Nat. Hazards Earth Syst. Sci., 26, 3541–3558, 2026
https://doi.org/10.5194/nhess-26-3541-2026

Natural Hazards and Earth System Sciences
Open Access

A ground motion prediction model for the Italian region based on a mixture of experts framework

Jinfeng Dai[1,2], Zifa Wang[1,2], Dengke Zhao[1,2], Xiangying Wang[3], Jianming Wang[1,2], Zhaoyan Li[1,2], Zhaodong Wang[1,2], and Jintao Xiao[1,2]

[1]Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin, 150080, China
[2]Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin, 150080, China
[3]School of Architecture and Civil Engineering, Huangshan University, Huangshan, 245041, China

Correspondence: Zifa Wang (zifa@iem.ac.cn)

Received: 19 December 2025 – Discussion started: 30 December 2025
Revised: 14 June 2026 – Accepted: 22 June 2026 – Published: 29 July 2026

Abstract. Earthquake ground-motion prediction is critical for seismic design, risk assessment, and infrastructure resilience. Traditional empirical models often struggle to capture the complex, non-linear physical mechanisms of wave propagation and site amplification, leading to high aleatory variability and limited predictive accuracy. To address these limitations, this study proposes a ground-motion prediction model for the Italian region based on a Mixture-of-Experts (MoE) framework, utilizing XGBoost regressors as specialized experts. To ensure a robust assessment of predictive generalization, the model is developed and verified using a strict event-wise chronological splitting strategy on the ITACAext 2.0 dataset, which effectively eliminates event-level information leakage. Compared with empirical ground-motion prediction equations (GMPEs) and baseline machine-learning models, the MoE-XGB framework achieves superior predictive accuracy, significantly reducing both inter-event and intra-event variability across a wide spectral period range. Beyond statistical performance, we employ SHAP (SHapley Additive exPlanations) analysis to open the "black box" of the model, demonstrating that the MoE-XGB framework successfully learns and synthesizes physically sound seismological relationships, including magnitude scaling, geometric attenuation, and site amplification. As a high-fidelity foundational predictive engine, the MoE-XGB model provides robust, physically consistent, and fully decomposed statistical inputs that serve as a reliable basis for future regional seismic hazard assessments and performance-based engineering. The proposed framework exhibits strong transferability and scalability, offering a valuable reference for data-driven ground-motion prediction in other seismically active regions.

1 Introduction

Italy experiences approximately 40 000 earthquakes annually, with seismic activity affecting over 21 million residents in high-risk areas across the peninsula. The 2016 central Italy earthquake sequence caused nearly 300 deaths and billions of dollars in losses, highlighting the critical role of precise ground-motion prediction in hazard assessment and risk mitigation. With the advancement of urbanization in seismically active zones and the aging of infrastructure, the limitations of existing ground motion prediction equations (GMPEs) and ground motion prediction models (GMPMs) have become increasingly apparent.

GMPEs and GMPMs constitute core tools for modern seismic hazard assessment, providing support for earthquake engineering, emergency preparedness, and urban planning. Historically, empirical models have established fundamental relationships between source parameters, propagation paths, and site conditions. The NGA-West project in the United States initially proposed five attenuation relations (Abrahamson and Silva, 2008; Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008; Idriss, 2008), establishing a benchmark for subsequent research. Based on Italian strong-motion data, Scasserra et al. (2009) noted that while Italian data show consistency with NGA GM-

Published by Copernicus Publications on behalf of the European Geosciences Union.

PEs regarding magnitude and site shear-wave velocity scaling, short-period ground motions attenuate faster; thus, adjustments to the constant term, distance attenuation slope, and fictitious depth term were necessary to improve applicability. Building on this, Bindi et al. (2011) developed a set of GMPEs for the Italian strong-motion database, covering magnitudes 4.0 to 6.9 and distances up to 200 km for horizontal and vertical ground-motion parameters, with standard deviations between 0.34 and 0.38. However, this model exhibited large variability in predicting small-to-moderate magnitude events, indicating its limitations. To further improve the model, Lanzano et al. (2019) revised the ITA10 ground-motion prediction model, extending the applicable magnitude range to above 6.9 and vibration periods to 10 s, using RotD50 as the ground-motion parameter. They introduced fault distance and heteroscedastic variability models, significantly improving the accuracy of near-field strong-motion predictions while reducing the standard deviation at intermediate-to-long periods by approximately 20 %. On this basis, Michelini et al. (2019) utilized the GMPE of Bindi et al. (2011), combined with data from the National Institute of Geophysics and Volcanology (INGV), to develop post-earthquake damage distribution maps, providing important tools for post-disaster assessment. In recent years, the introduction of machine learning has opened new possibilities for ground-motion prediction, as it can capture nonlinear relationships and complex interactions that are difficult to model with traditional parametric methods. Mori et al. (2022) employed Gaussian process regression (GPR), combining approximately 16 000 accelerometer data points and 46 000 geological and geophysical data points, and considering stratigraphic and geomorphological conditions, to generate ground-motion prediction maps with a resolution of 50 m $\times$ 50 m. This method significantly improved the precision and accuracy of near-real-time prediction; notably, the GPR model outperformed traditional GMPEs when accounting for the significant influence of topographic features and local site conditions on ground-motion amplification patterns.

Although ground-motion prediction models designed specifically for Italy have been widely applied (Bindi et al., 2006, 2014), existing methods still possess significant limitations (Brando et al., 2020; Falcone et al., 2020a). First, these models are typically developed for a single earthquake type (mainly shallow crustal events), making it difficult to fully adapt to diverse source types and complex geological conditions (Agrawal and McCloskey, 2024). This restricts the ability of models to characterize the diversity of Italian seismic activity and regions where multiple source types jointly constitute the overall seismic risk, thereby affecting the comprehensive assessment of seismic hazard. Second, prediction accuracy is insufficient; particularly in complex scenarios involving topographic amplification, site-specific effects, and different source mechanisms, the models exhibit high variability, limiting their reliability. Furthermore, single-model approaches cannot fully leverage the complementary advantages of different modeling techniques, potentially missing opportunities to enhance prediction performance through ensemble methods or expert system architectures. Existing models are also limited in the scope of their output parameters. For example, the machine learning model of Mori et al. (2022) predicts only limited ground-motion parameters such as peak ground acceleration (PGA), peak ground velocity (PGV), spectral acceleration at 0.3 s (SA $T = 0.3$), spectral acceleration at 1 s (SA $T = 1$), and spectral acceleration at 3 s (SA $T = 3$), failing to provide comprehensive ground-motion prediction values. This is particularly problematic for comprehensive hazard analyses, which require determining multiple ground-motion parameters for different structural types and performance objectives; a single output parameter cannot meet the diverse needs of engineering applications and risk assessment frameworks. Additionally, due to regional differences in seismic wave propagation, site effects, and source characteristics, models developed based on a specific tectonic environment may fail in other regions, limiting the applicability and robustness of prediction results. This lack of validation across different tectonic environments and source characteristics further raises concerns regarding model transferability and reliability in international applications. In summary, future research needs to integrate multi-source data, expand the range of prediction parameters, and develop more universal and robust models to enhance the comprehensiveness and accuracy of ground-motion prediction.

This paper proposes a novel mixture of experts (MOE) framework, employing XGBoost as the expert model, aimed at addressing the fundamental limitations of existing ground-motion prediction models and providing comprehensive predictions for ground motion in Italy. We designed a unified modeling approach capable of adapting to multiple earthquake types and achieving prediction accuracy superior to traditional single models, thereby providing more reliable support for seismic hazard assessment. The mixture of experts model effectively captures the unique characteristics of earthquakes of different magnitudes by training specialized expert models based on earthquake data subsets partitioned by magnitude. To determine the optimal model configuration, we conducted systematic testing on the number of expert models, and the results indicated that employing three expert models performed best across multiple evaluation metrics. Specifically, we divided the optimized dataset into three categories – small, medium, and large – based on magnitude, and performed Bayesian optimization for each category to construct expert models targeting different magnitude ranges. This hierarchical optimization strategy enables the model to better adapt to magnitude-dependent ground-motion characteristics, thereby significantly improving prediction accuracy across the range of Italian seismic activity (Wang et al., 2026).

The main contributions of this study are as follows:

1. Development and validation of the MOE-XGB framework: A ground-motion prediction model based on a mixture of experts (MOE) framework is proposed, incorporating XGBoost as the expert submodel. Through comparison with the existing GPR model, the significant accuracy improvement of this model across multiple ground-motion intensity measures is validated.

2. Application in multi-type earthquake scenarios: Based on the ITACAext 2.0 dataset (Lanzano et al., 2024), which has a larger sample size and broader coverage of earthquake types, the MOE-XGB model was retrained and evaluated. The results verify the high-precision prediction capability of the model across diverse seismic scenarios. Comparisons with the existing GPR model and the traditional Bindi 2017 model demonstrate more significant improvements in accuracy.

3. Independent testing and generalization capability validation: Through testing on independent earthquake events, it is further proven that the MOE-XGB model maintains stable prediction performance when facing unknown earthquake events, reflecting its strong generalization capability and robustness.

The significance of the development of the MOE-XGB model is reflected not only in the improvement of prediction accuracy but also in the innovation of seismic hazard modeling methods, the deep understanding of magnitude-dependent prediction strategies, and the practical application in Italian seismic risk assessment. By overcoming the limitations of existing methods, this study aims to lay the foundation for the next generation of ground-motion prediction models and to develop a more universal prediction framework applicable to diverse seismic environments. In summary, the proposed framework exhibits good transferability and scalability, offering a useful reference for fusion-model-driven ground-motion prediction in Europe and other regions.

2 Data sources and overview

The ITACAext flatfile version 2.0 (Lanzano et al., 2024) dataset is a crucial resource for ground-motion prediction research in Italy, containing 52 706 records covering 2338 seismic events between 1972 and 2022, recorded by 2294 stations (Foti et al., 2011). This dataset integrates core data from the ITACAext 2.0 database and supplements small-magnitude velocity records from under-sampled regions in Italy (e.g. Liguria, Piedmont, and western Sicily), while also incorporating network records from neighboring countries (e.g. France, Switzerland, Slovenia, Albania, and Montenegro) to improve the spatial and azimuthal coverage of events. This study utilizes the SA flatfile, which contains the median spectral acceleration (RotD50) of the horizontal components (5 % damping) and relevant metadata. RotD50 represents the median spectral acceleration after the rotation of horizontal components; it is calculated by rotating the two horizontal components (north–south and east–west) of the seismic record, computing the resultant spectral acceleration at all possible angles, and taking the 50th percentile. It characterizes direction-independent ground-motion properties and is widely applied in earthquake engineering and ground-motion prediction models.

The MORI dataset, compiled by Mori et al. (2026), includes seismological parameters, geophysical data, and morphological data, which are used for the training of machine learning models and ground-motion prediction. The seismological parameters consist of 15 779 data points, including PGA, PGV, and SA at 0.3, 1.0, and 3.0 s. These data are obtained from Italian (ESM, ITACAext 2.0) and European databases, covering parameters such as moment magnitude (M), epicentral distance (R), and hypocentral depth (H), with a focus on shallow crustal events ($H < 35$ km). The geophysical data characterize site conditions using V_{S30} (the time-averaged shear-wave velocity in the top 30 m), combining field tests with the V_{S30} map from Mori et al. (2020) as supplementary data (Luzi et al., 2016, 2020). The morphological data include elevation (h) obtained from the ALOS World 3D-30m digital elevation model (DEM) and terrain gradient parameters (h_x, h_y, h_{xx}, h_{yy}) generated via GIS ALOS (2021).

Figure 1 clearly shows that the ITACAext 2.0 dataset significantly outperforms the MORI dataset in terms of data density and comprehensiveness. The ITACAext 2.0 dataset covers a wider range of source depths and magnitudes, and the data volume far exceeds that of the MORI dataset. Therefore, we select the ITACAext 2.0 dataset as the primary data source for developing a ground-motion prediction model applicable to the Italian region to ensure the robustness and applicability of the model. Meanwhile, due to its unique data distribution characteristics, the MORI dataset serves as an independent test dataset for verifying model performance.

Dataset splitting: After screening, 45 080 valid records are retained. To prevent information leakage between training, validation, and test sets, we adopt an event-wise chronological splitting strategy. All strong-motion records originating from the same earthquake (identified by a unique Event-ID) are assigned exclusively to a single subset. The events are first sorted by their origin time, and the cumulative record count is computed along this temporal axis. Split boundaries are then drawn so that approximately 70 % of all records fall into the training set, 10 % into the validation set, and 20 % into the test set, proceeding chronologically from the earliest to the most recent event. This ensures that the model is evaluated on future earthquakes relative to its training data, mimicking a realistic prospective forecasting scenario.

Figure 2 illustrates the resulting partition using the ITACAext 2.0 dataset. Panel (a) confirms that the three subsets are temporally coherent: train, validation, and test periods are

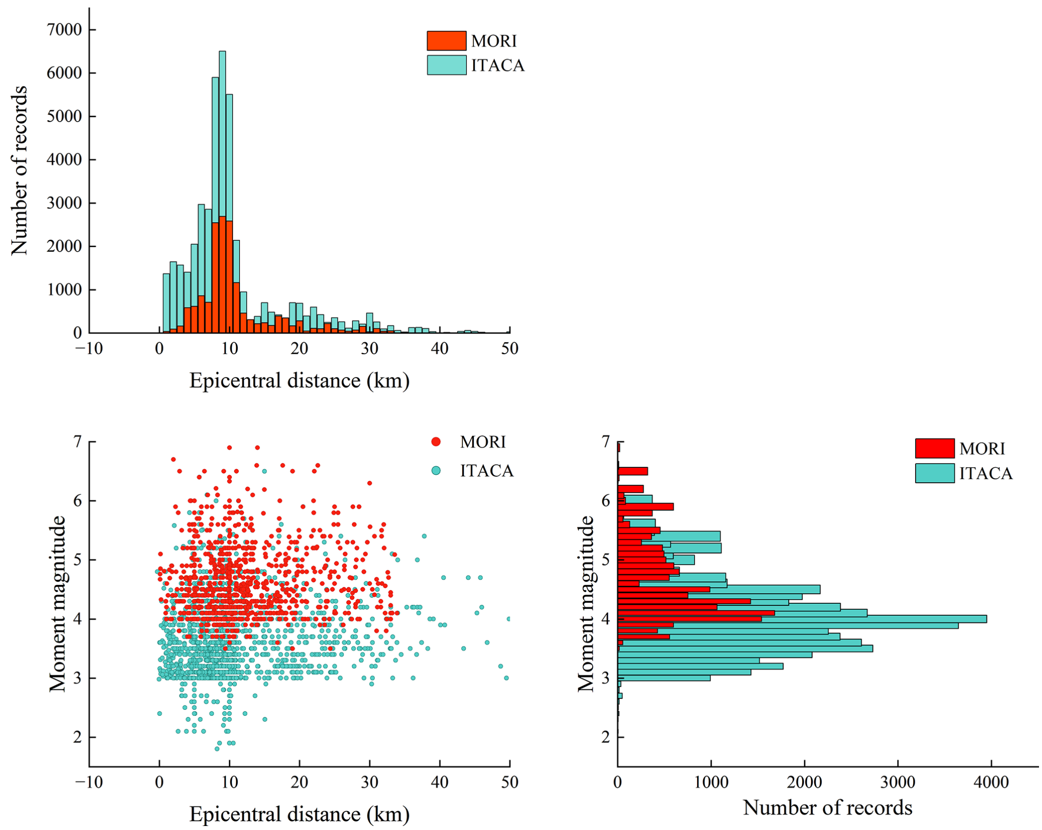

Figure 1. Comparison of data distribution for the ITACAext 2.0 (blue) and MORI (red) datasets.

contiguous in time, and no event is split across boundaries. Panel (b) reports the pairwise event-ID overlap counts – diagonal elements are the number of events in each subset, while all off-diagonal entries are zero, verifying the strict exclusivity of the event-wise assignment. Panel (c) further demonstrates that the three subsets exhibit comparable distributions of records per event (spanning approximately one to three orders of magnitude on a logarithmic scale), indicating that the natural heterogeneity of earthquake-event sizes is preserved across the train/validation/test partitions. Any model performance difference across subsets can therefore be attributed to temporal generalization rather than to a distributional mismatch in event sizes.

3 Model

The mixture of experts (MOE) model is a machine learning technique that collaboratively processes complex tasks through multiple specialized "expert" models; it has performed exceptionally well in the fields of natural language processing and computer vision in recent years. Given the nonlinearity of ground-motion data and its spatial heterogeneity, this study introduces the MOE framework into ground-motion prediction, aiming to enhance prediction accuracy and stability.

3.1 Model overview

The MOE architecture was first proposed by Jacobs et al. (1991) and has recently been fully utilized in the open-source large language model DeepSeek. DeepSeek (DeepSeek-AI et al., 2024) utilizes MOE to achieve efficient training and inference, enhancing performance and scalability. Its core concept is "divide and conquer", where a complex task is decomposed into subtasks processed by specialized expert models, and the results are then integrated through a gating mechanism. In earthquake engineering, ground-motion data possess high nonlinearity and spatial heterogeneity, making it difficult for traditional single models to comprehensively capture these characteristics. This study innovatively applies the MOE framework to ground-motion prediction, employing multiple expert models to capture multi-scale features and dynamically selecting and combining experts through a gating network. This design enables the model to adaptively handle seismic events with different tectonic environments, magnitudes, and propagation paths, providing a new paradigm for ground-motion prediction.

To achieve precise prediction of PGA and SA, this study proposes the MOE-XGB model, which combines the advantages of the XGBoost algorithm and neural networks, as shown in Fig. 3. The MOE-XGB model enhances predic-

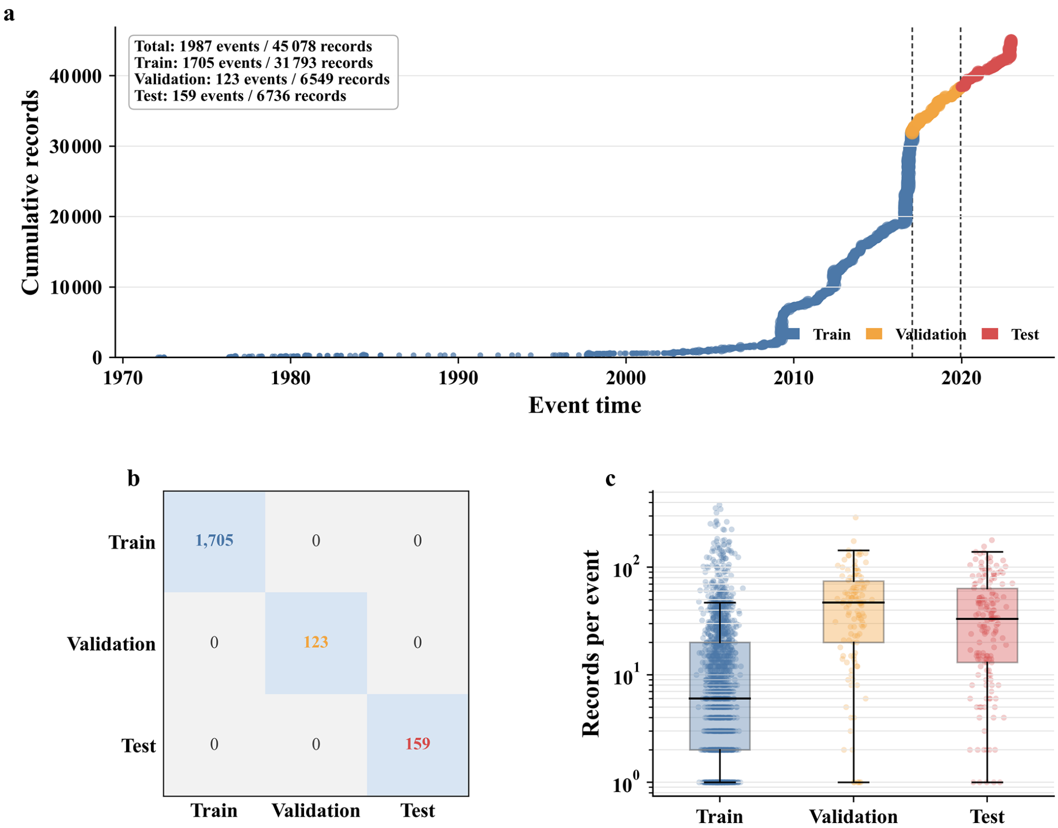

Figure 2. Chronological event-wise data split for the ITACA dataset. **(a)** Cumulative record count over time, with dashed lines and colors separating train (blue), validation (orange), and test (red) subsets. **(b)** Event-ID overlap matrix verifying strict isolation with zero leakage across subsets. **(c)** Records per event distribution confirming that event-size heterogeneity is consistently preserved across all partitions.

tive performance through collaborative division of labor, resolving the difficulty single models face in comprehensively capturing all information. The model consists of two main parts: expert models and a gating network. Through ablation experiments comparing the performance of 1, 2, 3, and 4 expert models, it was found that the test results of 3 expert models performed best across various metrics. Comprehensively considering prediction accuracy and computational efficiency, the number of expert models was finally determined to be 3. The expert models are based on the XGBoost algorithm, which excels at processing structured data such as earthquake magnitude, focal depth, and V_{s30}. This study assigns three XGBoost expert models to each SA period, employing three different parameter settings to capture diverse data features, ensuring adaptation to various seismic scenarios. The gating network is a compact neural network that assigns weights to each expert based on input earthquake features. Its processing procedure involves extracting feature representations $\boldsymbol{g}(\boldsymbol{x})$ through multi-layer linear transformations and non-linear activation functions (ReLU), and then calculating the weights $\boldsymbol{w}(\boldsymbol{x})$ using the Softmax function. The final prediction is the weighted sum of the expert outputs, expressed as the base-10 logarithm of the predicted value $\log_{10}\hat{y}$. The specific implementation is shown in Eqs. (1)–(3):

$$\boldsymbol{g}(\boldsymbol{x}) = \mathrm{Linear}_3(\mathrm{ReLU}(\mathrm{Linear}_2(\mathrm{ReLU}(\mathrm{Linear}_1 X)))) \quad (1)$$

$$\boldsymbol{w}(\boldsymbol{x}) = \mathrm{Soft\,max}(\boldsymbol{g}(\boldsymbol{x})) \quad (2)$$

$$\log_{10}\hat{y} = \sum_{k=1}^{K}\left[w(\boldsymbol{g}(\boldsymbol{x}))_k \cdot \boldsymbol{f}_k(\boldsymbol{x})\right] \quad (3)$$

where: Linear_1 transforms the input dimension to 64 dimensions; Linear_2 transforms 64 dimensions to 32 dimensions; Linear_3 transforms 32 dimensions to the number of experts; $\boldsymbol{x}$ represents the input feature vector; $\boldsymbol{g}(\boldsymbol{x})$ represents the latent feature representation; $\boldsymbol{w}_k(\boldsymbol{x})$ represents the weight assigned to the experts by the gating network; $\boldsymbol{f}_k(\boldsymbol{x})$ represents the prediction result of the kth expert model; K is the number of expert models; and $\hat{y}$ represents the predicted value of the ground-motion parameter.

This study selects 9 feature parameters, including moment magnitude, hypocentral depth, epicentral distance, V_{s30}, station elevation, and the latitude and longitude information of both the earthquake and the station. To provide a complete and unambiguous specification of the model inputs, Table 1 summarizes the full set of predictors used in the MOE-XGB

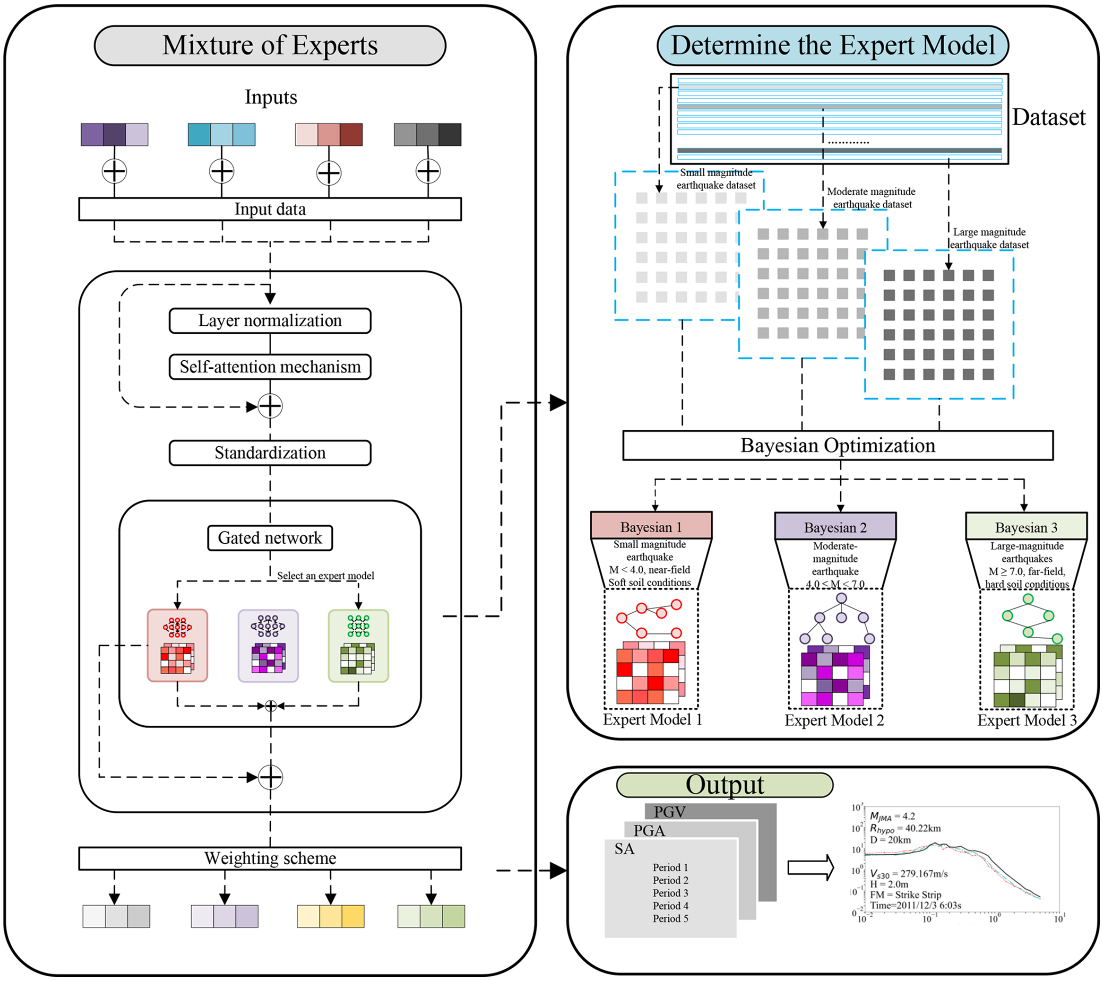

Figure 3. Schematic diagram of the MOE-XGB model structure.

framework, detailing their definitions, physical units, data sources, and the specific preprocessing steps applied.

Furthermore, the relative roles and physical consistency of these predictors in controlling the ground-motion predictions are systematically evaluated through a SHAP-based interpretability analysis in Sect. 4.4.

3.2 Bayesian optimization

To enhance the performance of the MOE-XGB model in ground-motion spectral acceleration prediction, this study adopts Bayesian optimization to determine optimal hyperparameters for each expert model, with the objective of minimizing the mean squared error (MSE) of the validation set. The goal is to generate a unique set of optimal hyperparameters for the XGBoost expert models at each SA period, ensuring parameter diversity through different dataset partitions.

The optimization process begins by loading the ground-motion dataset, containing 10 features (e.g. moment magnitude, hypocentral depth, and epicentral distance). To generate hyperparameters targeting different seismic scenarios, the training and validation sets are classified by magnitude into three parts: small earthquakes ($M < 4$), moderate earthquakes ($4 \leq M < 6$), and large earthquakes ($M \geq 6$). These sub-datasets are combined to form three groups of optimization datasets, targeting small, moderate, and large earthquake scenarios, respectively, to generate corresponding optimal hyperparameters. Bayesian optimization is conducted independently for each SA period, using the validation set MSE as the objective function, and calculating hyperparameter performance by training and evaluating the XGBoost model. The entire process uses only the training and validation sets, without the test set, thereby avoiding the risk of data leakage. Specifically, each optimization iteration explores a predefined hyperparameter search space to control model complexity and prevent overfitting. The explicitly optimized hyperparameters and their designated search ranges include the learning rate (0.01 to 0.3), maximum tree depth (3 to 10), number of estimators or boosting rounds (100 to 1000), subsample ratio of training instances (0.6 to 1.0), feature sampling ratio per tree (0.6 to 1.0), as well as the L1 and L2

Table 1. Summary of predictors used in the MOE-XGB model.

Predictor (Variable Name)	Definition and Unit	Data Source
M	Moment magnitude (dimensionless)	ITACAext 2.0
H	Hypocentral depth (km)	ITACAext 2.0
R_{epi}	Epicentral distance (km)	ITACAext 2.0
V_{s30}	Time-averaged shear-wave velocity in the top 30 m ($\mathrm{m\,s^{-1}}$)	ITACAext 2.0 / Field surveys
h	Station elevation (m)	ALOS World 3D-30m DEM
$\mathrm{Lat_{eq}}$	Earthquake latitude (°)	ITACAext 2.0
$\mathrm{Lon_{eq}}$	Earthquake longitude (°)	ITACAext 2.0
$\mathrm{Lat_{st}}$	Station latitude (°)	ITACAext 2.0
$\mathrm{Lon_{st}}$	Station longitude (°)	ITACAext 2.0

regularization terms on weights (0 to 10 and 1 to 10, respectively). This comprehensive search ensures that the three expert models obtain unique, well-regularized hyperparameter configurations, enhancing the model's adaptability and prediction accuracy for different seismic scenarios.

4 Results

4.1 Model evaluation framework

To rigorously evaluate the predictive performance and generalization capability of the proposed MoE-XGB model, a comprehensive statistical evaluation framework is established. This framework is grounded in standard residual diagnostics and mixed-effects regression, evaluated strictly under out-of-sample, event-wise independent testing conditions to prevent any event-level information leakage and guarantee unbiased verification.

Predictive performance is first assessed using total residuals. Consistent with the data preprocessing strategy and the reference empirical models (e.g., Bindi, 2017), the total residual R_{es} for the sth station recording of the eth earthquake event is calculated as the difference between the base-10 logarithms of the observed and predicted ground-motion intensity measures (IMs):

$$R_{es} = \log_{10}(Y_{\mathrm{obs},es}) - \log_{10}(Y_{\mathrm{pred},es}) \tag{4}$$

where $Y_{\mathrm{obs},es}$ represents the actual observed IM value and $Y_{\mathrm{pred},es}$ denotes the corresponding value predicted by the model based on the input features.

Furthermore, to align with standard ground-motion prediction equation (GMPE) evaluation protocols, the total residuals derived from the unseen test events are decomposed into event-specific and site/path-specific components using a standard mixed-effects linear regression:

$$R_{es} = c + \eta_e + \epsilon_{es} \tag{5}$$

where c represents the overall fixed model bias; η_e is the random inter-event residual capturing the average source-specific deviation for earthquake event e; and ϵ_{es} is the random intra-event residual capturing the site- and path-specific deviation for recording s. The random variables η_e and ϵ_{es} are assumed to be independent and normally distributed with a mean of zero. Their respective standard deviations are denoted as τ (inter-event variability) and ϕ (intra-event variability). The total aleatory variability σ of the prediction model is subsequently quantified as:

$$\sigma = \sqrt{\tau^2 + \phi^2} \tag{6}$$

Based on this mathematical framework, all subsequent performance metrics, residual scattering distributions, and aleatory variability spectra are computed and analyzed under strict event-wise independent conditions in the following subsections.

4.2 Comparison of model performance

To ensure a rigorously fair, apples-to-apples comparison with the baseline Gaussian Process Regression (GPR) study, the proposed MoE-XGB model is first evaluated under the identical benchmark protocol (designated as Track A). In this track, the 30 October 2016 Central Italy M_{w} 6.5 earthquake event, which comprises 241 strong-motion station recordings, is utilized as the strictly held-out independent test set. This significant event was entirely excluded from all model development, training, and cross-validation phases to eliminate any potential event-level information leakage.

As shown in Fig. 4 and Table 2, compared with the GPR model using an exponential kernel function, the proposed MOE-XGB model exhibits lower RMSE across all ground-motion parameters. Specifically, MOE-XGB reduced the RMSE by 12.9 % for PGA, 13.3 % for PGV, 15.2 % for SA (0.3 s), 16.7 % for SA (1.0 s), and 17.2 % for SA (3.0 s).

To compare the residual performance of the MOE-XGB and GPR (exponential) models in magnitude–distance space under a unified standard, we plotted residual comparison figures. Figure 5 displays magnitude on the horizontal axis; each subplot corresponds to the residuals of different prediction metrics, including PGA, PGV, SA (0.3 s), SA (1.0 s), and SA (3.0 s). In the figures, green scatter points represent the prediction residuals of MOE-XGB, and red scatter points

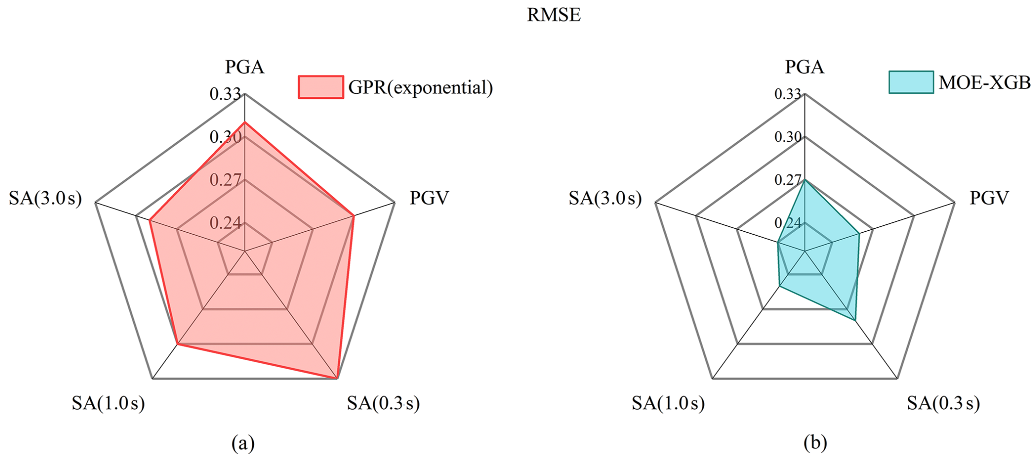

Figure 4. Comparison of RMSE between the GPR (exponential) model and the MOE-XGB model across different ground-motion parameters.

Table 2. RMSE performance of different models.

Model	PGA (Gal)	PGV ($\mathrm{cm\,s^{-1}}$)	SA ($T = 0.3$ s) (Gal)	SA ($T = 1.0$ s) (Gal)	SA ($T = 3.0$ s) (Gal)
GPR (exponential)	0.312	0.303	0.331	0.305	0.293
MOE-XGB	0.273	0.264	0.286	0.257	0.249

represent the prediction residuals of GPR. In this way, we demonstrate the spatial distribution characteristics of different models in ground-motion prediction from multiple dimensions, facilitating a more intuitive comparison of their predictive performance under different magnitude and source distance conditions.

In the overall statistical results of the five prediction metrics (PGA, PGV, SA (0.3 s), SA (1.0 s), and SA (3.0 s)), MOE-XGB and GPR (exponential) exhibit significant differences in residual distribution. For residuals with an absolute value greater than 1 ($|\delta| > 1$), the proportions corresponding to the five prediction metrics of MOE-XGB are 0.17 %, 0.13 %, 0.21 %, 0.08 %, and 0.08 %, respectively, with an overall average of only 0.13 %; whereas the corresponding proportions for GPR are 1.27 %, 0.51 %, 0.97 %, 0.59 %, and 0.68 %, respectively, with an overall average of 0.80 %. Compared to GPR, MOE-XGB reduces large residual points by approximately 83.8 %, demonstrating more robust predictive capability within the magnitude–distance space.

In the statistics for residuals with an absolute value greater than 0.5 ($|\delta| > 0.5$), the residual proportions for MOE-XGB are 2.95 %, 2.45 %, 3.80 %, 2.99 %, and 2.53 %, respectively, with an overall average of 2.94 %; whereas the corresponding proportions for GPR are 9.66 %, 7.79 %, 9.97 %, 7.46 %, and 7.09 %, respectively, with an overall average of 8.39 %. Evidently, MOE-XGB also possesses significant advantages in controlling moderate-amplitude residuals, reducing residual points by approximately 65.0 % compared to GPR.

In summary, MOE-XGB exhibits a more concentrated residual distribution, with the majority of residuals concentrated in the $|\delta| < 0.5$ interval, significantly outperforming GPR. Combined with the graphical results, it can be further observed that the residuals of MOE-XGB are distributed more tightly around the $\delta = 0$ axis overall, while the residuals of GPR present a wider distribution range and a higher proportion of large residual points. This fully demonstrates that MOE-XGB possesses more prominent advantages in both the accuracy and stability of ground-motion prediction.

To comprehensively evaluate the prediction residual characteristics of the MOE-XGB and GPR (exponential) models under five ground-motion intensity measures (PGA, PGV, and spectral acceleration at $T = 0.3$, 1.0, and 3.0 s), this paper selected two statistical metrics for comparative analysis: peak density and peak position. Among them, peak density reflects the degree of concentration of residuals around zero; a higher value indicates that the model prediction errors are more concentrated. Peak position characterizes the symmetry of the residual distribution; a value closer to zero indicates smaller systematic bias. The MOE-XGB model exhibits higher peak density across all intensity measures, indicating a more concentrated residual distribution. Meanwhile, the peak position of MOE-XGB is closer to zero, indicating that the model does not exhibit significant systematic bias. These features are also intuitively reflected in Fig. 6: the residual distribution of MOE-XGB (top row) is generally sharper and more symmetrical, whereas the GPR model (bottom row) presents flatter and more diffused characteristics.

In summary, the MOE-XGB model not only possesses higher prediction accuracy but also comprehensively outperforms the GPR model in terms of the concentration, symme-

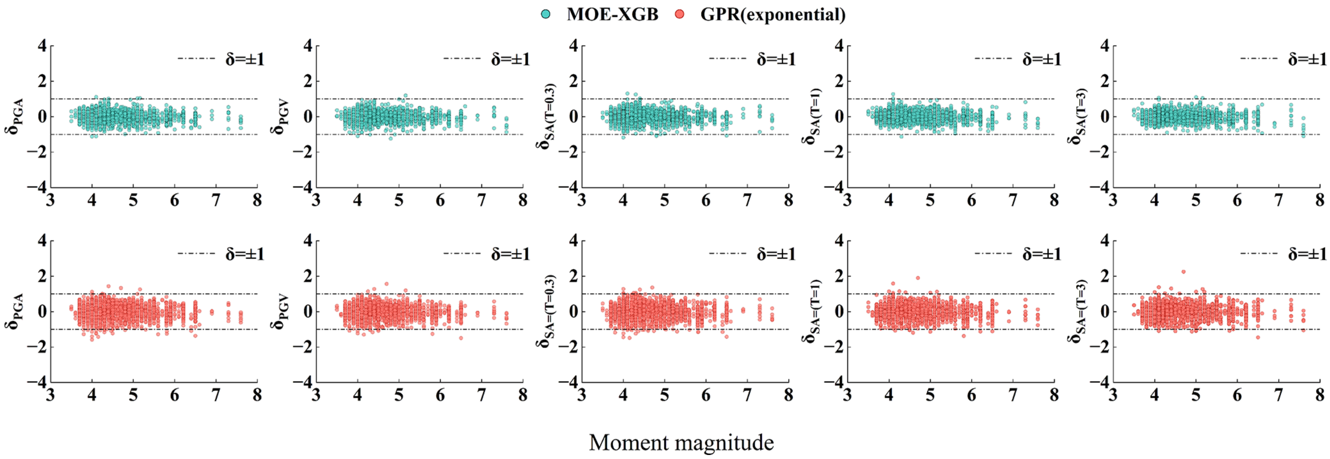

Figure 5. Variation of residuals with magnitude for the MOE-XGB and GPR (exponential) models based on the MORI dataset for PGA and different SA periods.

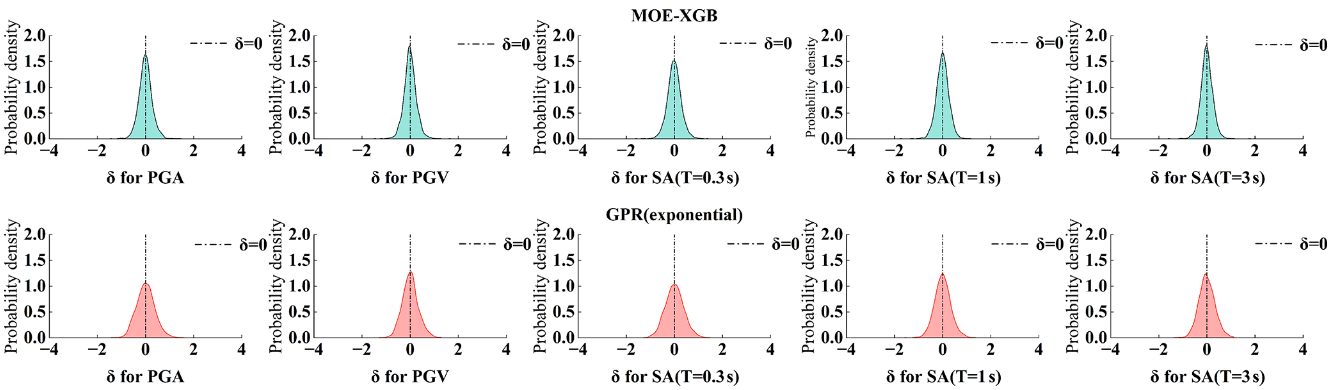

Figure 6. Comparison of residual probability density distributions for the MOE-XGB and GPR (exponential) models based on the MORI dataset for PGA and different SA periods.

try, and stability of residual distribution, demonstrating its potential advantages in ground-motion parameter modeling.

4.3 Comprehensive model evaluation

To evaluate the robustness, scalability, and temporal generalization of the proposed MoE-XGB model on a broader regional scale, the framework is trained and verified using the expanded official ITACAext 2.0 dataset (designated as Track B). In this track, model development and hyperparameter optimization are performed directly utilizing the chronologically partitioned training and validation subsets. Following this phase, the finalized MoE-XGB model is evaluated strictly on the held-out chronological test dataset. As established in Sect. 2, this test dataset consists entirely of unseen "future" earthquakes that occurred after the training timeline. Evaluating the framework exclusively on these chronologically subsequent events effectively eliminates any potential event-level information leakage, providing an unbiased verification of the model's true generalization capability.

Table 3. Performance comparison and sensitivity to data splitting strategy.

Model	Validation Strategy	RMSE	R	SD
MOE-XGB	Record-wise split	0.225	0.956	0.265
MOE-XGB	Event-wise evaluation	0.347	0.844	0.338
Bindi 2017	Event-wise evaluation	0.384	0.810	0.372
GPR	Event-wise evaluation	0.395	0.794	0.391

To explicitly quantify the impact of the validation design, a sensitivity analysis is conducted. Table 3 compares the performance metrics of the models under the previously adopted random record-wise split versus the new, strict event-wise chronological evaluation.

As expected, migrating from a random record-wise split to the strict event-wise split yields a decrease in the apparent performance of the MoE-XGB model (e.g., the Root Mean Square Error (RMSE) increases from 0.225 to 0.347, and the Pearson correlation coefficient (R) decreases from 0.956

to 0.844). This shift explicitly confirms that random record-wise partitioning suffers from event-level information leakage, which previously led to overly optimistic estimates. By removing this leakage, the updated event-wise metrics now objectively reflect the model's true predictive capacity for completely unknown earthquakes.

Crucially, as demonstrated in Table 3, even under this stringent, fully independent event-wise evaluation framework, the proposed MoE-XGB model consistently outperforms both the machine-learning baseline (GPR) and the empirical benchmark (Bindi 2017). For instance, the MoE-XGB model achieves an RMSE of 0.347, which remains significantly lower than that of Bindi 2017 (0.384) and the GPR model (0.395). Similarly, the MoE-XGB model maintains a higher correlation coefficient ($R = 0.844$) compared to the baselines (0.810 and 0.794, respectively), alongside a lower standard deviation (0.338 vs. 0.372 and 0.391). This rigorous sensitivity analysis explicitly confirms that, although the absolute metrics have been corrected to reflect true generalization, the relative superiority and robust predictive advantages of the proposed MoE-XGB framework remain strongly supported.

Building upon the performance metrics presented in Table 3, the total residual analysis further substantiates the predictive superiority of the proposed MoE-XGB model under strict event-wise independent testing. Figure 7 presents the total residual scatter plots for the three models across different spectral periods. Visually, the residuals of the MoE-XGB model are more densely concentrated around the zero mean, exhibiting no obvious systematic bias.

Rather than focusing solely on average metrics, it is crucial to evaluate a model's robustness against severe mispredictions. To quantitatively assess this, we analyze the frequency of large prediction deviations – specifically, instances where the absolute total residual exceeds 1.0 (which corresponds to an order-of-magnitude difference between the observed and predicted intensity). For the PGA period, the MoE-XGB model produces only 71 such large residuals, whereas the GPR and Bindi 2017 models produce 127 and 131 instances, respectively. This represents significant reductions of approximately 44.1 % and 45.8 %.

This enhanced capability to constrain extreme prediction deviations is consistently maintained across the spectral acceleration periods. For SA ($T = 0.3$ s), the number of large residuals for MoE-XGB (60 points) is reduced by 53.8 % compared to GPR (130 points), and by 3.2 % compared to Bindi 2017 (62 points). For SA ($T = 1.0$ s), MoE-XGB yields only 20 such deviations, corresponding to reductions of 64.9 % and 13.0 % against GPR (57 points) and Bindi 2017 (23 points). Similarly, at the long period of SA ($T = 3.0$ s), the MoE-XGB model exhibits a clear advantage with only 15 large residual occurrences, representing reductions of 83.5 % and 44.4 % compared to GPR (91 points) and Bindi 2017 (27 points).

To further investigate the physical consistency of the proposed model and its ability to isolate prediction uncertainty, the total residuals are decomposed into inter-event (η_e) and intra-event (ϵ_{es}) components using the mixed-effects linear regression defined in Eq. (5). Both residual components are plotted against earthquake magnitude (M) to evaluate source scaling and statistical decoupling.

Figure 8 illustrates the random inter-event residuals (η_e) across different spectral periods. As formulated in Eq. (5), η_e isolates the average source-specific deviation for each independent earthquake. Visually, the η_e values of the MoE-XGB model cluster tightly around the zero-bias baseline, exhibiting a remarkably flat trend line across the entire magnitude spectrum. In contrast, traditional empirical models like Bindi 2017, constrained by rigid functional forms, often exhibit distinct upward or downward bias trends at the lower and upper magnitude bounds. The superior flatness of the MoE-XGB trend demonstrates that the gating network effectively optimizes the weights of specialized experts for different magnitude regimes, thereby eliminating systematic under- or over-estimation in source scaling.

Furthermore, Fig. 9 presents the random intra-event residuals (ϵ_{es}) also plotted against magnitude (M). This specific visualization is designed to verify the statistical decoupling between the source term and the path/site effects. Ideally, intra-event residuals – which capture wave attenuation and site amplification – should be completely independent of earthquake size. Across all evaluated periods, the ϵ_{es} scatter of the MoE-XGB model exhibits a highly uniform, symmetric distribution around zero, maintaining consistent dispersion (homoscedasticity) from small to large magnitudes. This indicates a rigorous mathematical decoupling within the MoE-XGB architecture: the source information is entirely absorbed by the inter-event term, leaving no residual magnitude-dependent pollution in the intra-event variance. Such strict orthogonality further confirms the robust physical generalization of the framework.

To further validate the statistical properties of the prediction errors, Fig. 10 compares the probability density distributions of the total residuals for the three models across representative spectral periods. In standard ground-motion prediction modeling, residuals are theoretically required to follow a normal distribution to be compatible with probabilistic seismic hazard analysis (PSHA). Visually, the residual distributions of all three models conform well to this standard normal assumption, exhibiting classic bell-shaped curves centered near zero.

However, a distinct differentiation in predictive precision is evident from the shape of these density curves. Across all evaluated periods, the probability density curve of the MoE-XGB model is consistently taller and narrower than those of the GPR and Bindi 2017 models. Quantitatively, for the PGA, the MoE-XGB model achieves a lower standard deviation (0.377) compared to the baseline GPR (0.421) and empirical Bindi 2017 (0.423) models, representing signif-

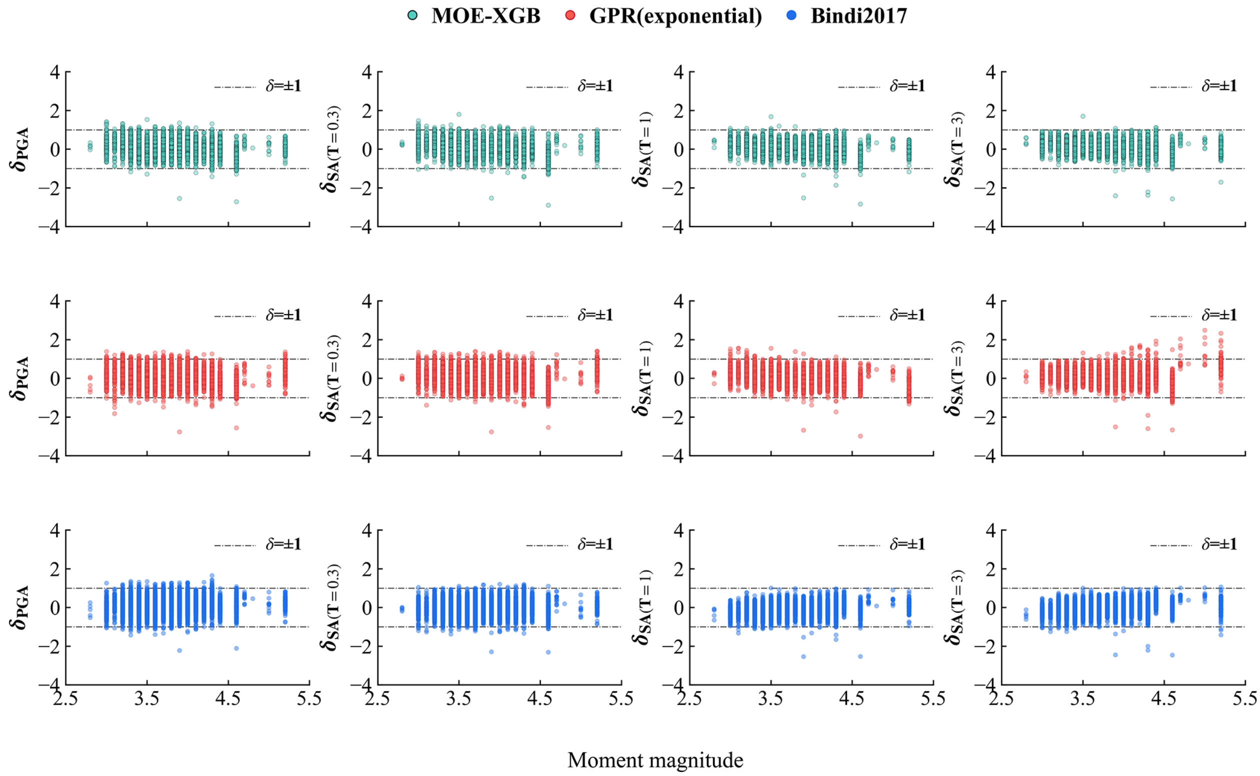

Figure 7. Variation of residuals with moment magnitude for different models based on the ITACAext 2.0 dataset for PGA and different SA periods.

icant reductions in prediction dispersion of approximately 10.5 % and 10.9 %, respectively. This pattern of reduced variance is strictly maintained at longer spectral periods. For instance, at SA ($T = 0.3$ s), the standard deviation of the MoE-XGB model is well constrained to 0.358, outperforming GPR (0.412) by 13.1 % and Bindi 2017 (0.388) by 7.7 %.

A narrower distribution with a higher peak density statistically confirms that the prediction errors of the MoE-XGB model are more densely concentrated around the mean (zero bias). This explicitly demonstrates that the proposed framework not only preserves the physical Gaussian assumption required for downstream hazard engineering but also significantly reduces the overall aleatory variability (uncertainty) of the predictions relative to conventional machine-learning and empirical approaches.

To intuitively demonstrate the practical engineering applicability of the models, Fig. 11 presents a direct comparison between the predicted response spectra and the actual measured strong-motion records. To rigorously prevent data leakage and evaluate true temporal generalization, we selected two recent, unseen earthquake events from the strictly held-out chronological test dataset for this demonstration. The key physical parameters of these two events, including magnitude, source-to-site distance, and focal depth, are detailed in the subplots of Fig. 11.

As illustrated in the two panels of Fig. 11, the continuous response spectra predicted by the proposed MoE-XGB model exhibit remarkable consistency with the actual ground-truth recordings across the entire period range (from 0.01 to 4.0 s). Through an in-depth comparison across different spectral bands, a distinct differentiation in the predictive capabilities of the models can be clearly observed:

In the high-frequency, short-period region (approximately 0.1–0.5 s): Due to local site amplification effects and complex non-linear source mechanisms, actual response spectra typically generate severe resonance peaks in this interval. Within this band, the traditional empirical Bindi 2017 model, constrained by its rigid functional form, fails to adaptively capture these highly non-linear spectral peak features, resulting in obvious overestimation. The baseline GPR model also deviates from the true spectral shape due to degraded generalization capabilities on unseen earthquakes. In sharp contrast, the prediction curve of the MoE-XGB model is tightly coupled with the actual strong-motion records in this range, reproducing the predominant periods and peak amplitudes with exceptional precision. This holds critical engineering significance for accurately assessing the seismic demand on rigid structures and low-rise buildings.

In the mid-to-long-period region (1.0–4.0 s): Ground motion gradually transitions to being controlled by low-frequency wave propagation and attenuation. In this interval,

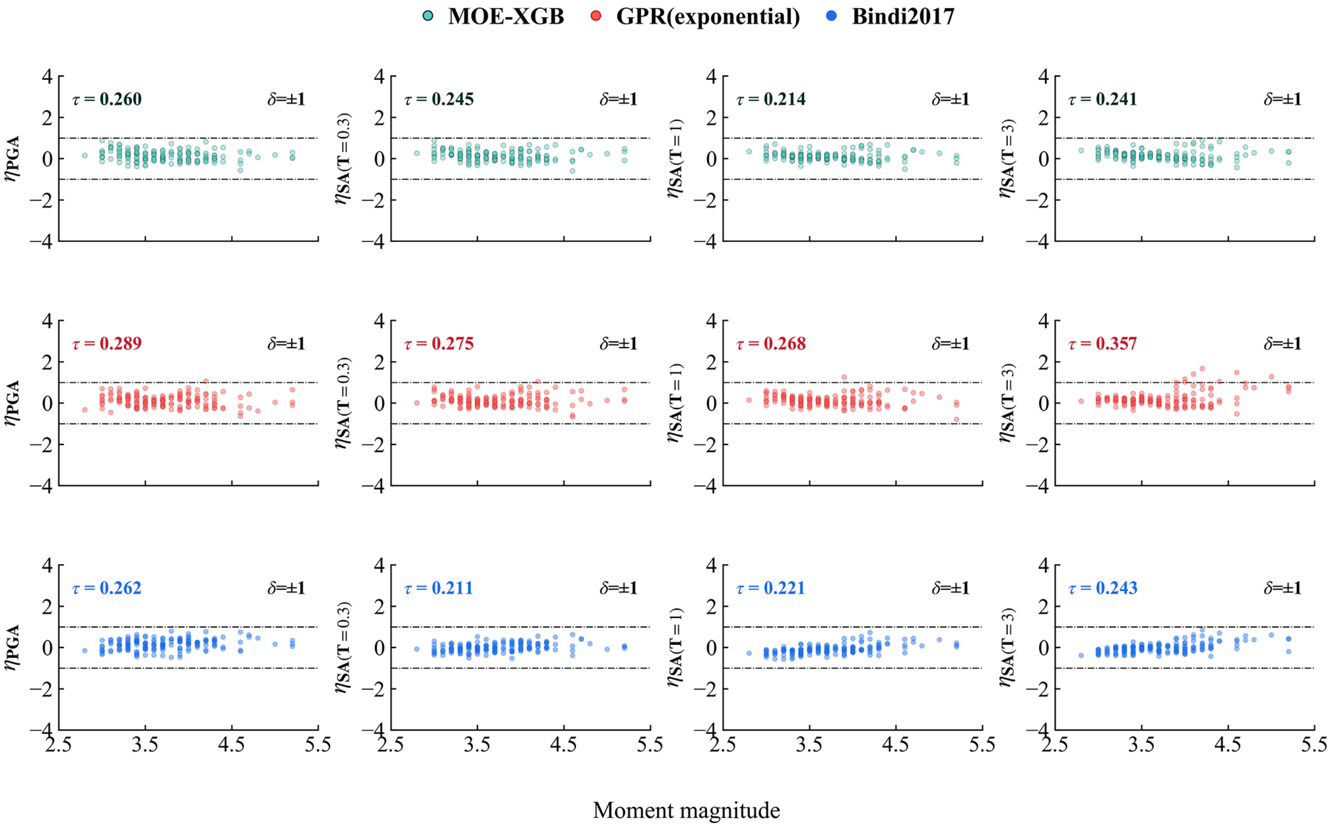

Figure 8. Inter-event residual analysis for different models based on the ITACAext 2.0 dataset for PGA and different SA periods.

the prediction curves of the traditional empirical model and the standard machine-learning model exhibit noticeable deviations and gaps from the actual records. However, the MoE-XGB model manages to smoothly and stably trace the natural attenuation trend of the actual records in this low-frequency band.

As established in the mixed-effects regression framework in Sect. 4.1 (Eqs. 5 and 6), the total aleatory variability (σ) of the ground-motion models can be decomposed into inter-event (τ) and intra-event (ϕ) standard deviations. Evaluating the continuous evolution of these variability components across the entire spectral period range provides essential insights into the physical interpretability and engineering reliability of the models.

Figure 12 illustrates the period-dependence of these variability components (σ, τ, and ϕ in natural logarithm (ln) scale) for the proposed MoE-XGB model, alongside the GPR and Bindi 2017 reference models. All MoE-XGB residuals presented here are computed exclusively from the strictly held-out chronological test events, ensuring that the reported variabilities reflect true out-of-sample uncertainty without any event-level information leakage.

As shown in Fig. 12a, the MoE-XGB model achieves a significant reduction in the total standard deviation (σ) across almost the entire period spectrum. For instance, at PGA ($T = 0.01$ s), σ is reduced from 0.973 (Bindi 2017) and 0.969 (GPR) down to 0.868 for the MoE-XGB model. This overall reduction is fundamentally driven by the framework's ability to constrain specific variance components. Figure 12b presents the inter-event variability (τ). The MoE-XGB model exhibits highly competitive performance in modeling source-specific deviations. While the empirical Bindi 2017 model performs slightly better in certain intermediate periods (e.g., at $T = 0.3$ s, $\tau = 0.485$ for Bindi vs. 0.564 for MoE-XGB), the MoE-XGB model effectively lowers the inter-event uncertainty at both short-period and long-period ranges (e.g., at PGA, τ drops from 0.604 for Bindi to 0.599 for MoE-XGB; at $T = 1.0$ s, from 0.509 to 0.492). This indicates that the mixture-of-experts gating mechanism is highly adaptable to varying earthquake magnitudes and rupture complexities across broad frequency bands. Most crucially, as depicted in Fig. 12c, the MoE-XGB model demonstrates a predominant advantage in constraining intra-event variability (ϕ). The MoE-XGB model yields significantly lower ϕ values compared to both the GPR and Bindi 2017 models across all evaluated spectral periods. At PGA, the MoE-XGB model achieves a ϕ value of 0.675, massively outperforming the GPR ($\phi = 0.796$) and the empirical baseline ($\phi = 0.816$). This structural decrease explicitly confirms that the independent tree-based experts within the framework capture the

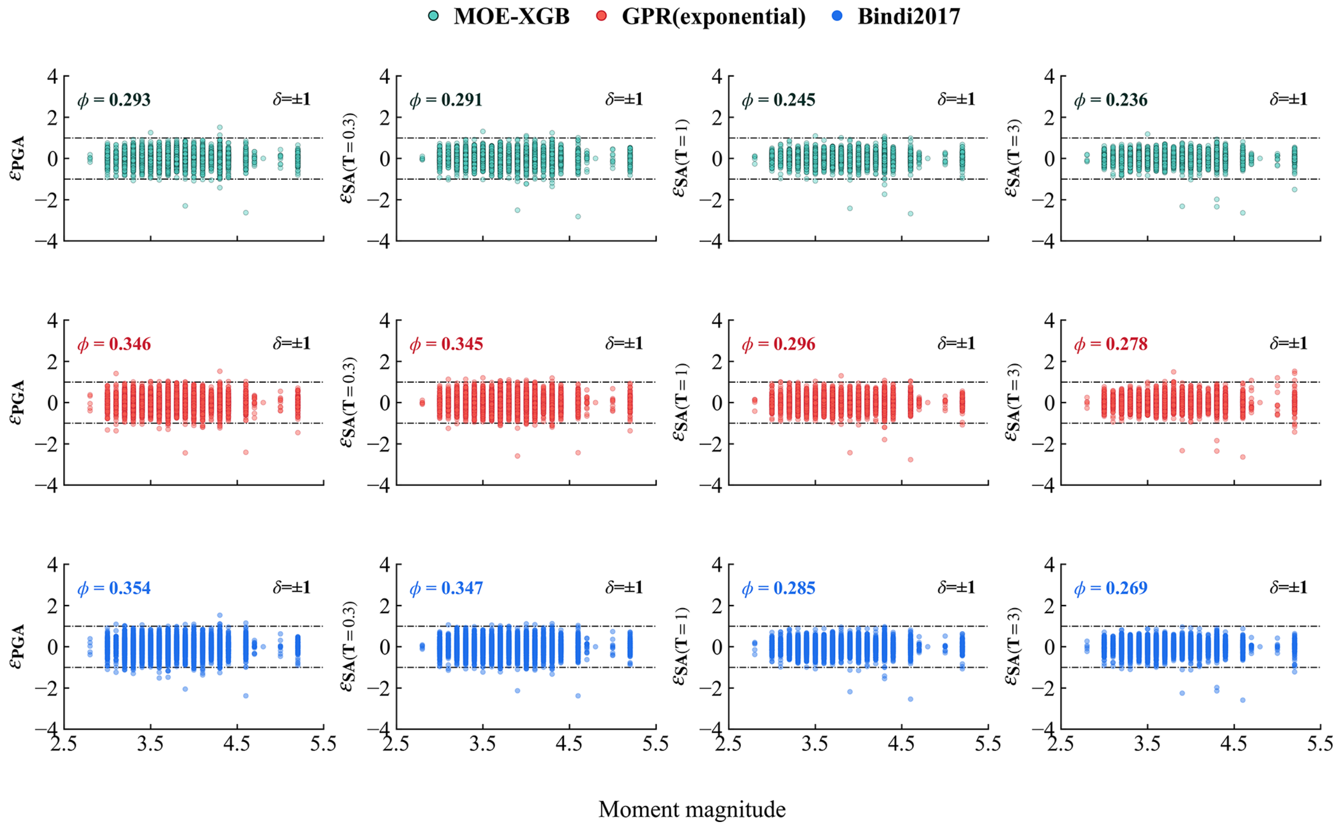

Figure 9. Intra-event residual analysis for different models based on the ITACAext 2.0 dataset for PGA and different SA periods.

highly non-linear geometric spreading and local site amplification effects much more effectively than rigid empirical equations. Ultimately, this substantial reduction in intra-event variance drives the overall drop in aleatory variability, translating directly to narrower and more precise confidence intervals for Probabilistic Seismic Hazard Analysis (PSHA).

4.4 Model interpretability and physical consistency

While the MoE-XGB framework serves as an effective optimization of current mainstream machine learning models, evaluating its physical consistency is equally crucial to ensure that the framework captures meaningful seismological relationships rather than purely data-driven artifacts. To this end, a global interpretability analysis was conducted using SHapley Additive exPlanations (SHAP). Figure 13 presents a composite visualization of the SHAP analysis across four representative intensity measures: PGA, and spectral accelerations (SA) at $T = 0.3$, 1, and 3 s. The top row (Fig. 13a–d) displays SHAP beeswarm plots illustrating the distribution of feature effects, while the bottom row (Fig. 13e–h) presents the corresponding SHAP summary bar plots representing the mean absolute global importance of the predictors.

The SHAP bar plots reveal the relative roles of the predictors in controlling the ground-motion outputs. Consistently across all IMs, source-to-site distance (R_{hypo}) and magnitude (M_w) dominate the predictions, aligning with fundamental ground-motion attenuation principles. Furthermore, the quantitative analysis successfully captures the frequency-dependent evolution of feature importance. For high-frequency IMs, distance plays the most critical role; in the PGA model, R_{hypo} ranks first with a mean absolute SHAP value of 0.42, closely followed by M_w (0.31). Conversely, as the spectral period increases, the global importance of magnitude expands significantly. For SA ($T = 1$ s) and SA ($T = 3$ s), M_w surpasses R_{hypo} to become the primary driving feature, reaching a peak mean SHAP value of 0.42 at $T = 3$ s (compared to 0.28 for R_{hypo}). This physically consistent shift reflects the seismological fact that large-magnitude earthquakes excite proportionally more low-frequency (long-period) seismic energy. Additionally, spatial parameters such as station latitude and longitude demonstrate notable importance, particularly at short periods (ranking third and fourth for PGA, with SHAP values of 0.11 and 0.09, respectively), indicating the model's ability to implicitly capture unmodeled regional path or local site effects.

The SHAP beeswarm plots (Fig. 13a–d) further validate the physical directionality of these learned features. In these plots, each point represents a single seismic record, with its color denoting the original feature value (red for high, blue for low) and its horizontal position indicating the SHAP

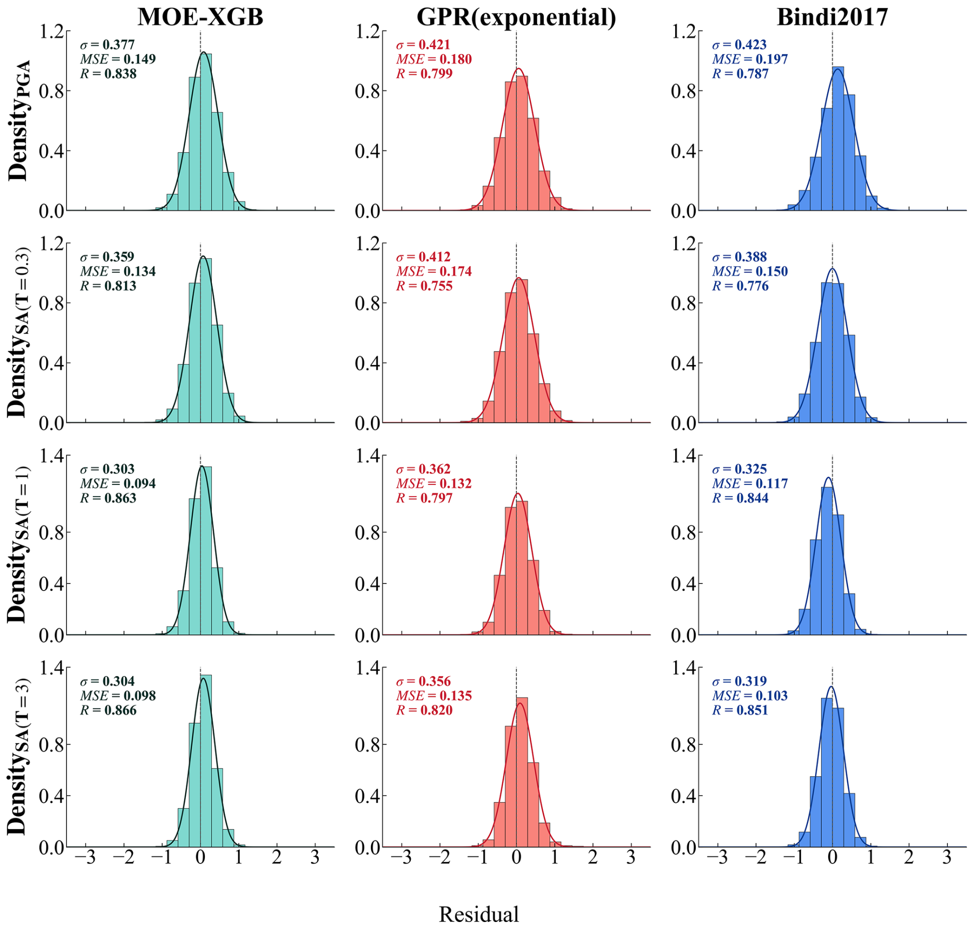

Figure 10. Comparison of residual probability density distributions for different models based on the ITACAext 2.0 dataset for PGA and different SA periods.

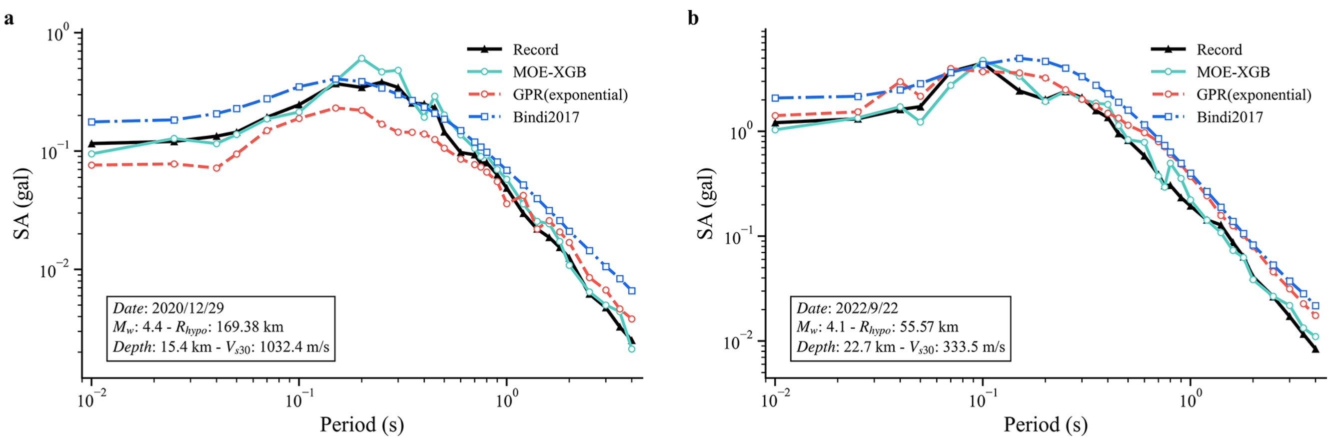

Figure 11. Comparison of predicted values and measured records for different mod.

impact on the model output. The relationships learned by the MoE-XGB model strictly adhere to expected physical laws across all analyzed parameters. For magnitude (M_w), high feature values consistently yield substantial positive SHAP values across all periods, demonstrating that larger earthquakes correctly drive higher predicted ground motions, which reflects expected magnitude scaling. Regarding distance (R_hypo), high feature values representing long distances cluster strongly on the negative side of the SHAP axis, while short distances show positive impacts, accurately capturing

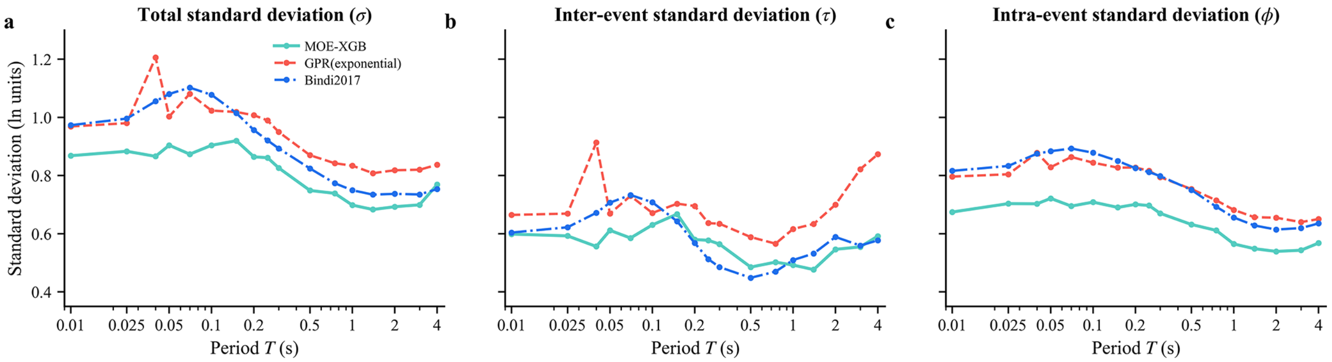

Figure 12. Comparison of the aleatory variability components across spectral periods for the proposed MOE-XGB model, the GPR model, and the Bindi 2017 empirical GMPE. Panels **(a)**–**(c)** show the total standard deviation (σ), inter-event standard deviation (τ), and intra-event standard deviation (ϕ), respectively. All values are computed under a strict event-wise cross-validation framework. The horizontal axis represents the spectral period T on a logarithmic scale, with Peak Ground Acceleration (PGA) plotted at $T = 0.01$ s.

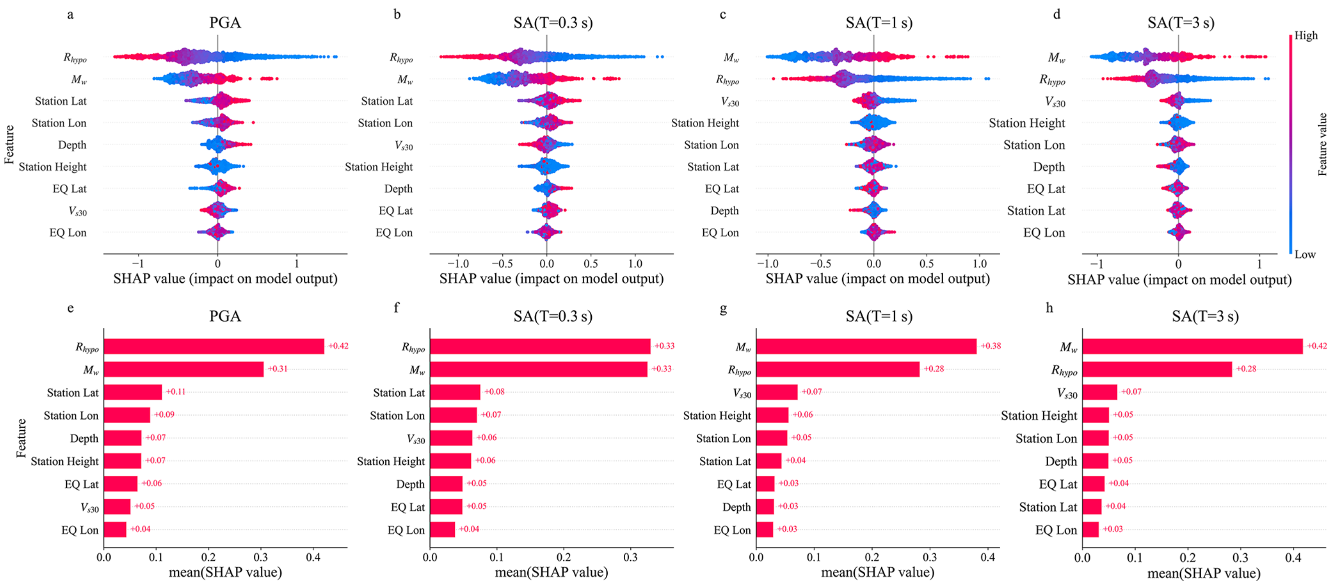

Figure 13. SHAP-based global interpretability analysis of the MoE-XGB model for predicting ground motion intensity measures. The top row **(a–d)** displays SHAP beeswarm plots for PGA, SA ($T = 0.3$ s), SA ($T = 1$ s), and SA ($T = 3$ s), respectively, illustrating the distribution and direction of feature effects on the model output (red indicates high feature values, while blue indicates low feature values). The bottom row **(e–h)** presents the corresponding SHAP summary bar plots, ranking the global importance of each predictor based on its mean absolute SHAP value.

the geometric spreading and anelastic attenuation of seismic waves. In terms of site conditions (V_{S30}), lower values corresponding to softer soils generally extend into the positive SHAP region, effectively modeling local site amplification effects. Conversely, high V_{S30} values representing stiff soil or rock correspond to negative or neutral SHAP impacts, indicating de-amplification. By explicitly mapping these dependencies, the quantitative SHAP analysis confirms that the MoE-XGB framework achieves its robust predictive capabilities by correctly learning and synthesizing complex, yet physically sound, source, path, and site effects across diverse spectral periods.

5 Discussion

To explicitly position the proposed MoE-XGB framework within the current landscape of Italian ground-motion modeling, it is essential to compare its performance with recent advanced approaches. Table 4 summarizes the methodological differences and quantitative residual variabilities (σ in $\log_{10}$ units) between our proposed framework and the recent XGBoost-based model developed by Mori et al. (2026).

As shown in Table 4, Mori et al. (2026) achieved an impressive total residual variability (e.g., $\sigma = 0.294$ at PGA) by employing complex finite-fault distance metrics (e.g., R_{rup}, R_x) and high-resolution DEM-based geomorphomet-

Table 4. Performance comparison and sensitivity to data splitting strategy.

indicator	Mori et al. (2026)	This study	
	XGBoost	MoE-XGB (Independent single-event evaluation)	MoE-XGB (Chronological regional evaluation)
σ_{PGA}	0.294	0.273	0.377
σ_{PGV}	0.280	0.264	0.331
$\sigma_{\mathrm{SA}\,(T=0.3\,\mathrm{s})}$	0.308	0.286	0.358
$\sigma_{\mathrm{SA}\,(T=1\,\mathrm{s})}$	0.299	0.257	0.303

ric predictors on a highly filtered subset of 90 major events. When evaluated under a conceptually comparable baseline – our strictly independent single-event evaluation targeting the M_{w} 6.5 Central Italy earthquake – the proposed MoE-XGB model yields a highly competitive σ of 0.273 at PGA and 0.257 at SA ($T = 1.0\,\mathrm{s}$), demonstrating its strong predictive accuracy.

Furthermore, the broader capability of the MoE-XGB framework is demonstrated through our chronological regional evaluation. While the total variability in this scenario is numerically higher (PGA $\sigma = 0.377$), this explicitly reflects the massive scale and extreme heterogeneity of the testing conditions. This evaluation encompasses 1987 highly heterogeneous events, including a vast number of complex small-magnitude tremors with inherently high aleatory variability, evaluated on completely unseen “future” events. Most importantly, unlike models dependent on detailed rupture characterizations, the MoE-XGB framework is deliberately designed to rely exclusively on readily accessible parameters (e.g., point-source distances and scalar V_{s30}). Obtaining accurate finite-fault rupture geometries in the immediate aftermath of a seismic event is notoriously difficult. Thus, relying on this fundamental predictor subset provides the MoE-XGB model with a distinct operational advantage for rapid post-earthquake emergency response and automated ShakeMap generation. Nevertheless, the findings of Mori et al. (2026) offer a valuable reference; integrating their advanced geomorphometric features into the MoE routing mechanism represents a highly promising direction to further enhance the model’s predictive precision without compromising its regional scalability.

6 Conclusions

The MOE-XGB model proposed in this study demonstrates significant advantages in ground-motion prediction compared to the traditional GPR and Bindi 2017 models. Through multiple evaluation metrics, such as RMSE, the R value, and standard deviation (SD), we verified the superiority of MOE-XGB in terms of prediction accuracy and stability. Particularly in the residual analysis of short and long periods, the MOE-XGB model exhibited more precise and consistent prediction results under different seismic event and station conditions. Compared with the GPR and Bindi 2017 models, the prediction residuals of MOE-XGB are smaller, and across multiple periods, the model can better capture ground-motion characteristics, displaying strong adaptability and stability.

Furthermore, this study further verified the generalization capability of the MOE-XGB model in various ground-motion scenarios through intra-event and inter-event residual analyses. The intra-event residual analysis indicates that MOE-XGB provides more precise prediction results under different station conditions for the same seismic event, excluding the risk of data leakage and ensuring the independence and reliability of model evaluation. The inter-event residual analysis further indicates that MOE-XGB can effectively handle variations between different seismic events, demonstrating stronger model consistency and stability.

By validating on the latest ground-motion records in the ITACAext 2.0 dataset, we ensured that the model still possesses high prediction accuracy when facing unknown data. This design avoids the limitations of the dataset, particularly against the backdrop where the traditional dataset is updated only until 2022, MOE-XGB demonstrated good adaptability and performance when facing the latest seismic events.

Overall, the MOE-XGB model exhibits significant advantages by virtue of its excellent predictive capability, stability, and high accuracy, especially in practical engineering applications in the field of ground-motion prediction. Furthermore, high-precision ground-motion prediction models form the foundation for Probabilistic Seismic Hazard Analysis (PSHA) (Weatherill et al., 2024) and serve as a cornerstone for assessing future seismic risks. Future research can further explore the application of MOE-XGB in different seismic scenarios, especially in complex ground-motion environments and broader seismically active regions, to verify its applicability and scalability in a wider range of engineering problems.

This study proposes a ground-motion prediction model for the Italian region based on a Mixture-of-Experts (MoE-XGB) framework. To ensure rigorous evaluation and address potential event-level information leakage, the model was developed and verified using a strict event-wise chronological splitting strategy on the expanded ITACAext 2.0 dataset.

The comprehensive statistical evaluation and physical interpretability analysis yield the following primary conclusions:

First, under strictly independent testing conditions on unseen "future" earthquakes, the MoE-XGB model consistently outperforms the baseline Gaussian Process Regression (GPR) and the empirical Bindi 2017 models. By adaptively routing inputs to specialized tree-based experts, the framework accurately captures both high-frequency peak characteristics and low-frequency attenuation trends, exhibiting a remarkably tight fit with actual measured strong-motion records.

Second, mixed-effects residual decomposition confirms that the MoE-XGB architecture achieves a physically sound decoupling of source and path/site effects. The framework significantly reduces both inter-event (τ) and intra-event (ϕ) variabilities across all evaluated spectral periods. Compared to traditional models, the substantial reduction in the overall aleatory variability (σ) explicitly translates to a higher concentration of prediction errors around the zero-bias mean, successfully maintaining the critical Gaussian distribution assumption. Furthermore, global interpretability analysis using SHAP successfully opens the "black box" of the machine-learning regressor, quantitatively mapping the physically consistent directional impacts and the frequency-dependent evolution of magnitude scaling, geometric spreading, and local site amplification.

Finally, while the MoE-XGB model demonstrates superior predictive accuracy, we explicitly acknowledge that transitioning from a predictive ground-motion engine to a comprehensive Probabilistic Seismic Hazard Analysis (PSHA) workflow involves specific complexities, such as modeling spatial cross-correlation and extended rupture geometries. Accordingly, the current framework is proposed as a high-fidelity foundational predictive engine rather than an immediate, off-the-shelf PSHA tool. Nevertheless, by providing robust median predictions alongside explicitly decomposed τ and ϕ standard deviations, the essential outputs of the MoE-XGB model are structurally compatible with standard open-source hazard engines (e.g., OpenQuake). The practical feasibility of integrating this MoE architecture into simulation-based PSHA workflows has already been successfully demonstrated in prior operational assessments in other regions, such as Chile. Future research will focus on extending the framework to explicitly incorporate complex rupture geometries and spatial correlation models, ultimately advancing its full integration into regional seismic hazard and risk mapping in Italy.

Code and data availability. The MORI ground-motion dataset was downloaded from https://data.ingv.it/dataset/404#additional-metadata (last access: 8 December 2025). The ITACAext 2.0 dataset was downloaded from https://itaca.mi.ingv.it/ItacaNet_40/#/home (last access: 12 December 2025). The model code is available from the corresponding author upon request.

Author contributions. JD designed the model and prepared the original draft. ZW supervised the project and secured funding. DZ and XW contributed to data curation. JW and ZL performed the formal analysis and validation. ZDW and JX assisted with visualization. All authors contributed to the review and editing of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Financial support. This research has been supported by the National Natural Science Foundation of China, Key Programme (grant nos. 52378544 and 52378543) and the Institute of Engineering Mechanics, China Earthquake Administration (grant no. 2023A01).

Review statement. This paper was edited by Marleen de Ruiter and reviewed by two anonymous referees.

References

Abrahamson, N. and Silva, W.: Summary of the Abrahamson & Silva NGA ground-motion relations, Earthq. Spectra, 24, 67–97, https://doi.org/10.1193/1.2924360, 2008.

Agrawal, H. and McCloskey, J.: Estimating ground motion intensities using simulation-based estimates of local crustal seismic response, Nat. Hazards Earth Syst. Sci., 24, 3519–3536, https://doi.org/10.5194/nhess-24-3519-2024, 2024.

ALOS: Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)", Japan Aerospace Exploration Agency (JAXA), https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm, last access: 24 May 2021.

Bindi, D.: The predictive power of ground-motion prediction equations, B. Seismol. Soc. Am., 107, 735–742, https://doi.org/10.1785/0120160251, 2017.

Bindi, D., Luzi, L., Pacor, F., Franceschina, G., and Castro, R. R.: Ground-motion prediction from empirical attenuation relationships versus recorded data: the case of the 1997–1998 Umbria-Marches, Central Italy, strong motion data set, B. Seismol. Soc. Am., 96, 984–1002, https://doi.org/10.1785/0120050102, 2006.

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., and Paolucci, R.: Ground motion prediction equations derived from the Italian strong motion database, B. Earthq. Eng., 9, 1899–1920, https://doi.org/10.1007/s10518-011-9313-z, 2011.

Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Puglia, R., and Augliera, P.: Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 %-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset, B. Earthq. Eng., 12, 391–430, https://doi.org/10.1007/s10518-013-9525-5, 2014.

Boore, D. and Atkinson, G.: Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 %-damped PSA at spectral periods between 0.01 s and 10.0 s, Earthq. Spectra, 24, 99–138, https://doi.org/10.1193/1.2830434, 2008.

Brando, G., Pagliaroli, A., Cocco, G., and Di Buccio, F.: Site effects and damage scenarios: The case study of two historic centers following the 2016 Central Italy earthquake, Eng. Geol., 272, 105647, https://doi.org/10.1016/j.enggeo.2020.105647, 2020.

Campbell, K. and Bozorgnia, Y.: NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5 % damped linear elastic response spectra for periods ranging from 0.01 to 10 s, Earthq. Spectra, 24, 139–171, https://doi.org/10.1193/1.2857546, 2008.

Chiou, B. and Youngs, R.: An NGA model for the average horizontal component of peak ground motion and response spectra, Earthq. Spectra, 24, 173–215, https://doi.org/10.1193/1.2894832, 2008.

DeepSeek-AI, Liu, A., Feng, B., Wang, B., Wang, B., Liu, B., Zhao, C., Dengr, C., Ruan, C., Dai, D., Guo, D., Yang, D., Chen, D., Ji, D., Li, E., Lin, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Xu, H., Yang, H., Zhang, H., Ding, H., Xin, H., Gao, H., Li, H., Qu, H., Cai, J. L., Liang, J., Guo, J., Ni, J., Li, J., Chen, J., Yuan, J., Qiu, J., Song, J., Dong, K., Gao, K., Guan, K., Wang, L., Zhang, L., Xu, L., Xia, L., Zhao, L., Zhang, L., Li, M., Wang, M., Zhang, M., Zhang, M., Tang, M., Li, M., Tian, N., Huang, P., Wang, P., Zhang, P., Zhu, Q., Chen, Q., Du, Q., Chen, R. J., Jin, R. L., Ge, R., Pan, R., Xu, R., Chen, R., Li, S. S., Lu, S., Zhou, S., Chen, S., Wu, S., Ye, S., Ma, S., Wang, S., Zhou, S., Yu, S., Zhou, S., Zheng, S., Wang, T., Pei, T., Yuan, T., Sun, T., Xiao, W. L., Zeng, W., An, W., Liu, W., Liang, W., Gao, W., Zhang, W., Li, X. Q., Jin, X., Wang, X., Bi, X., Liu, X., Wang, X., Shen, X., Chen, X., Chen, X., Nie, X., Sun, X., Wang, X., Liu, X., Xie, X., Yu, X., Song, X., Zhou, X., Yang, X., Lu, X., Su, X., Wu, Y., Li, Y. K., Wei, Y. X., Zhu, Y. X., Xu, Y., Huang, Y., Li, Y., Zhao, Y., Sun, Y., Li, Y., Wang, Y., Zheng, Y., Zhang, Y., Xiong, Y., Zhao, Y., He, Y., Tang, Y., Piao, Y., Dong, Y., Tan, Y., Liu, Y., Wang, Y., Guo, Y., Zhu, Y., Wang, Y., Zou, Y., Zha, Y., Ma, Y., Yan, Y., You, Y., Liu, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Huang, Z., Zhang, Z., Xie, Z., Hao, Z., Shao, Z., Wen, Z., Xu, Z., Zhang, Z., Li, Z., Wang, Z., Gu, Z., Li, Z., and Xie, Z.: DeepSeek-V2: a strong, economical, and efficient mixture-of-experts language model, arXiv [preprint], https://doi.org/10.48550/arXiv.2405.04434, 2024.

Falcone, G., Romagnoli, G., Naso, G., Mori, F., Peronace, E., and Moscatelli, M.: Effect of bedrock stiffness and thickness on numerical simulation of seismic site response. Italian case studies, Soil Dyn. Earthq. Eng., 139, 106361, https://doi.org/10.1016/j.soildyn.2020.106361, 2020a.

Foti, S., Parolai, S., Bergamo, P., Di Giulio, G., Maraschini, M., Milana, G., Picozzi, M., and Puglia, R.: Surface-wave surveys for seismic site characterization of ITACA accelerometric stations, B. Earthq. Eng., 9, 1797–1820, https://doi.org/10.1007/s10518-011-9306-y, 2011.

Idriss, I. M.: An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, Earthq. Spectra, 24, 217–242, https://doi.org/10.1193/1.2924362, 2008.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E.: Adaptive mixtures of local experts, Neural Comput., 3, 79–87, https://doi.org/10.1162/neco.1991.3.1.79, 1991.

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., and D'Amico, M.: A revised ground-motion prediction model for shallow crustal earthquakes in Italy, B. Seismol. Soc. Am., 109, 525–540, https://doi.org/10.1785/0120180210, 2019.

Lanzano, G., Vitrano, L., Felicetta, C., Russo, E., D'Amico, M., Mascandola, C., Sgobba, S., Brunelli, G., Ramadan, F., Pacor, F., and Luzi, L.: TACAext flatfiles: parametric tables of metadata and strong motion intensity measures, Istituto Nazionale di Geofisica e Vulcanologia (INGV), https://doi.org/10.13127/itaca.4.0/itacaext_flatfile.2.0, 2024.

Luzi, L., Puglia, R., Russo, E., and ORFEUS WG5: Engineering Strong Motion Database, version 1.0, Observatories & Research Facilities for European Seismology, Istituto Nazionale di Geofisica e Vulcanologia [data set], https://doi.org/10.13127/ESM, 2016.

Luzi, L., Lanzano, G., Felicetta, C., D'Amico, M. C., Russo, E., Sgobba, S., Pacor, F., and ORFEUS Working Group 5: Engineering Strong Motion Database (ESM) (Version 2.0), Istituto Nazionale di Geofisica e Vulcanologia (INGV) [data set], https://doi.org/10.13127/ESM.2, 2020.

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R., and Luzi, L.: The new ShakeMap in Italy: progress and advances in the last 10 yr, Seismol. Res. Lett., 91, 317–333, https://doi.org/10.1785/0220190130, 2019.

Mori, F., Mendicelli, A., Moscatelli, M., Romagnoli, G., Peronace, E., and Naso, G.: A new Vs30 map for Italy based on the seismic microzonation dataset, Eng. Geol., 275, 105745, https://doi.org/10.1016/j.enggeo.2020.105745, 2020.

Mori, F., Mendicelli, A., Falcone, G., Acunzo, G., Spacagna, R. L., Naso, G., and Moscatelli, M.: Ground motion prediction maps using seismic-microzonation data and machine learning, Nat. Hazards Earth Syst. Sci., 22, 947–966, https://doi.org/10.5194/nhess-22-947-2022, 2022.

Mori, F., Naso, G., and Fiorentino, G.: Geomorphometry-informed ground-motion modeling for earthquake-induced landslides, Remote Sens., 18, 1169, https://doi.org/10.3390/rs18081169, 2026.

Scasserra, G., Stewart, J. P., Bazzurro, P., Lanzo, G., and Mollaioli, F.: A comparison of NGA ground-motion prediction equations to Italian data, B. Seismol. Soc. Am., 99, 2961–2978, https://doi.org/10.1785/0120080133, 2009.

Wang, Z., Dai, J., Zhao, D., Wang, X., Wang, J., Li, Z., Feng, Y., and Wang, Z.: A mixture of experts model for shallow crustal earthquake ground motion prediction in Japan, Comput. Geosci., 208, 106095, https://doi.org/10.1016/j.cageo.2025.106095, 2026.

Weatherill, G., Kotha, S. R., Danciu, L., Vilanova, S., and Cotton, F.: Modelling seismic ground motion and its uncertainty in different tectonic contexts: challenges and application to the 2020 European Seismic Hazard Model (ESHM20), Nat. Hazards Earth Syst. Sci., 24, 1795–1834, https://doi.org/10.5194/nhess-24-1795-2024, 2024.