Hindawi

Research Article

Machine Learning-Based Probabilistic Seismic Demand Model of Continuous Girder Bridges

Wenshan Li ⓘ,[1] Yong Huang ⓘ,[1] and Zikai Xie[2]

[1]*Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150000, China*
[2]*School of Computing and Information Systems, University of Melbourne, Melbourne 3010, Australia*

Correspondence should be addressed to Yong Huang; huangyong@iem.ac.cn

Probabilistic seismic demand model (PSDM) is one of the critical components of performance-based earthquake engineering frameworks. The aim of this study is to propose a procedure to generate PSDMs for a typical regular continuous-girder bridge subjected to far and near-fault ground motions (GMs) utilizing machine-learning methods. A series of nonlinear time history analyses (NTHAs) is carried out to calculate the damage caused by the far and near-fault GMs for four different site conditions, and 21 seismic intensity measures (IMs) are considered. Subsequently, PSDMs are established for the IMs and engineering demand parameters based on the existing NTHA data using machine-learning methods, which include linear regression, Bayesian regression (BR), and a tree-based model. The results indicated that random forest (RF) is the most suitable model to predict the longitudinal and transverse curvature at the bottom of the four piers from the coefficients of determination. More specifically, the relative importance of each parameter in the model is evaluated, and peak ground velocity (PGV), peak spectral velocity (PSV), Arias intensity (AI), and Fajfar intensity (FI) are found to be the critical factors for the RF-based PSDM. Finally, all of these parameters, except AI, are correlated with velocity. The research results explore a new method for establishing the seismic demand model of continuous-girder bridges, which can provide suggestions for seismic damage prediction and seismic insurance risk evaluation.

1. Introduction

Recent earthquakes have highlighted that continuous girder bridges, as key components of transportation networks, are one of the most vulnerable infrastructure components [1, 2]. As a key step in performance-based seismic design (PBSD) [3, 4], probabilistic seismic demand model provides a means to describe probabilistically the relationships between seismic intensity measures (IMs) and engineering demand parameters for bridges subjected to potential earthquakes.

Several studies investigated seismic demand models for continuous girder bridges in recent years. Mackie et al. [5, 6] established a PSDM for columns of California highway bridges using nonlinear time history analysis and incremental dynamic analysis. Nielson et al. [7, 8] established PSDMs for columns, bearings, and abutments by analyzing the seismic fragility curves of typical bridges in the central

and southeastern United States. Padgett et al. [9] investigated the PSDMs for a class of retrofitted multispan continuous concrete girder bridges using two suites of synthetic GMs. Pan et al. [10] focused on multispan simply supported steel highway bridges in New York, USA, and established PSDMs for columns and bearings based on nonlinear time history analysis. Pan et al. [11] used the demand/capacity ratio as a variable and adopted quadratic instead of linear regression model to predict the demand, while analyzing the fragility of bridges in New York, USA. Padgett et al. [12] proposed criteria, such as efficiency, practicality, proficiency, sufficiency, and hazard computability, to select the optimal IMs for highway bridges. Zheng et al. [13] established PSDMs for columns, bearings, and abutments by analyzing the seismic fragility curves of typical simply supported girder bridges in Wenchuan, China. The results showed that peak ground acceleration and engineering demand parameters of

components were not ideal in logarithmic space. Ma et al. [14] compared the near-fault damage mechanism of continuous bridges with far-fault earthquakes and established a PSDM using an intensity parameter. The results showed that Housner intensity had the best correlation with the bridge pier drift ratio. The strategy currently employed to establish PSDMs for continuous girder bridges is based on linear regression utilizing only a single IM and engineering demand parameters. However, a single IM cannot account for all relevant characteristics of a ground motion. Therefore, it is essential to include more IMs to improve the capability of regression models to fit data. Several methods have been proposed to improve the outcomes of the evaluation. Jiang et al. [15] investigated the Optimal IMs of PSDMs for isolated bridges subjected to pulse-like GMs. Similarly, Wang et al. [16] proposed a multidimensional fragility evaluation methodology considering multiple performance limit states and seismic demand parameters, indicating that the uncertainty and dependence between seismic demand parameters are dispensable in the fragility analysis process. Taking the dependence of the seismic demands on ground motion characteristics and the prevailing uncertainties into consideration, Huang et al. [17] constructed the probabilistic demand models for reinforced concrete highway bridges with one single-column bent.

The prevalence of machine learning (ML) methods is attributed to their high accuracy and efficiency. Compared to the traditional linear regression, ML methods have advantages for analysis complex and uncertain problems, facilitating decision-making, and propagating [18]. However, the application of ML methods to PSDMs of regular continuous girder bridges with multiple IMs has rarely been investigated in the past, which provides the motivation for the present study. This paper proposes a general procedure to establish PSDMs for continuous girder bridges based on ML models. A series of NTHAs were performed on a bridge finite element model in OpenSees, and 21 seismic IMs were considered. Various ML models, such as linear regression, Bayesian regression, and tree-based model were used to establish a PSDM for the continuous girder bridge and compared to the conventional linear regression model. From the comparison, random forest (RF) was found to be the most suitable model, and the relative importance of each input IM was elucidated.

2. Overview of ML Regression Methods

ML is an important branch of artificial intelligence. It is an approach to optimize the performance of a computational process using available data or previous experience. Various ML regression methods, shown in Figure 1, such as linear regression, BR, and tree-based models [19] will be used to establish a PSDM for continuous girder bridges in this paper.

2.1. Linear Regression. Linear regression uses a linear function to fit the data based on the mean square error (MSE) between the observed and the predicted values to calculate the penalty function. It adopts the gradient descent method to find a set of weights to minimize the MSE. A linear regression model is established by minimizing the MSE. The model [19] takes the following form:

$$
y_{\text{true}_i} = y_{\text{predict}_i} (\boldsymbol{\omega}, \mathbf{x}_i) + e,
$$
$$
y_{\text{predict}_i} (\boldsymbol{\omega}, \mathbf{x}_i) = \omega_0 + \omega_1 x_{i1} + \cdots + \omega_n x_{in},
\tag{1}
$$

where $\omega = [\omega_1, \omega_2, \ldots, \omega_n]^T$ is the regression coefficients vector, $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ is the ith input variables vector, y_{true_i} is the ith observed response, y_{predict_i} is the ith predicted response, and e is the error term.

Lasso regression (LR) and ridge regression (RR) add L1 and L2 norm regularization terms, respectively, to the standard linear regression penalty function. For LR, the penalty function [19] becomes

$$
f(\boldsymbol{\omega}) = \sum_{i=1}^{m} \left(y_{\text{true}_i} - \mathbf{x}_i^T \boldsymbol{\omega} \right)^2 + \lambda \sum_{i=1}^{n} |\omega_i|,
\tag{2}
$$

while the penalty function [19] for RR can be written as follows:

$$
f(\boldsymbol{\omega}) = \sum_{i=1}^{m} \left(y_{\text{true}_i} - \mathbf{x}_i^T \boldsymbol{\omega} \right)^2 + \lambda \sum_{i=1}^{n} \omega_i^2.
\tag{3}
$$

The regularization terms in LR and RR are different, which lead to different results.

Elastic net (EN) is a linear regression model utilizing both L1 and L2 norm regularization terms. The penalty function [19] of EN is as follows:

$$
f(\boldsymbol{\omega}) = \sum_{i=1}^{m} \left(y_{\text{true}_i} - \mathbf{x}_i^T \boldsymbol{\omega} \right)^2 + \lambda_1 \sum_{i=1}^{n} |\omega_i| + \lambda_2 \sum_{i=1}^{n} \omega_i^2.
\tag{4}
$$

Support vector regression (SVR) is another linear regression model based on the concept of support vector machine (SVM). Given a training dataset for binary classification, SVM constructs a hyperplane to divide the data into two classes, where the term 'support vector' refers to the data point nearest to the hyperplane. The hyperplane is chosen to maximize its distance to the support vectors, i.e., the classification margin is maximized by SVM. Similar to support vector machine, SVR finds a hyperplane as the regression result, such that the distance between the sample points farthest from the hyperplane is the shortest. The optimization problem [19] to find the regression hyperplane can be written as follows:

$$
\min_{w,b} \frac{1}{2} \|\boldsymbol{\omega}\|^2,
$$
$$
\text{s.t. } \left| y_{true_i} - (\boldsymbol{\omega} \mathbf{x_i} + \mathbf{b}) \right| \leq \varepsilon, \quad \forall i,
\tag{5}
$$

where ε is the hyperparameter that determines the width of the interval boundary.

2.2. Bayesian Regression. BR is a linear regression model solved by statistical Bayesian inference. BR regards the parameters of a linear model as random variables and calculates the posterior estimates using the prior values of

```
           ┌─────────────────────────┐
           │  Regression used machine │
           │      learning methods    │
           └─────────────────────────┘
```

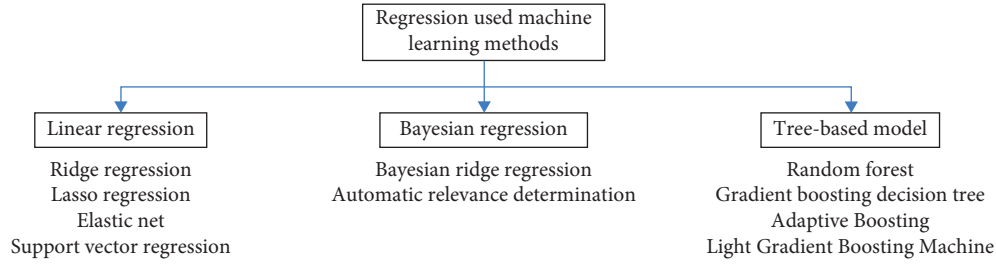| Linear regression	Bayesian regression	Tree-based model
Ridge regression	Bayesian ridge regression	Random forest
Lasso regression	Automatic relevance determination	Gradient boosting decision tree
Elastic net		Adaptive Boosting
Support vector regression		Light Gradient Boosting Machine

FIGURE 1: Some commonly used ML regression methods.

the model parameters. Compared to other models, BR can avoid under- and overfitting caused by choosing too simple or too complex a model. At the same time, it can make full use of available data, so as to avoid data underutilization. In order to obtain a fully probabilistic BR model, the output y is assumed to be Gaussian distributed around $X\omega$ [18]:

$$P(y|X, \omega, \alpha) = N(y|X\omega, \alpha), \tag{6}$$

where α is treated as a random variable that is to be estimated from the data, ω is the normal distribution weight, and X is the input data.

BR can be used for parameter regularization during the prediction stage. The regularization parameters are not selected automatically, but rather manually adjusted. In Bayesian ridge regression (BRR), we do not manually adjust the regularization parameters, but let it be estimated from the data as a variable. The BRR assumes that the prior of the coefficient ω follows a spherical Gaussian distribution [18]:

$$P(\omega|\lambda) = N\left(\omega|0, \lambda^{-1}I_P\right), \tag{7}$$

where the priors over α and λ are chosen from a gamma distributions, λ controls the variance of the Gaussian distribution, and I_P is the p-dimensional identity matrix.

Automatic relevance determination (ARD), also known as sparse Bayesian learning, is very similar to BRR, but it will typically lead to sparser weights ω. Specifically, it weakens the assumption that the Gaussian distribution is spherical and instead assumes a different a priori hypothesis for ω, which is an elliptical Gaussian distribution for noncorrelated variables [18]:

$$P(\omega|\lambda) = N\left(\omega|0, A^{-1}\right), \tag{8}$$

where $\text{diag}(A) = \lambda = \{\lambda_1, \ldots, \lambda_p\}$.

2.3. Tree-Based Models. Tree-based models have the advantages of strong interpretability, convenience, and high accuracy. Regression decision tree mainly refers to the classification and regression tree algorithm. The values output by internal nodes are "yes" or "no," which constitute a binary tree structure. Regression tree divides the feature space into several units, each with a specific output. Because each node outputs "yes" or "no," the boundary is parallel to the coordinate axis. For the test data, the corresponding output can be obtained as long as we classify it into a unit according to the division of the tree nodes.

RF represents an improvement over bagged trees, which mainly reduces the correlation of multiple trees. In the training stage, RF uses bootstrap sampling to collect multiple different subtraining data sets from the input training data set to train multiple different decision trees in turn. In the prediction stage, the prediction results of multiple decision trees in the RF are averaged to obtain the final results.

Gradient boosting decision tree (GBDT), which adds the gradient boosting method to the ordinary decision tree, evolves from a single decision tree to multiple decision trees to gradually improve the learning accuracy. Adaptive boosting (AdaBoost) [20] strengthens the importance of the samples wrongly assigned by the previous application of the basic classifier and uses all the weighted samples to retrain the basic classifier. At the same time, a new weak classifier is added in each round until a predetermined small error rate or maximum number of iterations are reached. Light gradient boosting machine (LightGBM) [20] is a model based on decision tree algorithms, in which the model is generated leafwise rather than depthwise (as in other decision tree-based methods). Such a leafwise generation leads to more complex but also more accurate trees.

3. Example Bridge and Its Modeling

A three-span reinforced concrete continuous girder bridge in China was used to investigate the capacity of the proposed ML-based demand models. The girder had a box cross section, 1.6 m high, and 10.5 m wide. Each of the two piers consisted of two solid circular columns with a diameter of 1 m. The diameter of longitudinal reinforcements was 25 mm with a total of 20 reinforcement bars arranged at equal intervals. The diameter of stirrups was 16 mm, and the distance between stirrups was 10 cm. Pot rubber bearings were used at the bent beams and abutments as girder supports.

A 3D finite element model of the bridge was constructed in the OpenSees platform [21] (Figure 2). The girder was simulated using the elastic beam-column elements, as it usually remained in the elastic state under earthquake excitation. Nonlinear-beam-column elements were adopted to simulate the piers due to expectation that plastic hinges will develop in them. The cross sections were fiber-defined cross sections composed of concrete fibers and steel fibers. Concrete02 material model was used for both the confined and unconfined concrete, but different material parameters were assumed for the two. Reinforcing-steel material model
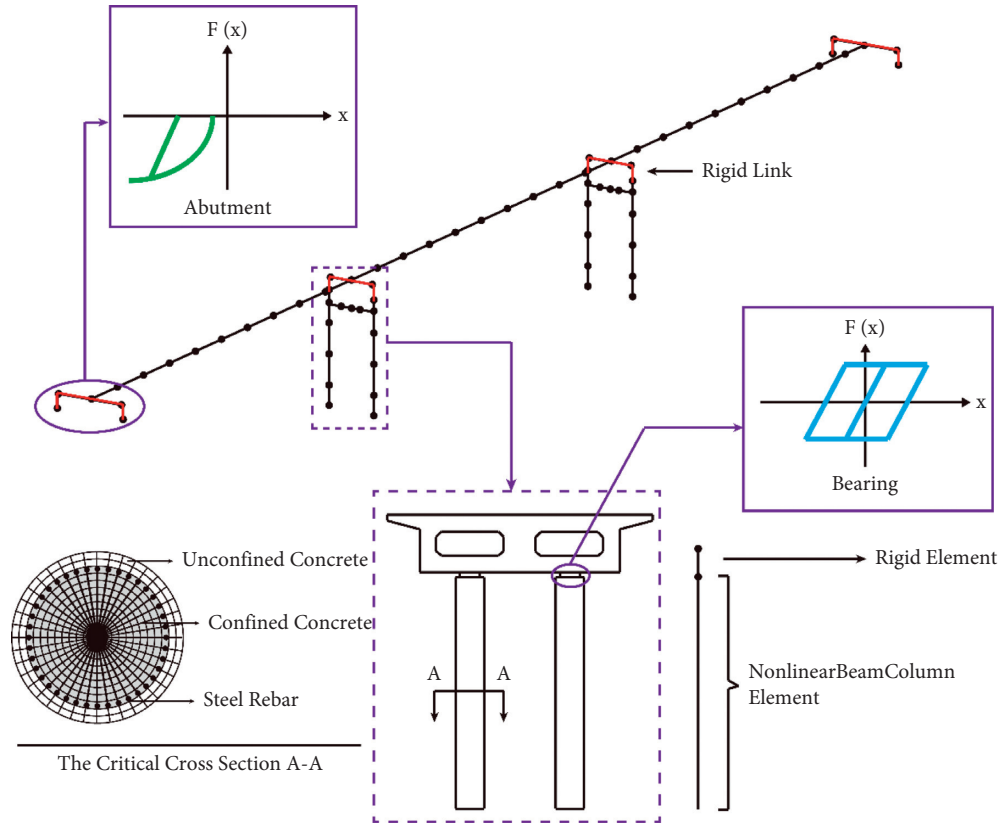
FIGURE 2: Numerical model of example bridge.

was used for longitudinal reinforcement. The bridge bearings were modeled with flat slider bearing elements. Hyperbolic gap model proposed by Wilson and Elgamal [22] was used to simulate the nonlinear deformation characteristics and resistance of abutments. In order to simplify the model and reduce the calculation time, soil-structure interaction was ignored.

4. Ground Motion Selection

Ground motion is a time varying process, whose main characteristics are the amplitude, spectrum, and duration, which are referred to as the three key characteristics of an earthquake. In recent years, earthquake engineering researchers have proposed many ground motion parameters to describe the intensity of ground motion. Different IMs correlate differently with the seismic response of the same bridge structure. Therefore, it is necessary to use multiple IMs to evaluate the seismic response of bridges. The current study used the suite of GMs developed by Baker et al. [23], which was established as part of the PEER Transportation Research Program for the seismic risk assessment of infrastructure systems in California. Four sets of GMs were selected (set 1: $M = 6$, $R = 25$ km, soil site; set 2: $M = 7$, $R = 10$ km, rock site; set 3: $M = 7$, $R = 10$ km, soil site; set 4: pulse-like GMs). Figure 3 presents the response spectra for each ground motion set. A total of 21 IMs were considered as given in Table 1. In this paper, the bridge was excited in the longitudinal direction.

5. Probabilistic Seismic Demand Models

In this study, the horizontal (longitudinal and transverse) components of each ground motion [23] excited the bridge model in the longitudinal direction, and the seismic responses were calculated. The seismic responses refer to longitudinal pier curvatures (Figure 2) (at the bottom of pier #1 ($Lcur1$), pier #2 ($Lcur2$), pier #3 ($Lcur3$), and pier #4 ($Lcur4$)) and transverse pier curvatures (at the bottom of pier #1 ($Tcur1$), pier #2 ($Tcur2$), pier #3($Tcur3$), and pier #4 ($Tcur4$)). In this study, the IMs and observed responses were transformed by taking their natural logarithm.

As a key step in PBSD, the PSDMs are established to describe the probabilistic relationships between engineering demand parameters and IMs. Previous studies on PSDMs mainly used the model proposed by Cornell et al. [31]. Cornell et al. [31] assumed that the median seismic demand (S_d) and the intensity parameter (IM) satisfy the following exponential relationship:

$$S_d = a\,(\mathrm{IM})^b, \tag{9}$$

where a and b are the regression coefficients. Taking natural logarithm of both sides of (5) yields:

$$\ln\,(S_d) = \ln\,a + b \ln\,(\mathrm{IM}). \tag{10}$$

In the present work, the developed model has been assessed using the coefficient of determination (R^2):
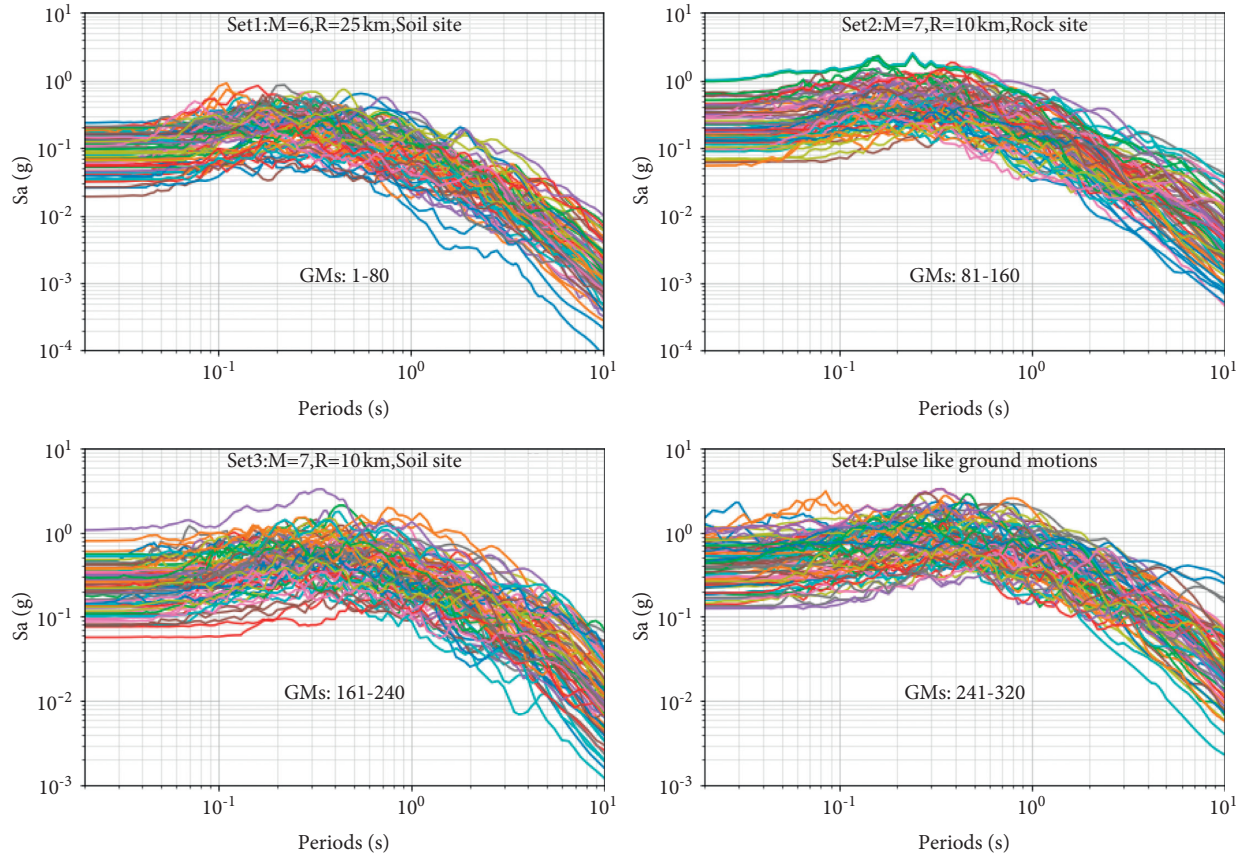
FIGURE 3: Response spectra for each ground motion set.

TABLE 1: Description of IMs.

IM	Full name	Definition	Unit
PGA	Peak ground acceleration	$\text{PGA} = \max|\ddot{u}_g(t)|$	g
PGV	Peak ground velocity	$\text{PGV} = \max|\dot{u}_g(t)|$	cm/sec
PGD	Peak ground displacement	$\text{PGD} = \max|u_g(t)|$	cm
(v_{max}/a_{max})	Peak velocity and acceleration ratio	$(v_{max}/a_{max}) = (\max|\dot{u}_g(t)|/\max|\ddot{u}_g(t)|)$	sec
(d_{max}/v_{max})	Peak displacement and velocity ratio	$(d_{max}/v_{max}) = \max|u_g(t)|/\max|\dot{u}_g(t)|$	sec
Arms24	Root-mean-square of acceleration	$\text{Arms} = \sqrt{1/t_{tot} \int_0^{TD} [\ddot{u}_g(t)]^2 dt}$	g
Vrms24	Root-mean-square of velocity	$\text{Vrms} = \sqrt{1/t_{tot} \int_0^{TD} [\dot{u}_g(t)]^2 dt}$	cm/sec
Drms24	Root-mean-square of displacement	$\text{Drms} = \sqrt{1/t_{tot} \int_0^{TD} [u_g(t)]^2 dt}$	cm
TD25	Significant duration	$\text{TD} = t(95\%\text{AI}) - t(5\%\text{AI})$	sec
AI26	Arias intensity	$\text{AI} = 1/t_{tot} \int_0^{TD} [\dot{u}_g(t)]^2 dts$	m/sec
SED	Specific energy density	$\text{SED} = \int_0^{TD} [\dot{u}_g(t)]^2 dt$	cm²/sec
CAV27	Cumulative absolute velocity	$\text{CAV} = \int_0^{TD} |\dot{u}_g(t)| dt$	cm/sec
CAD28	Cumulative absolute displacement	$\text{CAD} = \int_0^{TD} |\ddot{u}_g(t)| dt$	cm
CAI28	Cumulative absolute energy	$\text{CAD} = \int_0^{TD} |u_g(t)| dt$	cm/sec
PSA	Peak spectral acceleration	$\text{PSA} = \max(S_a(T_i))$	g
PSV	Peak spectral velocity	$\text{PSV} = \max(S_v(T_i))$	cm/sec
PSD	Peak spectral displacement	$\text{PSD} = \max(S_d(T_i))$	cm
$Ia29$	Compound acceleration-related IM	$\text{Ia} = \text{PGA} \times \text{TD}^{1/3}$	$g.\text{sec}^{1/3}$
$Iv29$	Compound velocity-related IM	$\text{Iv} = \text{PGV}^{2/3} \times \text{TD}^{1/3}$	$\text{cm}^{2/3}/\text{sec}^{1/3}$
$Id29$	Compound displacement-related IM	$\text{Id} = \text{PGD} \times \text{TD}^{1/3}$	$\text{cm.sec}^{1/3}$
FI30	Fajfar intensity	$\text{FI} = \text{PGV} \times \text{TD}^{1/4}$	$\text{cm/sec}^{3/4}$

	Lcur1	Tcur1	Lcur2	Tcur2	Lcur3	Tcur3	Lcur4	Tcur4
PGA	0.64	0.63	0.64	0.64	0.66	0.63	0.65	0.62
PGV	0.81	0.79	0.8	0.79	0.81	0.79	0.8	0.8
PGD	0.57	0.54	0.55	0.54	0.58	0.58	0.58	0.56
vmax/amax	0.17	0.12	0.17	0.16	0.13	0.14	0.16	0.13
dmax/vmax	0.017	0.015	0.0073	0.0083	0.042	0.0087	0.012	0.027
Arms	0.69	0.67	0.67	0.65	0.69	0.67	0.67	0.67
Vrms	0.72	0.71	0.73	0.71	0.73	0.72	0.72	0.73
Drms	0.49	0.48	0.48	0.45	0.49	0.49	0.47	0.49
TD	0.059	0.027	0.0042	0.046	0.02	0.046	0.034	0.0078
AI	0.73	0.75	0.75	0.75	0.77	0.75	0.74	0.73
SED	0.69	0.65	0.69	0.68	0.7	0.69	0.67	0.7
CAV	0.61	0.61	0.6	0.63	0.64	0.62	0.62	0.63
CAD	0.51	0.51	0.48	0.5	0.53	0.53	0.49	0.53
CAI	0.39	0.36	0.37	0.36	0.37	0.37	0.36	0.37
PSA	0.55	0.53	0.56	0.52	0.58	0.55	0.57	0.56
PSV	0.82	0.8	0.81	0.8	0.81	0.81	0.81	0.81
PSD	0.5	0.51	0.5	0.5	0.49	0.52	0.51	0.5
Ia	0.66	0.64	0.67	0.66	0.66	0.65	0.66	0.66
Iv	0.73	0.73	0.73	0.74	0.74	0.73	0.74	0.72
Id	0.5	0.52	0.53	0.52	0.54	0.52	0.54	0.53
FI	0.77	0.77	0.78	0.77	0.79	0.78	0.78	0.77

(a)

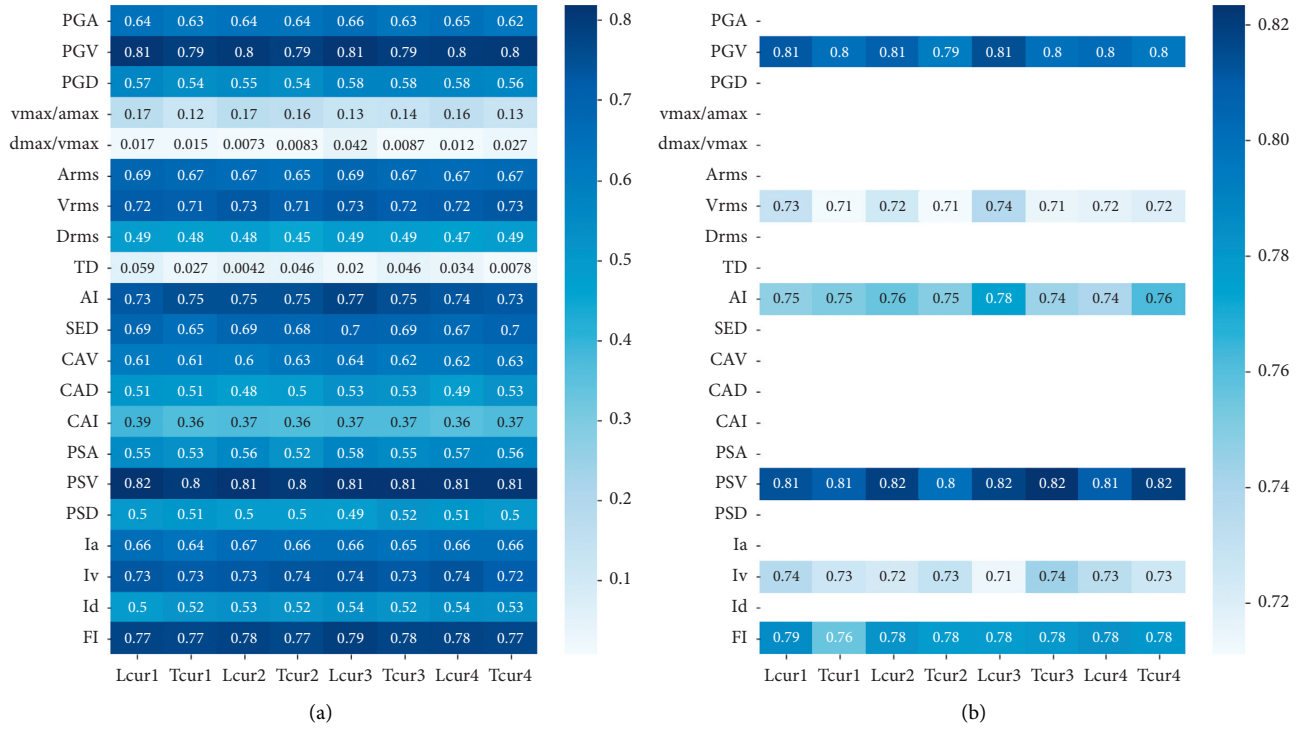|	Lcur1	Tcur1	Lcur2	Tcur2	Lcur3	Tcur3	Lcur4	Tcur4
PGV	0.81	0.8	0.81	0.79	0.81	0.8	0.8	0.8
Vrms	0.73	0.71	0.72	0.71	0.74	0.71	0.72	0.72
AI	0.75	0.75	0.76	0.75	0.78	0.74	0.74	0.76
PSV	0.81	0.81	0.82	0.8	0.82	0.82	0.81	0.82
Iv	0.74	0.73	0.72	0.73	0.71	0.74	0.73	0.73
FI	0.79	0.76	0.78	0.78	0.78	0.78	0.78	0.78

(b)

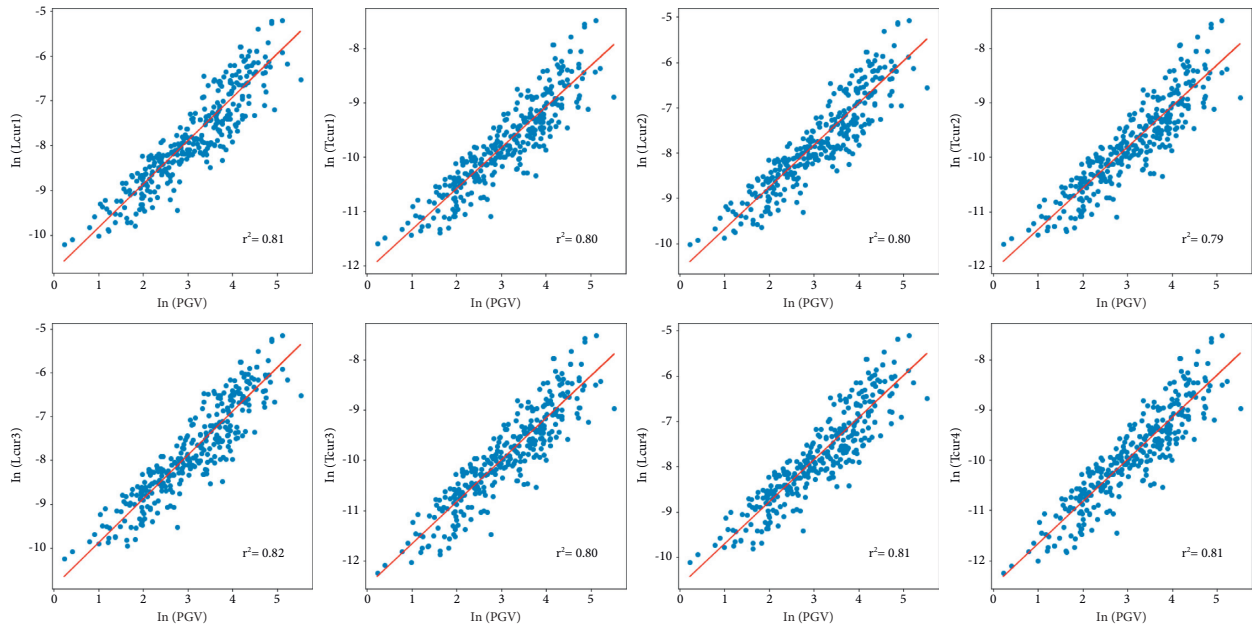FIGURE 4: Coefficients of determination (R^2) for seismic responses with respect to IMs.

FIGURE 5: PSDM results for longitudinal (*Lcur1*, *Lcur2*, *Lcur3*, and *Lcur4*) and transverse (*Tcur1.*, *Tcur2*, *Tcur3*, and *Tcur4*) pier curvatures at pier bottom vs. PGV.

$$R^2 = 1 - \frac{\sum_1^n \left(y_{\text{predict}_i} - y_{\text{true}_i}\right)^2}{\sum_1^n \left(y_{\text{trueMean}_i} - y_{\text{true}_i}\right)^2}, \quad i \in (1, n). \quad (11)$$

The larger the R^2 value, the better the fitting of the regression curve to the data.

The coefficients of determination (R^2) for the considered seismic responses with respect to the 21 IMs listed in Table 1

are summarized in Figure 4(a), and values of R^2 greater than 0.7 are shown in Figure 4(b). It can be observed that the PGV and PSV fit the data better than the other IMs. Figures 5 and 6 present the PSDM results of the seismic responses versus PGV and PSV. It can be seen that the prediction ability of the regression model generated by fitting only one IM (PGV or PSV) required improvement mainly because using PGV or PSV only cannot guarantee to capture all the main
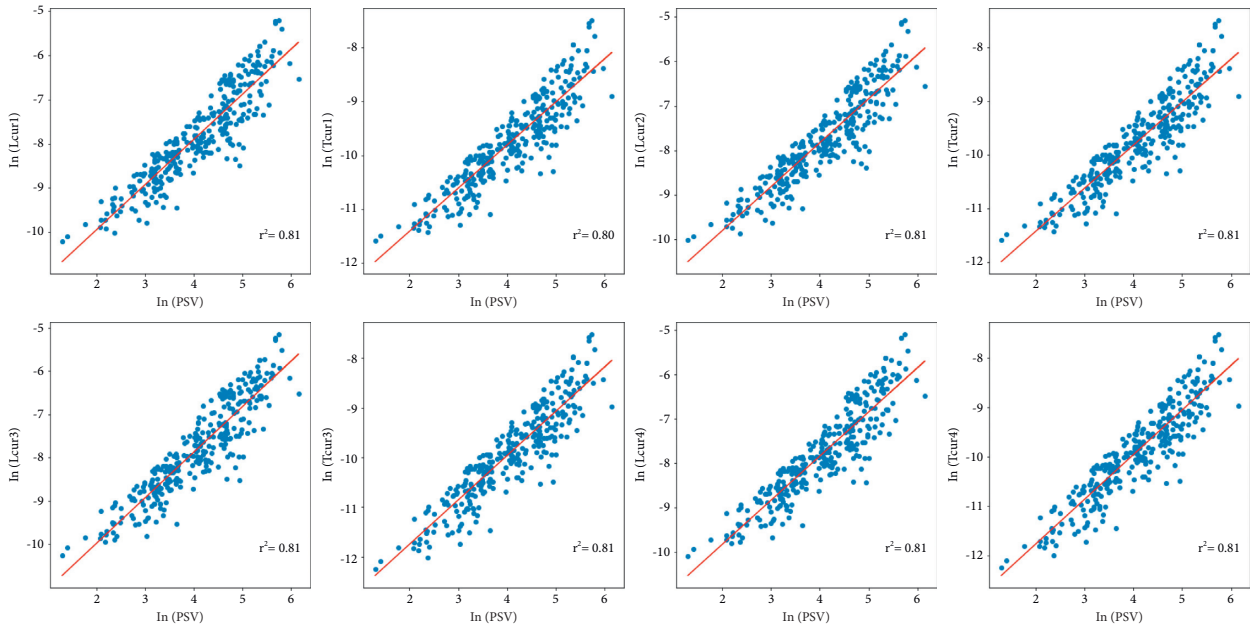
FIGURE 6: PSDM results for longitudinal (*Lcur*1, *Lcur*2, *Lcur*3, and *Lcur*4) and transverse (*Tcur*1, *Tcur*2, *Tcur*3, and *Tcur*4) pier curvatures at pier bottom vs. PSV.

TABLE 2: Coefficients of determination of traditional and various ML demand models.

|	Methods	*Lcur*1	*Tcur*1	*Lcur*2	*Tcur*2	*Lcur*3	*Tcur*3	*Lcur*4	*Tcur*4
Traditional method	Only PGV	0.81	0.80	0.81	0.79	0.81	0.80	0.80	0.80
	Only PSV	0.81	0.81	0.82	0.80	0.82	0.82	0.81	0.82
Linear regression	RR	0.86	0.85	0.86	0.85	0.86	0.86	0.86	0.86
	LR	0.47	0.31	0.46	0.31	0.46	0.40	0.44	0.40
	EN	0.65	0.57	0.63	0.59	0.66	0.61	0.65	0.62
	SVR	0.88	0.87	0.88	0.87	0.88	0.87	0.87	0.87
Bayesian regression	BRR	0.86	0.85	0.86	0.86	0.86	0.86	0.85	0.86
	ARD	0.85	0.85	0.86	0.85	0.86	0.86	0.86	0.84
Tree-based model	RF	**0.89**	**0.88**	**0.89**	**0.87**	**0.88**	**0.87**	**0.88**	**0.88**
	GBDT	0.88	0.86	**0.89**	0.86	**0.88**	0.86	0.87	0.87
	AdaBoost	0.87	0.86	0.87	0.86	**0.88**	0.85	0.87	0.87
	LightGBM	**0.89**	0.86	0.88	0.86	**0.88**	0.86	0.87	0.85
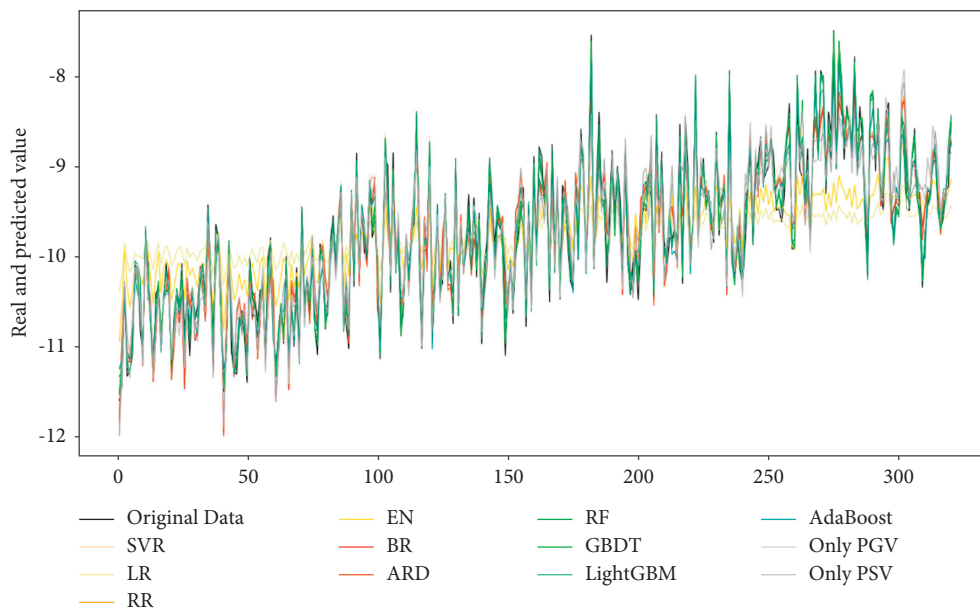
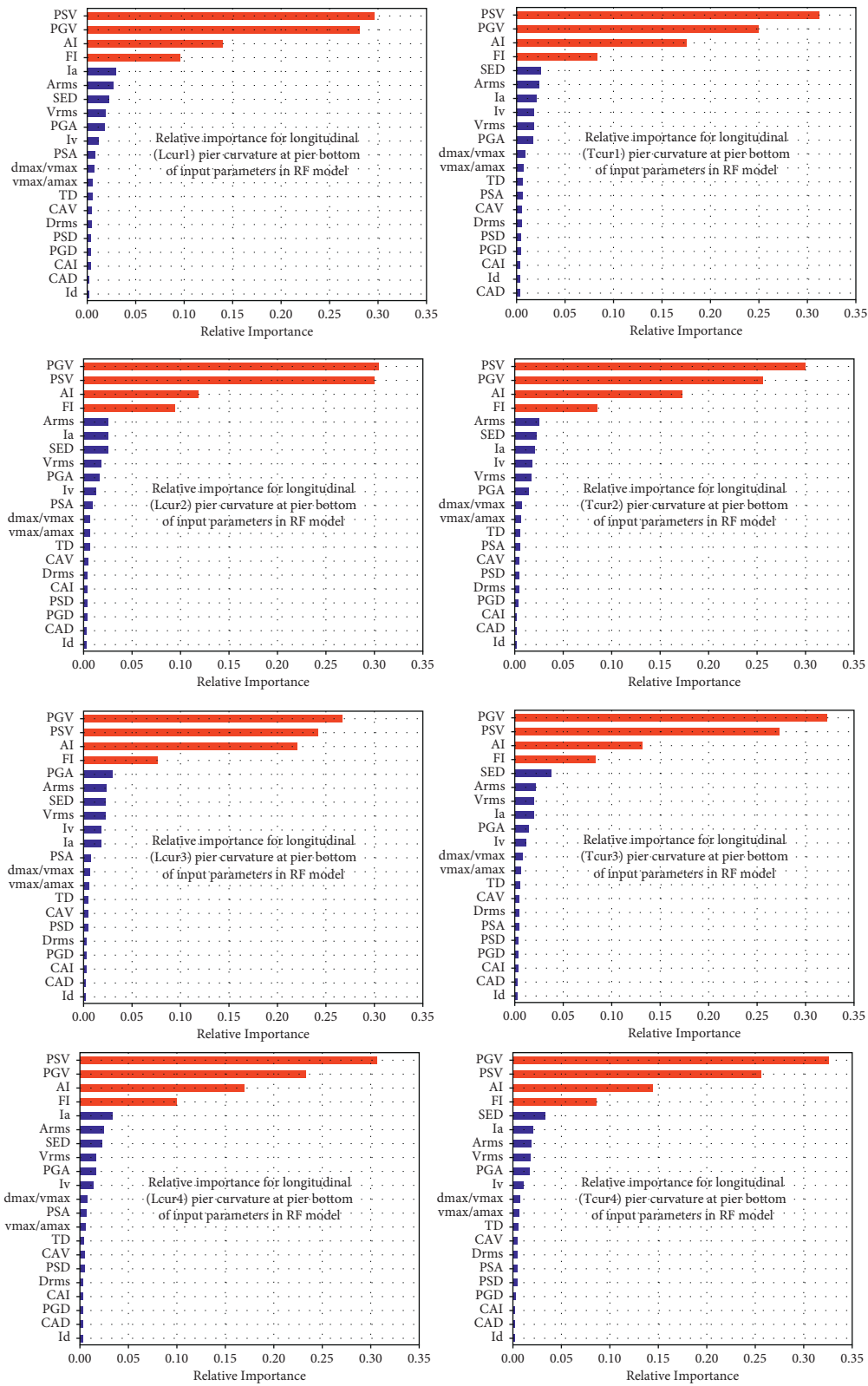FIGURE 7: Comparison of regression results of traditional and various ML methods.

FIGURE 8: Relative importance of input parameters in RF model.

characteristics of a ground motion. In order to improve the ability of the regression models to fit the data and consider more IMs, ML methods were adopted to establish the PSDMs in this study.

The ML methods described in Section 2 were used to establish the machine-learning-based PSDMs. 90% of the available data were used to establish the prediction model (training set), and the remaining 10% to evaluate the

performance of the prediction model (test set). The partitioning of the entire dataset into the training set and testing set was random, and the performance of the model on the testing set was taken as an indication of the performance of the model on the unknown data. The ML codes for the models mentioned in Section 2 were developed using the open-source Python package scikit-learn [32]. The performance of each ML model was evaluated using the coefficient of determination (R^2) (Table 2). The maximum values of each column in Table 2 are marked by bold font. As shown in Table 2, the results of the tree-based model are better than those of the linear regression and BR, while the results of the different tree-based models (RF, GBDT, AdaBoost, and LightGBM) are close to one another. It can be concluded from Table 2 that the tree-based models had a higher accuracy than the traditional methods, linear regression, and BR. It can be seen from Figure 7 that the fitting degree of RF (green) was closest to the original data (black), which is consistent with the results listed in Table 2. Thus, RF was deemed to be the most suitable model to establish the PSDM in this research.

To evaluate how the performance of the RF model is affected by the input parameters, a further analysis was carried out to identify the importance of the input parameters, as shown in Figure 8. The importance of a feature [32] was computed as the (normalized) total reduction of the Gini importance [33] of that feature. The higher the Gini importance, the more important the feature. Note that the sum of all the values above the vertical bars in Figure 8 is 100%. As seen in Figure 8, PGV, PSV, AI, and FI are all critical factors of the PSDMs, while all other IMs have much less influence on the demand models. It can also be seen that, based on their definitions, the critical IMs (PGV, PSV, AI, and FI) are correlated with the velocity except for AI.

6. Conclusions

The main purpose of this paper was to develop a PSDM to predict the seismic response of regular continuous girder bridges based on ML methods. The efficiency of various ML regression models, such as lasso regression, ridge regression, elastic net, supported vector regression, Bayesian ridge regression, automatic relevance determination, random forest, gradient boosting decision tree, adaptive boosting, and light gradient boosting machine, for generating of PSDMs for a regular continuous girder bridge was first evaluated. A three-dimensional numerical model of a bridge was generated using OpenSees. A total of 320 GMs attributed to four site conditions were selected to excite the bridge. The coefficient of determination (R^2) was used to assess the ML models developed. Several main conclusions are drawn as follows:

(i) PGV and PSV are more appropriate than other IMs for establishing the PSDMs for the regular continuous girder bridges subjected to the selected GMs. However, the machine-learning-based demand models utilizing more than one IMs are better than the traditional demand models based on a single IM.

(ii) This paper evaluated three mainstream ML regression methods (linear regression, BR, and tree-based model), incorporating 10 ML models: LR, RR, EN, SVR, BRR, ARD, RF, GBDT, AdaBoost, and LGBM. The tree-based models had a higher accuracy than traditional methods, linear regression, and a significantly improved coefficient of determination. The obtained results indicated that RF can be used for predicting the seismic behavior of regular continuous girder bridges subjected to far and near-fault GMs.

(iii) This study identified the critical factors for the RF-based PSDMs of the bridge for various engineering demand parameters. PGV, PSV, AI, and FI were found to be the critical factors of the RF-based PSDMs, and they were all correlated with velocity, except for AI.

Although the findings of this study are based on the case study of a three-span regular continuous girder bridge in China, the methodology is applicable to other bridges. As the accurate probabilistic seismic demand analysis is a common challenge in performance-based earthquake engineering frameworks, the proposed approach will help to improve establishing the PSDMs.

Data Availability

The data used to support the findings of this study were supplied by Li under license and so cannot be made freely available. Requests for access to these data should be made to lwshan1995@gmail.com.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] L. S. Chen, W. L. Zhuang, H. Q. Zhao, and Z. Wan, *Report on Highways' Damage in the Wenchuan Earthquake*, China Communications Press, Beijing, China, 2012.

[2] Y. Nakao and S. Kataoka, "Statistical study on damage to highway bridges due to the 2011 off the pacific coast of Tohoku Earthquake," *Journal of Japan Society of Civil Engineers*, vol. 73, no. 4, pp. 628–I, 2017.

[3] C. A. Cornell and H. Krawinkler, *Progress and Challenges in Seismic Performance Assessment*, Peer Center News, 2000, https://www.researchgate.net/publication/237131473_Progress_and_Challenges_in_Seismic_Performance_Assessment.

[4] J. Moehle and G. G. Deierlein, "A Framework Methodology for Performance-Based Earthquake Engineering," in *Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, Canada, August 2004.

[5] K. Mackie and B. Stojadinović, "Probabilistic seismic demand model for California highway bridges," *Journal of Bridge Engineering*, vol. 6, no. 6, pp. 468–481, 2001.

[6] KR. Mackie, *Fragility-based Seismic Decision Making for Highway Overpass Bridges*, University of California, Berkeley, CA, USA, 2004.

[7] BG. Nielson, *Analytical Fragility Curves for Highway Bridges in Moderate Seismic Zones*, Georgia Institute of Technology, Atlanta, GA, USA, 2005.

[8] B. G. Nielson and R. Desroches, "Seismic fragility methodology for highway bridges using a component level approach," *Earthquake Engineering & Structural Dynamics*, vol. 36, no. 6, pp. 823–839, 2007.

[9] J. E. Padgett and R. Desroches, "Methodology for the development of analytical fragility curves for retrofitted bridges," *Earthquake Engineering & Structural Dynamics*, vol. 37, no. 8, pp. 1157–1174, 2010.

[10] Y. Pan, A. K. Agrawal, and M. Ghosn, "Seismic fragility of continuous steel highway bridges in New York state," *Journal of Bridge Engineering*, vol. 12, no. 6, pp. 689–699, 2007.

[11] Y. Pan, A. K. Agrawal, M. Ghosn, and S. Alampalli, "Seismic fragility of multispan simply supported steel highway bridges in New York state. II: fragility analysis, fragility curves, and fragility surfaces," *Journal of Bridge Engineering*, vol. 15, no. 5, pp. 462–472, 2010.

[12] J. E. Padgett, B. G. Nielson, and R. Desroches, "Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios," *Earthquake Engineering & Structural Dynamics*, vol. 37, no. 5, pp. 711–725, 2010.

[13] K. F. Zheng, L. B. Chen, W. L. Zhuang, and H. S. Ma, "Bridge vulnerability analysis based on probabilistic seismic demand models," *Engineering Mechanics*, vol. 30, no. 5, pp. 165–171, 2013.

[14] H.-B. Ma, W.-D. Zhuo, D. Lavorato et al., "Probabilistic seismic response analysis on continuous bridges under near-fault ground motions," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, no. 3, pp. 491–500, 2019.

[15] L. Jiang, J. Zhong, M. He, and W. Yuan, "Optimal seismic intensity measure selection for isolated bridges under pulse-like ground motions," *Advances in Civil Engineering*, vol. 2019, Article ID 3858457, 22 pages, 2019.

[16] Q. Wang, Z. Y. Wu, and S. W. Liu, "Multivariate probabilistic seismic demand model for the bridge multidimensional fragility analysis," *KSCE Journal of Civil Engineering*, vol. 22, 2008.

[17] Q. Huang, P. Gardoni, and S. Hurlebaus, "Probabilistic seismic demand models and fragility estimates for reinforced concrete highway bridges with one single-column bent," *Journal of Engineering Mechanics*, vol. 136, no. 11, pp. 1340–1353, 2010.

[18] M. Mohammed, M. Badruddin Khan, E.. Bashier, and M.. Bashier, *Machine Learning: Algorithms and Applications*, CRC Press, Boca Raton, FL, USA, 2016.

[19] F. Pedregosa, G. Varoquaux, and A. Gramfort, *Scikit-learn: Machine Learning in Python*, 2012, https://www.researchgate.net/publication/305386496_Scikit-Iearn_Machine_learning_in_python.

[20] T. Hastie, R. Tibshirani, and J. H. Friedman, "The elements of statistical learning," *Springer, Elements*, vol. 1, no. 3, pp. 267-268, 2009.

[21] F. McKenna, G. L. Fenves, and M. H. Scott, *Open System for Earthquake Engineering Simulation*, Pacific Earthquake Engineering Research Center, Berkeley, CA, USA, 2010.

[22] P. Wilson and A. Elgamal, "Large scale measurement of lateral earth pressure on bridge abutment back-wall subjected to static and dynamic loading," in *Proceedings of the New Zealand Workshop on Geotechnical Earthquake Engineering*, Christchurch, New Zealand, January 2006.

[23] J. W. Baker, T. Lin, and S. K. Shahi, *New Ground Motion Selection Procedures and Selected Motions for the PEER Transportation Research Program*, Pacific Earthquake Engineering Research Center, Berkeley, CA, USA, 2011.

[24] S. Kramer, *Geotechnical Earthquake Engineering*, Prentice-Hall, Hoboken, NJ, USA, 2008.

[25] TD. Trifunac and AG. Brady, "A study on the duration of strong earthquake ground motion," *Bulletin of the Seismological Society of America*, vol. 65, no. 3, pp. 581–626, 1975.

[26] A. Arias, "A Measure of Earthquake Intensity," *Seismic Design for Nuclear Power Plants*, Massachusetts Inst. of Tech. Press, Cambridge, MA, USA, 1970.

[27] J. W. Reed and R. P. Kassawara, "Criterion for determining exceedance of the operating basis earthquake," *Nuclear Engineering and Design*, vol. 123, no. 2, pp. 387–396, 1990.

[28] J. Hu, W. Wu, and L. Xie, "Review and analysis of cumulative absolute velocity related parameters of ground motion," *Journal of Earthquake Engineering & Engineering Vibration*, 2013, https://www.researchgate.net/publication/288274519_Review_and_analysis_of_cumulative_absolute_velocity_related_parameters_of_ground_motion.

[29] R. Riddell and J. E. Garcia, "Hysteretic energy spectrum and damage control," *Earthquake Engineering & Structural Dynamics*, vol. 30, no. 12, pp. 1791–1816, 2001.

[30] P. Fajfar, T. Vidic, and M. Fischinger, "A measure of earthquake motion capacity to damage medium-period structures," *Soil Dynamics and Earthquake Engineering*, vol. 9, no. 5, pp. 236–242, 1990.

[31] C. A. Cornell, F. Jalayer, and R. O. Hamburger, "The probabilistic basis for the 2000 SAC/FEMA steel moment frame guidelines," *Journal of Structural Engineering*, vol. 128, no. 4, 2002.

[32] F. Pedregosa, G. Varoquaux, and A. Gramfort, *Scikit-learn: Machine Learning in Python*, 2012.

[33] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.