

Ground motion prediction model for shallow crustal earthquakes in Japan based on XGBoost with Bayesian optimization

Haotian Dang^a, Zifa Wang^{b,c,*}, Dengke Zhao^b, Xiangqi Wang^a, Zhaoyan Li^b, Dongliang Wei^c, Jianming Wang^b

^a School of Civil and Architecture Engineering, Henan University, Kaifeng, Henan, 475000, China

^b Institute of Engineering Mechanics, China Earthquake Administration, Harbin, Heilongjiang, 150000, China

^c CEAJK ADPRHexa Inc, Shaoguan, Guangdong, 512000, China

ARTICLE INFO

Keywords:

Machine learning
Ground motion prediction model
SHAP
Bayesian optimization

ABSTRACT

Ground motion prediction is an important and complex research subject in earthquake engineering and traditional approaches based on statistical regression have much room for improvement in prediction accuracy. Utilizing the 67,164 ground motion records from KiK-net and K-Net for 777 shallow crustal earthquakes between 1997 and 2019 in Japan, this paper proposes a ground motion prediction model XGBoost-SC based on the machine learning algorithm of eXtreme Gradient Boosting (XGBoost) for Japan. Magnitude, focal depth, hypocentral distance, V_{S30} , site altitude, and focal mechanism were used as the feature parameters, and XGBoost, Random Forest, and Deep Neural Networks (DNN) algorithms were selected for model training while Bayesian optimization was used to search for optimized hyperparameters to improve the prediction accuracy. XGBoost algorithm was selected for further study based on the comparison of results from the three algorithms. Residual change with magnitude and hypocentral distance, the probability distribution of residuals, residual standard deviation (σ), residual mean squared error (MSE), and Pearson correlation coefficient (R) were used as the evaluation parameters, and a comparison study was performed against the ground motion prediction equation based on traditional approaches. Actual earthquake events were selected to compare the prediction results against the observation records. To further validate and explain the proposed model, SHapley Additive exPlanations (SHAP) analysis was performed to explain the impact of selected feature parameters on the proposed model. The results demonstrate that the proposed XGBoost-SC model has good prediction stability for all periods, and its residual errors are smaller than those of other models. Therefore, the proposed model can better reflect the ground motion attenuation for shallow crustal earthquakes in Japan and can serve as a better model for ground motion prediction in future aseismic design and earthquake disaster mitigation efforts.

1. Introduction

The proper aseismic design has been proven to be effective for reducing the casualty and economic loss of earthquakes. The ground motion level is an important parameter in the design, and it is dependent on many parameters. Richter and Gutenberg [1] found that ground acceleration depended on both the magnitude and focal distance of the earthquakes based on the records from California earthquakes, and the result was the pioneering research on ground motion attenuation. Ground motion attenuation is often represented by the Ground Motion Prediction Equations (GMPE), and its latest improvement was further

referred to as Ground Motion Prediction Models (GMPM) [2], which represent the impact on prediction results by the parameters of source characteristics, propagation path, and site conditions based on the analysis of ground motion records. Parameters such as Peak Ground Acceleration (PGA) or Spectral Acceleration (SA) were often selected as the predicted results which were further used in various earthquake disaster mitigation efforts [3].

Traditional GMPEs were built based on regression analysis of observed ground motion records and the GMPE may be dependent on the tectonics of the earthquakes. The National Seismic Hazard Mapping Project (NSHMP) team of the United States Geological Survey (USGS)

* Corresponding author. Institute of Engineering Mechanics, China Earthquake Administration, Harbin, Heilongjiang, 150000, China.

E-mail addresses: heroic98@163.com (H. Dang), zifa@iem.ac.cn (Z. Wang), denco666@163.com (D. Zhao), 1341088139@qq.com (X. Wang), hkjlizhaoyan@163.com (Z. Li), wei_dl1921@163.com (D. Wei), jwang780@gmail.com (J. Wang).

divided the United States (US) into three zones: the Western US (WUS) where the seismicity is dominated by shallow crustal earthquakes; the Central and Eastern US (CEUS) where the earthquakes are mostly stable continental ones, and the Pacific Northwest (PNW) where earthquakes are mostly interface types [4]. In 2003, with the collaboration from USGS and Southern California Earthquake Center (SCEC), Pacific Earthquake Engineering Research Center (PEER) initiated the Next Generation Attenuation (NGA) project to develop the Next Generation GMPE [5]. The second phase of the project NGA-West2 [6] expanded the ground motion database with additional records from California, Japan, New Zealand and Chinese Taipei. A number of GPMEs [7–11] were developed using this database for the WUS with varying parameters and site conditions. Zhao et al. [12] developed a GMPE (hereafter referred to as Zhao2016) using the Japanese records for shallow crustal earthquakes and earthquakes in the upper mantle, with improved prediction accuracy considering even the impact from volcanoes. However, traditional approaches based on regression analysis are limited in their prediction accuracy because of the statistical property of the approaches. For example, a fixed function has to be defined for the traditional approaches, and this fixed function form may limit the realization of the traditional approaches to deal with complex nonlinear relationships.

The machine learning approach has become a popular one adopted in many areas, and it solves the classification and optimization problems by automatically improving the results via complex computer models or algorithms [13], such as XGBoost [14], Random Forests [15], and Deep Learning models [16,17]. With the rapid development of machine learning and the accumulation of observed ground motion records, machine learning has also been introduced in GMPE or GMPM development. Dhanya and Raghukanth [18] proposed a shallow crustal GMPE based on an artificial neural network for the city of Shimla in the Himalaya area to predict the response spectrum at given sites with acceptable accuracy. Hamze-Ziabari and Bakhshpoori [19] proposed a GMPE based on an efficient bagging ensemble model of M5' and Classification and Regression Trees (CART) algorithms to predict PGA, Peak Ground Velocity (PGV) and Peak Ground Displacement (PGD) with improved accuracy. Derakhshani and Foruzan [20] proposed a GMPE for PGA, PGV and PGD using the NGA-West2 database based on Deep Neural Networks (DNN) to improve the learning capability for complex nonlinear features of the model via added hidden layers and neurons. Kubo et al. [21] constructed a GMPE based on machine learning and found that the model performed well for weak ground motion while underpredicted for strong ground motion, and the reason was probably that few records for the strong motion were used and that there was a distribution bias in the dataset used. The authors proposed that a combination of traditional and machine learning based GMPE should be used to address these issues. There are still a number of issues with machine learning models even though the machine learning based model can better predict the ground motion than traditional approaches. These issues are listed in the following.

1. Machine learning based models using a limited number of samples limit their generalization capability;
2. Hyperparameters impact the learning capability and accuracy of machine learning models. For instance, regularization parameters are used to reduce the risk of over-fitting. Manual adjustment of the hyperparameters is difficult to not only find the optimum combination but also implement in practice for the case of many hyperparameters.
3. Machine learning models are often considered “black-box” models, and because there is no clear explicit relationship between input and outputs, it is hard to explain the results.

To address the issues listed above, this paper assembled 67,164 Japanese shallow crustal earthquake ground motion records with feature parameters of magnitude (M_{JMA}), hypo-central distance (R_{hypo}), focal depth(D), travel-time averaged shear wave velocity of the top 30 m

under surface (V_{s30}), site altitude (H), and focal mechanism (FM). The dataset was divided into training, validation and testing subsets using a ratio of 7:1.5:1.5. XGBoost algorithm was selected after comparison study of results for XGBoost, Random Forest and DNN algorithms. The hyperparameters were optimized via the Bayesian approach, and SHAP analysis was performed to explain the impact of feature parameters on the prediction results.

2. Data

2.1. Data sources

KiK-net and K-Net are the Japanese strong motion observation networks built and maintained by NIED (National Research Institute for Earth Science and Disaster Resilience) (see **Data availability**). The observation network has been deployed nationwide in Japan and the location distribution of network stations can be found on the website of NIED. Detailed soil data such as shear wave velocity and layer depth are provided for each station. These datasets, together with the ground motion records observed, which are abundant in quantity and high in quality, are valuable for researchers worldwide.

2.2. Dataset selection

Based on the database for both the KiK-net and K-Net, 727,205 strong motion records were initially selected for earthquakes between 1997 and 2019. Additional data filtering was performed using the following criteria.

1. For KiK-net stations, records on the ground surface were selected;
2. Based on the earthquake classification scheme in Zhao et al. [22] and combining the subduction model of Slab2.0 [23], the shallow crustal earthquakes were selected;
3. Magnitude $M_{JMA} \geq 4$;
4. Hypo-central distance $R_{hypo} \leq 300$ km.
5. Baseline adjustment was performed for the accelerograms from KiK-net and K-Net, and fourth-order Butterworth filter with a low-pass and high-pass cut-off frequency of 30 Hz and 0.2 Hz, respectively [24] was applied to filter the accelerograms and then integration was performed to obtain velocity and displacement time-history records. Records with obvious baseline drifting were excluded in this study.

After applying the above filtering criteria, 67,164 strong motion records were chosen for 777 shallow crustal earthquakes, and the

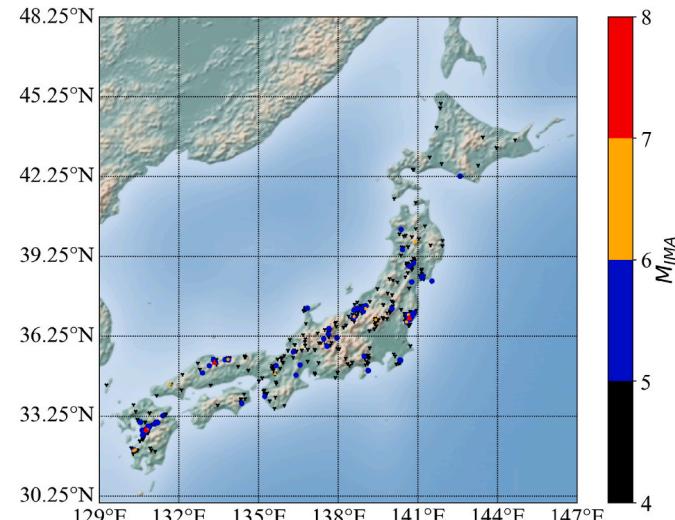


Fig. 1. Location distribution of the selected earthquakes in this study.

distribution of the earthquake locations is shown in Fig. 1.

2.3. Data processing

Additional processing was performed to derive the information for the feature parameters as explained in the following.

1. The focal mechanism of the selected earthquakes was obtained from the active fault database compiled by the Japanese National Institute of Advanced Industrial Science and Technology (see **Data availability**);
2. The travel-time averaged shear wave velocity of the top 30 m for soil layers V_{s30} was calculated from the borehole data provided by NIED according to the following equation [25].

$$V_{s30} = 30 \sqrt{\sum_1^n \frac{d_i}{V_{si}}} \quad (1)$$

where d_i is the depth of the i -th soil layer, V_{si} is the shear wave velocity for the i -th layer, and n is the number of soil layers in the top 30 m. When the depth of the borehole is less than 30 m, the last layer is expanded to 30 m deep in calculating V_{s30} [26].

3. Two widely popular traditional models are selected for comparison, and these two models are Zhao2016 and BSSA14 [8], which was developed by the NGA-West2 team. Since both the Zhao2016 and BSSA14 GMPE models use moment magnitude M_w , therefore for comparison purpose M_{JMA} in this study is converted to M_w using Equation (2) as in the following [27] if there no matching record with Magnitude M_w found in the Global Centroid-Moment-Tensor (GCMT) database (see **Data availability**):

$$M_w = 0.923 \times M_{JMA} + 0.370 \quad (2)$$

4. The geometric mean values for PGA and SA of the two horizontal components, i.e., the north-south and the east-west components were calculated the spectral acceleration with a 5 % damping ratio at 34 period points between 0.01s and 5s.
5. Hypo-central distance was selected to represent the path effect, but not the other parameters relating to rupture depth [28–30]. The reason is that a large number of earthquakes need to be used for machine learning training and there is no sufficient information on the rupturing data for all the earthquakes. On the other hand, the

information of hypo-central distance is readily available, and it is chosen in this study.

The distribution of magnitude versus hypo-central distance and magnitude versus focal depth is shown in Fig. 2.

3. Algorithms and hyperparameter optimization

3.1. Algorithms

Different machine learning algorithms have their own accuracy in solving a specific problem with a given set of data. Three machine learning algorithms, XGBoost, Random Forest and DNN, were used to construct the proposed GMPMs for comparison study to determine the best algorithm for further study. It should be noted that there is substantial difference between XGBoost and Random Forest, although they both belong to the ensemble learning. Random Forest constructs all the decision trees in one batch and the final result is the weighted average from all the decision trees, while the XGBoost approach constructs the decision tree one at a time, and subsequent decision trees are constructed based on the learning results of previous decision trees.

The same training, validation and testing datasets were used for the model training, hyperparameter optimization and testing. The feature parameters were magnitude, focal depth, hypo-central distance, V_{s30} , site altitude, and focal mechanism, and the logarithm of PGA and SA were used as the output. The proposed GMPM can be represented as in the following:

$$\ln(\hat{Y}) = f(M_{JMA}, R_{hypo}, D, V_{s30}, H, FM) \quad (3)$$

where \hat{Y} is the predicted PGA or SA in the unit of gal, and $f(\cdot)$ represents the machine learning algorithm as either XGBoost, Random Forest, or DNN.

3.2. Bayesian optimization of hyperparameters

Machine learning algorithms often use many hyperparameters which are important in tuning the performance of the algorithms, and different settings of the hyperparameters can result in different performance of the proposed models. For example, for XGBoost, lambda and alpha are regularization parameters used to reduce the risk of over-fitting. For Random Forest, n_estimators is the number of decision trees and it impact the learning capability of the model. For DNN, activation function is used to introduce the nonlinear property into the model so that

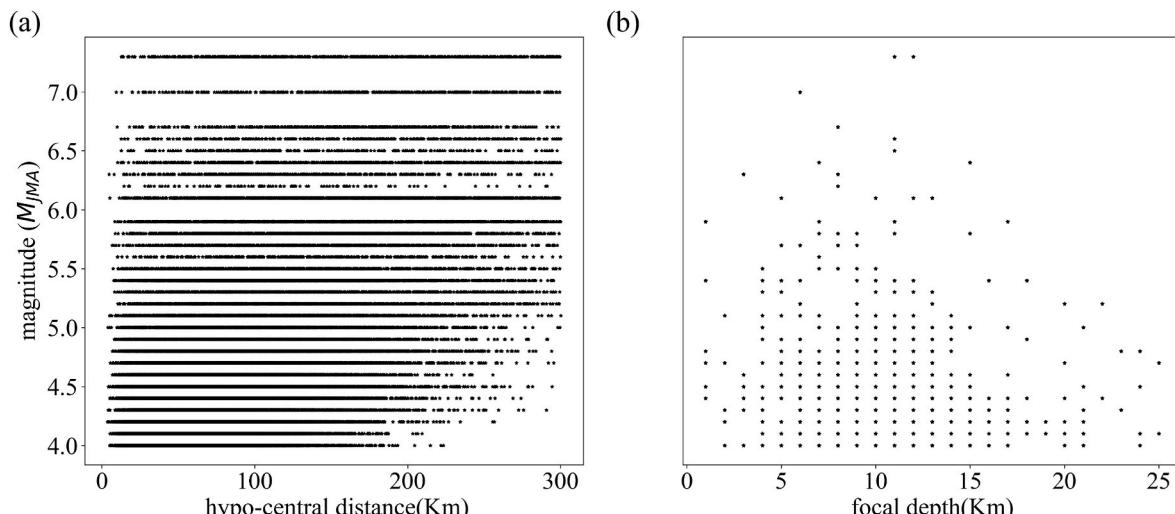


Fig. 2. Distribution of the selected datasets in this study: (a) magnitude versus hypo-central distance, (b) magnitude versus focal depth.

the model can deal with complex nonlinear challenges. Therefore, it is important to use the optimum hyperparameters for the best performance of the algorithm.

Hyperparameter optimization in machine learning algorithms is often realized through manual adjustment, grid searching, or Bayesian optimization [31]. Manual adjustment usually utilizes empirical approaches to search for the best hyperparameters and it does not guarantee the best result while requiring a lot of effort. The grid searching approach uses the grid style to determine the search area and search step, which is also time consuming when the result can still be less satisfactory.

In this study, the optimization framework based on Optuna [32] was selected. The optimization objective function (MSE_{obj}) is the average value of the mean squared error of actual $\ln(PGA)$ and $\ln(SA)$ versus the predicted corresponding value, which is calculated as in Equations (4) and (5). The objective of the optimization is to find the best hyperparameter combination corresponding to the minimum MSE_{obj} .

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$MSE_{obj} = \frac{1}{35} \left(MSE_{PGA} + \sum_{j=1}^{34} MSE_{SA_j} \right) \quad (5)$$

where y is the actual value of either $\ln(PGA)$ or $\ln(SA)$, \hat{y} is the predicted value of the corresponding parameter, n is the total number of data samples, j represents the j^{th} period in the selected 34 periods points between 0.01s and 5s. Table 1 shows the computation time and the MSE_{obj} for manual adjustment, grid searching, and Bayesian optimization of hyperparameters. As can be seen from Table 1, the Bayesian approach is the best among the three approaches.

Computer hardware configuration: Intel(R) Core(TM) i9-10900 K CPU @ 3.70 GHZ, NVIDIA GeForce RTX 2080 Ti.

The Bayesian optimization process starts with an initial set of randomly selected hyperparameters, and the probability distribution of the result is used to adjust the hyperparameter for continuous iteration until a satisfactory set of hyperparameters is found. After many rounds of trial and error, considering the time and accuracy requirement, the iteration number for hyperparameter optimization is set to be 200, and the iteration process can be illustrated in Fig. 3. As can be seen from the figure, the optimized hyperparameters gradually move to the lower end as the iteration proceeds while yielding the smallest MSE_{obj} .

For comparison of different machine learning algorithms, Table 2 shows the performance of XGBoost, Random Forest and DNN. As can be seen from Table 2, XGBoost exhibits the best performance, whose MSE_{obj} is the smallest among the three. Therefore, XGBoost algorithm is selected as the one for further study to construct the GMPM, hereafter referred to as XGBoost-SC.

4. Results

4.1. Models and assessment

The hyperparameters after optimization in the previous section are adopted in the XGBoost, Random Forest, and DNN algorithms and model prediction results are compared with those by Zhao2016 and BSSA14. The testing dataset is used to study the residual change with magnitude

Table 1

Computation time and MSE_{obj} of XGBoost under different hyperparameter optimization approaches for the validation dataset.

Optimization approach	Computing time(h)	MSE_{obj}
Manual adjustment	/	0.3698
Grid searching	8	0.3509
Bayesian optimization	3	0.3341

and hypo-central distance, and the probability density distribution of the residuals for different periods. The standard deviation, MSE, and Pearson correlation coefficient are used to study the prediction results by various models.

4.2. Residual analysis

Three periods ($T = 0.01\text{s}$, $T = 0.3\text{s}$, and $T = 1\text{s}$) were selected to study the residual change for the three machine learning based models and the two traditional models. The residual can be expressed as in Equation (6). In addition, using the same approach as in Abrahamson and Youngs [33], the residual can be divided into inter-event and intra-event residuals as in Equation (7).

$$\delta = \ln(y) - \ln(\hat{y}) \quad (6)$$

$$\ln y_{ij} = \ln \hat{y}_{ij} + \eta_i + \varepsilon_{ij} \quad (7)$$

Where η_i is the inter-event residual for event-i and ε_{ij} is the intra-event residual for event-i and record j.

At the selected periods ($T = 0.01\text{s}$, $T = 0.3\text{s}$, and $T = 1\text{s}$), the residual change with magnitude and hypo-central distance for all models is shown in Fig. 4, Fig. 5 respectively. As can be seen from the figures, for the three machine learning based approaches, the residual distribution between $[-2, 2]$ was all above 99 %, and the residual between $[-1, 1]$ was 92.20 %, 85.30 %, 81.42 % for XGBoost-SC, Random Forest, and DNN, respectively, while the residual distribution between $[-2, 2]$ for Zhao 2016 and BSSA14 was 93.61 % and 91.09 % and it between $[-1, 1]$ was 62.34 % and 59.95 %, respectively. In addition, with the change of magnitude and hypo-central distance, the residual of the machine learning models even distributes along the $\delta = 0$ line, but the residual of the traditional models exhibits some bias along the $\delta = 0$ line. Fig. 6 is the probability density distribution of the residual, and the shape of the distribution for machine learning models conforms to a normal distribution with 0 average while it for the traditional models shows a skewed distribution rather than a normal one with 0 average. As can be seen from Fig. 7, inter-event residuals for XGBoost-SC scatter along both sides of 0-axis, indicating no prediction bias when magnitude changes while the residuals for both Zhao2016 and BSSA14 show clear deviation from 0-axis. Intra-event residuals for all models are shown in Fig. 8 with no deviation from 0-axis, but the residual for XGBoost-SC is smaller than it for other models.

From the above it can be concluded that the proposed GMPM does not show systematic bias with the changes of magnitude and hypo-central distance, and the normal probability distribution of the residual demonstrates that the proposed model has good prediction stability for Japanese shallow crustal earthquakes.

4.3. The assessment and comparison study of residuals

Another way to assess the performance of the proposed GMPM is to compare its result with the one via traditional approaches. The following three parameters are used as the evaluation parameters which can be defined as in the following.

1. The MSE, defined in Equation (4), for all periods is used to assess the overall error. The smaller this value, the smaller the error of the model;
2. The standard deviation of the residual, as defined in Equation (8), for all periods is used to assess the scattering of residuals. The smaller this value, the smaller the scattering of the predicted results by the model;

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \mu)^2} \quad (8)$$

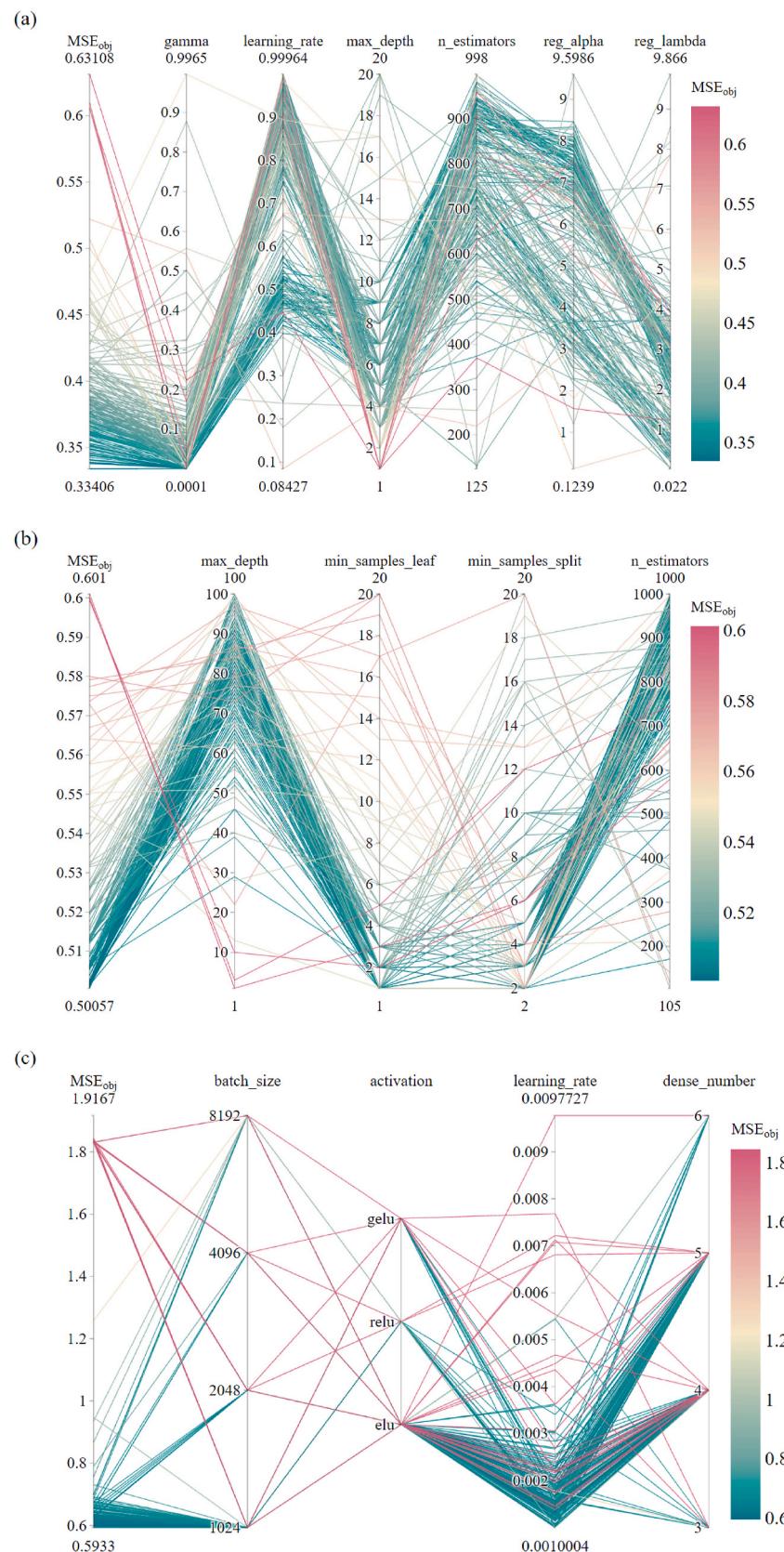


Fig. 3. Evolution of hyperparameter combination and its corresponding MSE_{obj} during the Bayesian optimization process when the iteration number is set to 200: (a) XGBoost, (b) Random Forest, (c) DNN.

Table 2Hyperparameters and MSE_{obj} of three algorithms for the validation dataset.

Algorithm	Hyperparameter	Hyperparameter value	MSE_{obj}
XGBoost	gamma	0.0002	0.3341
	learning_rate	0.4398	
	max_depth	5	
	n_estimators	877	
	reg_alpha	6.7831	
	reg_lambda	2.9461	
Random Forest	n_estimators	986	0.5006
	max_depth	83	
	min_samples_leaf	1	
	min_samples_split	2	
DNN	batch_size	256	0.5933
	activation_function	elu	
	learning_rate	0.0011	
	dense_number	5	

3. The Pearson correlation coefficient, as defined in Equation (9), for all periods is used to assess the correlation between the actual and predicted values. The bigger the value, the higher the accuracy of the predicted results.

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2}} \quad (9)$$

where μ is the mean of the residual, \bar{y}_i is the average of actual $\ln(PGA)$ or $\ln(SA)$, and $\bar{\hat{y}}_i$ is the average predicted $\ln(PGA)$ or $\ln(SA)$.

The comparison results for both the machine learning and traditional GMPE models are shown in Figs. 9 and 10.

As can be seen from Fig. 9, XGBoost-SC is the best among the 5

models in estimating ground motion from Japanese shallow crustal earthquakes. From Fig. 9(a), the standard deviation for XGBoost-SC is the smallest, and it shows little variation for different periods. The standard deviation for both Random Forest and DNN is smaller than the traditional ones, and it also shows little variation for different periods. Although Zhao2016 has larger variation than the machine learning algorithms, the standard deviation for different periods is pretty stable, while BSSA14 has the highest standard deviation with a decreasing trend when the period increases. Similarly, MSE in Fig. 9(b) for the three machine learning algorithms shows little variation with different periods, and it for the traditional models shows large variation for different periods, especially the one for BSSA14. Further study of the NGA-West2 database demonstrates that the percentage of Japanese records is small, and this may be the main cause of the largest MSE and its variation with different periods for BSSA14. As can be seen from both Figs. 4 and 5, the prediction error of Zhao2016 is relatively large for small events, resulting in less satisfactory prediction results by the model. The Pearson correlation coefficient in Fig. 9(c) shows an increasing trend with the period, but the increment for XGBoost-SC is the smallest. Incorporating the trends observed in both Fig. 9(a) and (b), it can be concluded that all models predict ground motion with better accuracy at long periods than at short periods. Fig. 10 shows the performance improvement of the three machine learning models over the two traditional models. While all three machine learning models improved the prediction performance, the performance by XGBoost-SC is the highest.

Using the earthquake records from Oklahoma, Kansas, and Texas, Zalachoris and Rathje [34] developed a GMPE based on the traditional approach, referred to as ZR18 in this study. Based on the same dataset, Khosrovikia and Clayton [35] developed three machine models using Artificial Neural Networks (ANN), Random Forest (hereafter referred as to “Random Forest-KC2021”), Support Vector Machine (SVM) and compared its performance against ZR18 using standard deviation σ and Pearson correlation coefficient R . Listing the performance improvement

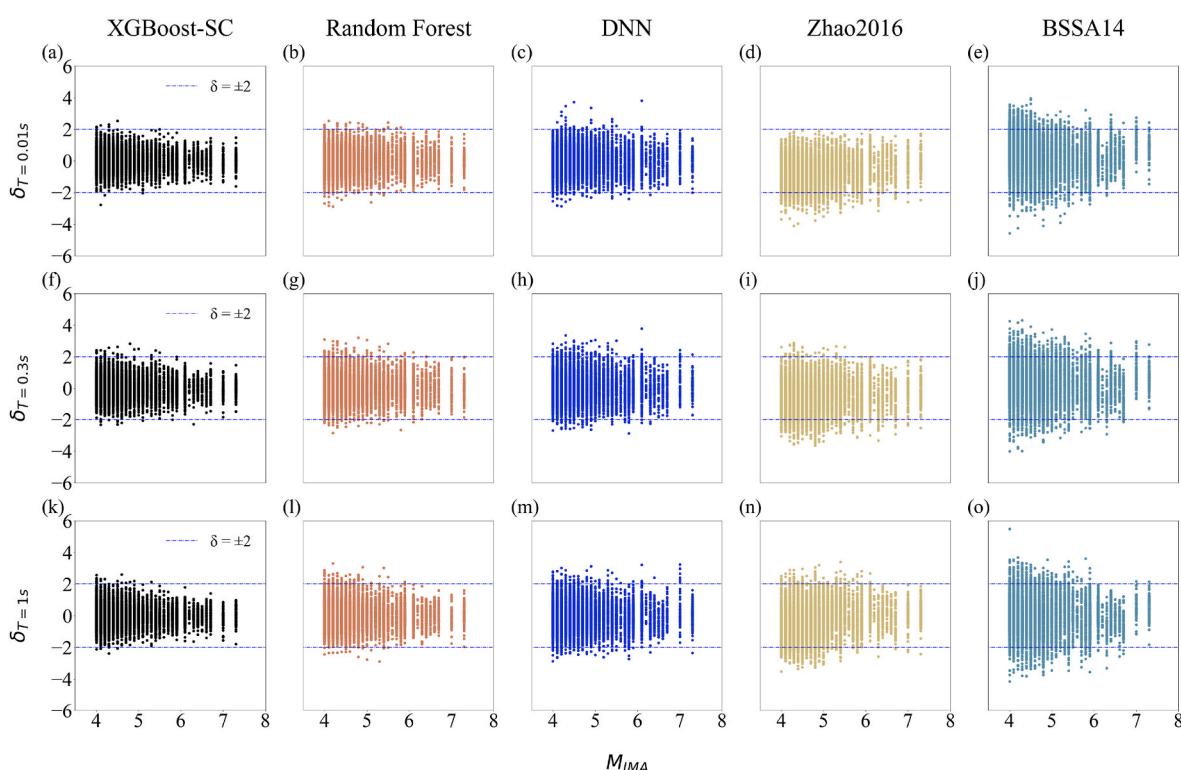


Fig. 4. The residual change with the magnitude of XGBoost-SC, Random Forest, DNN, Zhao2016, and BSSA14 at the three periods selected for the testing dataset. Each column represents the residual by the same model, while each row represents the residual for the same period: (a)–(e) $T = 0.01s$, (f)–(j) $T = 0.3s$, and (k)–(o) $T = 1s$.

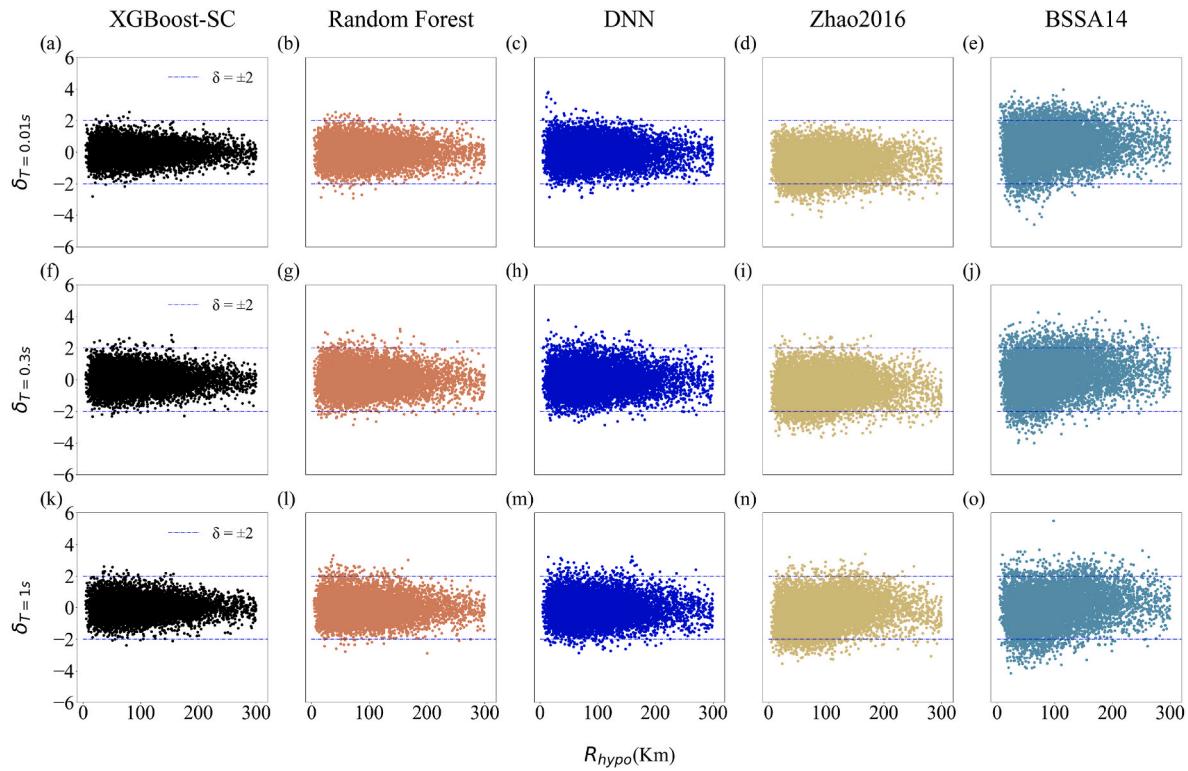


Fig. 5. The residual change with the hypo-central distance of XGBoost-SC, Random Forest, DNN, Zhao2016, and BSSA14 at the three periods selected for the testing dataset. Each column represents the residual by the same model, while each row represents the residual for the same period: (a)–(e) $T = 0.01s$, (f)–(j) $T = 0.3s$, (k)–(o) $T = 1s$.

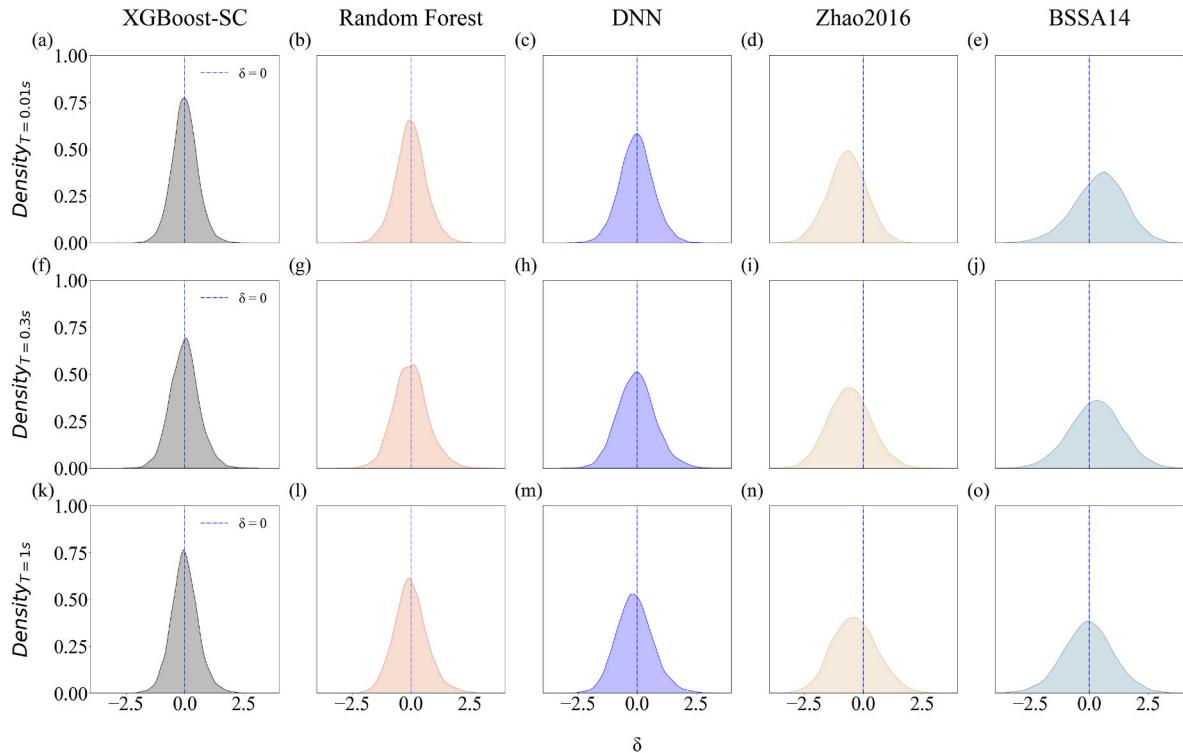


Fig. 6. The residual probability density distribution of XGBoost-SC, Random Forest, DNN, Zhao2016, and BSSA14 at the three periods selected for the testing dataset. Each column represents the residual by the same model, while each row represents the residual for the same period: (a)–(e) $T = 0.01s$, (f)–(j) $T = 0.3s$, (k)–(o) $T = 1s$.

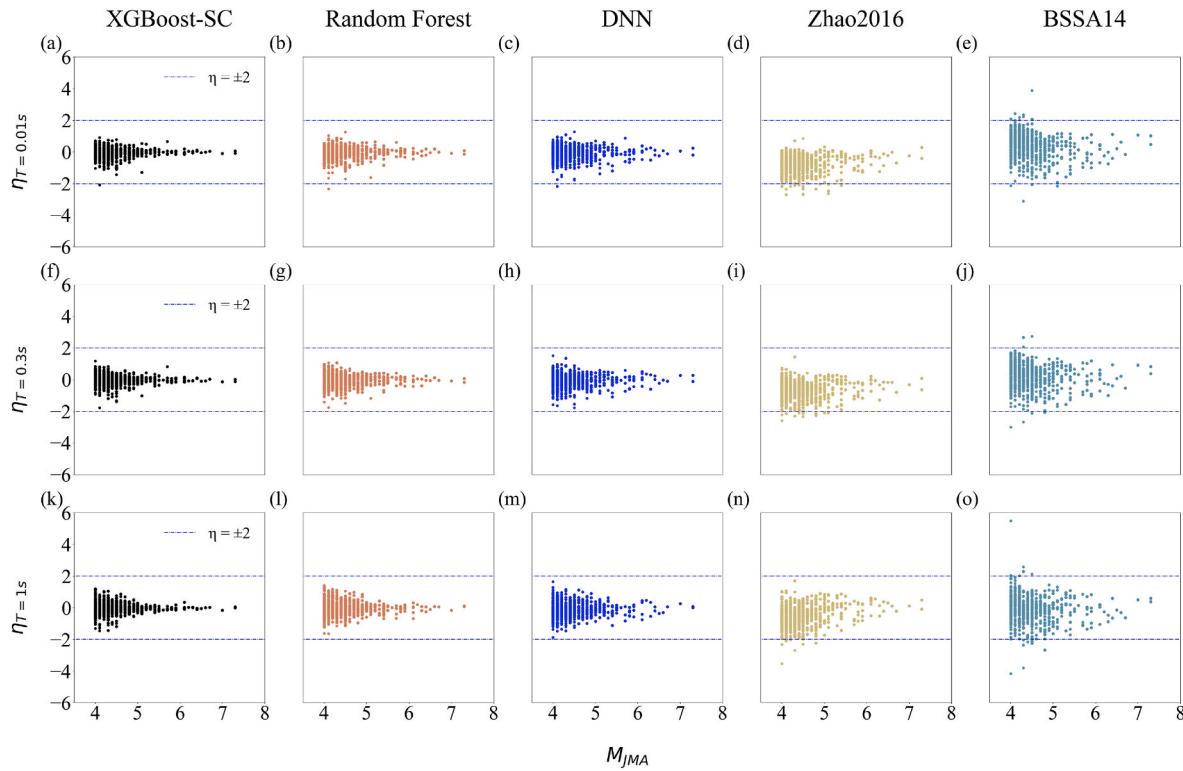


Fig. 7. The inter-event residual change with the magnitude of XGBoost-SC, Random Forest, DNN, Zhao2016 and BSSA14 at the three periods selected for the testing dataset. Each column represents the residual by the same model, while each row represents the residual for the same period: (a)–(e) $T = 0.01\text{s}$, (f)–(j) $T = 0.3\text{s}$, and (k)–(o) $T = 1\text{s}$.

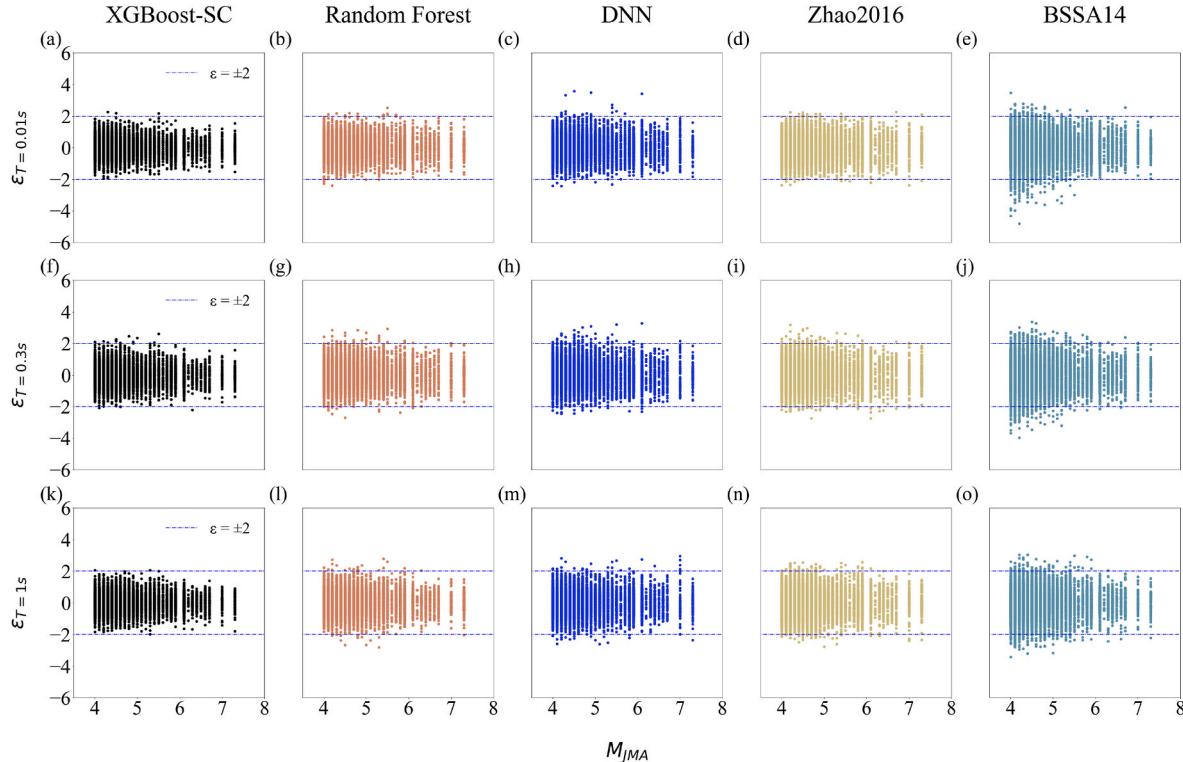


Fig. 8. The intra-event residual change with the magnitude of XGBoost-SC, Random Forest, DNN, Zhao2016 and BSSA14 at the three periods selected for the testing dataset. Each column represents the residual by the same model, while each row represents the residual for the same period: (a)–(e) $T = 0.01\text{s}$, (f)–(j) $T = 0.3\text{s}$, and (k)–(o) $T = 1\text{s}$.

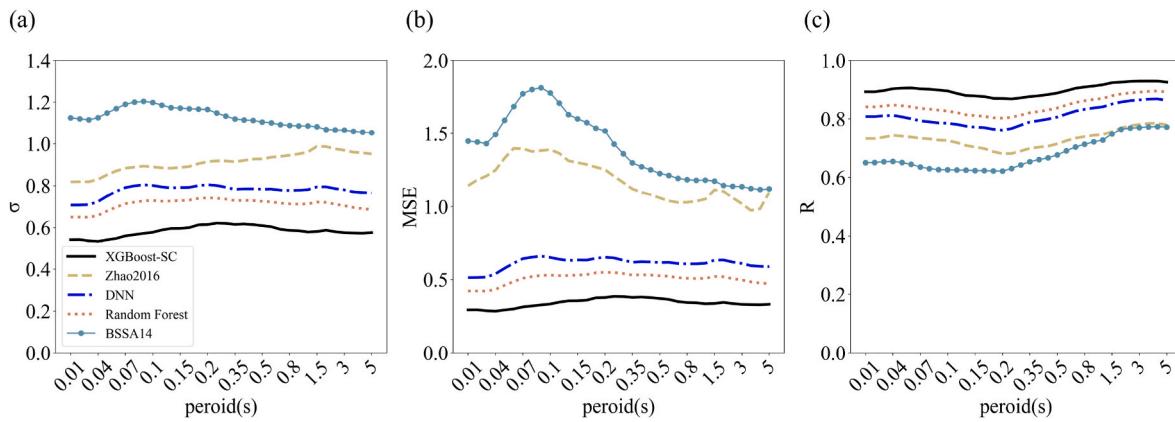


Fig. 9. Comparison of results for all periods by XGBoost-SC, Random Forest, DNN, Zhao2016, and BSSA14: (a) Standard deviation, (b) Mean Squared Error, (c) Pearson correlation coefficient.

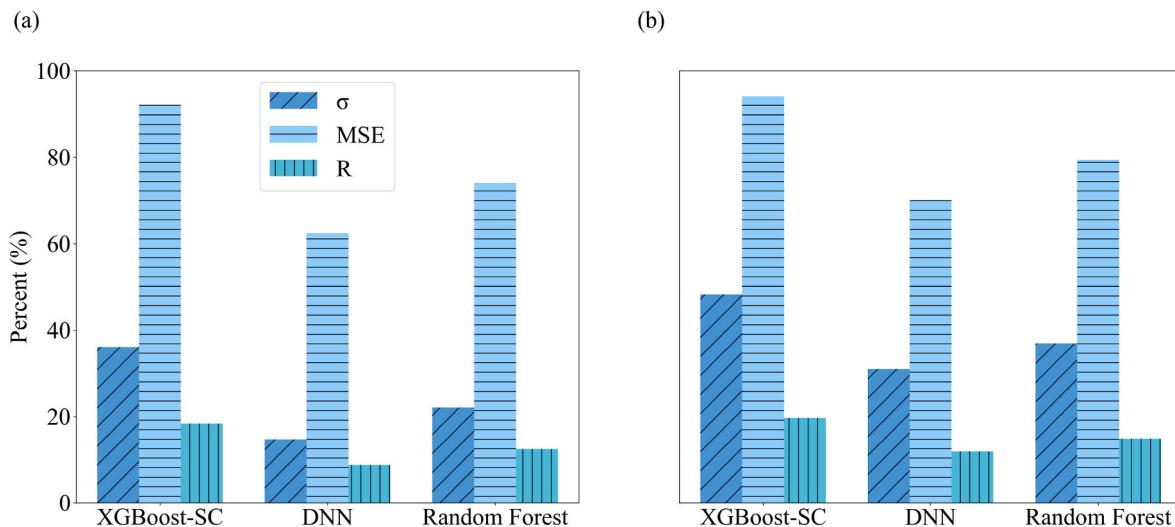


Fig. 10. Performance improvement of σ , MSE, R for the three machine learning algorithms over the two traditional approaches: (a) Improvement over Zhao2016, and (b) Improvement over BSSA14.

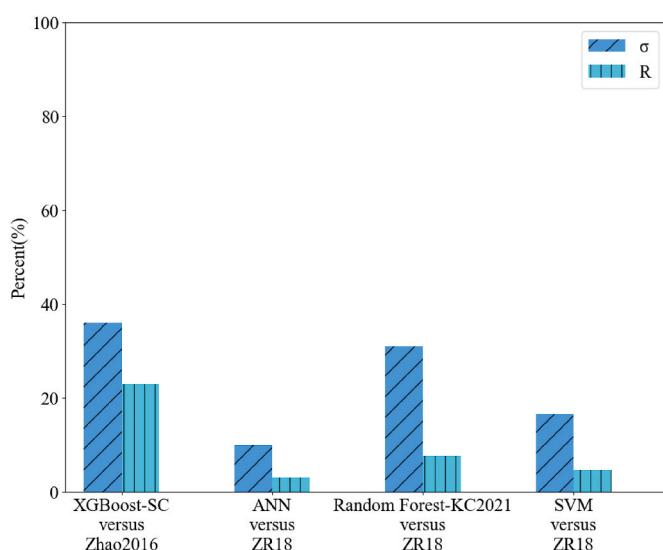


Fig. 11. From left to right: XGBoost-SC's model improvement from Zhao2016; Khosravikia and Clayton (2021)'s model improvement from ZR18 using ANN, Random Forest-KC2021, and SVM.

by Khosravikia and Clayton together with the improvement by XGBoost-SC in this study, the result can be seen in Fig. 11. As seen in Fig. 11, using σ and R as the evaluation parameters, the improvement by XGBoost model over traditional approaches is much better than that by the other three models, indicating the best performance by XGBoost.

4.4. Comparison with observed records

To further verify the performance of the proposed model, the comparison is performed for actual records in different ranges of magnitude and hypo-central distance. In Fig. 12, the average SA is shown for XGBoost-SC, Random Forest, DNN, Zhao2016, and BSSA14 and the records for different combinations of magnitude and hypo-central distance bins. As can be seen from Fig. 12 and stated previously, machine learning models outperform traditional ones in all cases. As shown in Fig. 12(a)–(d), both Zhao2016 overestimates and BSSA14 underestimates ground motion at short periods for small magnitude and hypo-central distance. It can be generally concluded that traditional approaches perform better for longer periods than for short periods, a phenomenon already exhibited in Fig. 9. In contrast, machine learning models perform well for both short and long periods, demonstrating the superior prediction capability of the machine learning models.

In addition to compare performance for the average results, the comparison is also carried out for selected events. As shown in Fig. 13,

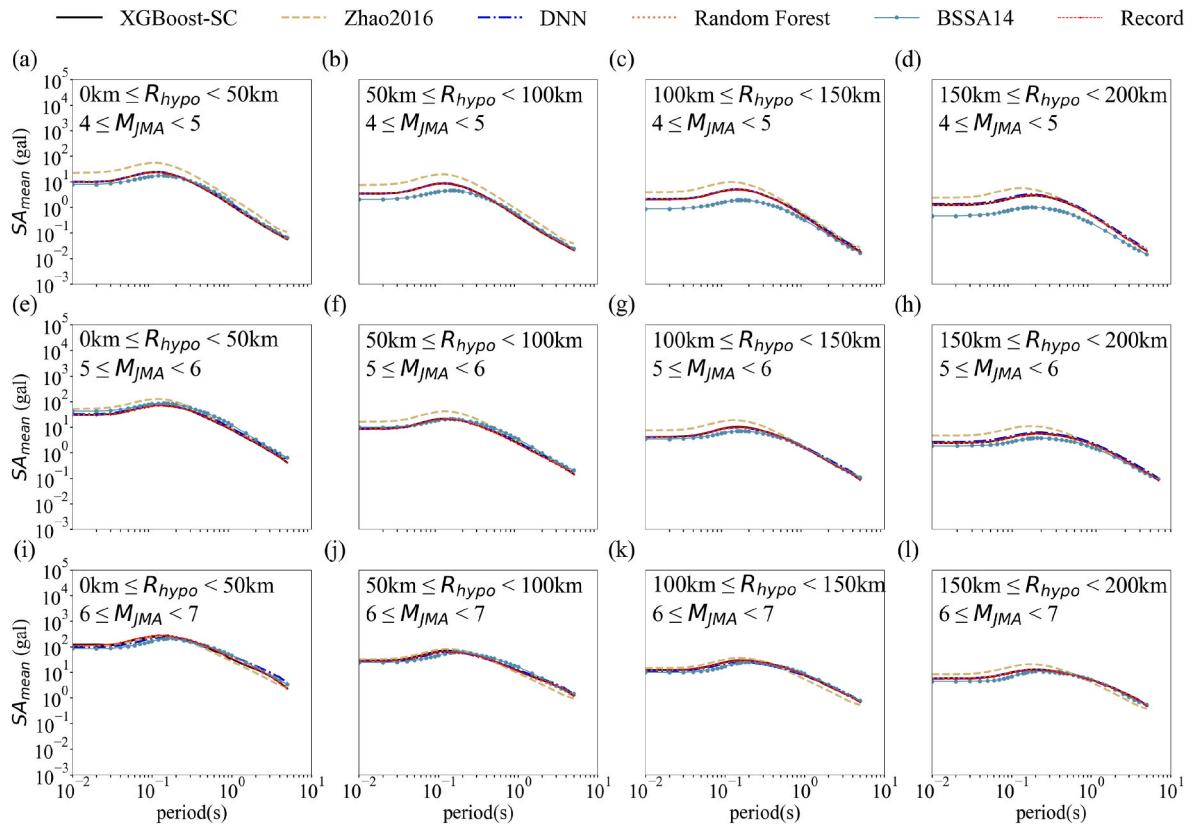


Fig. 12. The average SA comparison for different magnitude and hypo-central distance ranges.

the prediction by XGBoost-SC is very close to the actual observed records for all periods, while all other models show prediction variation at different places. Further study is needed to identify the cause of this deviation.

To test the generalization capability of the proposed model, actual events outside the database in this study are used to validate the performance. Two recent earthquakes are selected for the SA prediction by the proposed model, and the result is shown in Fig. 14. Again, as shown in Fig. 14, XGBoost-SC demonstrates excellent generalization capability and stability in predicting SA for these two earthquakes.

Summarizing the model comparison results, it can be concluded that XGBoost-SC performs the best for Japanese shallow crustal earthquakes, and it can represent the ground motion attenuation better than the other two machine learning models and the two traditional GMPE models. From the comparison, the prediction by machine learning models does show small variation at some periods, while the prediction by traditional GMPE is much smoother for different periods. This is partly because the traditional GMPEs use fixed equations and adopted a smoothing technique in the process [36]. It should be noted that the SA for actual records also display a small variation, and other researchers also noticed similar behavior in their machine learning models as in the case by Seo et al. [37]. We may attribute this outperformance to the superior learning capability of machine learning models but further study is needed to confirm this theory.

4.5. SA comparison for various ranges and different periods

Because the distribution of SA values in the data used in this paper is relatively inhomogeneous, they are divided into different value intervals for assessment. Three periods, namely 0.01s, 0.3s, and 1s, and three SA ranges (0-10gal, 10-100gal, and 100-1000gal) are chosen for the comparison study. The Pearson correlation coefficient is used to assess the performance of the model for the selected cases, as shown in Fig. 15.

As can be seen from the Figure, when $T = 0.01s$, there is not much difference in prediction accuracy in regard to the SA range. The prediction accuracy at $T = 0.3s$ increases and at $T = 1s$ decreases with increasing SA range value. This trend is partly due to the change of number of records in different SA range values. As shown in Figs. 12–14, SA values tend to be highest at medium periods around 0.3 s. In summary, there is a lot of data in the small SA range for both the short (0.01s) and the long (1.0s) period, so XGBoost-SC performs well. For the medium period (0.3s), there is more data for mid and high SA values, so XGBoost-SC also performs relatively well.

5. Model explanation

5.1. Model performance analysis

As explained above, XGBoost-SC model performs better than both the two machine learning models and the two traditional models, and it even predicts a small variation of ground motion, which may lead to concerns on whether the model is over-fitted because over-fitting can limit its prediction capability for future earthquakes [38]. To study this issue, the performance of the model for training, validation, and testing dataset is compiled, as in Table 3. As seen from Table 3, although the σ , MSE, and R for both validation and testing datasets are worse than those for the training dataset, the difference is not significant to reach a conclusion that the model is over-fitted.

Furthermore, machine learning models are in essence “black-box” models, and the results are hard to explain via traditional theory and framework. The once popular approach was to quantify the feature importance in order to explain the model results but such approach has much limitations. Lundberg and Lee [39] proposed a unified approach based on additive feature importance measures to explain the model from different angles using SHAP value as the evaluation parameter. SHAP value can express the contribution of each feature parameter to

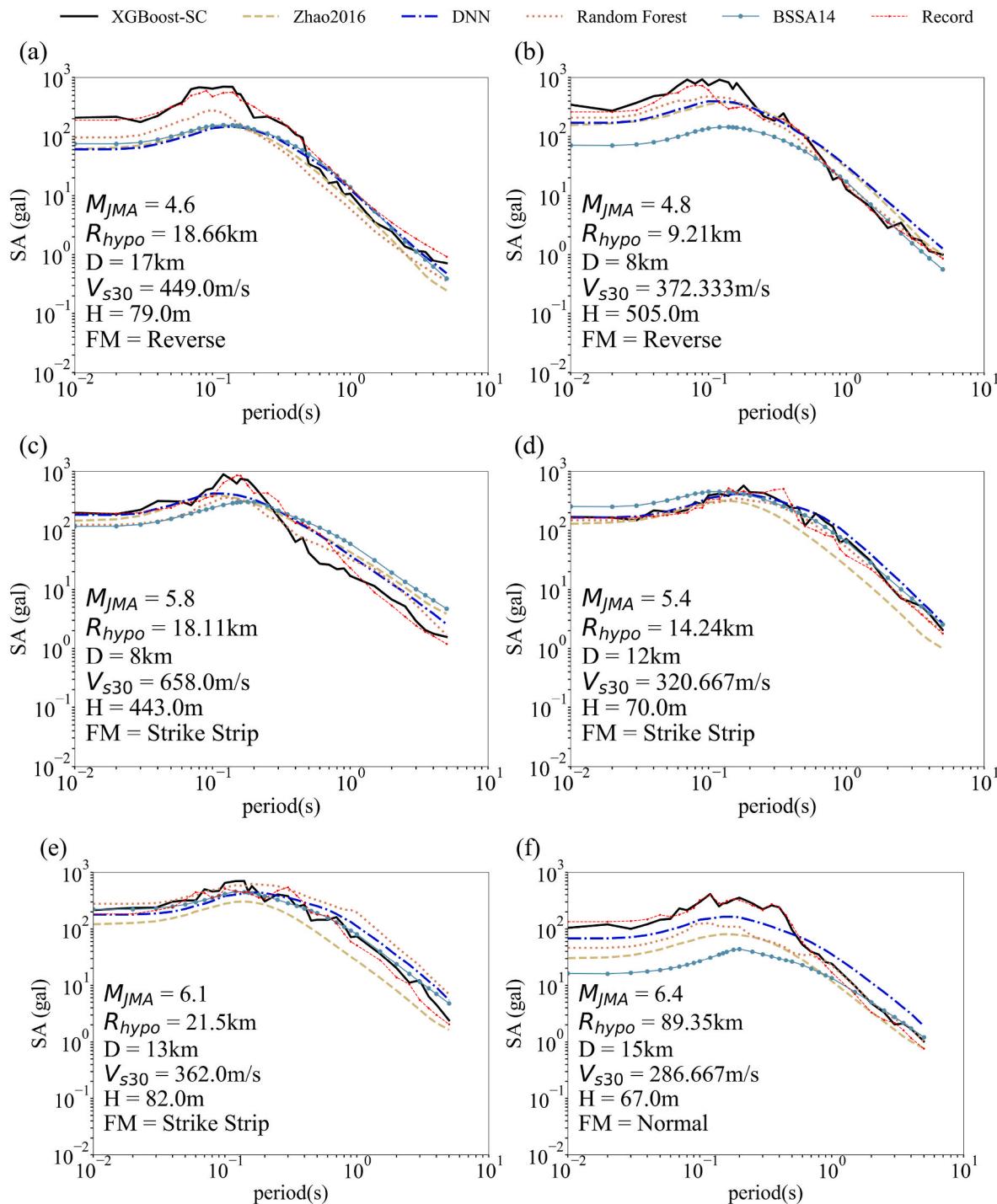


Fig. 13. Comparison of model performance for single earthquake events.

prediction results via either a positive or negative value to indicate the direction of contribution.

5.2. SHAP value for feature parameters

Fig. 16 shows the SHAP value for all feature parameters of the XGBoost-SC model when $T = 0.01\text{s}$, $T = 0.3\text{s}$, and $T = 1\text{s}$. The higher the SHAP value, the larger contribution of the feature parameter to prediction results. The positive or negative values indicate the positive or negative contribution. The color expresses the order of feature parameter value, and red shows a larger value while blue is for smaller values. From analyzing the combination of feature parameters and their

corresponding SHAP values, the following can be summarized.

1. The SHAP value for M_{JMA} and R_{hypo} is much larger than the value for the other four feature parameters. SHAP is bigger for larger M_{JMA} , indicating the positive contribution of M_{JMA} to the prediction results of SA. On the other hand, SHAP becomes a negative value when R_{hypo} increases, indicating negative contribution of R_{hypo} to the prediction of SA. The trend change conforms to the trend in traditional GMPMs.
2. The SHAP value for V_{s30} does not show a unilateral trend, and the correlation can be both positive and negative, which is consistent with the observed results by Wang et al. [40] when studying the site amplification using machine learning algorithms.

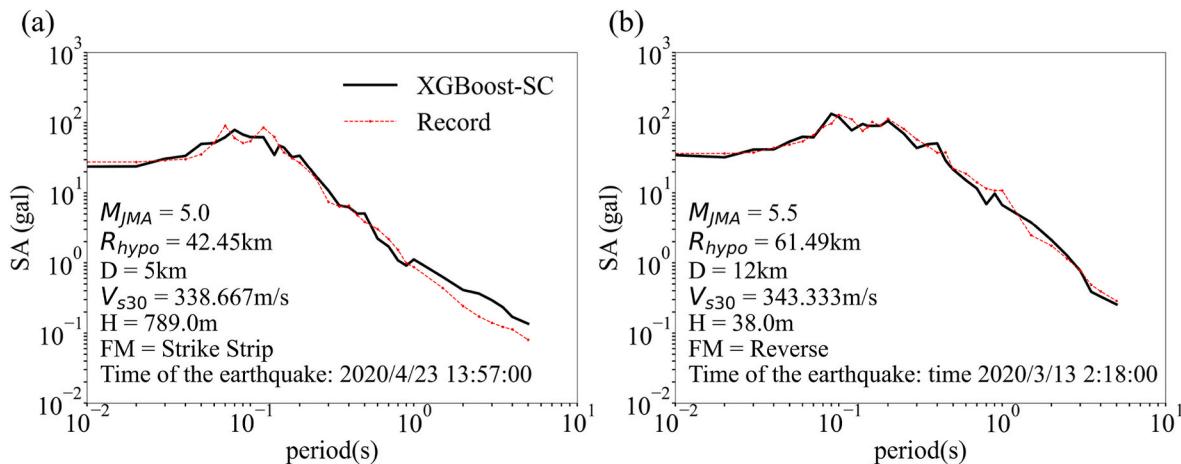


Fig. 14. Generalization capability testing of the proposed model for two recent earthquakes.

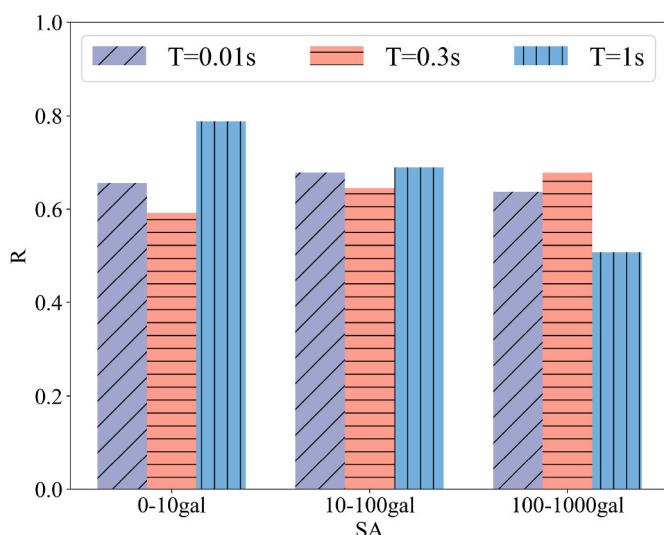


Fig. 15. Comparison of Pearson correlation coefficient R for different periods and different SA range values.

Table 3

Standard deviation, measure squared error, and Pearson correlation coefficient for training, validation, and testing datasets.

	training dataset	validation dataset	testing dataset
σ	0.4582	0.5773	0.5826
MSE	0.2102	0.3341	0.3402
R	0.9396	0.9017	0.8979

3. The SHAP value for D slightly increases with increasing D , indicating relatively larger impact for deeper D . This is because most of the records used in this study are for far fields, and deeper earthquakes may exert larger impact for sites at far fields when other feature parameters are constant.
4. The SHAP value for H does not show clear trend when the period is short, but when the period increases, the lower H , the larger SHAP value appears, indicating potential basin effect for long period ground motion [41].
5. SHAP value for FM is relatively small for both short and long periods, and there is no clear trend observed with changing FM . This result shows that the impact from focal mechanism is not significant in predicting ground motion.

5.3. Interdependence of feature parameters

Interdependence of feature parameters can be used to assess the model from multi-dimensions. The two most important feature parameters, the hypo-central distance and magnitude, are used to study the interdependence and their impact in different ranges. Fig. 17 shows the SHAP value changes with hypo-central distance and magnitude for $T = 0.01s$, $T = 0.3s$, and $T = 1s$, respectively. It can be seen that magnitude affects the contribution of hypo-central distance on predicted results, and this trend is similar for three periods. The SHAP value decreases rapidly for a hypo-central distance shorter than 50 km, and the magnitude exhibits a positive impact on the results, but the SHAP value's decrease slows after the hypo-central distance surpasses 50 km, and when it is over 150 km, SHAP value turns into a negative value, indicating a negative impact on the prediction results even for earthquakes with large magnitude.

As can be seen from both Fig. 16, which shows the impact on prediction by each feature parameter, and Fig. 17, which shows the combined contribution of two feature parameters, XGBoost-SC's prediction is consistent with the commonly understood domain knowledge, such as the positive correlation for ground motion with magnitude given hypo-central distances. The above analysis confirms that the prediction by the proposed GMPM is both reasonable and with better accuracy.

5.4. Application of the model results

Traditional models have explicit functions to represent the ground motion prediction, and the explicit function expression can be easily implemented in any application system. As machine learning models such as XGBoost-SC do not have an explicit function expression, extra computer programming efforts are needed to implement the models.

For the verification and application of the proposed XGBoost-SC model, the model, source code, and explanation document are deposited into GitHub (see **Data availability**) with additional tools based on Python provided for computation and graph plotting. For simpler implementation, the model is further refined to require only feature parameters such as magnitude and distance (as shown in Equation (3)) to compute PGA or SA directly. It should be noted that the model uploaded to GitHub is a refined model, which can be used directly without modification to obtain the computational results.

For implementation in probabilistic seismic hazard assessment (PSHA) when multiple GMPM candidates are available for the target area, there is a need to evaluate the GMPM candidates to construct the GMPM logic tree for PSHA in the target area [42]. Scherbaum et al. [43] proposed a general theoretical framework to select the most appropriate GMPM based on information-theoretic approach, and the framework is

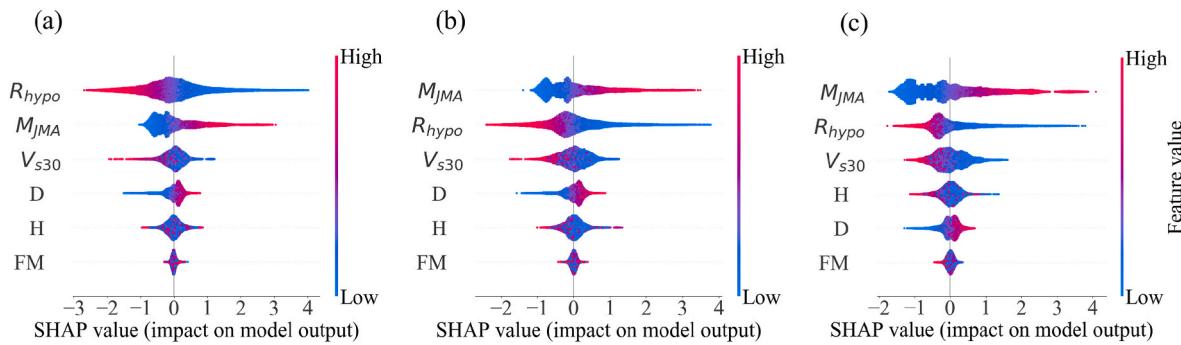


Fig. 16. SHAP value for all 6 feature parameters at the selected periods: (a) $T = 0.01s$, (b) $T = 0.3s$, (c) $T = 1s$.

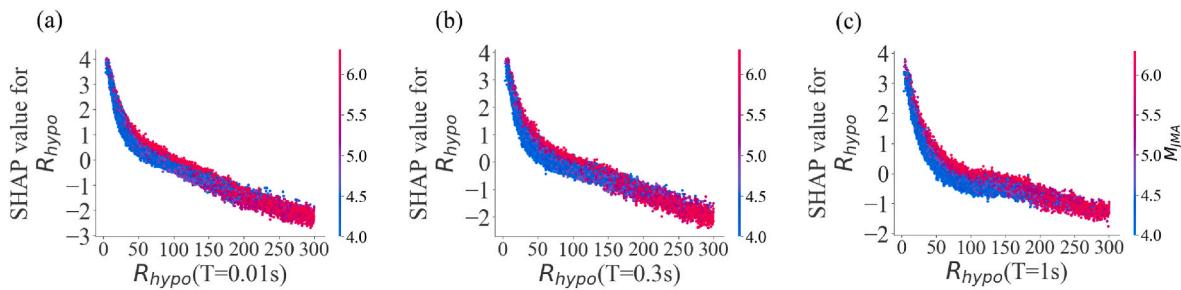


Fig. 17. SHAP value for hypo-central distance with changes of magnitude at selected periods: (a) $T = 0.01s$, (b) $T = 0.3s$, (c) $T = 1s$.

used to calculate Average Sample Log-Likelihood Differences (LLH) as the evaluation parameter for GMPM selection. The LLH based approach does not need ad hoc assumptions, so it is more versatile than other evaluation approaches. XGBoost-SC proposed in this paper can be used to quickly predict the ground motion for any target area in Japan, and LLH can be easily derived from the model outputs which can improve both the efficiency and accuracy in PSHA projects.

6. Conclusion

Utilizing the shallow crustal earthquake ground motion records from KiK-net and K-Net in Japan, a GMPM was proposed based on XGBoost algorithm with hyperparameters optimized by Bayesian approach. Six feature parameters were used and the model was constructed for PGA and SA at 34 different periods. Residual analysis and comparison study were used to confirm the performance of the proposed model and SHAP value was used to explain the impact of feature parameters. The main conclusions from this study were summarized in the following.

1. Comparison studies demonstrate that the XGBoost algorithm is best suited for constructing machine learning based GMPM for Japan among DNN, Random Forest and XGBoost algorithms.
2. The XGBoost based model proposed using 6 feature parameters outperforms the traditional approach using σ , R , and MSE as the evaluation parameters.
3. The comparison against historical events and the generalization test against recent events further validate the outperformance of the proposed model.
4. Bayesian optimization of the hyperparameters improved the speed and performance of the proposed GMPM.
5. SHAP analysis of the 6 feature parameters explains the impact of each feature parameter indicating stronger influence from both the magnitude and the hypo-central distance while the contribution from the focal mechanism is minimal, and the result is consistent with domain expert knowledge.
6. Given the outstanding performance of the proposed GMPM, it is anticipated that XGBoost-SC can help better predict the earthquake

ground motion for shallow crustal earthquakes in Japan, and subsequently help mitigating the future earthquake impacts in Japan and other applicable regions.

CRediT authorship contribution statement

Haotian Dang: Visualization, Writing – original draft. **Zifa Wang:** Conceptualization, Supervision. **Dengke Zhao:** Writing – review & editing. **Xiangqi Wang:** Data curation. **Zhaoyan Li:** Resources. **Dongliang Wei:** Validation. **Jianming Wang:** Investigation, Software.

Declaration of competing interest

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

Data availability

The ground motion data was downloaded from National Research Institute for Earth Science and Disaster Resilience (<https://www.kyoshin.bosai.go.jp>), and the last visited date was July, 2023). The active fault database was downloaded from the Japanese National Institute of Advanced Industrial Science and Technology (<https://gbank.gsj.jp/activefault>), and the last visited date was July, 2023). Moment magnitude data is from Global Centroid-Moment-Tensor (<https://www.globalcmt.org/>), and the last visited date was July, 2023). The model, source code and explanation document have been deposited into GitHub at <https://github.com/getSomeChip/XGBoost-SC.git>.

Acknowledgments

This study was supported by the Scientific Research Fund of Institute of Engineering Mechanics, China Earthquake Administration (Grant No. 2021B09) and National Natural Science Foundation of China (51978634).

References

- [1] Gutenberg B, Richter CF. Earthquake magnitude, intensity, energy, and acceleration. *Bull Seismol Soc Am* 1942;32(3):163–91.
- [2] Atik LA, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F, Kuehn N. The variability of ground-motion prediction models and its components. *Seismol Res Lett* 2010;81(5):794–801.
- [3] 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(1):99–138.
- [4] Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Rukstales KS. Documentation for the 2008 update of the United States national seismic hazard maps (No. 2008-1128). US Geological Survey; 2008.
- [5] Power M, Chiou B, Abrahamson N, Bozorgnia Y, Shantz T, Roblee C. An overview of the NGA project. *Earthq Spectra* 2008;24(1):3–21.
- [6] Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Donahue JL, et al. NGA-West2 database. *Earthq Spectra* 2014;30(3):989–1005.
- [7] Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. *Earthq Spectra* 2014;30(3):1025–55.
- [8] Boore DM, Stewart JP, Seyhan E, Atkinson GM. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq Spectra* 2014;30(3):1057–85.
- [9] Campbell KW, Bozorgnia Y. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq Spectra* 2014;30(3):1087–115.
- [10] Chiou BSJ, Youngs RR. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra* 2014;30(3):1117–53.
- [11] Idriss IM. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq Spectra* 2014;30(3):1155–77.
- [12] Zhao JX, Zhou S, Zhou J, Zhao C, Zhang H, Zhang Y, Irikura K, et al. Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions. *Bull Seismol Soc Am* 2016;106(4):1552–69.
- [13] Mitchell TM. Machine learning. N Y: McGraw-hill 1997;1(No. 9).
- [14] Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K. Xgboost: extreme gradient boosting. R package version 0.4-2 2015;1(4):1–4.
- [15] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
- [16] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [17] Wu L, Ma D, Wang Z, Zhang J, Zhang B, Li J, Tong J. A deep CNN-based constitutive model for describing of statics characteristics of rock materials. *Engineering Fracture Mechanics*; 2023, 109054.
- [18] Dhanya J, Raghukanth STG. Ground motion prediction model using artificial neural network. *Pure Appl Geophys* 2018;175(3):1035–64.
- [19] Hamze-Ziabari SM, Bakhshpoori T. Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5' and CART algorithms. *Appl Soft Comput* 2018;68:147–61.
- [20] Derakhshanl A, Foruzan AH. Predicting the principal strong ground motion parameters: a deep learning approach. *Appl Soft Comput* 2019;80:192–201.
- [21] 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(1):1–12.
- [22] Zhao JX, Zhou S, Gao P, Long T, Zhang Y, Thio HK, Rhoades DA. An earthquake classification scheme adapted for Japan determined by the goodness of fit for ground-motion prediction equations. *Bull Seismol Soc Am* 2015;105(5):2750–63.
- [23] Hayes GP, Moore GL, Portner DE, Hearne M, Flamme H, Furtney M, Smoczyk GM. Slab2, a comprehensive subduction zone geometry model. *Science* 2018;362(6410):58–61.
- [24] Converse A, Brady AG. Basic strong-motion accelerogram processing software. USGS Open-file Report; 1992.
- [25] Boore DM. Estimating vs (30)(or NEHRP site classes) from shallow velocity models (depths < 30 m). *Bull Seismol Soc Am* 2004;94(2):591–7.
- [26] Boore DM, Thompson EM, Cadet H. Regional correlations of VS 30 and velocities averaged over depths less than and greater than 30 meters. *Bull Seismol Soc Am* 2011;101(6):3046–59.
- [27] Tsampas AD, Scordilis EM, Papazachos CB, Karakasis GF. Global-magnitude scaling relations for intermediate-depth and deep-focus earthquakes. *Bull Seismol Soc Am* 2016;106(2):418–34.
- [28] Cotton F, Scherbaum F, Bommer JJ, Bungum H. Criteria for selecting and adjusting ground-motion models for specific target regions: application to central Europe and rock sites. *J Seismol* 2006;10:137–56.
- [29] Delavaud E, Cotton F, Akkar S, Scherbaum F, Danciu L, Beauval C, Theodoulidis N. Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *J Seismol* 2012;16:451–73.
- [30] Danciu L, Kale Ö, Akkar S. The 2014 Earthquake Model of the Middle East: ground motion model and uncertainties. *Bull Earthq Eng* 2018;16:3497–533.
- [31] Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE* 2015;104(1):148–75.
- [32] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining; 2019, July. p. 2623–31.
- [33] Abrahamson NA, Youngs RR. A stable algorithm for regression analyses using the random effects model. *Bull Seismol Soc Am* 1992;82(1):505–10.
- [34] Zalachoris G, Rathje EM. Ground motion model for small-to-moderate earthquakes in Texas, Oklahoma, and Kansas. *Earthq Spectra* 2019;35(1):1–20.
- [35] Khosravikia F, Clayton P. Machine learning in ground motion prediction. *Comput Geosci* 2021;148:104700.
- [36] Gülerce Z, Kamai R, Abrahamson NA, Silva WJ. Ground motion prediction equations for the vertical ground motion component based on the NGA-W2 database. *Earthq Spectra* 2017;33(2):499–528.
- [37] 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(3):1549–64.
- [38] Bindi D. The predictive power of ground-motion prediction equations. *Bull Seismol Soc Am* 2017;107(2):1005–11.
- [39] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;30.
- [40] Wang X, Wang Z, Wang J, Miao P, Dang H, Li Z. Machine learning based ground motion site amplification prediction. *Front Earth Sci* 2023;11:1053085.
- [41] Fan G, Liu F, Zhang J, Wen R. Influence of topography on ground motion. *China Earthquake Engineering Journal* 2014;36(4):1039–46.
- [42] Kale Ö, Akkar S. A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): the Euclidean distance-based ranking (EDR) method. *Bull Seismol Soc Am* 2013;103(2A):1069–84.
- [43] Scherbaum F, Delavaud E, Riggelsen C. Model selection in seismic hazard analysis: an information-theoretic perspective. *Bull Seismol Soc Am* 2009;99(6):3234–47.