18**: 205–220.
41. Matias L, Dias NA, Morais I, Vales D, Carrilho F, Madeira J, *et al.* The 9th of July 1998 Faial Island (Azores, North Atlantic) seismic sequence. *Journal of Seismology* 2007; **11**: 275–298.
42. Senos ML, Gaspar JL, Cruz J, Ferreira T, Nunes JC, Pacheco J, *et al.* O terramoto do Faial de 9 de Julho de 1998. *Proceedings do*, vol. 1, 1998.

43. Madeira J, Brum da Silveira A. Active tectonics and first paleoseismological results in Faial, Picoand S. Jorge islands (Azores, Portugal). *Annals of Geophysics* 2003.
44. Wells DL, Coppersmith KJ. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 1994; **84**(4): 974–1002.
45. Hirn A, Haessler H, Trong PH, Wittlinger G, Victor LAM. Aftershock sequence of the January 1st, 1980, earthquake and present-day tectonics in the Azores. *Geophysical Research Letters* 1980; **7**(7): 501–504.
46. Silva M. Caracterização da sismicidade histórica dos Açores com base na reinterpretação de dados de macrossísmica: contribuição para a avaliação do risco sísmico nas ilhas do Grupo Central. *University of Azores, Ponta Delgada (MSc Thesis)* 2005.
47. Gaspar JL, Queiroz G, Ferreira T, Medeiros AR, Goulart C, Medeiros J. Earthquakes and volcanic eruptions in the Azores region: geodynamic implications from major historical events and instrumental seismicity. *Volcanic Geology of Sao Miguel Island (Azores Archipelago) Geological Society, London, Memoirs* 2015; **44**: 33–49.
48. Boore DM. Strong-motion seismology. *Reviews of Geophysics* 1983; **21**(6): 1308–1318.
49. Brune JN. Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research* 1970; **75**(26): 4997–5009.
50. Boore DM. Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics* 2003; **160**: 635–676.
51. Beresnev IA, Atkinson GM. FINSIM--a FORTRAN program for simulating stochastic acceleration time histories from finite faults. *Seismological Research Letters* 1998; **69**(1): 27–32.
52. Motazedian D, Atkinson GM. Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America* 2005; **95**(3): 995–1010.
53. Boore DM. Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bulletin of the Seismological Society of America* 2009; **99**(6): 3202–3216.
54. Atkinson GM, Assatourians K. Implementation and validation of EXSIM (a stochastic finite-fault ground-motion simulation algorithm) on the SCEC broadband platform. *Seismological Research Letters* 2015; **86**(1): 48–60.
55. Atkinson GM, Goda K, Assatourians K. Comparison of nonlinear structural responses for accelerograms simulated from the stochastic finite-fault approach versus the hybrid broadband approach. *Bulletin of the Seismological Society of America* 2011; **101**(6): 2967–2980.
56. Karimzadeh S, Lourenço PB. Stochastic Ground Motion Simulation of the 9th of July 1998 Faial Earthquake (Azores, North Atlantic). *Authorea Preprints* 2021.
57. Boore DM, Joyner WB, Fumal TE. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seismological Research Letters* 1997; **68**(1): 128–153.
58. Kubo H, Kunugi T, Suzuki W, Suzuki S, Aoi S. Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Scientific Reports* 2020; **10**(1). DOI: 10.1038/s41598-020-68630-x.
59. Frazier PI. A tutorial on Bayesian optimization. *ArXiv Preprint ArXiv:180702811* 2018.
60. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems* 2012; **25**.
61. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems* 2011; **24**.
62. Withers KB, Moschetti MP, Thompson EM. A machine learning approach to developing ground motion models from simulated ground motions. *Geophysical Research Letters* 2020; **47**(6): e2019GL086690.

Appendix A. Creating a Web-Based Software Application

In this study, Streamlit was employed to construct a graphical user interface (GUI) tool that enables easy access to the GMM developed by XGBoost. The code for the GMM can be found at

https://github.com/amirxdbx/GMM_Azores. The user interface of the tool is illustrated in Figure A, and it can be accessed at https://amirxdbx-gmm-azores-deploy-36glao.streamlit.app/.
As depicted in Figure A, the tool allows users to define the characteristics of a scenario earthquake in terms of M_W, R_{JB}, and FD. The software provides the predicted values of PGA, PGV, and PSA for periods between 0 and 2.0 seconds. Overall, this web-based application software provides a user-friendly interface for estimating ground motion parameters in the bedrock using the proposed XGBoost-GMM for the Azores Plateau.

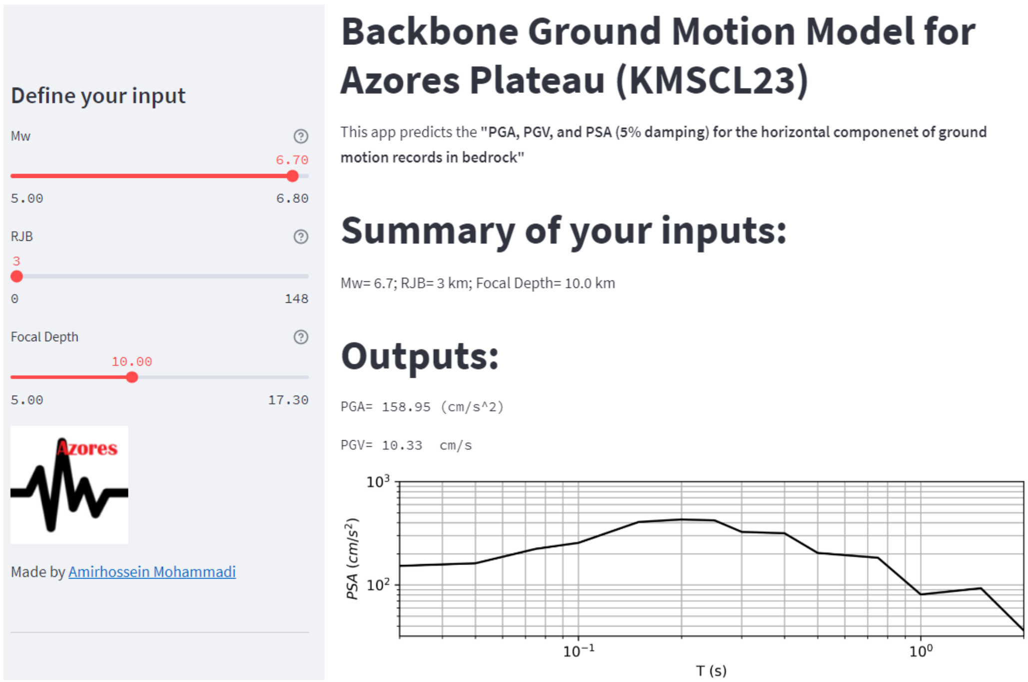


Figure A GUI of the XGBoost-GMM

Authors Contributions

The authors' contributions to the paper are as follows: Shaghayegh Karimzadeh performed the ground motion simulations, analysed, and interpreted the results, and prepared the initial manuscript. She also collaborated with the second author to establish the ground motion model and supervised the third author. Amirhossein Mohammadi developed the XGBoost-GMM using Python, created all the figures and prepared the initial manuscript. Usman Salahuddin assisted with preparing the first draft by creating figures depicting the geology and neotectonics of the region. Alexandra Carvalho contributed to defining the tectonic features of the region, as well as manuscript review and editing. Paulo B Lourenço primarily contributed to funding acquisition, supervision, manuscript reviewing, and editing. All authors reviewed the findings and approved the final version of the manuscript.

Data availability

The data of this study are available on request from the authors.