

Backbone ground motion model through simulated records and XGBoost machine learning algorithm: An application for the Azores plateau (Portugal)

Shaghayegh Karimzadeh¹  | Amirhossein Mohammadi¹  | Usman Salahuddin² | Alexandra Carvalho³ | Paulo B. Lourenço¹ 

¹Department of Civil Engineering, ISISE, ARISE, University of Minho, Campus de Azurém, Guimarães, Portugal

²Department of Civil Engineering, University of Camerino, Camerino, Italy

³Structural Department, LNEC, Lisbon, Portugal

Correspondence

Shaghayegh Karimzadeh, Department of Civil Engineering, ISISE, ARISE, University of Minho, Campus de Azurém, Guimarães, Portugal.
Email: shaghkn@civil.uminho.pt

Funding information

Foundation for Science and Technology (FCT), Grant/Award Numbers: 2020.08876.BD, UIDB/04029/2020, LA/P/0112/2020; H2020 European Research Council, Grant/Award Number: No.833123

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 results undergo validation against the 1998 Faial event and comparison with empirical models for volcanic and Pan-European datasets. 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 (M_w), Joyner and Boore distance (R_{JB}), and focal depth (FD). The quantification of GMM uncertainty is accomplished by analyzing the residuals, providing insight into inter- and intra-event uncertainties. The outcomes demonstrate the effectiveness of the suggested model in simulating physical phenomena.

KEYWORDS

Azores plateau (Portugal), ground motion model (GMM), XGBoost, stochastic finite-fault ground motion simulation

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 urbanization. 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 utilized to develop GMMs. If such data is unavailable, ground motion 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} Douglas⁵ 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} Nonetheless, the incorporation of more intricate source, site, and path behaviors into the regression-based approaches with pre-defined mathematical forms remains a challenging task. To overcome this, machine learning (ML) algorithms can be used to capture the nonlinear and complex behavior 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 of⁵ summarizes 39 nonparametric GMMs worldwide.¹⁹

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 by²⁸ 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 effectiveness of the simulations is first assessed through a validation process, where the simulated records of the 1998 Faial event are compared with real seismic records. Subsequently, the simulated dataset is compared against well-established empirical models, specifically those by,^{30–33} in order to underscore the importance of developing region-specific models. Finally, to develop the region-specific GMM in the study area, the XGBoost algorithm is implemented.³⁴ The performance of the model is evaluated through the coefficient of determination (R^2), Pearson correlation 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 behavior of earthquake ground 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.³⁵ 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

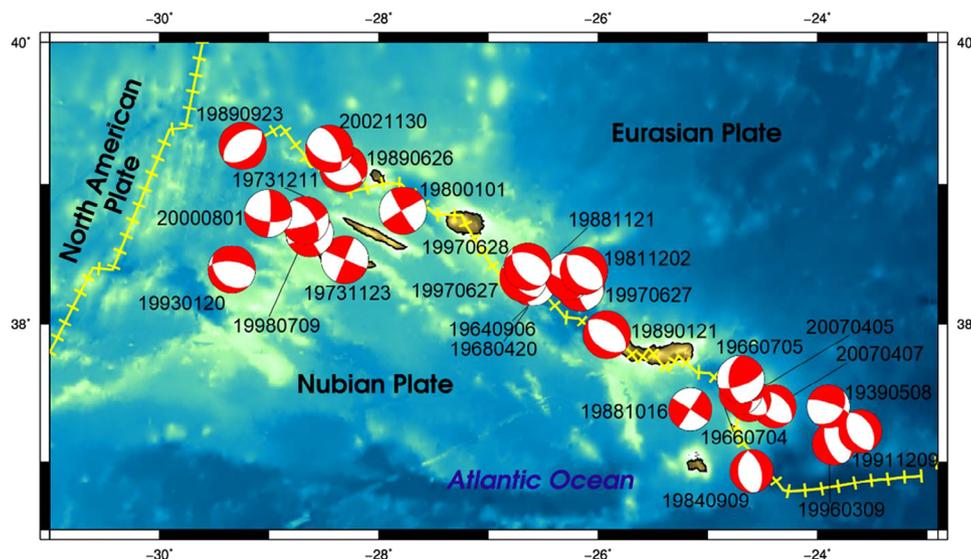


FIGURE 1 Map of central and eastern Azores with tectonic plates and records.²⁹

indicate the existence of a plume in the region.³⁶ The Azores archipelago's central and eastern island groups are located near the western section of the Eurasian-African plate boundary.³⁷ Azores Gibraltar Fracture Zone is the Atlantic part of this plate boundary having three sections with distinctive morphology and seismotectonic regimes.³⁷ 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,^{35,38} and is considered to migrate towards north from east fracture zone.³⁹

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.^{40,41} 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.⁴² 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,⁴³ and are examples of large magnitude events that inflicted significant damage. The 1998 event had an epicenter located offshore, about 10 km NE of Faial Island. The maximum recorded intensity was VIII in NE Faial towards the epicenter, and significant local amplification effects were observed.⁴⁴ 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 1500 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.⁴³

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

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.⁴⁵ A WNW-ESE trending graben structure named the

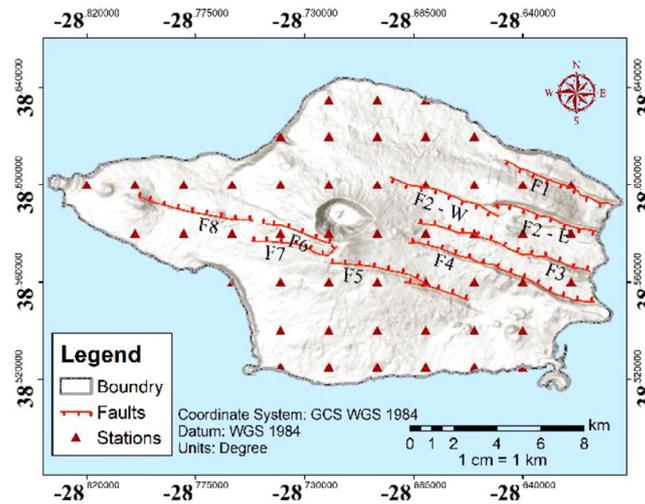


FIGURE 2 Active tectonics in Faial Island with the stations used for simulations shown by triangular symbols.

TABLE 1 Information on the active faults of Faial Island.³⁷

No	Fault name	Fault rupture length (km)	M_w -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

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 summarizes information on the active faults in Faial Island. It is evident that the fault mechanism is predominantly normal for all, with the maximum expected M_w of 6.6 due to the potential rupture of the Espalamaca fault.

It should be emphasized 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.

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 trending 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.⁴⁵ Figure 3 presents the tectonic map of Pico Island, while Table 2 provides details on the main active faults. Using the correlations of,⁴⁶ the maximum expected M_w for this island ranges from 6.1 to 6.6 with a normal faulting mechanism.

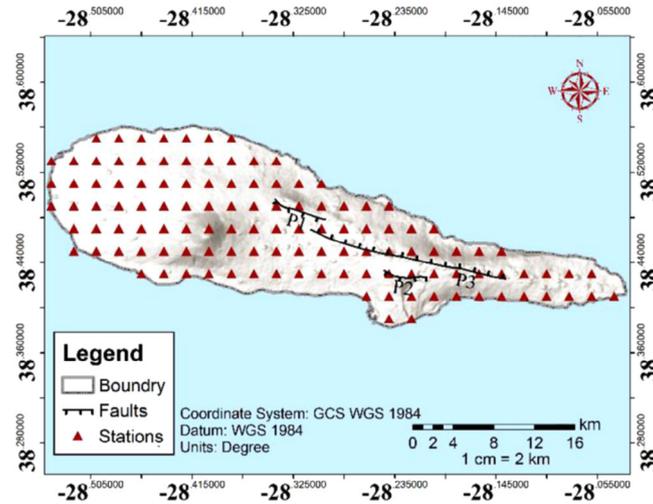


FIGURE 3 Active tectonics in Pico Island with the stations used for simulations shown by triangular symbols.

TABLE 2 Information on the active faults of Pico Island.³⁷

No	Fault name	Fault rupture length (km)	M_w -max	Fault mechanism	Strike (°)	Dip (°)
P1	Lagoa do Capitaio	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	-

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.

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 characterize 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.⁴⁵ 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.⁴⁵ 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 6.8.⁴⁶

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

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 epicenters 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.^{47,48} 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

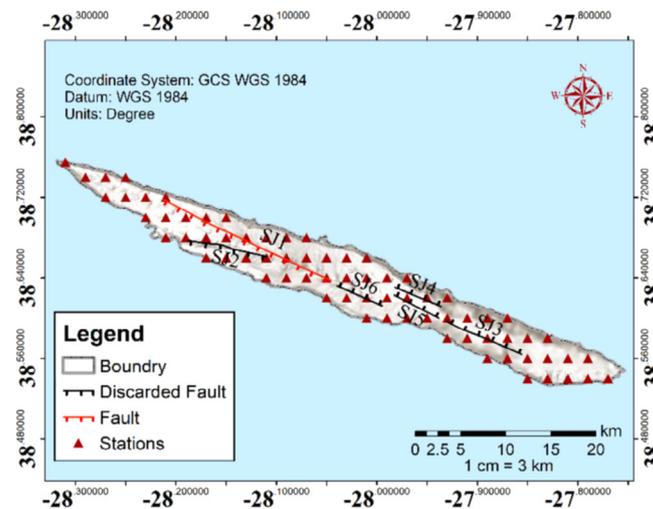


FIGURE 4 Active tectonics in São Jorge Island with the stations used for simulations shown by triangular symbols.

TABLE 3 Information on the active faults of Sao Jorge Island.³⁷

No	Fault name	Fault rupture length (km)	M_{w-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	-

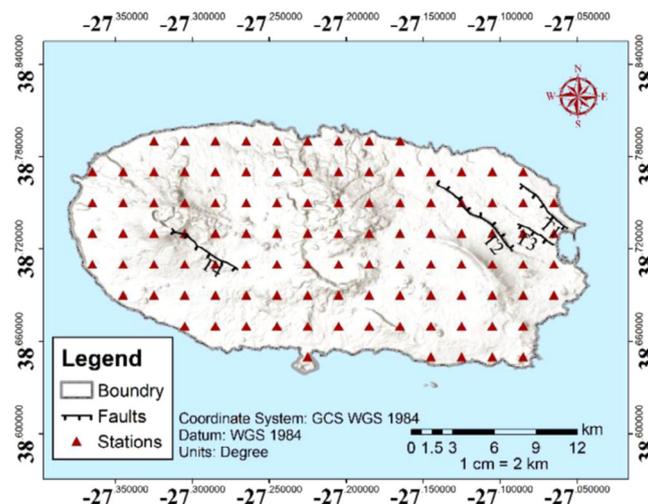


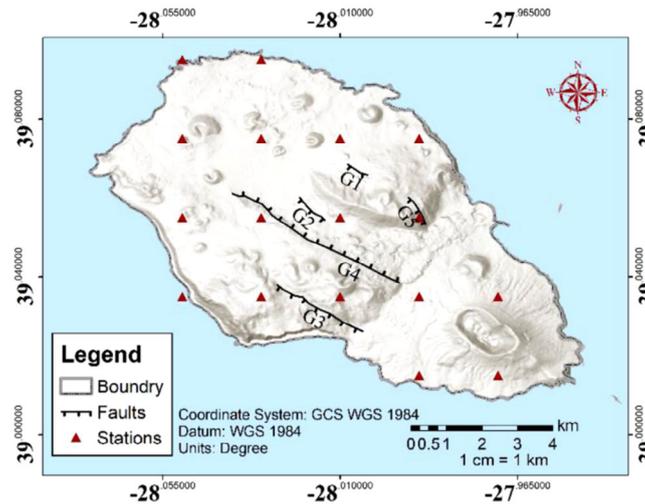
FIGURE 5 Active tectonics in Terceira Island with the stations used for simulations shown by triangular symbols.

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.⁴⁶

It is noteworthy that Figure 5 illustrates a total of 104 stations that have been utilized for simulations in the subsequent section of the study, and these stations are marked by triangular symbols. Table 4 also provides details on the main active faults.

TABLE 4 Information on the active faults of Terceira Island.³⁷

No	Fault name	Fault rupture length (km)	M_{w-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 6** Active Tectonics in Graciosa Island with the stations used for simulations shown by triangular symbols.**TABLE 5** Information on the active faults of Graciosa Island.³⁷

No	Fault name	Fault rupture length (km)	M_{w-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	-

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.^{40,49} 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 symbolized 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.⁴⁶

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 also lists details on the main active faults.

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 modeled 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 color in Figures 2 and 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,⁵⁰ includes the following shear wave acceleration spectrum at an observation point:

$$A(f) = CM_o \left[(2\pi f)^2 / \left(1 + (f/f_c)^2 \right) \right] G(R) e^{-\pi f R / (Q(f)\beta)} e^{(-\pi f \kappa_o)} S(f) \quad (1)$$

where M_o 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, 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.⁵¹ $G(R)$ 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.^{52–54} In subsequent developments, Boore⁵⁵ enhanced the approach proposed by.⁵⁴ 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,⁵⁶ this study employs the latest version of the stochastic finite-fault ground motion simulation methodology to model acceleration time series of scenario earthquakes.⁵⁶ The algorithm proposed by,⁵⁴ which was developed based on the original FINSIM code by,⁵³ is enhanced in this technique by adding the improvements suggested by.⁵⁵ 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 recognize 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.⁵¹ Depending on how far a sub-source is from the hypocenter, 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:

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

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,^{52,55} Δt_i is the sum of the fracture initiation and time delay due to the distance of the i^{th} sub-source from the hypocenter, the term T_i relates to a fraction of rise time considered for additional randomization and finally, the term H_i resembles the normalization factor of the i^{th} sub-source introduced for the conservation of energy with the following formula:

$$H_i = \frac{M_o}{M_{oi}} \sqrt{\sum_j \left(\frac{f_{0j}^2 f_j^2}{f_0^2 + f_j^2} \right)^2 / N \sum_j \left(\frac{f_{0i}^2 f_j^2}{f_{0i}^2 + f_j^2} \right)^2} \quad (3)$$

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

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:

$$M_{0i} = \frac{M_0 \cdot s_i}{\sum_{i=1}^N s_i} \quad (4)$$

$$f_{ci} = 4.9 \times 10^6 \beta_s \left(\frac{\Delta\sigma}{p \times M_0} \right)^{\frac{1}{3}} \text{ where } p = \begin{cases} \frac{N_R}{N} & \text{if } N_R < N \times PP \\ PP & \text{if } N_R \geq N \times PP \end{cases} \quad (5)$$

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 summarized in Table 6. Karimzadeh and Lourenço⁵⁷ 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 Karimzadeh and Lourenço⁵⁷ and Karimzadeh et al.⁵⁸ 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 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 ²⁹ Depth = 14.0 km \rightarrow 2.93
Shear wave velocity (km/s)	Depth = 0.0 km \rightarrow 3.1 Depth = 2.5 km \rightarrow 3.7 ²⁹ Depth = 8.0 km \rightarrow 4.2 Depth = 14.0 km \rightarrow 4.6
Shear wave velocity/crustal velocity	0.8
Geometric spreading	$R^{-1.0}$ $R \leq 1.5D$ km $R^{0.0}$ $1.5D$ km $< R \leq 2.5D$ 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
Hypocenter 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	$(76 \pm 11)f^{0.69 \pm 0.09}$ ²⁹	Lognormal

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 by²⁹ is used to implement regional models with their PDFs. Parameters of Table 8 are utilized 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 Figures 2–6.

3.3 | Ground motion database

Simulations of this study result in 247,710 ground motion records for the entire AP. All simulated data undergo filtering using baseline correction and a 4th-order Butterworth filter with a frequency range between 0.1 to 25 Hz. 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 normalized 5% damped Pseudo Spectral Acceleration (PSA) of the Azores dataset for different magnitude intervals is shown in Figure 8. This figure displays the normalized 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. The observed variations primarily stem from the distance effect, with the plot incorporating data from all stations where ground motion simulations are conducted. Another source of variation relates to the uncertainty in region-specific seismotectonic parameters. This consideration is addressed using

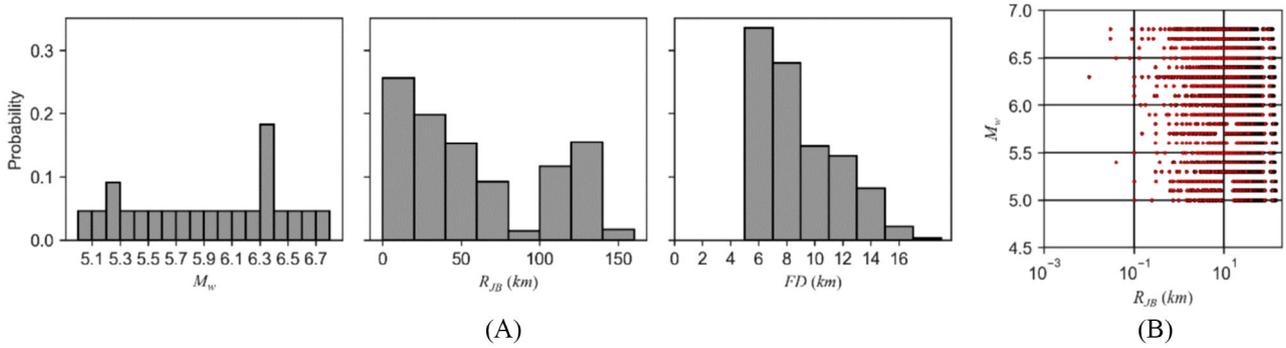


FIGURE 7 (A) Histograms of seismological characteristics of the Azores ground motion dataset, and (B) distribution of M_w versus R_{JB} for the Azores ground motion dataset.

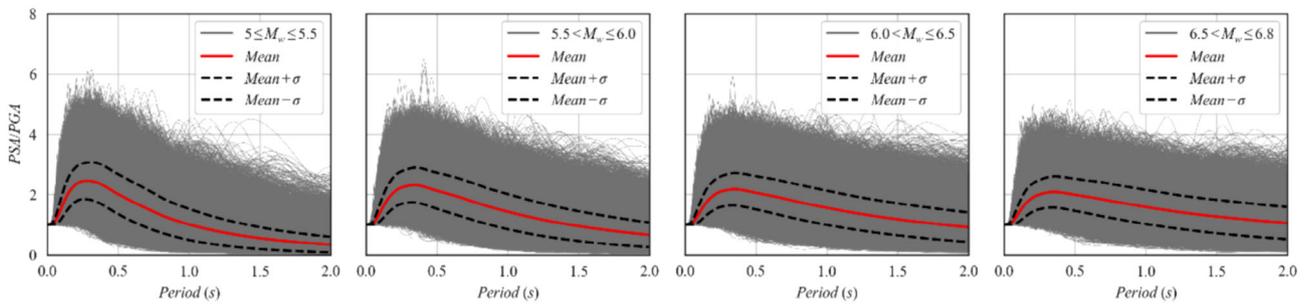


FIGURE 8 Normalized 5% damped PSA of the Azores dataset.

the MCS method, allowing for a comprehensive exploration of parameter uncertainties. It is essential to highlight that Carvalho et al.²⁹ determined uncertainties using real records from the Azores region. Nonetheless, a direct comparison of simulations with real records is not feasible due to the discrepancy in recording depths. The recorded motions correspond to past events at the surface level, whereas the simulations herein are at the bedrock, leading to distinct datasets that cannot be directly compared. Finally, 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.

3.4 | Comparison of simulations with recorded motions and existing ground motion models

Simulations are typically validated by comparing their ability to estimate observed records of past earthquakes.^{59–61} In cases where recorded motions are not available, their trend is compared with appropriate existing GMMs.^{62–65} In this study, we begin by validating the model through simulations of the 1998 earthquake, comparing the results with recorded motions from that event. Subsequently, we conduct a comparative evaluation with the most relevant existing GMMs, highlighting the imperative for the development of region-specific models.

3.4.1 | Model validation

The input-model parameters for simulations used in this research, such as tectonic information, geometric and anelastic attenuation models, rupture velocity, etc., were previously developed based on the Azores ground motion dataset.²⁹ The study of Carvalho et al.²⁹ estimated source and attenuation parameters of the Azores region based on the past events recorded in the AP through records obtained by the Portuguese digital seismic and accelerometer network. These parameters have already been validated for the 1998 Faial event with a magnitude of 6.2 in.^{57,58} The authors utilized alternative sets of input-model parameters to optimize the alignment with recorded motions of the 1998 Faial event. Their analysis determined that “Set2” yielded the most favorable fit, as evidenced by the highest Goodness of Fit (GOF) scores.⁶⁶ Notably,

TABLE 9 Goodness of Fit (GOF) scores for the real and simulated records of the 1998 Faial earthquake ($M_w = 6.2$).

Station	GOF	Fit type
GZC	68	Very good
PVI	67	Very good
HOR	72	Very good
SEB	57	Fair

TABLE 10 Comparison of the real and simulated ground motion parameters of the 1998 Faial earthquake at the available four stations with their respective information.

Station Code	Latitude (°)	Longitude (°)	R_{epi} (km)	$PGA-EW$ (cm/s ²)	$PGA-NS$ (cm/s ²)	PGA -simulated (cm/s ²)	$PGV-EW$ (cm/s)	$PGV-NS$ (cm/s)	PGV -simulated (cm/s)
GZC	39.084	-28.006	72	17.76	14.62	16.04	0.90	1.02	1.25
PVI	38.726	-27.057	132	8.32	10.06	9.09	0.76	0.73	1.42
HOR	38.530	-28.630	11	418.10	399.16	422.64	31.94	37.38	21.71
SEB	38.668	-27.088	129	17.17	21.72	10.72	1.40	1.98	1.14

the parameters associated with “Set2” align with those proposed by.²⁹ The authors consistently reported very good GOF scores for most available stations that recorded the 1998 Faial earthquake ($M_w = 6.2$). Table 9 displays the GOF scores calculated by^{57,58} for the four available stations that recorded the Faial 1998 event, along with their respective information. It is noted that in their study, a wide range of IMs were employed to calculate GOF scores. These metrics include PGA , Peak Ground Velocity (PGV), peak ground displacement, PGV/PGA , Arias intensity, cumulative absolute velocity, acceleration spectrum intensity, modified acceleration spectrum intensity for the period range of 0.1 to 2.5 seconds,⁶⁷ velocity spectrum intensity, Housner intensity, bracketed duration, Fourier Amplitude Spectra (FAS) within the frequency range of 0.1 to 25 Hz, and PSA within the period range of 0 to 4 seconds using 5% damping ratio. The evidenced GOF scores for a wide range of ground motion intensity parameters affirm the validation of the input-model parameters specific to the region. The authors, as mentioned in,⁵⁸ further confirmed the accuracy of their simulations in evaluating the seismic response of prototype masonry structures, as well as in assessing the damage incurred by a church in the area affected by the 1998 Faial event.

For a detailed examination of the simulations at the stations, Table 10 lists the real and simulated peak ground motion parameters in terms of PGA and PGV , highlighting their concordance. Additionally, Figure 9 schematically compares the full time series of the real and recorded motions at these stations along with FAS and PSA using a 5% damping ratio. This graphical representation reinforces the earlier findings across time, frequency, and spectral domains. Consequently, this substantiates the regional specificity of the chosen input-model parameters. Thus, during the ground motion simulation process, the utilization of region-specific input-model parameters, which have been derived from real ground motion datasets and further validated for the 1998 Faial event, instils confidence in the model’s ability to generate a region-specific ground motion dataset. Nonetheless, it is of utmost importance to underscore that the validation process, involving a comparison with the 1998 Faial event, inherently incorporates soil effects, primarily due to the absence of bedrock stations. In contrast, this study focuses exclusively on bedrock simulations, intentionally excluding soil effects, a decision grounded in the confidence of validation within the bedrock domain.

Finally, it is important to note that the development of stochastic-based ground motion synthetics associated with seismological finite-fault modeling is a worldwide approach, allowing the reproducing of specific source, path and crustal effects, as long as the input-model parameters are calibrated and verified with earthquakes recorded in the studied region. The study by⁶⁸ also highlighted that the stochastic model employed by⁶⁹ demonstrates a strong match in describing the attenuation of PGA in seismic records from both south Iceland and the Azores earthquakes. The similarity in the simulation approach underscores the method’s effectiveness in capturing attenuation characteristics within such regions when appropriate region-specific parameters are utilized.

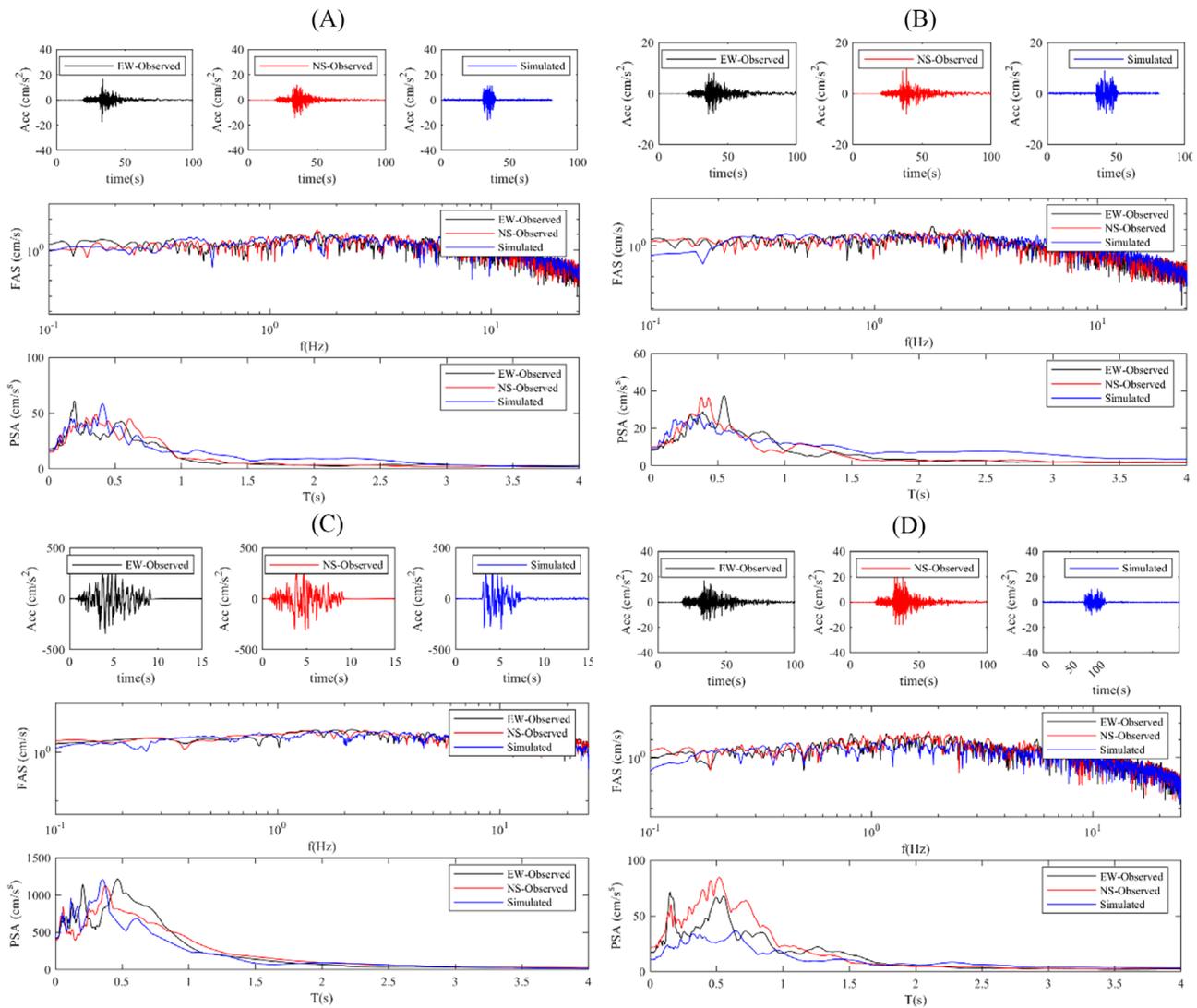


FIGURE 9 Comparison of the real and simulated records at the stations which recorded the 1998 Faial event ($M_w = 6.2$), including (A) GZC, (B) PVI, (C) HOR, and (D) SEB.

3.4.2 | Comparison of simulated dataset against existing ground motion models

As a subsequent step, the simulated ground motions are compared to the expected values generated by well-established existing GMMs. This comparison assesses the accuracy and reliability of the simulated data and determines if it generally aligns with the predicted results from the equations. This evaluation process is commonly employed in seismology research and engineering to assess the performance of ground motion simulations and their suitability for practical applications.^{60,63,65} However, it should be noted that the attenuation of various ground motion IMs is expected to be region-specific, and each IM may exhibit a distinct trend compared to the available GMMs, emphasizing the need for proposing region-specific models.

In the context of a comparative analysis aimed at identifying the most suitable GMMs, the study by⁷⁰ observed that the attenuation characteristics of the model proposed by³⁰ closely resemble those observed in the Azores region. However, it should be noted that that model exhibits a slightly faster attenuation rate on the Azores Islands. As a result, we have chosen to employ this particular model for our comparison, as it is considered the most appropriate for the region. Moreover, the selection of GMMs for this study takes into account their compatibility with the unique tectonic characteristics of the region. This includes the consideration of models explicitly tailored for volcanic areas, such as Atkinson,³¹ which was specifically developed for volcanic earthquakes in Hawaii, and the model by,³² designed for earthquakes occurring within the Taupo Volcanic Zone, where anelastic attenuation plays a more significant role. Additionally, we include a

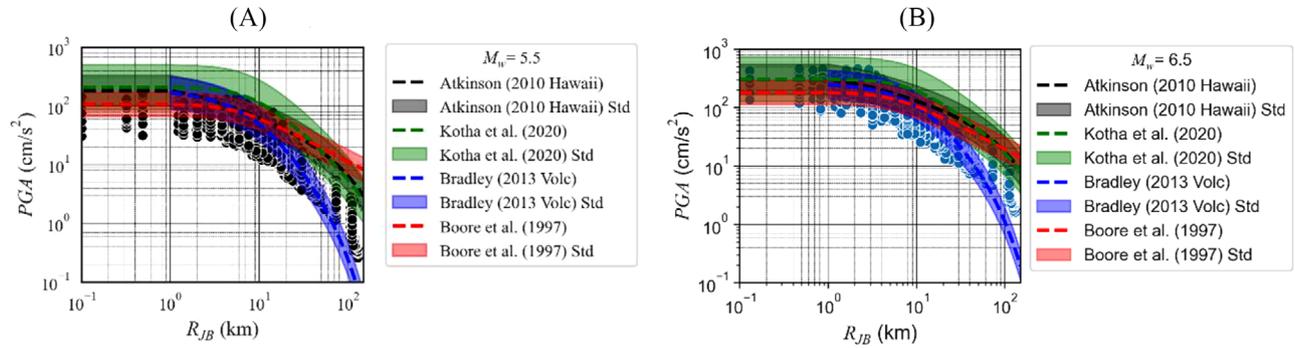


FIGURE 10 Attenuation of simulated ground motions in terms of PGA for scenario events of (A) $M_w = 5.5$ and (B) 6.5 against empirical GMM.

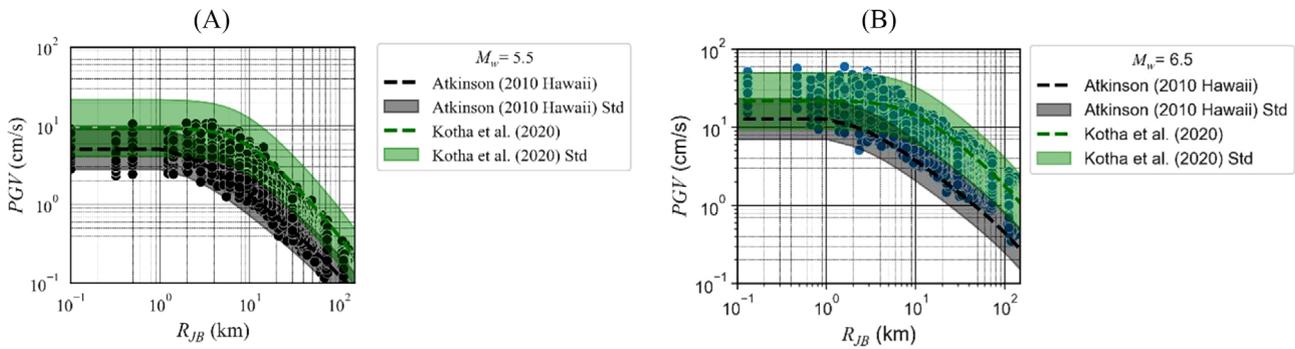


FIGURE 11 Attenuation of simulated ground motions in terms of PGV for scenario events of (A) $M_w = 5.5$ and (B) 6.5 against empirical GMM.

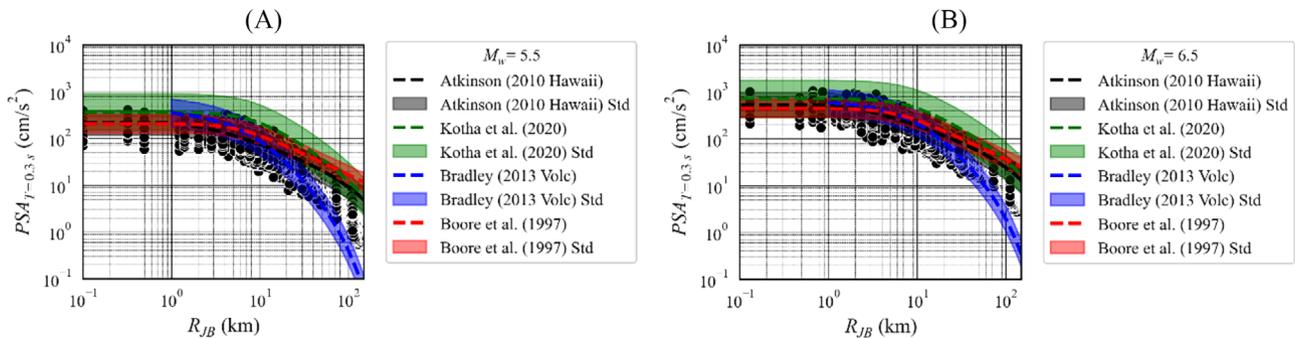


FIGURE 12 Attenuation of simulated ground motions in terms of PSA ($T = 0.3$ s) for scenario events of (A) $M_w = 5.5$ and (B) 6.5 against empirical GMM.

recent model introduced by,³³ which relies on European datasets. The inclusion of these models significantly enhances our ability to perform a comprehensive comparison, encompassing a diverse range of model variations.

The comparisons of the simulated dataset for different magnitudes against the GMMs in terms of PGA , PGV , PSA ($T = 0.3$ s), and PSA ($T = 1.5$ s) are plotted in Figures 10, 11, 12, and 13, respectively. It is noted that the models by³² and³⁰ are not given for PGV . The comparisons herein have been limited to just two representative magnitude levels, specifically $M_w = 5.5$ and $M_w = 6.5$. The observations revealed that when compared to Boore et al.,³⁰ the simulated attenuation in the Azores region in terms of PGA is indeed faster, thereby corroborating the findings reported in.⁷⁰ At larger distances, the outcomes align more consistently with Bradley³² for PGA , irrespective of the magnitude. Concerning PGV and spectral ordinates at periods of 0.3 and 1.5 seconds, the proposed model exhibits a closer alignment with the model by,³³ irrespective of the magnitude under consideration. However, it's important to highlight that the proposed model exhibits a faster attenuation

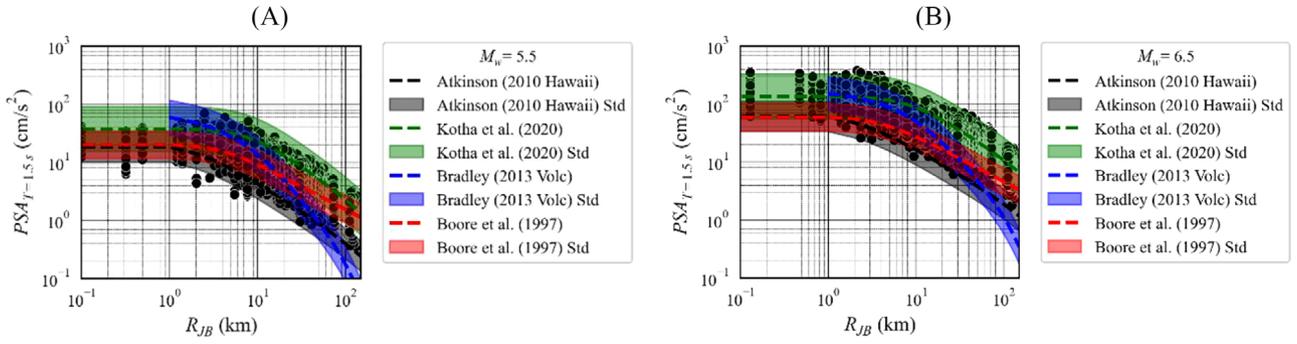


FIGURE 13 Attenuation of simulated ground motions in terms of PSA ($T = 1.5$ s) for scenario events of (A) $M_w = 5.5$ and (B) 6.5 against empirical GMMs.

rate. In conclusion, this comparison underscores the significance of customizing simulations to suit the distinct characteristics of the region. Given the diverse attenuation behaviors observed, it reinforces the essential requirement for the creation of models finally tuned to the region's unique attributes.

4 | GROUND MOTION MODELING METHODOLOGY

The common method for predicting ground motion IMs, such as PGA , PGV or PSA , is to utilize GMMs. These models are typically developed through empirical approaches that involve statistical regression analysis of large datasets of ground motion intensities.³⁰ 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:

$$\ln(y_{ij}) = \begin{pmatrix} \ln(PGA) \\ \ln(PGV) \\ \ln(PSA_{0.03s}) \\ M \\ \ln(PSA_{2s}) \end{pmatrix} = f(M_w, R_{JB}, FD) + \eta_i + \varepsilon_{ij} \quad (6)$$

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 modeled using XGBoost algorithm. Two components of residuals in GMMs, namely inter- 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- 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:

$$\phi = \sqrt{\sigma^2 + \tau^2} \quad (7)$$

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:

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

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:

$$\eta_i = \frac{\tau^2 \sum_1^{n_i} \delta_{ij}}{n_i \tau^2 + \sigma^2} \approx \frac{\sum_1^{n_i} \delta_{ij}}{n_i} \quad (n_i \tau^2 \gg \sigma^2) \quad (9)$$

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.⁷¹ Finally, the intra-event residuals can be obtained as follows:

$$\varepsilon_{ij} = \delta_{ij} - \eta_i \quad (10)$$

The GMM of this study is developed using the XGBoost algorithm. XGBoost³⁴ 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.⁷² This approach has gained popularity in fine-tuning hyperparameters of ML algorithms due to its flexibility in optimizing derivative-free functions.⁷³ In comparison to generic optimization techniques such as grid and random search, Bayesian optimization is considered to perform better.^{73,74} 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 14 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.

The performance indicators of the developed model, including R^2 , r , MAPE, and RMSE are presented in Figure 15. 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.^{20,75,76}

The model's potential bias is evaluated by examining the inter- and intra-event uncertainties in relation to source- and site-related parameters, M_w and R_{JB} , respectively, as shown in Figures 16 and 17. 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- and intra-event residuals are consistent with previous research,⁷⁷ 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.

Figure 18 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

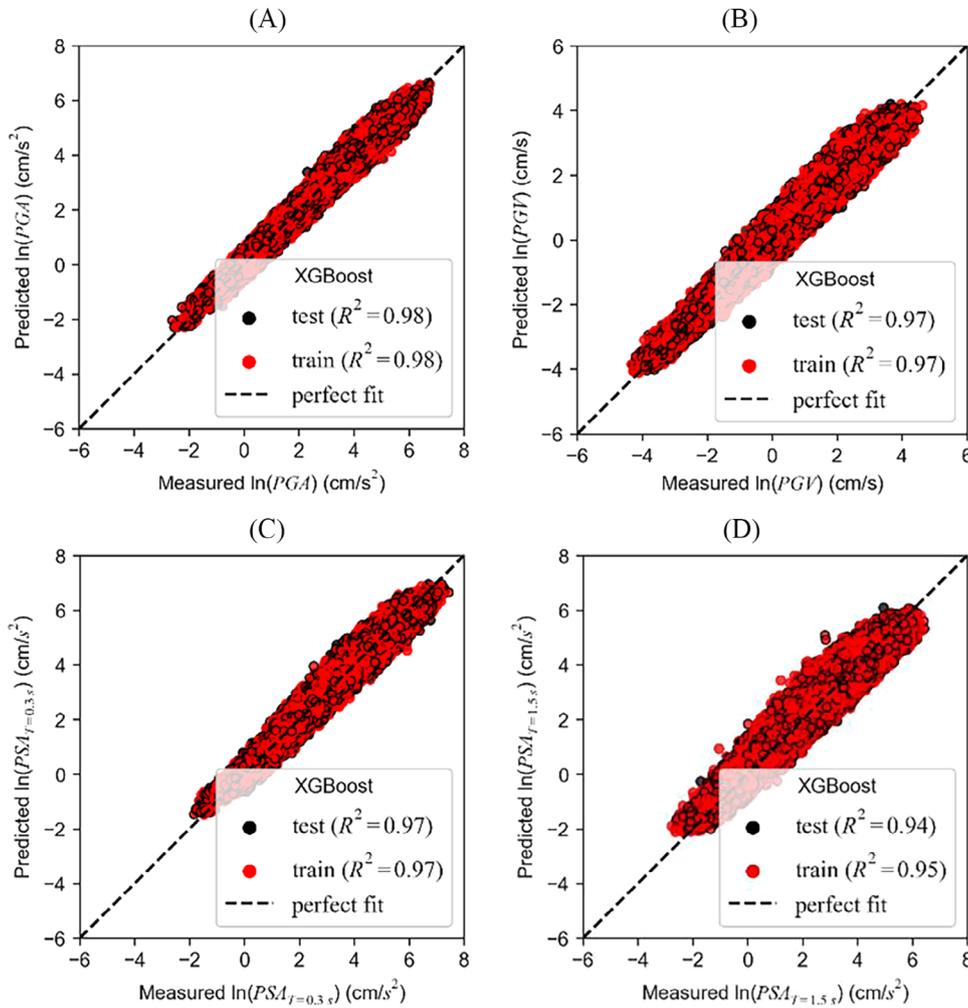


FIGURE 14 Observed versus predicted values of the developed GMM for the selected IMs including (A) $\ln(PGA)$, (B) $\ln(PGV)$, (C) $\ln(PSA_{T=0.3s})$, and (D) $\ln(PSA_{T=1.5s})$.

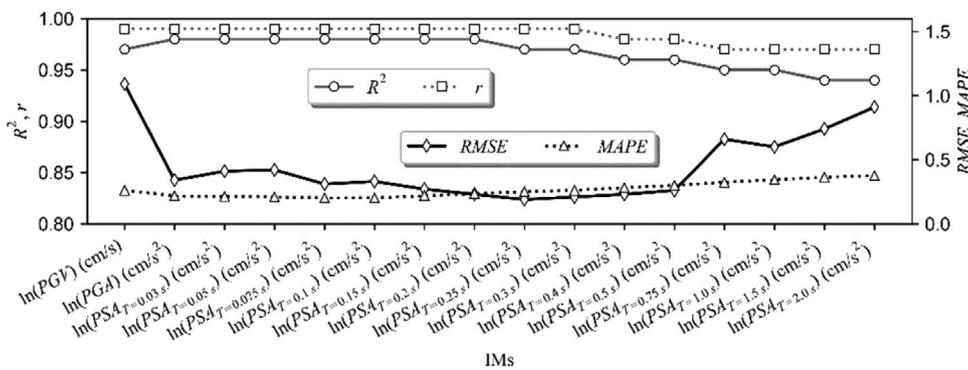


FIGURE 15 Model performance indicators for different IMs.

previous findings. Furthermore, the XGBoost-GMM model exhibited an acceptable uncertainty range and performed well when compared to existing models.⁷⁷

In the existing literature, there is an ongoing challenge regarding the capability of ML algorithms to replicate the underlying physics of real ground motion datasets. Therefore, the proposed GMM is subjected to further evaluation to determine its ability to represent physics-based phenomena regarding the behavior of real earthquakes. To this end, the results for various magnitude and distance combinations, utilizing the FD of 8.0 km, are compared. Figure 19 illustrates the estimated

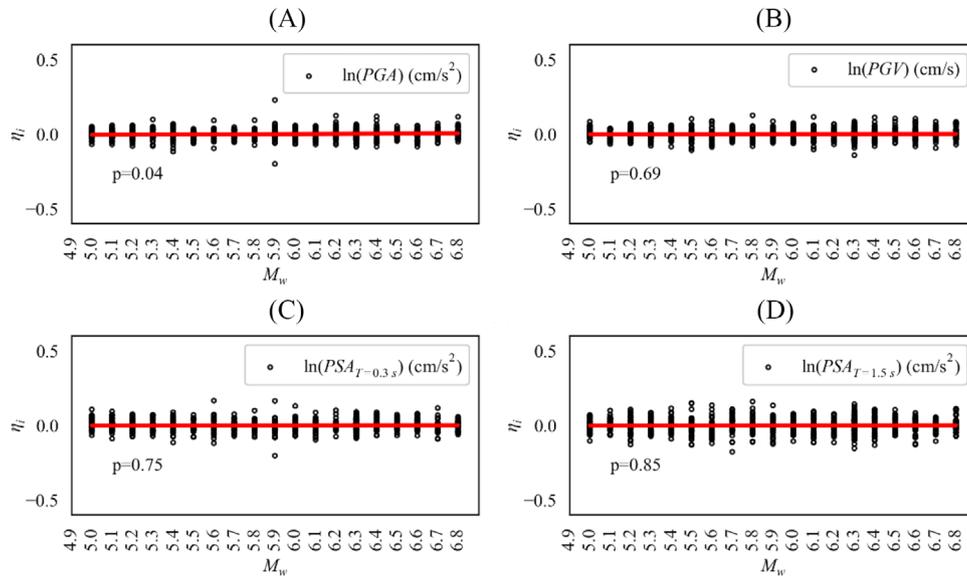


FIGURE 16 Distribution of the inter-event residuals with respect to M_w for the selected IMs including (A) $\ln(PGA)$, (B) $\ln(PGV)$, (C) $\ln(PSA_{T=0.3s})$, and (D) $\ln(PSA_{T=1.5s})$.

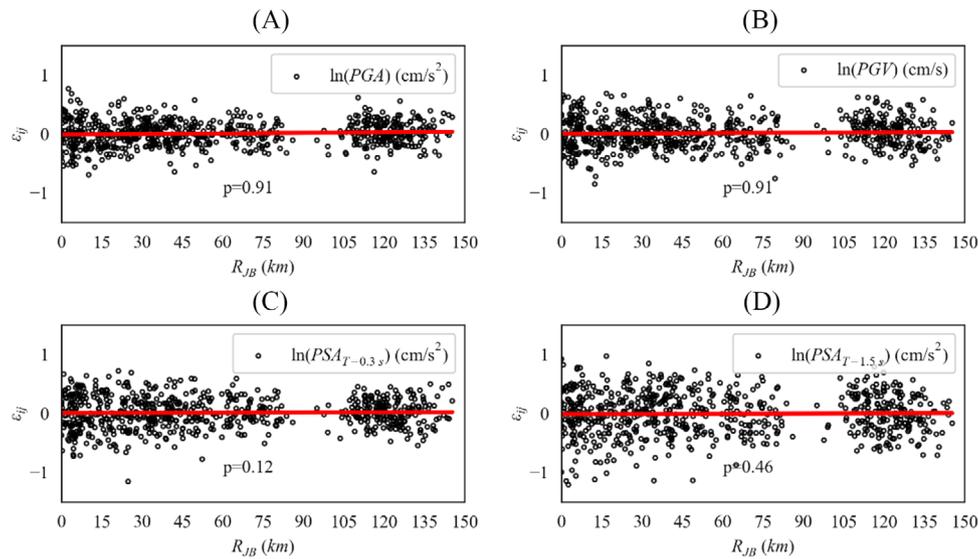


FIGURE 17 Distribution of the intra-event residuals with respect to R_{JB} for the selected IMs including (A) $\ln(PGA)$, (B) $\ln(PGV)$, (C) $\ln(PSA_{T=0.3s})$, and (D) $\ln(PSA_{T=1.5s})$.

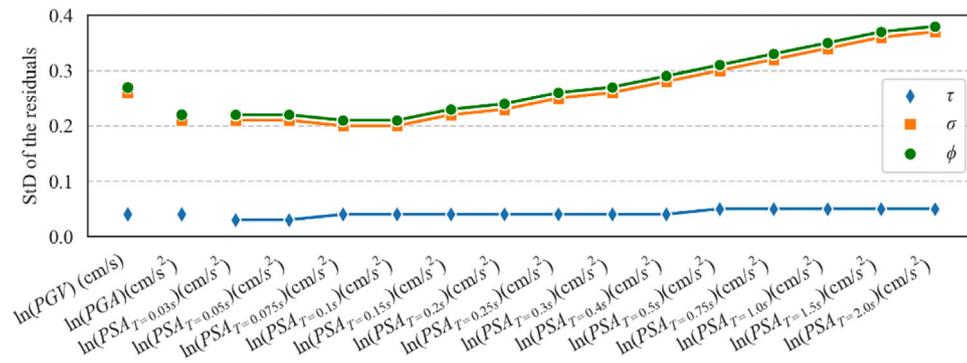


FIGURE 18 Standard deviation of the inter-event, intra-event, and total residuals for PGA , PGV , and all spectral ordinates.

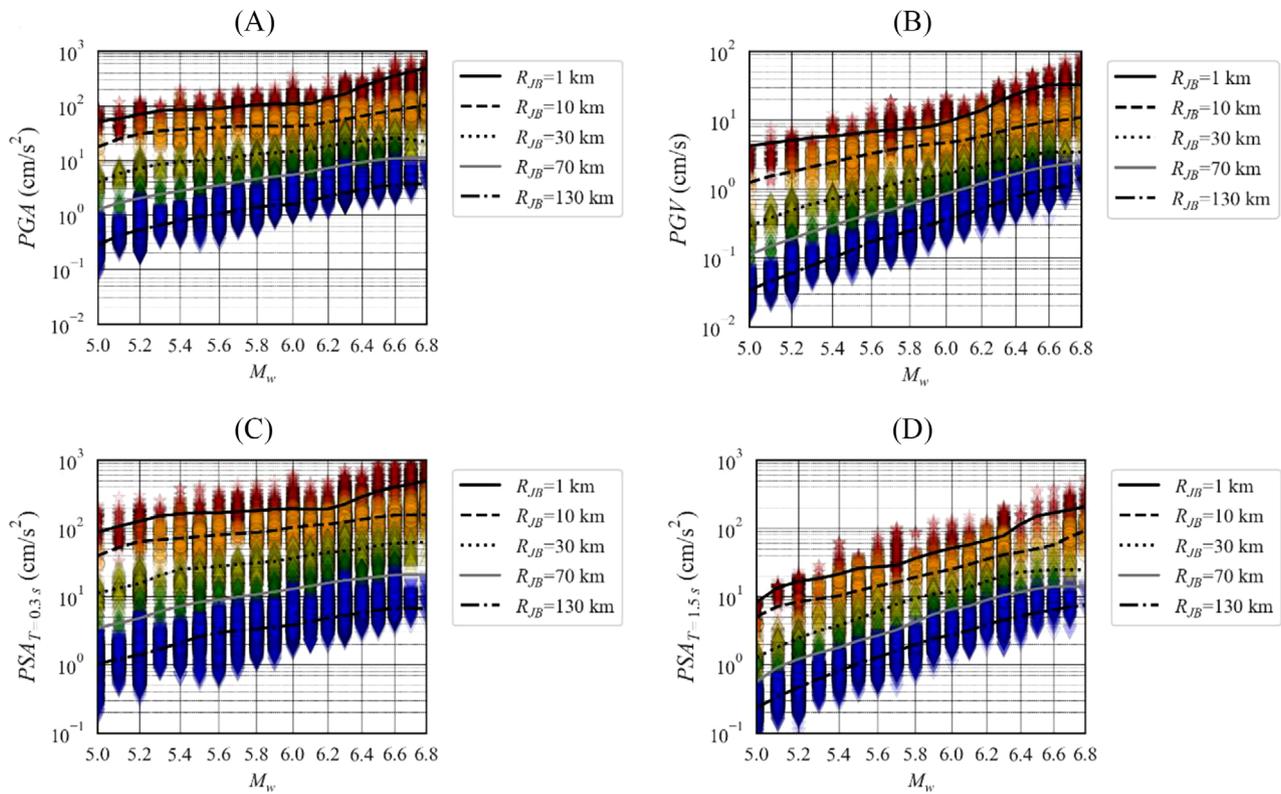


FIGURE 19 Variation of the selected IMs including (A) $\ln(PGA)$, (B) $\ln(PGV)$, (C) $\ln(PSA_{T=0.3s})$, and (D) $\ln(PSA_{T=1.5s})$ with respect to M_w using R_{JB} of 1, 10, 30, 70, and 130 km and FD of 8.0 km.

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, 10, 30, 70, and 130 km. Our observation indicates that the proposed GMM effectively captures the patterns present in earthquake records. Specifically, we note that as earthquake magnitude increases and the distance from the source decreases, there is a corresponding increase in PGA , PGV , and PSA levels across all period ranges. This outcome aligns with expectations, as it doesn't account for local amplification resulting from site effects.

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 and 150 km. The results are plotted in Figure 20. 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. Therefore, across all examined magnitudes and distances, the utility of the ML algorithm in replicating the behavior of the simulations is apparent.

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 21. The results reveal that when comparing the same distance but with different magnitudes, the peak of spectral ordinates shifts towards longer periods as the magnitude increases. This phenomenon is attributed to the acceleration spectral source model. Alternatively, when considering the same magnitude of an earthquake, the peak value of the PSA shifts towards longer periods as the distance increases, 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 behavior 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 AP, 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 AP.

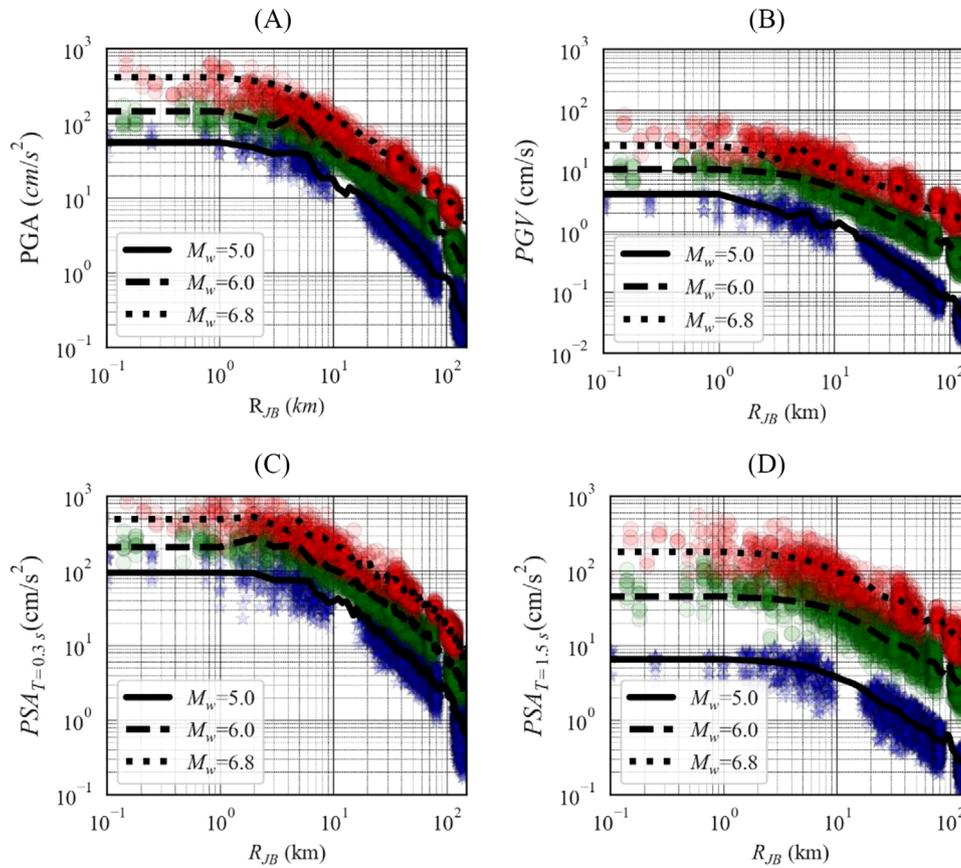


FIGURE 20 Variation of the selected IMs including (A) $\ln(PGA)$, (B) $\ln(PGV)$, (C) $\ln(PSA_{T=0.3s})$, and (D) $\ln(PSA_{T=1.5s})$ with respect to R_{JB} using M_w of 5.0, 6.0, and 6.8 and FD of 8.0 km.

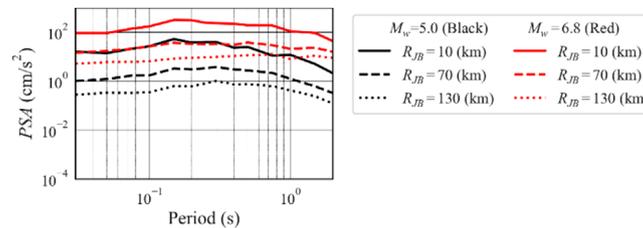


FIGURE 21 Variation PSA with respect to R_{JB} (including 10, 70, and 130 km) using M_w of 5.0 and 6.8 and FD of 8.0 km.

6 | EVALUATION OF THE PERFORMANCE AND CONSTRAINTS OF THE DEVELOPED GROUND MOTION MODEL

Despite the inherent model uncertainty, especially in the context of region-specific investigations, the GMM proposed by this study diligently addresses region-specific uncertainties in the modeling of seismotectonic parameters. Specifically, this model is developed using the region-specific parameters by²⁹ which meticulously estimated source and attenuation parameters for the Azores region employing data obtained from the Portuguese digital seismic and accelerometer network. In the simulations within this region, all input-model parameters are thoughtfully bound to the regional context and are derived from empirical relationships established by²⁹ based on the real region-specific dataset.

By including variability, the simulation incorporates stochastic behavior. Parameters exhibiting significant variability, as determined by²⁹ from regional earthquakes, are treated as random variables with PDFs in ground-motion simulations. Conforming to the methodology outlined by⁷⁸ we employ lognormal or uniform distributions to express the uncertainty, depending on the specific parameter being modelled. These parameters encompass source characteristics and propagation path attributes. Additionally, uncertainty in source stress and quality factor is represented by a lognormal distribution,

while the uncertainty in kappa values, hypocenter location and pulsing percent is modeled with a uniform distribution. The mean and uncertainty values for these parameters (stress, quality factor, kappa value) were estimated by.²⁹

This comprehensive approach significantly enhances the potential of the model to produce accurate and dependable results within the region. Recognizing the importance of addressing epistemic uncertainty in seismic hazard assessment, the proposed model complements other models within the logic tree framework effectively. Notably, the study by⁷⁸ underscores the necessity of incorporating a variety of GMMs within the logic tree framework to address epistemic uncertainty in probabilistic seismic hazard analyses, as previously demonstrated by.^{4,79} However, it is worth noting that determining appropriate weights for each model within this framework remains an ongoing research endeavor.

Furthermore, it is well-documented that hazard estimates often exhibit significant variability due to inconsistencies in the GMMs used, particularly in regions where no region-specific GMMs exist. To mitigate this challenge, we adopt a backbone approach, where a median GMM (referred to as the backbone model) is systematically scaled up and down. As emphasized by,⁴ the backbone approach offers several advantages, including transparency regarding the level of uncertainty implied by the ground-motion model, a clearer understanding of the meaning of logic tree weights (as each branch is mutually exclusive and collectively exhaustive), and the ability to tailor the model to specific site conditions.

The proposed model effectively predicts hazard levels in the region while diligently considering the uncertainties associated with the source and path-related parameters, as derived from the recorded motions in the AP by.²⁹ Therefore, it can serve as a robust backbone model for seismic hazard assessments in this region. However, this model appears to exhibit superior performance for *PGA* and spectral ordinates of shorter periods compared to the longer periods.

Ultimately, the proposed model serves as a backbone model exclusively developed for application within the AP. It is specifically tailored for shallow seismic events falling within the M_w range of 5.0 to 6.8 and at R_{JB} of up to 150 km in bedrock. Yet, it is worth noting that, as underscored by,⁵⁴ stochastic methodologies have encountered difficulties in accurately representing the coherent, long-period pulses that exert an influence on the characteristics of ground motions in near faults, particularly in the context of *PGV* and *PSA* at longer periods.

7 | CONCLUSIONS

This paper proposes a ML-based backbone GMM for the AP 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 AP. A simulated database is generated by assuming the rupture of ten faults, nine on Faial Island and one on São Jorge Island, with a maximum magnitude boundary of 6.8.

The initial evaluation of the simulations involves a validation procedure in which the simulated records of the 1998 Faial event are compared against the real seismic records.^{57,58} The input-model parameters have been further calibrated herein for the simulation of scenario earthquakes at the bedrock. Next, the simulation results of scenario events are compared against the well-known empirical GMMs suitable for the tectonic characteristics of the region. Finally, the eXtreme Gradient Boosting (XGBoost) algorithm is used to develop a region-specific GMM, which is known for its accuracy, flexibility, and computational speed in regression problems.

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

- The developed simulated dataset accurately replicates the real behavior of earthquake ground motions, as demonstrated through a comparison with the recorded ground motion data from the 1998 Faial earthquake at the relevant monitoring stations. The attainment of higher GOF values, which are considered seismologically acceptable, along with parameter calibration based on region-specific characteristics accounting for aleatory uncertainty in input-model parameters, adds further validation to the precision and reliability of the simulations.
- This study assesses the conformity of the simulated dataset with selected GMMs mainly compatible with the tectonic characteristics of the region, including but not limited to³⁰ Boore et al.,³⁰ Atkinson,³¹ Bradley,³² and Kotha et al.³³ Our comparative analysis of the simulated dataset against various GMMs for different magnitudes, specifically focusing on Peak Ground Acceleration (*PGA*), Peak Ground Velocity (*PGV*), Pseudo Spectral Acceleration (*PSA*) ($T = 0.3$ s), and *PSA* ($T = 1.5$ s) offers several trends against GMMs. The findings underscore that the simulated attenuation within the

Azores region, particularly with respect to *PGA*, exhibits a faster decay rate when compared to Boore et al.,³⁰ aligning with the conclusions drawn in.⁷⁰ Additionally, at larger distances, our observations consistently favor Bradley³² for *PGA*, regardless of the magnitude under consideration. With regard to *PGV* and *PSA* at $T = 0.3$ s and $T = 1.5$ s, our proposed model demonstrates a closer match with Kotha et al.,³³ regardless of the magnitude being analyzed. However, it's crucial to note that our proposed model exhibits a more rapid attenuation rate. In summary, this comparative analysis underscores the imperative need for region-specific simulations that align with the distinctive characteristics of the Azores region. The observed variations in attenuation behavior further emphasize the necessity for the development of finely tuned models specifically calibrated to the unique attributes of the region.

- The study evaluates the potential bias of the developed GMM by analyzing the inter- and intra-event uncertainties with respect to source and site parameters. The residuals are found to be unbiased across all considered IMs 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 IMs 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 a homogeneous simulated ground motion dataset over all magnitude and distance ranges.
- 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 longer periods of the *PSA* compared to *PGA* and *PSA* with shorter 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.
- 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. The total uncertainty varies, starting at 0.2 for $\ln(PGA)$ and $\ln(PSA)$ in shorter periods, increasing to 0.4 for $\ln(PSA)$ in longer periods, and estimated at 0.3 for $\ln(PGV)$. These results should be taken into consideration when applying the model in practice.
- 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 findings of this study indicate that variations in earthquake magnitude while maintaining a constant distance, lead to a shift in the peak of spectral ordinates towards longer periods, influenced by the acceleration spectral source model. Similarly, when keeping the magnitude constant and altering the distance, the peak value of *PSA* shifts towards longer periods in accordance with the physical principles governing distance-dependent damping of ground motions. Additionally, it is observed that the magnitude of the earthquake directly influences the extent of this peak shift, aligning with established earthquake physics.
- Recognizing the significance of addressing epistemic uncertainty in seismic hazard assessment, our proposed model effectively complements other models within the logic tree framework. Nevertheless, determining appropriate weights for models in this framework remains an ongoing research pursuit. Additionally, it is evident that hazard estimates often exhibit substantial variability due to inconsistencies in the choice of GMMs, particularly in regions lacking region-specific GMMs. To address this challenge, we employ a backbone model approach, systematically scaling a median GMM. This approach provides advantages such as transparency in representing model-related uncertainty, a clearer interpretation of logic tree weights, and adaptability to specific site conditions.

Overall, the findings of this study have significant implications for seismology research and engineering. This study addresses the scarcity of records from past events, especially for large-magnitude events, by presenting a substantial ground motion dataset comprising complete time-series data in bedrock. This extensive dataset plays a crucial role in bolstering the development of a more robust backbone GMM customized for the region's characteristics. The XGBoost-GMM is limited in its ability to extrapolate beyond the input range of predictor variables due to the ML algorithm's lack of adherence to underlying physical formulations. Therefore, it is advisable to utilize the model that has been developed for the AP 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. The validation of simulations against the 1998 Faial earthquakes, along with the model's robust performance, inspires a reasonable level of confidence in its potential applicability across a diverse range of magnitudes and distances. However, it is important to acknowledge the constraints of stochastic ground motion simulations in reproducing low-frequency components of ground motion records in near-field events. Future research endeavors should prioritize the integration of physics-based simulations within stochastic simulations to address the representation of low-frequency content. This would enhance the accuracy and reliability of the GMM for seismic hazard assessment in the future.

AUTHOR CONTRIBUTIONS

The authors' contributions to the paper are as follows: Shaghayegh Karimzadeh performed the ground motion simulations, analyzed, 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.

ACKNOWLEDGEMENTS

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 partially financed by national funds through FCT – Foundation for Science and Technology, under grant agreement 2020.08876.BD attributed to the second author.

DATA AVAILABILITY STATEMENT

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

ORCID

Shaghayegh Karimzadeh  <https://orcid.org/0000-0003-3753-1676>

Amirhossein Mohammadi  <https://orcid.org/0000-0002-2940-109X>

Paulo B. Lourenço  <https://orcid.org/0000-0001-8459-0199>

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. Pesaresi M., Ehrlich D., Kemper T., Siragusa A., Florczyk A., Freire S., Corban, C. Atlas of the Human Planet 2017: Global Exposure to Natural Hazards. 2017, EUR 28556 EN, Publications Office of the European Union, Luxembourg, 2017.
3. McGuire RK. Deterministic vs. Probabilistic earthquake hazards and risks. *Soil Dyn Earthq Eng.* 2001;21:377-384.
4. Douglas J. Calibrating the Backbone Approach for the Development of Earthquake Ground Motion Models. *Best Pract. Physics-based Fault Rupture Model. Seism. Hazard Assess. Nucl. Install. Issues Challenges Toward Full Seism. Risk Anal.* 2018.
5. Douglas J. Ground motion prediction equations 1964–2021. *Dept Civ Environ Eng, Univ Strat Glas UK.* 2022:670.
6. Rezaeian S, Sun X. Stochastic ground motion simulation. In: Beer M, Kougoumtzoglou IA, Patelli E, Au SK. (eds) *Encyclopedia of Earthquake Engineering.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35344-4_239
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 J Earth Sci.* 2015;24:529-565.
8. Karimzadeh S, Ozsarac V, Askan A, Erberik MA. Use of simulated ground motions for the evaluation of energy response of simple structural systems. *Soil Dyn Earthq Eng.* 2019;123:525-542.
9. 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 Struct.* 2020;18:249-261.

10. Tsioulou A, Taflanidis AA, Galasso C. Modification of stochastic ground motion models for matching target intensity measures. *Earthq Eng Struct Dyn*. 2018;47:3-24. doi:10.1002/eqe.2933
11. Vetter CR, Taflanidis AA, Mavroeidis GP. Tuning of stochastic ground motion models for compatibility with ground motion prediction equations. *Earthq Eng Struct Dyn*. 2016;45:893-912. doi:10.1002/eqe.2690
12. 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.
13. Atkinson GM, Boore DM. Recent trends in ground motion and spectral response relations for North America. *Earthq Spectra*. 1990;6:15-35.
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. *Bull Seismol Soc Am*. 2003;93:1012-1033.
15. Megawati K, Pan T-C, Koketsu K. Response spectral attenuation relationships for sumatran-subduction earthquakes and the seismic hazard implications to Singapore and Kuala Lumpur. *Soil Dyn Earthq Eng*. 2005;25:11-25.
16. Douglas J. Ground Motion Estimation Equations 1964–2003' Intermediary Report. 2006.
17. Rietbrock A, Strasser F, Edwards B. A stochastic earthquake ground-motion prediction model for the United Kingdom. *Bull Seismol Soc Am*. 2013;103:57-77.
18. 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. *Bull Earthq Eng*. 2022;20:4919-4936. doi:10.1007/s10518-022-01417-9
19. Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. *Earthq Spectra*. 2014;30:1025-1055. doi:10.1193/070913EQS198M
20. Akkar S, Çağnan Z. A local ground-motion predictive model for turkey, and its comparison with other regional and global ground-motion models. *Bull Seismol Soc Am*. 2010;100:2978-2995. 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. *Comput Struct*. 2011;89:2176-2194. doi:10.1016/j.compstruc.2011.08.019
22. Khosravikia F, Clayton P. Machine learning in ground motion prediction. *Comput Geosci*. 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. *Nat Hazards*. 2021;107:423-446. 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. *Bull Seismol Soc Am*. 2022;112:1549-1564. doi:10.1785/0120210244
25. 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 Dyn Earthq Eng*. 2023;172:108008. doi:10.1016/j.soildyn.2023.108008
26. Sreenath V, Podili B, Raghukanth STG. A hybrid non-parametric ground motion model for shallow crustal earthquakes in Europe. *Earthq Eng Struct Dyn*. 2023. doi:10.1002/eqe.3845
27. Yerlikaya-Özkurt F, Askan A, Weber GW. An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method. *Eng Optim*. 2014;46:1651-1668. 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. *Bull Earthq Eng*. 2016;14:1885-1902.
30. 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. *Seismol Res Lett*. 1997;68:128-153.
31. Atkinson GM. Ground-motion prediction equations for hawaii from a referenced empirical approach. *Bull Seismol Soc Am*. 2010;100:751-761.
32. Bradley BA. A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models. *Bull Seismol Soc Am*. 2013;103:1801-1822.
33. Kotha SR, Weatherill G, Bindi D, Cotton F. A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *Bull Earthq Eng*. 2020;18:4091-4125.
34. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016;785-794. doi:10.1145/2939672.2939785
35. 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. *Mar Geophys Res*. 1998;20:141.
36. Silveira G, Stutzmann E, Davaille A, Montagner J-P, Mendes-Victor L, Sebai A. Azores hotspot signature in the upper mantle. *J Volcanol Geotherm Res*. 2006;156:23-34.
37. 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. *Geol Soc London, Mem*. 2015;44:15-32.
38. Vogt PR, Jung WY. The Terceira rift as hyper-slow, hotspot-dominated oblique spreading axis: a comparison with other slow-spreading plate boundaries. *Earth Planet Sci Lett*. 2004;218:77-90. doi:10.1016/S0012-821X(03)00627-7
39. Heezen BC, Tharp M, Ewing M. The floors of the oceans: I. The North Atlantic. Geological Society of America; 1959.
40. 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;56:1-18.
41. Carmo R, Madeira J, Ferreira T, Queiroz G, Hipólito A. Chapter 6 volcano-tectonic structures of São Miguel Island, Azores. *Geol Soc London, Mem*. 2015;44:65-86.

42. Bezzeghoud M, Adam C, Buforn E, Borges JF, Caldeira B. Seismicity along the Azores-Gibraltar Region and Global Plate Kinematics. *J Seismol*. 2014;18:205-220.
43. Matias L, Dias NA, Morais I, et al. The 9th of July 1998 Faial Island (Azores, North Atlantic) Seismic Sequence. *J Seismol*. 2007;11:275-298.
44. Senos ML, Gaspar JL, Cruz J, et al. O Terramoto Do Faial de 9 de Julho de 1998. Proceedings of the Proceedings do. 1998;1:61-68.
45. Madeira J, Brum da Silveira A. Active tectonics and first paleoseismological results in Faial, Pico and S. Jorge Islands (Azores, Portugal). *Ann Geophys*. 2003.
46. Wells DL, Coppersmith KJ. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull Seismol Soc Am*. 1994;84:974-1002.
47. 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. *Geophys Res Lett*. 1980;7:501-504.
48. 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. *Univ. Azores, Ponta Delgada (MSc thesis)* 2005.
49. 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. *Volcan Geol Sao Miguel Isl (Azores Arch Geol Soc London, Mem)*. 2015;44:33-49.
50. Boore DM. Strong-motion Seismology. *Rev Geophys*. 1983;21:1308-1318.
51. Brune JN. Tectonic stress and the spectra of seismic shear waves from earthquakes. *J Geophys Res*. 1970;75:4997-5009.
52. Boore DM. Simulation of ground motion using the stochastic method. *Pure Appl Geophys*. 2003;160:635-676.
53. Beresnev IA, Atkinson GM. FINSIM – a FORTRAN program for simulating stochastic acceleration time histories from finite faults. *Seismol Res Lett*. 1998;69:27-32.
54. Motazedian D, Atkinson GM. Stochastic finite-fault modeling based on a dynamic corner frequency. *Bull Seismol Soc Am*. 2005;95:995-1010.
55. Boore DM. Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bull Seismol Soc Am*. 2009;99:3202-3216.
56. 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. *Bull Seismol Soc Am*. 2011;101:2967-2980.
57. Karimzadeh S, Lourenço PB. Stochastic ground motion simulation of the 9th of July 1998 Faial Earthquake (Azores, North Atlantic). *Authorea Prepr*. 2021.
58. Karimzadeh S, Funari MF, Szabó S, Hussaini SMS, Rezaeian S, Lourenço PB. Stochastic simulation of earthquake ground motions for the seismic assessment of monumental masonry structures: source-based vs site-based approaches. *Earthq. Eng. Struct. Dyn*. 2023.
59. Taborada R, Bielak J. Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake. *Bull Seismol Soc Am*. 2013;103:131-156.
60. MC R, Somala SN. Simulation of strong ground motion for the 25 April 2015 Nepal (Gorkha) Mw 7.8 earthquake using the SCEC broadband platform. *J Seismol*. 2017;21:777-808.
61. Can G, Askan A, Karimzadeh S. 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*. 2021;150:106832.
62. Raghukanth STG, Kavitha B. Ground motion relations for active regions in India. *Pure Appl Geophys*. 2014;171:2241-2275.
63. Burks LS. *Ground Motion Simulations: Validation and Application for Civil Engineering Problems*. Stanford University; 2014; ISBN 9798662513914.
64. Nh H, Kumar A. Ground motion prediction equation for North India, applicable for different site classes. *Soil Dyn Earthq Eng*. 2020;139:106425.
65. Kotha SR, Weatherill G, Bindi D, et al. A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *Bull Seismol Soc Am*. 2017;21:4091-4125.
66. Olsen KB, Mayhew JE. Goodness-of-fit criteria for broadband synthetic Seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake. *Seismol Res Lett*. 2010;81:715-723.
67. Yakut A, Yılmaz H. Correlation of deformation demands with ground motion intensity. *J Struct Eng*. 2008;134:1818-1828. doi:10.1061/(ASCE)0733-9445(2008)134:12(1818)
68. Ólafsson S, Galluzzo D, Carvalho A, Langer H, Oliveira C, Sigbjörnsson R. Ground Motion Modelling Based on Data from the Azores Islands, Italy and South-Iceland. Proceedings of the Proceedings of the Second European Conference on Earthquake Engineering and Seismology, 2014;25-29.
69. Ólafsson S, Sigbjörnsson R. Ground Motion Prediction Equation for South Iceland. Proceedings of the second European conference on earthquake engineering and semiology, Istanbul; 2014.
70. Oliveira CS, Sigbjörnsson R, Ólafsson S. A Comparative Study on Strong Ground Motion in Two Volcanic Environments: Azores and Iceland. Proceedings of the Proceedings of the 13th World Conference on Earthquake Engineering; Citeseer, 2004:13.
71. 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. *Sci Rep*. 2020;10. doi:10.1038/s41598-020-68630-x
72. Frazier PI. A tutorial on Bayesian optimization. arXiv Prepr. arXiv1807.02811 2018. <https://doi.org/10.48550/arXiv.1807.02811>
73. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst*. 2012;25.
74. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. *Adv Neural Inf Process Syst*. 2011;24.

75. Akkar S, Sandikkaya MA, Bommer JJ. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull Earthq Eng*. 2014;12:359-387. doi:10.1007/s10518-013-9461-4
76. 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. *Earthq Spectra*. 2008;24:99-138.
77. Withers KB, Moschetti MP, Thompson EM. A machine learning approach to developing ground motion models from simulated ground motions. *Geophys Res Lett*. 2020;47:e2019GL086690.
78. Atkinson GM, Bommer JJ, Abrahamson NA. Alternative approaches to modeling epistemic uncertainty in ground motions in probabilistic seismic-hazard analysis. *Seismol Res Lett*. 2014;85:1141-1144.
79. Bommer JJ, Scherbaum F. The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthq Spectra*. 2008;24:997-1009. doi:10.1193/1.2977755

How to cite this article: Karimzadeh S, Mohammadi A, Salahuddin U, Carvalho A, Lourenço PB. Backbone ground motion model through simulated records and XGBoost machine learning algorithm: An application for the Azores plateau (Portugal). *Earthquake Engng Struct Dyn*. 2024;53:668–693. <https://doi.org/10.1002/eqe.4040>

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.

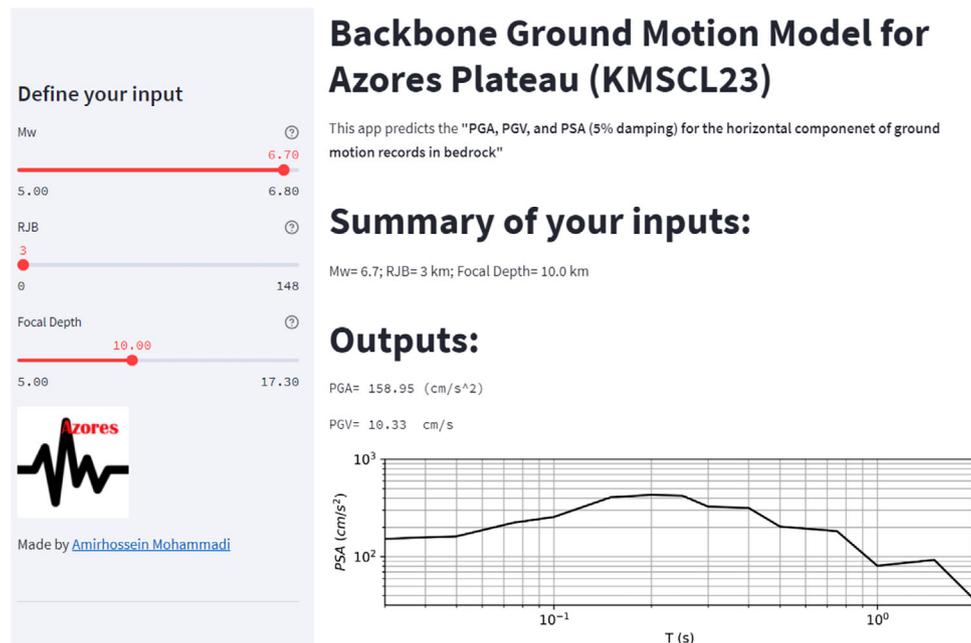


FIGURE A GUI of the XGBoost-GMM.