

XIX ANIDIS Conference, Seismic Engineering in Italy

Machine Learning-based Seismic Fragility Curves for RC Bridge Piers

Xuguang Wang^{a,b}, Cristoforo Demartino^{a,b,*},
Giorgio Monti^c, Giuseppe Quaranta^d, Alessandra Fiore^e

^aZhejiang University - University of Illinois at Urbana Champaign Institute, Haining 314400, Zhejiang, PR China

^bDepartment of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

^cDepartment of Structural Engineering and Geotechnics, Sapienza University of Rome, via A. Gramsci 53, 00197 Rome, Italy

^dDepartment of Structural Engineering and Geotechnics, Sapienza University of Rome, via Eudossiana 18, 00184 Rome, Italy

^eDepartment of Civil Engineering Sciences and Architecture, Polytechnic University of Bari, via Giovanni Amendola 126/B, Bari, Italy

Abstract

A significant part of ongoing studies in the field of earthquake engineering is directed toward the seismic risk assessment of buildings and infrastructures at a territorial scale. This task is usually accomplished by grouping the structures into homogenous classes in terms of typology, for which seismic fragility curves are then obtained for different limit states via numerical simulations or from the statistical analysis of observational data when available. Particularly, the development of typological fragility curves for bridges under earthquake is useful for assessing the reliability and resilience of transportation networks in seismic areas and can be also effective decision-making support. Within this framework, the proposed study establishes a machine learning-based paradigm for the closed-form prediction of the main statistical parameters required to obtain relevant seismic fragility curves for reinforced concrete bridge piers. Initially, a huge training dataset has been obtained by Monte Carlo simulations and displacement-based bridge pier assessments by assuming data representative of the Italian highway transportation network. Next, symbolic nonlinear regression formulae for estimating the main statistical parameters of seismic fragility curves have been generated. With the aid of those formulae, the effort of calculating the seismic fragility curves is greatly reduced since the corresponding main statistical parameters can be directly calculated from a set of commonly available attributes. Therefore, the proposed study provides a helpful tool for the rapid preliminary assessment of damage and risk level of existing highway transportation networks exposed to seismic hazards.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the XIX ANIDIS Conference, Seismic Engineering in Italy.

Keywords: Bridge Piers; Fragility Curve; Interpretable Data-driven Models; Seismic Assessment.

1. Introduction

Physical vulnerability can be defined as the susceptibility of an exposed asset to seismic impacts (damage) determined by the likelihood of the occurrence of certain damage levels caused by seismic action. In this context, fragility curves are important tools for evaluating the potential seismic hazard and its impact on civil infrastructure systems. Fragility curve is a statistical tool representing the probability of exceeding a given damage state (or performance) as a function of an engineering demand parameter that represents the ground motion (preferably spectral displacement at a given frequency). Performance evaluation of bridge piers using seismic fragility curves can provide valuable insight into the asset management process, such as maintenance scheduling and decision-making process under disaster.

Fragility curves can be assessed using historical records. For example, Thapa et al. (2022) constructed fragility curves for RC bridges based on historical records; Mosleh et al. (2020) generated fragility curves based on the numerical analysis results using parameters extracted from the pre-1990 highway concrete bridges; Gautam et al. (2021) established seismic fragility curves for structural and non-structural elements of RC buildings using the historical information of 2015 Gorkha Earthquake in Nepal. Fragility curves can also be generated analytically through seismic response analysis (Hwang et al. 1994; Mander et al. 1999; Karim et al. 2003). The approach is beneficial when historical data are not available, but the constructed fragility curves are usually for a specific structure. Another common approach is to apply Monte Carlo simulations to directly compute the failure probability at a specific limit state. The most significant drawback of such an approach is the high computational cost.

A machine learning-based framework to overcome the drawback and limitations in generating fragility curves is proposed in this study. The overall architecture of the proposed framework is presented in Fig. 1. A dataset of displacement capacity of hollow rectangular RC bridge piers is generated through Monte Carlo simulation. The symbolic equations are established via genetic programming with the generated dataset. The details of the dataset are described in Section 2, while the training and validation of the machine learning model are presented in Section 3.

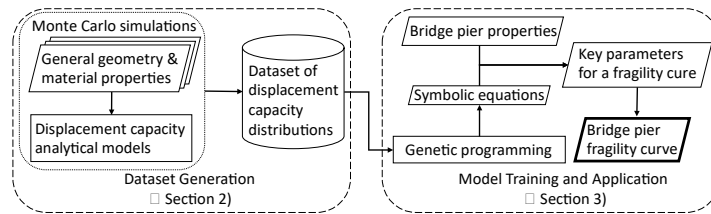


Fig. 1. Architecture of the proposed framework.

2. Dataset generation

The displacement of the pier under seismic action is selected as the Intensity Measure (IM) of the fragility curve. The fragility curve is presented as the cumulated probability function of the lognormal distribution, $\text{Lognormal}(\mu, \sigma^2)$, of the displacement capacity also including uncertainties. The two main parameters for generating a fragility curve are the mean, μ_X , and coefficient of variation, CoV , of the displacement capacity distribution at a given limit state (performance level). μ_X and CoV are converted to μ, σ^2 as follows:

$$\mu = \ln\left(\frac{\mu_X}{\sqrt{CoV^2 + 1}}\right) \quad \sigma^2 = \ln(CoV^2 + 1) \quad (1)$$

To train the symbolic models for those parameters through genetic programming, a dataset is generated by solving the analytical models. Different combinations of model input parameters are applied to consider the epistemic uncertainties. For each combination, Monte Carlo simulations are performed to take account of the aleatoric uncertainties by introducing randomness on certain parameters. The model considers bending and shear and takes the weaker between the two mechanisms as the governing one. The analytical model and its input parameters for preparing the dataset are described in detail in this section.

2.1. Bending capacity model

Simplified material models are employed for the bending capacity calculation. The relationship between steel stress, σ_s , and strain, ε_s , follows the Hooke's law within the elastic range, and the yield stress, f_y , occurs when the strain exceeds the elastic limit, $\varepsilon_{sy} = 0.2\%$. The concrete stress, σ_c , normalized with respect to the peak compressive strength, f_c , is given as function of the strain, ε_c , as follows:

$$\sigma_c = \begin{cases} 0 & \varepsilon_c \geq 0 \\ \left(1 - \frac{\varepsilon_c}{\varepsilon_{c0}}\right) - 1 & \varepsilon_{c0} < \varepsilon_c < 0 \\ -1 & \varepsilon_c \leq \varepsilon_{c0} \end{cases} \quad (2)$$

where $\varepsilon_{c0} = -0.2\%$ is the concrete strain when the peak strength (negative in compression) is reached.

The bending capacity is calculated for Damage Limit State (DLS) and Ultimate Limit State (ULS). The DLS occurs when the strain at the bottommost reinforcement steel is $\varepsilon_s = \varepsilon_{sy}$, while the ULS occurs when either the strain in concrete or steel reaches its ultimate strain (i.e., $\varepsilon_c = \varepsilon_{cu}$, with ε_{cu} the concrete ultimate strain, or $\varepsilon_s = \varepsilon_{su}$, with ε_{su} the steel ultimate strain). Thus, either ε_c or ε_s is known in a limit state, and the unknown strain can be found by solving the normalized equilibrium equation:

$$F_c + F_{s,top} + F_{s,bot} + F_{s,lat} + n_s = 0 \quad (3)$$

where F_c is the normalized force in the concrete, $F_{s,top}$ and $F_{s,bot}$ are the normalized forces in the top and bottom longitudinal steel, respectively, $F_{s,lat}$ is the normalized force in the lateral steel reinforcement, and n_s is the normalized axial load. The normalized forces in the concrete and reinforcement steel are calculated as:

$$F_c = \int_{\xi_{c,bot}}^{\xi_{c,top}} \sigma_c(\varepsilon(\xi)) \cdot \beta_c(\xi) d\xi \quad (4)$$

$$F_{s,bot} = \frac{\rho_{s,bot}}{2} \frac{f_y}{f_c} (\sigma_s(\varepsilon(\xi_{s,top1})) + \sigma_s(\varepsilon(\xi_{s,top2}))) \quad (5)$$

$$F_{s,top} = \frac{\rho_{s,top}}{2} \frac{f_y}{f_c} (\sigma_s(\varepsilon(\xi_{s,bot1})) + \sigma_s(\varepsilon(\xi_{s,bot2}))) \quad (6)$$

$$F_{s,lat} = \frac{\rho_{s,lat}}{\xi_{s,lat,top}} \frac{f_y}{f_c} \int_{\xi_{s,lat,bot}}^{\xi_{s,lat,top}} \sigma_s(\varepsilon(\xi)) d\xi \quad (7)$$

where β_c is the normalized width of the cross-section, ξ denotes the normalized location over the cross-section depth, and $\rho_{s,top}$, $\rho_{s,bot}$, $\rho_{s,lat}$ are the geometrical reinforcement ratio of the topmost, bottommost, and lateral steel. The longitudinal reinforcement is assumed to be evenly distributed on each side of the rectangular cross-section, so $\rho_{s,top} = \rho_{s,bot} = \rho_{s,lat} = \rho_s/4$, where ρ_s is total longitudinal reinforcement ratio. Based on the cross-section planarity assumption and strain compatibility, strain at any location ξ over the cross-section depth, is computed as:

$$\varepsilon(\xi, \varepsilon_c, \varepsilon_s) = \varepsilon_c + \frac{\varepsilon_s - \varepsilon_c}{1 + \delta} (\xi_{c,top} - \xi) \quad (8)$$

where δ is the concrete cover thickness, normalized to the cross-section effective depth.

The strain profile at each limit state also can be calculated using Eq. (8). The corresponding normalized moment for the given strain profile is calculated as follows:

$$\begin{aligned}
 m_R = & \int_{\xi_{c,bot}}^{\xi_{c,top}} \sigma_c(\epsilon(\xi)) \cdot \beta_c(\xi) \xi d\xi \\
 & + \frac{\rho_{s,top}}{2} \frac{f_y}{f_c} (\sigma_s(\epsilon(\xi_{s,top1})) \xi_{s,top1} + \sigma_s(\epsilon(\xi_{s,top2})) \xi_{s,top2}) \\
 & + \frac{\rho_{s,bot}}{2} \frac{f_y}{f_c} (\sigma_s(\epsilon(\xi_{s,bot1})) \xi_{s,bot1} + \sigma_s(\epsilon(\xi_{s,bot2})) \xi_{s,bot2}) \\
 & + \frac{\omega_{s,lat}}{\xi_{s,lat,top}} \int_{\xi_{s,lat,bot}}^{\xi_{s,lat,top}} \sigma_s(\epsilon(\xi)) \xi d\xi
 \end{aligned} \quad (9)$$

The dimensional moment M_R is calculated by multiplying the normalized moment by the section width b , by the squared distance d between topmost and bottommost reinforcement steel, and by the concrete strength f_c :

$$M_R = m_R b d^2 f_c \quad (10)$$

2.2. Shear capacity model

The shear capacity model suggested in Eurocode 2 (2004) is adopted for calculating the pier shear capacity V_R , as follows:

$$V_R = 0.9d(2t_w)\rho_w f_y \cot(\theta_{max}) \quad (11)$$

where t_w is the web thickness of the hollow section, ρ_w is the transverse reinforcement geometrical ratio, and θ_{max} is the concrete strut angle where shear is balanced between transverse reinforcement and concrete strut:

$$\cot(\theta_{max}) = \sqrt{0.5\alpha_n \frac{1}{\rho_w} \frac{f_c}{f_y} - 1} \in [1, 2.5] \quad (12)$$

where α_n is a coefficient accounting for the normalized axial load n_s , as follows:

$$\alpha_n = \begin{cases} \min(1 + n_s, 1.25) & 0 < n_s < 0.5 \\ 2.5(1 - n_s) & 0.5 < n_s < 1 \end{cases} \quad (13)$$

2.3. Pier displacement capacity

The pier displacement capacity Δ_c is calculated using the smaller between the shear capacity and the equivalent force corresponding to the moment capacity:

$$\Delta_c = Q \left[\frac{\min(M_R, V_R H)}{K_E H} + \phi L_p \left(H - \frac{L_p}{2} \right) \right] \quad (14)$$

where H is the height of the pier, K_E is the elastic stiffness of the pier, ϕ is the curvature capacity of the section, L_p , the base plastic hinge length, is a function of pier height and the cross-section depth, h , and width, b :

$$L_p = 0.1H + 0.17 \frac{b+h}{2} \quad (15)$$

, and Q is a factor accounting for the P-Delta effect:

$$Q = \left(1 + \frac{n_s b d f_c}{K_E H} \right)^{-1} \quad (16)$$

2.4. Input parameters for data generation

A dataset is generated for $H = 15$ m piers for a preliminary appraisal of the proposed framework. 20,737 combinations are created for input parameters that listed in Table 1. Monte Carlo simulations with 1,000 iterations are performed in each combination. The uncertainties in the Monte Carlo simulations are introduced by randomizing f_y , f_c , and L_p . The values of those three parameters are drawn from normal distributions that have coefficient of variations of 10% for f_y , 20% for f_c , and 33% for L_p . μ_X and CoV of the displacement capacity to pier height ratio (drift ratio) for each combination are estimated through the Maximum Likelihood Estimation (MLE).

Table 1. Input parameters for generating dataset.

Parameter	Symbol	Range	Parameter	Symbol	Range
Width	b	1 - 4 m	Web thickness	t_w	0.2 - 0.4 m
Depth	h	1 - 4 m	Yield stress of steel	f_y	300 - 500 MPa
Concrete cover thickness	c	30 mm	Yield stress of concrete	f_c	15 - 45 MPa
Topmost to bottommost steel distance	d	0.97 - 3.97 m	Normalized axial load	n_s	1 - 4 %
Longitudinal reinforcement ratio	ρ_s	0.25 - 1.25%	Transverse reinforcement ratio	ρ_w	0.04 - 0.16%

3. Machine learning-based seismic fragility curves

The symbolic classification and regression via genetic programming are employed for generating equations that estimate the two main parameters of fragility curves, μ_X and CoV . When performing the symbolic classification and regression, genetic programming uses a tree-based representation of the candidate solutions which include the variables, constants and the operations. Through the crossover and mutation in each generation, an optimized model for estimating the target value is established. In this study, the parameters listed in Table 1 are used as the features for training the classification model and regression models, while μ_X and CoV of the drift ratio are the targets. The analysis via genetic programming is performed by means of an open-source code, HeuristicLab (Wagner et al. 2005).

3.1. Classification model

Due to the nonlinearity introduced by the material properties, using only the regression equations to represent the relationship between the features and targets is too complex and hard to apply in practice. Therefore, a classification is performed to simplify the regression equations in each class. The entities in the dataset are labelled as classes 0, 1 and 2 based on the relationship between the median moment capacity (Eq. (10)) and median shear capacity (Eq. (11)) in the limit states shown in Table 2.

The dataset is shuffled and divided into the training set and test set with a 7 to 3 ratio. A classification equation is obtained after 15,000 generations:

$$L = \log \left(c_1 \cdot \frac{\sqrt{t_w} \cdot \rho_w^{c_2} \cdot \left(\frac{1}{2} \cdot \frac{\rho_s}{4} \right)^{c_3}}{h^{c_4} \cdot b} \right), \text{ Class} = \begin{cases} 0 & L < 0.28843 \\ 1 & 0.28843 \leq L < 1.1028 \\ 2 & L \geq 1.1028 \end{cases} \quad (17)$$

where $c_1 = 61.413$, $c_2 = 0.625$, $c_3 = 0.2024$, and $c_4 = 0.75$.

Table 2. Classification based on the moment and shear capacity relationship.

Assigned label	Relationship
0	$H \cdot V_R < M_{R,DLS}$
1	$M_{R,DLS} \leq H \cdot V_R < M_{R,ULS}$
2	$M_{R,ULS} \leq H \cdot V_R$

The training accuracy and the test accuracy are 88.8% and 88.2%, respectively. The Receiver Operator Characteristic (ROC) of the classification model is shown in Fig. 2. The prediction quality for Class 0 and Class 2 are similar, but the accuracy for Class 1 is slightly lower than the other two classes due to the higher variance caused by interfering of the governing mechanism.

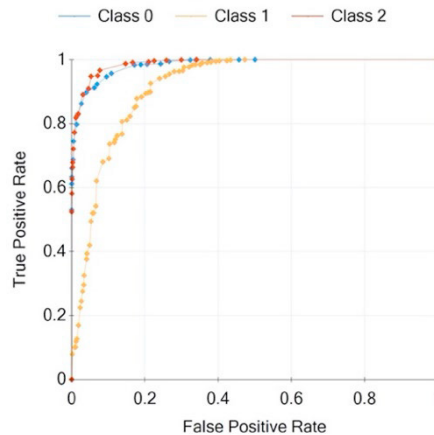


Fig. 2. ROC curve of the classification model.

3.2. Regression models

Regression models for estimating μ_X and CoV in the DLS and ULS are trained via genetic programming for all three classes. The established equations for each class show similar accuracy, so only the equations for class 0 are presented herein to be concise.

$$\mu_{DLS} = \frac{c_1 \cdot f_y \cdot \rho_w \cdot t_w \cdot b}{c_2 \cdot h \cdot t_w \cdot b^2 + c_3 \cdot h \cdot b^2 + c_4 \cdot t_w \cdot \rho_w \cdot f_y} + c_5 \quad (18)$$

where $c_1 = 40.104$, $c_2 = 1.4242$, $c_3 = 0.80905$, $c_4 = 19.476$, and $c_5 = 0.05429$.

$$CoV_{DLS} = c_1 \cdot \left[\log \left(\frac{1}{2} \cdot \frac{c_2 \cdot \rho_s}{h \cdot 4} \right) - \log(c_3 \cdot t_w) - \frac{c_4 \cdot h - c_5 \cdot b^3}{c_6 \cdot b^2} - c_7 \frac{h}{b} \right] + c_8 \quad (19)$$

where $c_1 = -0.05336$, $c_2 = 0.4061$, $c_3 = 1.1587$, $c_4 = 1.1275$, $c_5 = 0.6971$, $c_6 = 15.383$, $c_7 = 0.0492$, and $c_8 = -0.1983$.

$$\mu_{ULS} = \frac{c_1 \cdot \rho_w}{\log \left(c_2 \cdot t_w \right) \left(\frac{c_3 \cdot b \cdot h}{f_y} + c_4 \cdot \rho_w + c_5 \cdot h \cdot \frac{1}{2} \cdot \frac{\rho_s}{4} \right) + \frac{1}{2} \cdot \frac{c_6}{b} \cdot \frac{\rho_s}{4}} + c_7 \quad (20)$$

where $c_1 = -12.291$, $c_2 = 0.5373$, $c_3 = 0.4776$, $c_4 = 1.0914$, $c_5 = 1.0921$, $c_6 = 2.4477$, $c_7 = 0.01606$.

$$CoV_{ULS} = c_1 \log \left(c_2 \cdot t_w + c_3 \cdot \frac{t_w \cdot h}{\frac{1}{2} \cdot \frac{\rho_s}{4}} + c_4 \cdot \frac{t_w \cdot h}{\frac{1}{2} \cdot b \cdot \frac{\rho_s}{4}} + c_5 \cdot \frac{b}{t_w} \right) + c_6 \quad (21)$$

where $c_1 = 0.05716$, $c_2 = 1.6620$, $c_3 = 0.9727$, $c_4 = 1.0788$, $c_5 = 1.2587$, and $c_6 = -0.1883$.

The targets and estimations from the training set and test set are plotted in Fig. 3. The R^2 scores are all higher than 0.8, which indicate the established equations are acceptable.

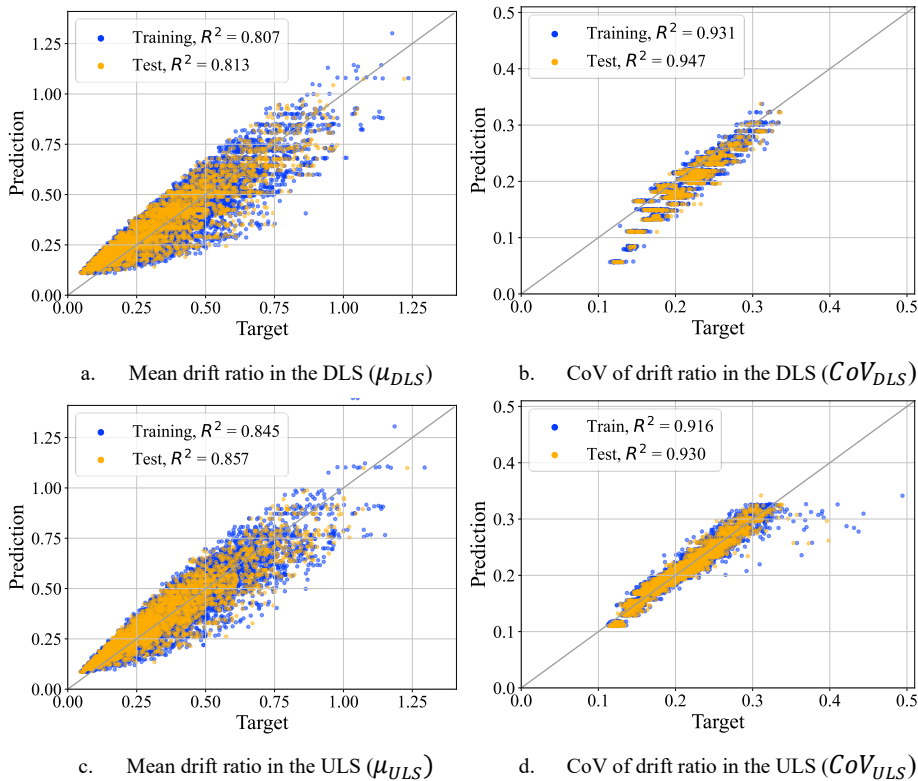


Fig. 3. Scatter plots of the regression models for predicting the mean drift ratios and their CoV .

3.3. Seismic fragility curve generation

With the classification and regression models generated via genetic programming, the main parameters for generating the fragility curve can be quickly estimated for a given set of input parameters. For instance, a set of input parameters is randomly selected from the generated dataset: $b = 2$ m, $h = 4$ m, $t_w = 0.4$ m, $f_y = 400$ MPa, $f_c = 35$ MPa, $\rho_s = 0.75\%$, $\rho_w = 0.04\%$, and $n_s = 2\%$. Two sets of μ and σ^2 , as listed in Table 3, are calculated by substituting μ_X and CoV from the generated dataset (analytical model) and the established equations (Eqs. (17), (20), and (21)) into Eq. (1). The fragility curves generated by fitting lognormal distribution with μ and σ^2 and the one directly generated from the Monte Carlo simulations are compared in Fig. 4. The 3 curves are well matched, showing that the regression model provides good accuracy with a much lower computational cost.

Table 3. Calculated μ and σ_x^2 based on μ_X and CoV from different sources

From the generated dataset				From the equations			
μ_X	CoV	μ	σ^2	μ_X	CoV	μ	σ^2
0.2642	0.2615	-1.3642	0.06614	0.2558	0.2609	-1.3963	0.06586

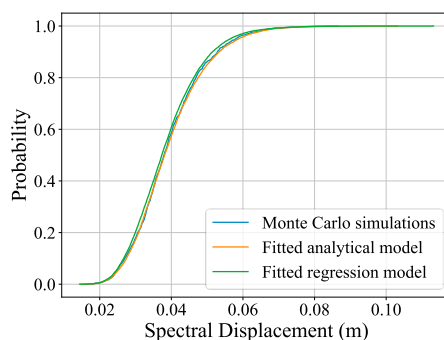


Fig. 4. Displacement-based seismic fragility curve generated from analytical model and regression models in ULS

4. Conclusions

This study proposed the preliminary framework for generating machine learning-based seismic fragility curves for RC bridge piers. The workflow includes:

1. preparing dataset via Monte Carlo simulation of analytical models,
2. training of symbolic classification and regression models via genetic programming, and
3. estimating parameters, μ and CoV through the trained models and generating seismic fragility curve.

Compared to the conventional approaches, the machine learning-based approach has a higher initial cost for the phase of dataset preparation and model training, but once the models are trained, the computational cost for generating fragility curves for bridge piers with different materials and geometry properties is much lower.

The classification and regression models demonstrate good accuracy in general. However, the accuracy of the established models is sacrificed as a trade-off to the model complexity. For future work, a more comprehensive dataset for piers with different heights will be generated, and other machine learning techniques, for example, random forest and neural network, can be applied to increase accuracy and generality.

Acknowledgements

This work was supported by ZJU-UIUC Joint Research Center Project No. DREMES202001, funded by Zhejiang University. ReLUI5 2019–2021 project, research line 4, is also acknowledged for the financial support given to the present research.

References

- Thapa, S., Shrestha, Y., and Gautam, D., 2022. Seismic fragility analysis of RC bridges in high seismic regions under horizontal and simultaneous horizontal and vertical excitations. *Structures* 37, 284–294.
- Mosleh, A., Jara, J., Razzaghi, M. S., and Varum, H., 2020. Probabilistic seismic performance analysis of RC bridges. *Journal of Earthquake Engineering* 24, no. 11, 1704–1728.
- Gautam, D., Adhikari, R., and Rupakhety, R., 2021. Seismic fragility of structural and non-structural elements of Nepali RC buildings. *Engineering Structures* 232, 111879.
- Hwang, H., and Huo, J., 1994. Generation of hazard-consistent ground motion. *Soil Dynamics and Earthquake Engineering* 13, no. 6, 377–386.
- Mander, J. B., and Basöz, N., 1999. Seismic fragility curve theory for highway bridges. *Optimizing post-earthquake lifeline system reliability*, 31–40.
- Karim, K. R., and Yamazaki, F., 2003. A simplified method of constructing fragility curves for highway bridges. *Earthquake engineering & structural dynamics* 32, no. 10, 1603–1626.
- European Committee for Standardization, 2004. EN 1992-1-1 Eurocode 2: Design of concrete structures—Part 1-1: General rules and rules for buildings. CEN, Brussels
- Wagner, S., Affenzeller, M., 2005. HeuristicLab: A Generic and Extensible Optimization Environment. In: Ribeiro, B., Albrecht, R.F., Dobnikar, A., Pearson, D.W., Steele, N.C. (Ed.). *Adaptive and Natural Computing Algorithms*. Springer, Vienna. https://doi.org/10.1007/3-211-27389-1_130