Research Article

Fragility curves accounting for uncertainties in material parameters and ground motion characteristics using a data driven surrogate model

Supratik Bose [ID], **Andreas Stavridis, Panagiotis Ch. Anastasopoulos, Kallol Sett**

Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY 14260, USA.

Correspondence to: Dr. Supratik Bose, Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY 14260, USA. E-mail: supratik@buffalo.edu

Abstract

This study uses a statistical surrogate model to develop fragility curves for an infilled reinforced concrete frame building, considering uncertainties in both material properties and ground motion parameters. The focal point of this study is a school building in Nepal damaged during the 2015 Gorkha earthquake. The school was instrumented, and its seismic response was simulated using a nonlinear numerical model. The model, developed following a recently proposed framework and extensively validated with the field data, is used in a parametric study conducted to identify the most influential material parameters (MPs). The model is then used in incremental dynamic analyses conducted to provide data for the calibration of a surrogate model. The three-staged least square statistical modeling approach is adopted to relate the influential MPs and ground motion intensity measures with important response quantities related to the peak and residual first-story drift ratios. The surrogate model is employed to generate fragility curves accounting for the two sources of uncertainty. The results indicate that accounting for uncertainties associated with the MPs can alter the fragility curves, causing a shift in the prediction of the median and dispersion of intensity measures.

Keywords: Modeling uncertainty, record-to-record uncertainty, surrogate model, fragility curves, infilled frame, three-staged least square approach

INTRODUCTION

Reinforced concrete (RC) frames with masonry infills are commonly found in seismically active regions worldwide. The infills are generally considered non-structural elements and are typically ignored in analysis and design despite their interaction with the bounding RC frame under seismic loads. In fact, their presence can drastically increase the lateral stiffness and strength of the structure, but it can lead to catastrophic brittle failure once the strength is reached. Past earthquakes have demonstrated the vulnerability of this type of structures to even moderate ground excitations, causing casualties and high economic losses[1,2]. Therefore, predicting the seismic behavior and assessing the vulnerability of these structures are necessary, yet challenging, tasks for practicing engineers. The variability in material properties and the complexities involved with the nonlinear behaviors and interactions between brick, mortar, and the bounding RC frame add to the uncertainties involved with the infilled RC frames.

The tools often used to address this challenge include limit analysis methods (among others[3,4]), strut models (among others[5-8]), and detailed finite element (FE) models (among others[9,10]). The limit analysis methods, though efficient, require the prediction of the failure pattern, which may not always be possible, given the number of possible failure patterns of actual structures. The strut models, which are commonly used in practice, may not capture the shear-dominated failure patterns and, hence, may lead to inaccurate results. The FE method, on the other hand, can provide the most accurate results and capture the failure mechanisms. However, an FE model is computationally expensive and time-consuming for large and complex buildings. Nowadays, an emerging tool to assess the seismic vulnerability of structures is the use of surrogate models (among others[11-13]). Such models, once calibrated, can account for the important structural features and excitation characteristics, and efficiently predict key features of the structural response.

Besides the numerical or statistical models, a significant effort has been made recently to develop fragility curves for classes of buildings and other structures. These curves can provide the probability of a structure reaching or exceeding a limit state or level of response when a ground motion parameter reaches a certain value. This is a quick and efficient tool for the prediction of the state of a structure as a function of the intensity of a potential earthquake. The accuracy of fragility curves depends on the information used to derive them, as they can be derived based on test, field, and/or numerical data (among others[14-20]). In most cases, these curves are developed based on numerical analyses of the behavior of a structure under different seismic excitations. Hence, they account for the uncertainty of the potential ground motion, but the possible inaccuracies and uncertainties related to the numerical models used are typically not accounted for.

To account for the modeling uncertainties, Narinder *et al.* propose a computationally efficient machine learning (ML)-based surrogate model framework for seismic fragility prediction of RC bridge piers[11]. Liel *et al.* assess the seismic collapse risk of buildings incorporating modeling uncertainties using a simplified procedure that combines aspects of the response surface and first-order second moment (FOSM) methods[16]. These studies generate surrogate models that can predict the mean and standard deviation of the fragility functions without predicting the structural response. Other studies[17-19] develop ML-based surrogate functions to predict structural response that can rapidly generate response statistics, along with fragility curves. However, these studies typically use linear models and/or 2D frames to implement their frameworks. Hence, there is a lack of studies implementing ML-based frameworks to predict the nonlinear seismic response of actual buildings.

The objective of this study is to develop fragility curves for a school building in Nepal consisting of infilled RC frames, accounting for the uncertainties associated with the material parameters (MPs) and the ground

motion records. For that purpose, an extensively validated nonlinear numerical model of the structure is first used in a sensitivity study to identify the influential model parameters[21]. In a second parametric study, incremental dynamic analyses are conducted to estimate the influence of the MPs and ground motion characteristics on the seismic response of the building. The generated data is used for the calibration of a surrogate model using the three-staged least square (3SLS) approach. The surrogate model is cross validated and then used in Monte Carlo (MC) simulations, which facilitate the development of fragility curves accounting for the uncertainties in the modeling parameters and the ground motion characteristics.

CASE STUDY BUILDING

This study focuses on the four-story school building shown in Figure 1, which is located in Sankhu, Nepal. The building consists of a masonry-infilled RC frame with seven bays in one direction and two bays in the perpendicular direction. It suffered extensive damage in the infills and RC members during the 2015 Gorkha earthquake. The damage in the first story included the shear failure of a beam-column joint in SE corner of the building, dominant cracks and spalling of the RC columns, and horizontal and diagonal cracks in the infills, while no damage was observed in the upper stories. The damaged columns revealed inadequate reinforcement detailing, with stirrups spaced at distances larger than the column dimensions. The concentration of damage in the south end of the ground story indicates that the frame exhibited a torsional response to the ground excitation. More information on the design details and the observed damage can be found in the study by Bose *et al.*[22].

DEVELOPMENT OF FINITE ELEMENT MODEL

The numerical model of the school building is developed based on the methodology proposed by Stavridis *et al.*[4] and Martin and Stavridis[23], which is adopted by ASCE 41-17[24]. According to this methodology, the relative stiffness of the infill and frame is used to classify the infill as strong or weak, while the shear and flexural strengths of the RC columns are used to classify the frame as ductile or non-ductile. The classification of the expected failure patterns allows the estimation of the lateral force *vs.* displacement envelop curve for each bay using analytical equations. The obtained backbone curves are used to calibrate the struts so that when added to the model of the bare single-bay RC frame, the combined response matches the analytically obtained backbone curve. The accuracy of this modeling methodology has been validated with data from actual buildings and large-scale test structures[21,22,25].

The numerical model of the building, consisting of 428 elements, is developed in OpenSEES[26] and utilizes displacement-based inelastic beam-column elements[27] for the RC members and diagonal truss elements for the struts. The RC members are divided into fibers to simulate the potential development of plastic hinges. The steel reinforcement is represented by the uniaxial material model proposed by Menegotto-Pinto[28] and extended by Filippou *et al.* to include isotropic strain hardening effects[29]. The Kent-Scott-Park material model with linear tension softening[30] is used for concrete and masonry, as it can represent the nonlinear behavior of quasi-brittle materials.

The comparison with the observations and data from the field indicates that the model can accurately simulate the torsional response of the actual structure and the concentration of damage in the south end of the first story. Moreover, the modal frequencies and shapes for the first three modes of the numerical model, after it is subjected to the nearby-recorded horizontal ground motions, match those extracted from the ambient vibration recordings obtained during the reconnaissance trip. More details on the numerical model and the calibrated properties can be found in the study by Bose *et al.*[22].
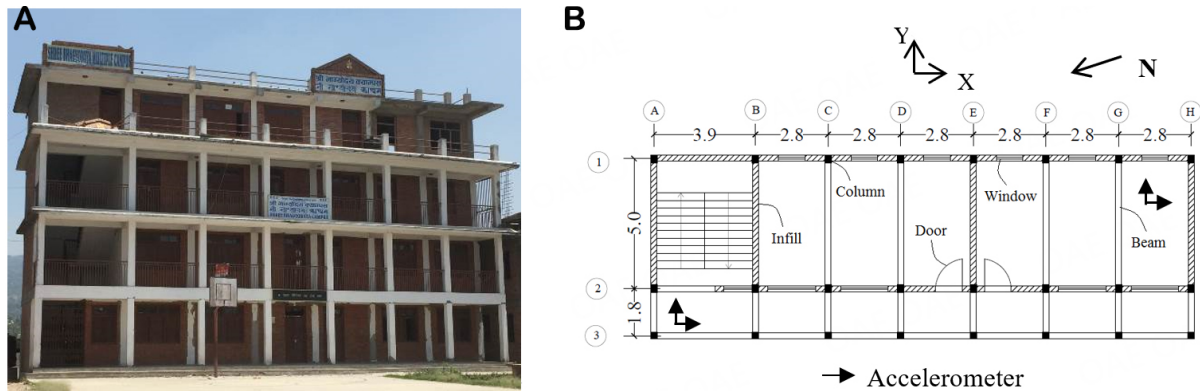
Figure 1. A four-story school building at Sankhu. (A) west elevation view from the entrance; (B) plan view (dimensions in m).

DEVELOPMENT OF DATA-DRIVEN SURROGATE MODEL

The validated FE model is used here to train a statistical surrogate model to efficiently represent the seismic response of the building. To this end, a sensitivity analysis is conducted with the FE model to investigate the effects of MP selection, as well as the ground motion parameters on the seismic response of the structure.

Material parameters

The material model adopted here for the RC members and diagonal masonry struts is defined by seven parameters for concrete and masonry: the peak compressive strengths (f_c', f_m'); the strains at the peak strength (ε_{1c}, ε_{1m}); the residual strengths (f_{rc}, f_{rm}); the strains at the onset of the residual strength (ε_{2c}, ε_{2m}); the parameters lambda (λ_c, λ_m) defining the ratio of the unloading stiffness at the onset of the residual strength to the initial stiffness; the tensile strengths (f_{tc}, f_{tm}); and the tension softening stiffnesses (E_{tsc}, E_{tsm}). The damping ratio, ξ, of the structure is numerically modeled using the mathematically convenient Rayleigh formulation, considering 3% of the critical damping for the first two modes. To incorporate the effect of the uncertainty associated with damping, it is also considered as one of the variables.

Sensitivity analyses are conducted to quantify the effect of each of these 15 parameters on the structural response. In the sensitivity study, one parameter is varied at a time; once to a lower and once to a higher value. The variation range of each variable is selected to reflect the level of confidence in the selected value. The strength, deformation, and stiffness parameters are perturbed by 20%, 30%, and 40%, respectively, from the calibrated values of the numerical model, as those values are selected based on information available from material tests in Nepal[31,32]. However, no information is available on the calibration of the parameter lambda (λ_c, λ_m) and the damping ratio; therefore, these are varied by 80% of the calibrated values to investigate their effects on the structural response. The calibrated properties, along with their percentage variations in the sensitivity analysis, are presented in Table 1. A total of 30 nonlinear models are created and, in each case, a nonlinear model is subjected to the two components of the ground motion recorded at the Univ Grants Comm., Sanothimi, Bhaktapur (THM) station, the closest station to the school building[33,34]. The modeling uncertainties associated with the parameters used to represent reinforcing steel are neglected in this study, as they are not expected to vary considerably.

The results of the sensitivity analysis are summarized in Figure 2. The figure demonstrates the variability in the peak first-story drift values in the two orthogonal directions for the response history analyses compared to the values obtained from the baseline model. The first story drift is selected for this comparison, as it is the most critical engineering demand parameter for these structures[10]. The results indicate that the

Table 1. Calibrated properties of the baseline numerical model[17]

Panels	Peak compressive strength	Residual strength	Tensile strength	Strain at peak strength	Strain at residual strength	Tension softening stiffness	Lambda*
	MPa (ksi)	MPa (ksi)	MPa (ksi)	-		MPa (ksi)	
% Variation	20	20	20	30	30	40	80
Concrete	9.71 (1.40)	0.97 (0.14)	1.38 (0.20)	0.0030	0.0080	276 (40)	0.1
Panel A	3.44 (0.50)	0.35 (0.05)	0.55 (0.08)	0.0020	0.0040	117 (17)	0.1
Panel B			0.55 (0.08)	0.0017	0.0042	117 (17)	
Panel C			1.72 (0.25)	0.0012	0.0025	345 (50)	
Panel D			1.72 (0.25)	0.0012	0.0025	345 (50)	
Panel E			1.72 (0.25)	0.0012	0.0024	345 (50)	
Panel F			0.55 (0.08)	0.0016	0.0029	117 (17)	
Panel G			0.55 (0.08)	0.0016	0.0029	117 (17)	

*Ratio of the unloading slope to the initial slope at the onset of the residual strength.

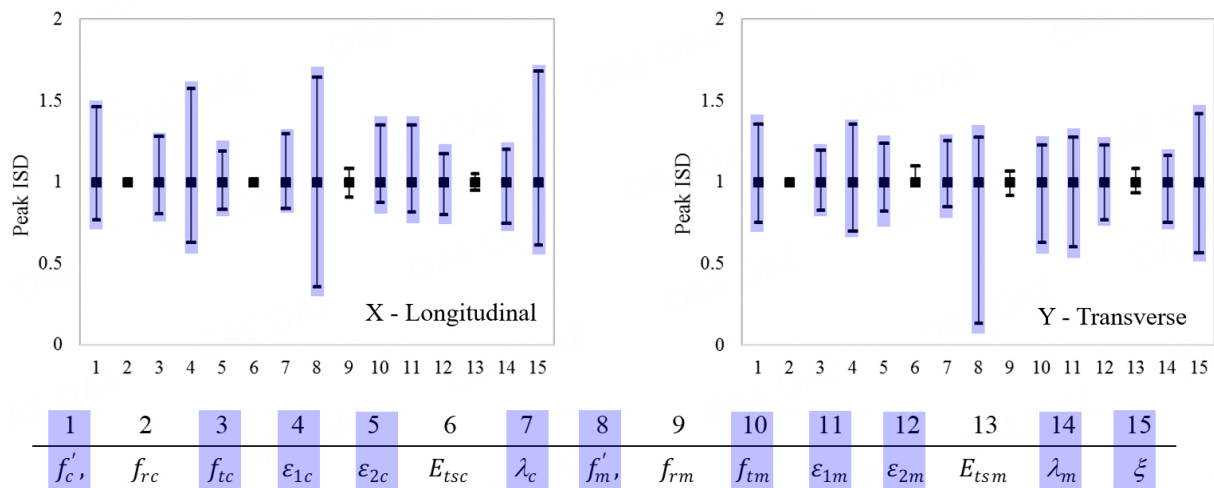| 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
f_c',	f_{rc}	f_{tc}	ε_{1c}	ε_{2c}	E_{tsc}	λ_c	f_m',	f_{rm}	f_{tm}	ε_{1m}	ε_{2m}	E_{tsm}	λ_m	ξ

Figure 2. Variations in the peak ISD model in two orthogonal directions as compared to the values obtained from the calibrated model during the sensitivity analysis.

response is sensitive to eleven parameters, which are shaded in Figure 2. These parameters are considered as input variables in the surrogate model developed in this study.

Ground motion selection and parameters

To incorporate the ground motion uncertainties, the bi-directional ground motion time-histories from the mainshock and the three major aftershocks of the 2015 Gorkha earthquake[28,29] recorded at five stations in the Kathmandu valley [KATNP (Kanti Path, Kathmandu, Nepal), PTN (Pulchowk Campus, Tribhuvan Univ., Patan), THM, KTP (Municipality Office, Kirtipur), and TVU (Dept. Geology, Tribhuvan Univ., Kirtipur)] are considered to obtain a set of 20 ground motions. Due to the lack of other recordings from Nepal, FEMA P695[35] is used to obtain additional motions in this study. FEMA P695 provides 22 far-field pairs of ground motions representing magnitudes in the range of 6.5-7.6 recorded on firm soil. Five of these 22 pairs of recorded motions and five from the set of 20 ground motions recorded in Nepal, summarized in Table 2, are employed here to run dynamic analyses to obtain data for the calibration of the surrogate model. These ground motions are selected to represent different soil types, frequency contents, amplitudes, *etc.*

Table 2. Ground motions selected in the study

EQ index	Earthquake			Recording station	Epicentral distance (km)	Vs_30 (m/s)	PGA (g)	Scale factor
	M_w	Year	Name					
1	7.9	2015	Gorkha	Municipality Office, Kirtipur (KTP)	75.8	-	0.260	0.24
2				Dept. Geology, Tribhuvan Univ., Kirtipur (TVU)	77.1	-	0.234	0.18
3				Kanti Path, Kathmandu, Nepal (KATNP)	59.9	305	0.163	0.26
4				Pulchowk Campus, Tribhuvan Univ., Patan (PTN)	79.3	-	0.154	0.30
5				Univ Grants Comm., Sanothimi, Bhaktapur (THM)	83.7	205	0.154	0.18
6	6.7	1994	Northridge	Beverly Hills - Mulhol	13.3	356	0.516	0.07
7	7.1	1999	Duzce, Turkey	Bolu	41.3	326	0.822	0.08
8	6.5	1979	Imperial Valley	Delta	33.7	275	0.351	0.17
9	7.3	1992	Landers	Yermo Fire Station	86.0	354	0.245	0.20
10	7.6	1999	Chi-Chi, Taiwan	CHY101	32.0	259	0.440	0.11
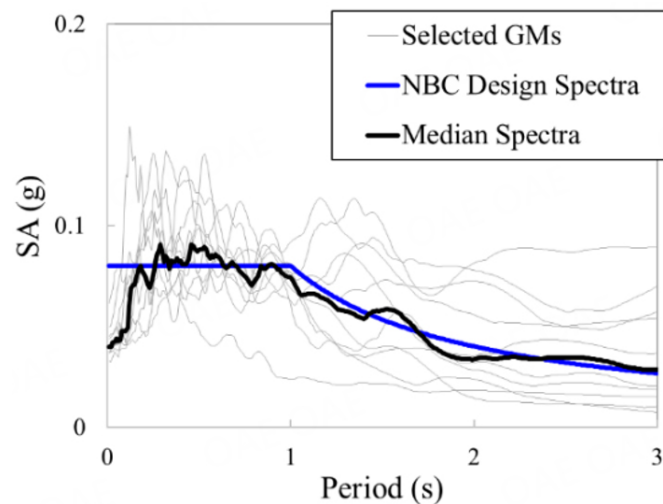
Figure 3. Acceleration response spectra of the selected ground motions.

The individual ground motions are first normalized by their peak ground velocities to remove unwarranted variability between records due to inherent differences in event magnitude, distance to the source, source type, and site conditions, while still maintaining the inherent record-to-record variability. To achieve statistical stability, the records are scaled to minimize the difference with the response spectrum used in Nepal (NBC 1994). The selected scale factors also ensure that the median of the geometric mean response spectra of all ground motions matches the response spectrum used in Nepal. Figure 3 shows the comparison between the NBC 1994 design spectrum and the median spectrum of the ten selected ground motions, scaled to the design hazard level considered in Nepal.

The selected records are used here to conduct incremental dynamic analysis (IDA)[36] to correlate a performance indicator (PI), such as the peak inter-story drift (ISD), with an intensity measure (IM) of the seismic excitation. The peak first story drift ratios in the two orthogonal dimensions are selected as PIs, as discussed in the sensitivity study. A number of IMs have been used in the literature[37-39] besides the

commonly used peak ground acceleration (PGA). Each of these IMs considers different characteristics of the ground motion with varying levels of influence on the structural response. Anastasopoulos *et al.* in their study on the response of bridge piers concluded that it is not possible to estimate the structural response satisfactorily using a single IM[40]. Hence, a series of statistically significant IMs are considered in this study, including the following. In all cases, the IMs are considered for the ground motions in both the orthogonal directions.

• the PGA, peak ground velocity (PGV), and peak ground displacement (PGD);

• the Arias Intensity, I_A, proportional to the integral of the squared acceleration *A(t)* time history defined according to Equation 1;

$$I_A = \frac{\pi}{2g} \int A^2(t)dt \tag{1}$$

• the spectral IMs in terms of acceleration (S_a), velocity (S_v), and displacement (S_d);

• the pseudo-spectral IMs in terms of acceleration (PS_a), velocity (PS_v), and displacement (PS_d);

• the Housner Intensity, I_H, integral of the pseudo-velocity response spectrum over the period range of 0.1 to 2.5 s[41] according to Equation 2;

$$I_H = \int_{0.1}^{2.5} PS_v(T, \xi = 5\%)dT \tag{2}$$

• the root-mean-square measures in terms of acceleration (A_{RMS}), velocity (V_{RMS}), and displacement (D_{RMS}) as the square roots of the mean values over the duration of the ground motion record;

• the characteristic Intensity, I_C, defined according to Equation 3;

$$I_C = (A_{RMS})^{3/2} \sqrt{D_{sig}} \tag{3}$$

• the sustained maximum measures in terms of acceleration (*SMA*) and velocity (*SMV*) defined as the third highest absolute peaks in the time histories[42];

• the spectrum IMs in terms of acceleration (*ASI*) and velocity (*VSI*), defined according to Equation 4 as the integral of the spectral acceleration (S_a) and velocity (S_v), respectively, over the period range of 0.1 to 0.5 s[43];

$$ASI = \int_{0.1}^{0.5} S_a(T, \xi = 5\%)dT, VSI = \int_{0.1}^{0.5} S_v(T, \xi = 5\%)dT \tag{4}$$

• the acceleration parameter, A95, defined as the level of acceleration which contains up to 95% of the Arias intensity[44];

• the strong motion duration or significant duration, D_{sig}, defined as the interval of time between the accumulations of 5% and 95% of Arias intensity[45];

• the Predominant Period, T_p, estimated using the 5% damped acceleration response spectrum at its maximum, as long as $T_p > 0.2$ s[46];

• the mean Period, T_{mean}, obtained from the Fourier amplitude spectrum, C_i for each frequency f_i within the range of 0.25 to 20 Hz, according to Equation 5[47];

$$T_{mean} = \frac{\sum(c_i^2/f_i)}{\sum c_i^2} \tag{5}$$

• the effective peak acceleration (*EPA*), defined as the mean of the spectral acceleration over the period range of 0.1 to 2.5 s according to Equation 6[48];

$$EPA = \frac{1}{2.5}\int_{0.1}^{2.5} S_a(T, \xi = 5\%)dT \tag{6}$$

• the spectral acceleration measure, S^*, defined in terms of pseudo-spectral accelerations $PS_a(T_1)$ and $PS_a(T_2)$, at the first and second mode periods, respectively, according to Equation 7[49];

$$S^* = PS_a(T_1)\left(\frac{PS_a(T_2)}{PS_a(T_1)}\right)^{0.5} \tag{7}$$

• the cumulative Absolute Velocity, *CAV*, defined in terms of the minimum and maximum accelerations according to Equation 8[50].

$$CAV = \sum_{i=1}^{N} H(PGA_i - A_{min})\int_{t_i}^{t_{i+1}} |A(t)|dt \tag{8}$$

where N is the number of 1-second windows in the time history, PGA_i and t_i are the maximum recorded acceleration (g) and start time of the window i, respectively, A_{min} is an acceleration threshold (user-defined, taken as 0.025 g typically) to exclude low-amplitude motions, and $H(x)$ is the Heaviside step function;

• the standardized CAV (*SCAV*), defined in terms of the CAV according to Equation 9[50];

$$SCAV = CAV_i + \int_{t_i}^{t_{i+1}} |A(t)|dt \tag{9}$$

• the normalized energy density (E_D), defined in terms of the velocity time history according to Equation 10[43];

$$E_D = \int V^2(t)dt \tag{10}$$

• the IM, I_z defined in terms of the acceleration history and *PGA*, *PGV* (Cosenza and Manfredi, 1998) according to Equation 11[46];

$$I_Z = \frac{\int A^2(t)dt}{PGA \times PGV} \tag{11}$$

• the uniform duration, D_U, the summation of the time windows where the ground acceleration is above a certain threshold, which is considered 0.05 g in this study[44];

• the number of peaks, defined as the number of times the acceleration time history crosses threshold acceleration values -0.025 g (N_{p025}), 0.05 g (N_{p050}), and 0.1 g (N_{p100}).

Statistical modeling

Nonlinear regression equations are developed herein to predict the seismic response of the four-story school building with functions of the material properties and ground motion characteristics. To this end, the 3SLS statistical modeling approach used by Anastasopoulos *et al.* to examine the effectiveness of various IMs in predicting the structural response of a bridge pier is adopted[40]. This method correlates the selected damage indices with a number of statistically important IMs and material properties. It also takes into account the errors associated with bias due to omissions of variables and the regression properties, such as heteroscedasticity, autocorrelation, and exogeneity of the regressors, among others[51].

Overview of the statistical modeling approach

In nonlinear regression analysis with multiple output parameters (PIs in this case), it is necessary to allow for cross-equation error correlation and endogeneity across the PIs that serve as dependent variables. The PIs would possibly be affected by unobserved characteristics that may be highly correlated with each other; therefore, cross-equation error correlation is anticipated. In addition, the endogeneity is underlying as the PIs are expected to be directly or indirectly affected by at least one or more of the other PIs. The 3SLS approach adopted here simultaneously accounts for both cross-equation error correlation and endogeneity across the PIs. More information on this econometric modeling approach can be found in the literature[40,52].

When the dependent variables (PIs) are endogenous, one PI is defined by a set of explanatory variables (IMs or MPs) that influence the variation of another PI and so on, and then the two-stage least squares (2SLS) approach is used for parameter estimation of the equations simultaneously. Nevertheless, the dependent variables (PIs) are influenced by similar unobserved factors because of which the random error terms are correlated. In those cases, the parameters are obtained using the seemingly unrelated regression equation (SURE) approach to account for the cross-equation error correlation. Thus, when the damage indices are endogenous with cross-correlated error terms, as in this study, the 3SLS approach, which combines the 2SLS and SURE methods, can be used to estimate the parameters of the equations simultaneously[40].

A number of response quantities can be used to represent the response of the structure. Since the damage in these structures tends to concentrate in the first story (Stavridis and Shing; Bose *et al.*,)[10,21,22], the first story drift ratio is considered in this study, as discussed earlier. More specifically, the peak and the residual ISDs along the X and Y directions, i.e., $PISD_X$, $PISD_Y$, $RISD_X$, and $RISD_Y$, where *P* and *R* stand for the peak and residual values, respectively, are selected here. Note that the absolute values of the PIs are considered here as the direction of the story displacement is not important. These are treated as dependent variables and are modeled using the 3SLS approach. Hence, the system uses one equation to estimate each of these variables. The PIs are related to one another, especially the peak and residual drifts along each direction, and

therefore, the parameters of the regression equations are estimated simultaneously so that the relations between the dependent variables can be captured accurately. Mathematically, the PIs are represented by the system of equations as follows:

$$PISD_x = f(IMs, MPs) + \varepsilon_1,$$
$$PISD_y = f(IMs, MPs) + \varepsilon_2,$$
$$RISD_x = f(IMs, MPs, PISD_x) + \varepsilon_3, \text{ and}$$
$$RISD_y = f(IMs, MPs, PISD_y) + \varepsilon_4$$

(12)

where IMs and MPs are the explanatory variables in the equations to define the selected PIs, and ε_1 to ε_4 are the correlated error terms.

The parameters of the simultaneous equations are estimated using the least square approach in Stage 1, where the instrumental variables are used as regressors to project the dependent variable. The correlation matrix of the error terms is calculated using the residuals of each equation in Stage 2. The projected values of the dependent endogenous variables, i.e., $PISD_x$ and $PISD_y$, from Stage 1 also replace the endogenous variables on the right-hand side of the equations for $RISD_x$ and $RISD_y$. Consequently, in Stage 3, the correlations of the error terms are used to improve the parameter estimates from Stage 2. Since the PIs can take only positive values, exponential relations are preferred between the dependent variables and the regressors (IMs and MPs) to avoid invalid negative values of the PIs. In addition, this exponential transformation further improves the statistical fit with increased forecasting accuracy[53]. The relations between the PIs and the instrumental variables are nonlinear; therefore, several transformations (power forms, logarithmic relationships, *etc.*) are considered to obtain better statistical fits. Dummy variables are also created for each earthquake (E1 to E10) to account for the fixed effects of the panel data, which are generated as a result of the earthquake scaling.

Model estimation using 3SLS
To train the surrogate model, the results of IDAs performed using the numerical model of the school building are used. In these analyses, the modeling parameters and ground motions are varied. The eleven sensitive MPs for concrete and masonry identified in Figure 2, i.e., the peak compressive strengths (f_c', f_m'), the tensile strengths (f_{tc}, f_{tm}), the strains at the peak strength (ε_{1c}, ε_{1m}), the strains at the onset of the residual strength (ε_{2c}, ε_{2m}), the lambdas (λ_c, λ_m) controlling the unloading for concrete and masonry, and the damping coefficient (ξ), are considered here. All eleven model parameters are varied, considering four values for each of them to develop 44 numerical models in total. The model parameters considered here are presented in Table 3. These parameters are varied one at a time in the parametric study. Hence, the possible correlations between them are neglected to limit the number of dynamic analyses and reduce the computation time. Moreover, there are no established relations to reliably describe the intra-dependencies of these parameters. The ten selected ground motions are scaled with a scale factor ranging from 0.4 to 4.0 at an increment of 0.4, resulting in ten levels of ground motions. The consideration of 44 models, ten ground motions, and ten excitation levels results in 4,400 dynamic analyses.

The results of the 4,400 dynamic analyses are used to train the surrogate model. The criterion for the inclusion of the input parameters in the models is their statistical significance and the statistically significant improvement of the overall model fit. In the calibration process, it was found that disregarding the variables with a level of confidence lower than 90% has no effect on the R-squared (R^2) value, which is used as a statistical measure of goodness of fit. Hence, a threshold of 90% level of confidence is selected here. The R^2 value is adjusted for the number of predictors in the model. The adjusted R^2 values are greater than 0.90 for

Table 3. Model parameters considered to develop the numerical models to fit the surrogate model

Model parameters	Variation (%)	Value-1	Value-2	Calibrated value	Value-3	Value-4
f_c' (ksi)	+/- 10 and 20	1.12	1.26	1.40	1.54	1.68
f_{tc} (ksi)	+/- 10 and 20	0.16	0.18	0.20	0.22	0.24
ε_{1c}	+/- 15 and 30	0.0021	0.00255	0.003	0.00345	0.0039
ε_{2c}	+/- 15 and 30	0.0056	0.0068	0.008	0.0092	0.00104
λ_c	x1/4, 1/2, 3.75, and 7.5	0.025	0.05	0.1	0.375	0.75
f_m' (ksi)	+/- 10 and 20	0.40	0.45	0.50	0.55	0.60
f_{tm} (ksi)	+/- 10 and 20	0.066	0.075	0.083	0.091	0.10
ε_{1m}	+/- 15 and 30	0.0014	0.0017	0.0020	0.0023	0.0026
ε_{2m}	+/- 15 and 30	0.0028	0.0034	0.0040	0.0046	0.0052
λ_m	x1/4, 1/2, 3.75, and 7.5	0.025	0.05	0.1	0.375	0.75
ξ	+/ 33 and 67	0.01	0.02	0.03	0.04	0.05

the four PIs, indicating good overall statistical fits, as the models account for at least 90% of the variance in the data. Tables 4 and 5 present the estimation results of the 3SLS surrogate models for the peak and residual ISDs along the orthogonal directions of the school building. In the tables, it can be observed that the signs of the coefficients might vary between equations in the 3SLS model. This indicates that the effect of an independent variable on the dependent variable could be affected by the inclusion of other independent variables, as the coefficients are calculated simultaneously through the system of equations. In addition, it can be observed that the interstory drift ratios along one direction can be dependent on the IMs along the other direction, as they are statistically significant with a high level of confidence.

Assessment of the surrogate model

To assess its accuracy, the statistical model is used to estimate the PIs for the models and ground motions used for its training. Figures 4 and 5 compare the interstory drift values obtained from the detailed nonlinear time history analyses and those estimated using the surrogate models. The comparisons indicate reasonably good matches between the statistical and the FE models. The fits are better for the higher drift levels, which are of interest when structural damage and/or collapse are estimated. This is particularly encouraging as the primary focus in this study is the estimation of the collapse potential of the structure, which occurs at larger drifts.

To further assess the confidence in the surrogate model, it is used to predict the dynamic responses of numerical models with all 11 MP values different than those presented in Table 3 used in the training process. These models are also subjected to ground motions not considered in the development of the surrogate model. These "unseen" ground motions are taken randomly from the set of 42 ground motions, which comprises the far-field suite of FEMA P695 and motions from the 2015 Gorkha earthquake, with the exclusion of those used to train the surrogate models. The performance of the surrogate model is evaluated for ten such cases using five "unseen" ground motions and two sets of "unseen" MPs. The two sets of MPs obtained by estimating the average of Values 1 and 2 and the average of Values 3 and 4 in Table 3, respectively for all MPs. For example, the values used for f_c' are 1.19 ksi and 1.61 ksi. The five "unseen" ground motions used here are presented in Table 6. The results of these analyses, marked as blue stars in Figures 4 and 5, are rather encouraging, considering the complex nonlinear behavior of the structure.

MONTE CARLO SIMULATIONS

The statistical models for the school building are used here to perform additional simulations needed to develop fragility curves incorporating modeling uncertainties. The MPs of the models for these simulations are selected with the MC method.

Table 4. Model estimation results for the peak ISDs of the building using the 3SLS approach

Variables	Coefficient	*t*-stat	Variables	Coefficient	*t*-stat
	Dependent variable = $\ln(PISD_x)$			Dependent variable = $\ln(PISD_y)$	
Constant	12.4621	6.17	$\ln(PGA_2)$	1.10084	8.96
PGA_1	-18.2248	-11.02	$(PGA_2)^2$	40.0635	10.18
$\ln(PGA_1)$	2.0090	10.57	$(PGV_1)^2$	0.00016	5.20
PGA_2	31.4898	9.49	$(PGV_2)^2$	0.00072	12.26
$\ln(PGA_2)$	3.49719	6.46	PGD_1	0.00832	11.68
$(PGA_2)^2$	46.6106	1.29	$(PGD_1)^2$	-2.98E-05	-7.19
PGV_1	-0.07481	-8.83	PGD_2	-0.01914	-11.62
$\ln(PGV_1)$	-3.63489	-1.62	$1/PGD_2$	0.91117	9.71
$(PGV_1)^2$	0.00073	9.25	$\ln(PGD_2)$	0.49799	4.07
$\ln(PGV_2)$	2.49558	7.83	I_{A1}	-0.02132	-5.50
PGD_1	0.01837	14.77	$(I_{A1})^2$	1.74E-05	5.85
$(PGD_1)^2$	-2.85E-05	-8.40	$\ln(I_{A2})$	-0.43301	-4.34
PGD_2	0.00755	4.19	$(S_{a1})^2$	-11.9568	-6.11
$(PGD_2)^2$	-1.03393	-8.52	S_{a2}	9.73939	19.46
$\ln(PGD_2)$	-1.59E-05	-2.92	I_{C1}	0.00694	13.54
I_{A1}	0.05540	3.30	$(I_{C1})^2$	-4.37E-07	-8.81
$(I_{A1})^2$	2.65080	11.14	I_{C2}	-0.00534	-7.65
$\ln(I_{A1})$	-2.18E-05	-6.78	D_{sig2}	0.00032	6.50
I_{A2}	-0.07171	-12.84	N_{p0252}	-0.01181	-11.45
$(I_{A2})^2$	-4.10723	-11.27	$(N_{p0252})^2$	2.54E-05	4.50
$\ln(I_{A2})$	2.75E-04	8.90	$(f_c')^{1/2}$	-0.48764	-2.15
$(S_{a1})^2$	-11.1860	-11.81	ε_{1c}	123.708	5.37
I_{C1}	0.00250	6.94	f_m'	-1.76156	-11.88
$(I_{C1})^2$	-2.14E-07	-2.74	f_{tm}'	-0.99478	-9.09
T_{mean1}	4.14491	4.92	ε_{1m}	382.273	13.07
$SCAV_1$	-0.00136	-6.36	ζ	-10.2329	-9.04
N_{p0251}	0.00233	5.76	E1	1.10084	-7.86
f_c'	-0.14210	-10.18			
ε_{1c}	63.8062	2.63			
f_m'	-1.74919	-17.06			
f_{tm}	-2.20036	-6.72			
ε_{1m}	305.414	10.02			
ζ	-11.0669	-9.73			
E7	-0.26461	-2.13			
R-square = 0.9535			R-square = 0.9549		
Adjusted R-square = 0.9531			Adjusted R-square = 0.9546		

Indices 1 and 2 for the intensity measures represent the two components of the ground motions. *t*-stat is equal to the coefficient divided by the standard error and gives the level of confidence that the coefficients are statistically different from zero (Washington *et al.*, 2011)[51].

Uncertainties in model parameters

The statistical models summarized in Table 4 indicate that only six of the eleven sensitive parameters identified in Figure 2 are statistically significant: the peak compressive strengths and corresponding strains of concrete and masonry, the tensile strength of masonry, and the damping coefficient. Therefore, these are considered in the MC simulations. The probability density functions of these parameters are estimated

Table 5. Model estimation results for the residual ISDs of the building using the 3SLS approach

Variables	Coefficient	*t*-stat	Variables	Coefficient	*t*-stat
Dependent variable = ln(RISD$_x$)			Dependent variable = ln(RISD$_y$)		
Constant	-1.24975	-5.93	Constant	-9.73778	-1.28
ln(PGA$_1$)	-0.38759	-8.01	PGA$_2$	7.29838	6.02
(PGA$_1$)^2	-9.62633	-15.45	ln(PGA$_2$)	-2.22195	-13.17
PGV$_1$	-0.0181	-9.39	(PGA$_2$)^2	-22.6409	-12.47
ln(PGV$_1$)	-0.51379	-7.74	PGV$_1$	-0.0601	-6.31
ln(PGV$_2$)	0.75928	12.71	ln(PGV$_1$)	1.95624	11.23
(PGD$_1$)^2	-8.00E-05	-7.77	PGV$_2$^{2}	0.00031	4.50
I$_{A1}$	-0.00402	-4.54	ln(PGV$_2$)	0.14066	2.27
(I$_{A1}$)^2	0.87201	17.38	PGD$_1$	0.00477	4.00
ln(I$_{A1}$)	7.43E-06	6.08	ln(PGD$_2$)	-0.90549	-16.86
I$_{A2}$	0.01232	12.25	(PGD$_2$)^2	-1.94E-05	-6.62
(I$_{A2}$)^2	-1.34569	-19.98	I$_{A1}$	-0.01196	-9.79
ln(I$_{A2}$)	-6.86E-06	-4.79	(I$_{A1}$)^2	2.82E-05	12.08
(S$_{a1}$)^2	5.27862	17.21	I$_{A2}$	0.02003	15.50
f_c'	-0.40747	-19.61	(I$_{A2}$)^2	-3.41E-05	-9.52
f_{tc}	-1.35843	-4.87	(S$_{a2}$)^2	6.39184	10.86
ε_{1c}	128.111	5.05	f_c'	-0.25552	-10.56
f_m'	1.40116	15.64	f_{tc}	-1.4873	-4.67
ε_{1m}	-564.748	-12.44	ε_{1m}	-487.725	-9.35
ln(PISD$_x$)	1.17262	11.16	λ_m	0.1844	3.21
E1	-0.32877	-8.82	ζ	-3.09035	-2.20
E2	-0.46238	-10.25	ln(PISD$_y$)	1.16615	18.63
E3	1.12909	15.80	E2	1.24972	16.41
			E5	-0.35982	-8.70
			E8	0.84758	9.74
R-square = 0.9114			R-square = 0.9055		
Adjusted R-square = 0.9109			Adjusted R-square = 0.9048		

t-stat is equal to the coefficient divided by the standard error and gives the level of confidence that the coefficients are statistically different from zero (Washington *et al.*, 2011)[51].

Table 6. "Unseen" ground motions selected to evaluate the surrogate model

Earthquake			Recording station	PGA (g)
M$_w$	Year	Name		
7.3	2015	Gorkha	Municipality Office, Kirtipur (KTP)	0.069
6.7			Univ Grants Comm., Sanothimi, Bhaktapur (THM)	0.097
6.7	1994	Northridge	Canyon County - WLC	0.482
6.9	1995	Kobe	Shin-Osaka	0.243
6.9	1989	Loma Prieta	Gilroy Array #3	0.555

based on data from material tests on concrete and masonry available in the literature[54-58]. In previous studies, the distributions are typically assumed to be lognormal[20,59]; however, to remove any possible bias due to the selection of the probability density functions and investigate which distribution is more suitable for these MPs, both Weibull and lognormal distributions are fitted here, as shown in Figure 6.
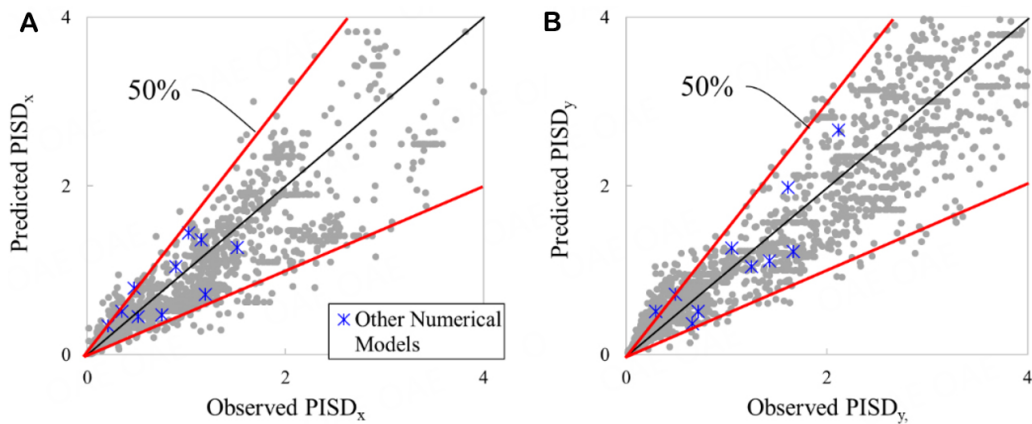
Figure 4. Observed and predicted values in terms of the peak ISD for the case-study building. (A) Peak ISD along the X direction; (B) Peak ISD along the Y direction.
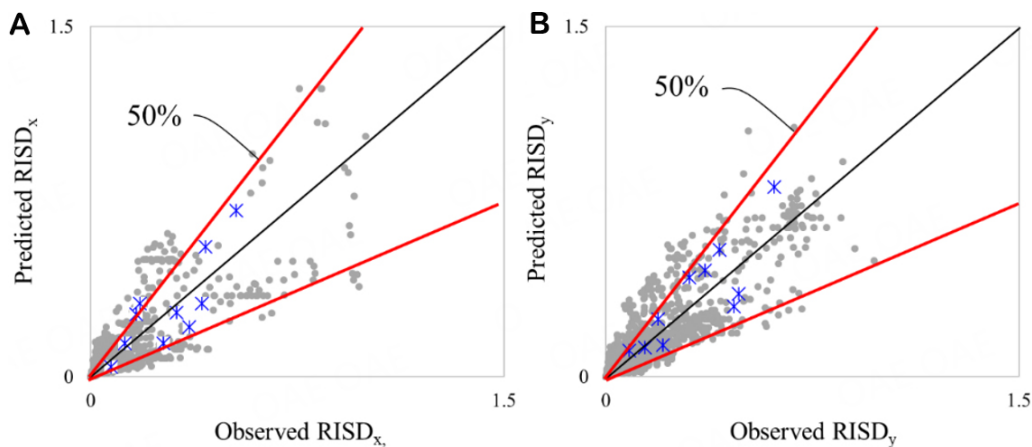

Figure 5. Observed and predicted values in terms of the residual ISD for the case-study building. (A) Residual ISD along the X direction; (B) Residual ISD along the Y direction.

It can be observed in Figure 6 that both distributions fit the test data equally well. Hence, a lognormal distribution is assumed in line with previous studies. Table 7 presents the median, e^μ, and standard deviation, σ for a lognormal distribution adopted for all the model parameters. In the case of the damping ratio, ξ, for which test data is very limited, the parameters are taken from the literature[20,54,60-62]. The standard deviation values of the datasets in Table 7 that have mean values closest to the parameters of the calibrated model summarized in Table 1 are employed here to generate samples for the MC simulations. These are indicated with bold text in the table.

Results

MC simulations are conducted here using the statistical models to investigate the effects of the uncertainties associated with the random variables (IMs and MPs) on the fragility curves. To generate the models for these analyses, the MPs are sampled N_{samp} times, generating a set of realizations consistent with the assumed lognormal distribution. Then, two approaches are considered to generate the cases analyzed with the statistical models.

Table 7. Estimated lognormal distribution statistical parameters of the experimentally obtained material properties

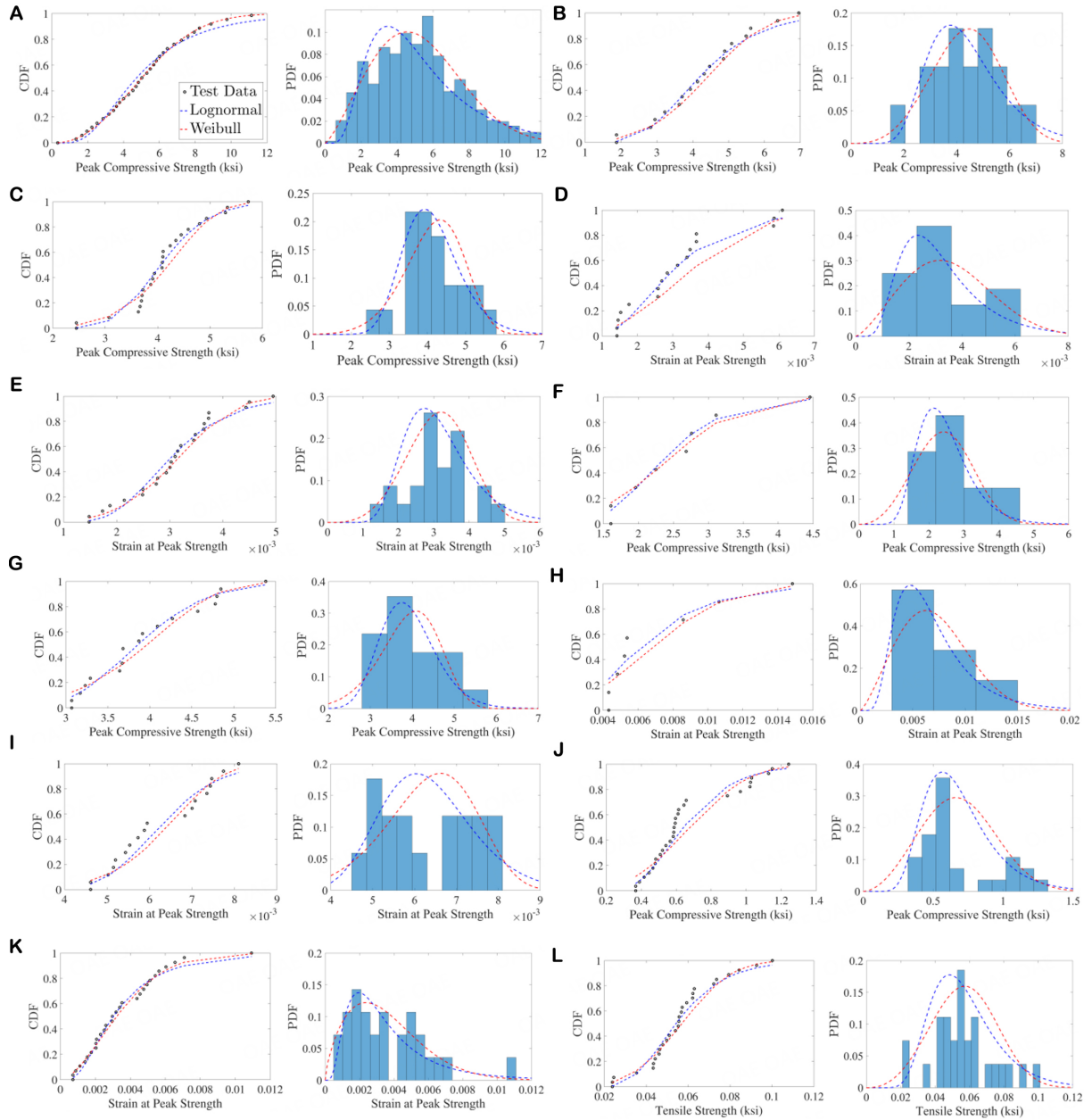|	Parameters			f_c' (ksi)	ε_{1c}	f_m' (ksi)	f_{tm} (ksi)	ε_{1m}	ξ (%)
Log normal distribution fitting parameters	Yeh, 1998[54]		Median (e^μ)	4.7	-	-	-	-	**COV = 0.4**
			Dispersion (σ)	0.55	-	-	-	-	(Porter *et al.*, 2003[62])
	Yousefianmoghadam *et al.*, 2015, 2018[25,56]		e^μ	4.2	**0.0029**	-	-	-	
			σ	0.33	**0.475**	-	-	-	
	Gao, 2021[58]		e^μ	4.1	0.003	3.9	-	0.0062	
			σ	0.18	0.3	0.17	-	0.18	
	Bose, 2013[55]		e^μ	-	-	**0.6**	**0.06**	**0.003**	
			σ	-	-	**0.3**	**0.34**	**0.6**	
	Cristofaro *et al.*, 2011[70]	Set 1	e^μ	**1.4**	-	-	-	-	
			σ	**0.45**	-	-	-	-	
		Set 2	e^μ	1.9	-	-	-	-	
			σ	0.57	-	-	-	-	
		Set 3	e^μ	2.4	-	-	-	-	
			σ	0.46	-	-	-	-	
	Schueremans and Dionys, 2006[71]	Set 1	e^μ	-	-	0.6	-	0.003	
			σ	-	-	0.23	-	0.55	
		Set 2	e^μ	-	-	0.7	-	0.0035	
			σ	-	-	0.18	-	0.4	
		Set 3	e^μ	-	-	0.9	-	-	
			σ	-	-	0.11	-	-	
	Shimizu *et al.*, 1999[72]	Set 1	e^μ	2.7	-	-	-	-	
			σ	0.3	-	-	-	-	
		Set 2	e^μ	3.4	-	-	-	-	
			σ	0.32	-	-	-	-	
		Set 3	e^μ	3.3	-	-	-	-	
			σ	0.23	-	-	-	-	
		Set 4	e^μ	1.6	-	-	-	-	
			σ	0.38	-	-	-	-	

Figure 6. Fitting cumulative and probability distributions to available test data. (A) f_c' from the lab tests in Taiwan[54]; (B) f_c' from the actual buildings in the USA[56,57]; (C) f_c' from the lab tests in UB, USA[58]; (D) ε_{1c} from the actual buildings in the USA[56,57]; (E) ε_{1c} from the lab tests in UB, USA[58]; (F) f_m' from an actual building in the USA[56]; (G) f_m' from the lab tests in UB, USA[58]; (H) ε_{1m} from an actual building in the USA[56]; (I) ε_{1m} from the lab tests in UB, USA[58]; (J) f_m' from the lab tests in IITK, India[55]; (K) ε_{1m} from the lab tests in IITK, India[55]; (L) f_{tm} from the lab tests in IITK, India[55].

Approach-1: The realizations are set up such that the first random variable is varied from 1 to N_{samp} while keeping the other random variables constant. Then, the second variable is varied from 1 to N_{samp} without changing the other variables, and so on. Thus, this approach can consider the interactions between the random variables. However, the number of models (N_{mod}) for each simulation increases exponentially as $N_{mod} = (N_{samp})^{N_{MP}}$, where N_{MP} is the number of input MPs varied.

Approach-2: With this approach, N_{samp} sets of random variables are generated, with each set representing one realization. In this case all MP values are different between the different models. Thus, solving the problem N_{samp} times gives information on the randomness in the output response of the system to each set of input variables. Thus, the number of models (N_{mod}) in this approach is equal to N_{samp}.

All the material properties considered here for developing the set of structural models are assumed to be uncorrelated. For each set of realizations of MPs, the statistical surrogate model is used to calculate the peak ISDs for the ten selected ground motions. The IMs are increased incrementally to perform the additional IDAs incorporating both modeling and record-to-record uncertainties. The geometric mean of the spectral acceleration at the first modal period of the building, $S_a[T_1]$ of the two components of each ground motion is selected as the IM, while the maximum drift ratio is selected as the PI to develop the IDA curves. The periods of the structure can differ from one model to another as some MPs can affect the stiffness; however, for consistency, the period of the calibrated model is used in all cases.

To examine the effects of modeling uncertainties at different levels of the structural behavior, distinct limit states in the form of damage grades (DGs) are considered for the case study building. For the damage classification, a critical aspect of post-earthquake damage assessment[63-65], the damage classification used by Brzev *et al.* to assess low-rise RC buildings in Nepal affected by the 2015 Gorkha earthquake is adopted here[65]. This damage classification scheme includes five DGs ranging from DG1 to DG5 based on the EMS-98 scale, to characterize the severity of damage in structural components. This damage classification is correlated to maximum drifts reached during an earthquake by using data from laboratory specimens[59,66] and by consulting with the field engineers from the National Society of Earthquake Technology (NSET), Nepal[67] who implemented it after the 2015 earthquake. As proposed by Hilly[67], the limit states, DG1 to DG5, correspond interstory drift ratios of 0.3%, 0.6%, 1.2%, 2.0%, and 3.0%, with DG5 representing severe damage and partial or total collapse of the building. The medians and standard deviations of $S_a[T_1]$ for the five DGs obtained from both approaches of MC simulations are summarized in Figures 7 and 8. It can be observed that the medians and standard deviations converge to the same values of $S_a[T_1]$, representing the five DGs for both approaches. However, the first approach needs 46,656 models for all median and standard deviation values to converge compared to the 100 analyses needed for the second approach. Therefore, the second approach is adopted here for further analyses.

To capture the uncertainties associated with record-to-record randomness but not the MPs, IDA is also performed for the baseline model and the ten ground motions considered here. The IDA curves of the deterministic model and the extended IDA plot for $N_{mod} = 100$ are presented in Figure 9.

Table 8 presents the medians and standard deviations obtained from the MC simulations, the baseline FE model, and the baseline surrogate models. For all DGs, the methods considered here result in similar trends for the median and dispersion values, thus further validating the use of the statistical models for MC simulations. However, one can observe that for all DGs, the median values of the spectral acceleration are lower in the case of the MC simulations, which account for the variability in the MPs. Moreover, one can note that the standard deviation increases by approximately 10% in the case of the surrogate model compared to the FE model. The increase is larger in the case of the MC simulations, and especially for the larger DGs. The substantially higher scatter can be expected, considering that in the case of the MC simulations, the uncertainties in the MPs are also included. This is evident in Figure 9, which demonstrates a higher scatter in the data points.

Table 8. Results of the Monte Carlo simulation compared to the base model

Damage grades	Median $S_a[T_1]$ (g)				Standard deviation $S_a[T_1]$ (g)			
	Baseline model		Monte carlo simulations		Baseline model		Monte carlo simulations	
	FE	3SLS	Approach-1	Approach-2	FE	3SLS	Approach-1	Approach-2
DG1	0.227	0.215	0.220	0.219	0.131	0.147	0.164	0.165
DG2	0.312	0.304	0.295	0.296	0.202	0.183	0.266	0.268
DG3	0.391	0.384	0.361	0.364	0.297	0.342	0.329	0.328
DG4	0.419	0.408	0.411	0.409	0.211	0.243	0.338	0.339
DG5	0.477	0.462	0.445	0.443	0.228	0.249	0.346	0.345
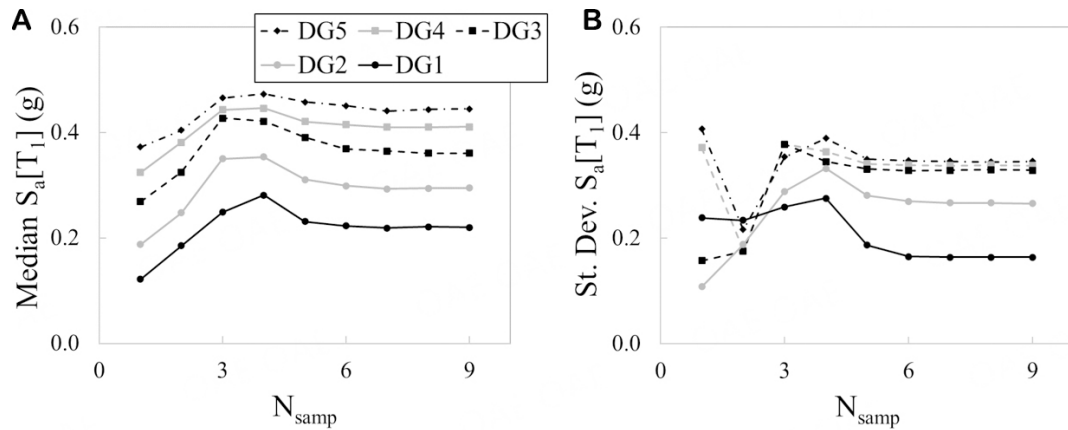
Figure 7. Medians and standard deviations of the $S_a[T_1]$ at various damage grades obtained from MCS - Approach 1. (A) Medians; (B) Standard deviations.
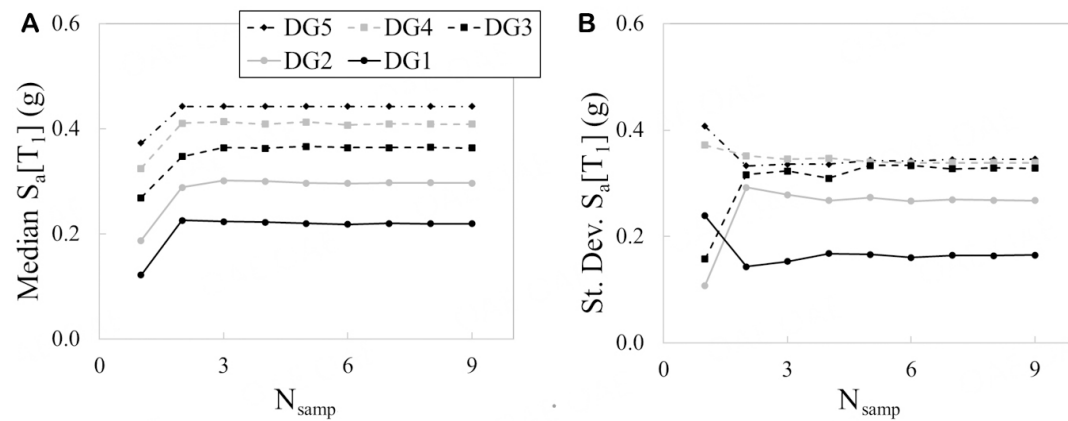
Figure 8. Medians and standard deviations of the $S_a[T_1]$ at various damage grades obtained from MCS - Approach 2. (A) Medians; (B) Standard deviations.

FRAGILITY CURVES

The fragility curves represent the expected probability of exceeding a DG as a function of an IM. In this study, the fragility curves are defined in terms of the spectral acceleration at the period of the first mode of the structure. Similar to previous studies[68,69], the fragility function parameters are obtained from the IDA plots for computing the median and standard deviation of the selected IM at each damage state, assuming a lognormal distribution.
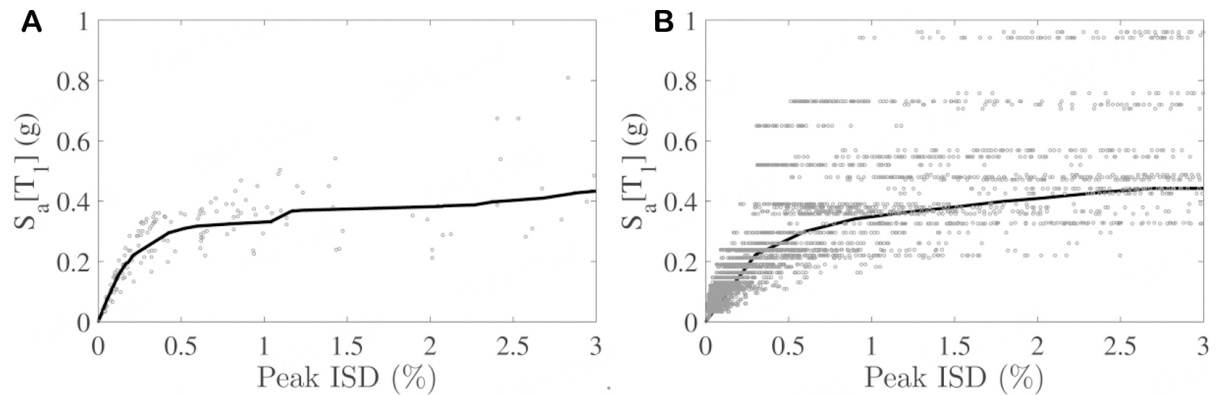
Figure 9. IDA curves for the calibrated model and from Monte Carlo Simulation using the surrogate model. (A) Baseline model; (B) Monte Carlo simulations Approach 2, N_{mod} = 100.

To assess the efficiency of this fitting approach for the fragility curves, IDA plots are generated for the baseline model using the entire set of 42 ground motions. The empirical cumulative distribution function for exceeding 1.0% drift obtained from the IDA data is compared to the fragility function fitted to the data generated by assuming lognormal distribution in Figure 10. It can be observed that the lognormal distribution provides a good estimate of the fragility data. Therefore, it is used in this study to develop the fragility curves.

Following this approach, the fragility curves are obtained for the five DGs from the MC simulations [Figure 11]. These curves incorporate the uncertainties associated with the MPs and the ground motion characteristics. Figure 11B demonstrates the effects of the uncertainties of the MPs by comparing the fragility curve of the baseline model, considering only record-to-record uncertainty with that obtained from the MC simulations for DG5. The difference between the two curves illustrates the importance of incorporating modeling uncertainties in vulnerability analysis. The baseline model, when not considering the modeling uncertainties, can be unconservative and underestimate the probability of collapse for spectral accelerations below 0.6 g.

EFFECT OF CORRELATIONS BETWEEN MATERIAL PARAMETERS

Possible correlations between the MPs are very difficult to quantify, particularly in the absence of field or test data for the material properties. However, it is important to identify the impact of the correlation assumptions on the effects of uncertainties associated with material properties. Three sets of MPs are considered here to examine the implications of correlation assumptions. In Set-1 the concrete properties are assumed to be correlated, but there are no correlations between the masonry properties. In Set-2 the masonry parameters are assumed to be correlated, while the concrete properties are not. Finally, in Set-3 all the random variables, including concrete and masonry parameters, are assumed to be correlated. The damping coefficient in all cases is assumed to have no correlation with the MPs. The correlations considered in Sets 1 and 2 are highly likely as the properties of concrete and masonry can be correlated among themselves, but the Set-3 correlation is considered here more for illustration purposes, and it does not correspond to a realistic or practical application. For each set of assumptions, the correlation coefficients are considered to be 0.2, 0.5, and 1.0 between the random variables. These correlation assumptions only affect the MC simulations for the generation of the input MPs for the statistical surrogate model. The statistical surrogate model is used here since varying the correlation assumptions using the FE model would require significant computational effort.

Table 9. Parametric study on the effects of the correlation assumptions on the fragility curves

Correlation assumptions	$S_a[T_1]$ for Set-1		$S_a[T_1]$ for set-2		$S_a[T_1]$ for set-3	
	Median	Std. dev.	Median	Std. dev.	Median	Std. dev.
0	0.443	0.346	0.443	0.346	0.443	0.346
0.2	0.443	0.339	0.444	0.339	0.443	0.340
0.5	0.444	0.338	0.444	0.336	0.445	0.331
1.0	0.445	0.338	0.446	0.331	0.449	0.327
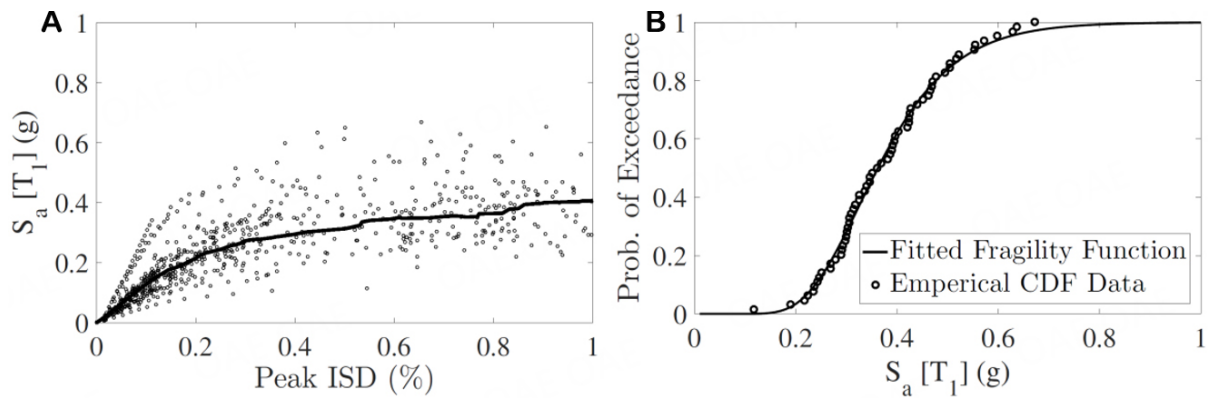
Figure 10. Incremental dynamic analysis results used to fit fragility curves assuming lognormal distribution. (A) IDA results of FE model; (B) Fitted fragility function.
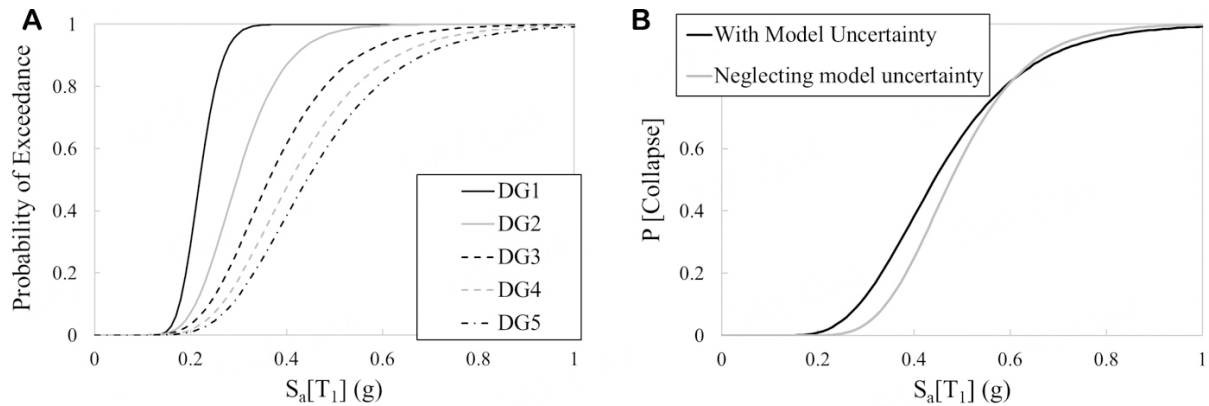

Figure 11. Computed collapse fragilities of the four-story case study building. (A) Fragility curves for all damage grades; (B) Collapse fragility for DG5.

The results for DG5 using N_{mod} = 1,000 are summarized in Table 9. It can be observed that the correlations between the parameters have no noticeable effects on the median $S_a[T_1]$, particularly for the first two sets. In fact, for Sets 1 and 2, correlation effects tend to be conservative if neglected. However, the standard deviations change slightly with the correlations between the variables.

CONCLUSIONS

In this study, a reliable data-driven surrogate model for a school building is obtained and used to generate fragility curves that consider uncertainties associated with the MPs and the ground motion characteristics. The surrogate model is developed using the 3SLS statistical approach to capture the relations among the

modeling parameters and IMs with selected structural response quantities. The 3SLS approach is employed as it accounts for two important misspecification issues when concurrently statistically modeling PIs as functions of IMs and MPs: namely, endogeneity and cross-equation error correlation. The obtained model is verified with results from nonlinear analyses not considered in the training process. The comparisons between the results obtained from the detailed nonlinear model and the surrogate statistical model indicate that the statistical model can accurately predict the first story drift ratios. Hence, it is further used in MC simulations.

The extended incremental dynamic analyses performed using the surrogate models lead to the development of fragility curves, which incorporate both the uncertainties associated with structural modeling parameters and the ground motions. Incorporating modeling uncertainties increases the dispersion and decreases the median in the response fragility. This indicates that neglecting the modeling uncertainties is unconservative and may lead to unsafe conclusions. Hence, it is important to incorporate the modeling uncertainties in performance-based earthquake engineering for the probabilistic seismic performance assessment of structures. Additional analyses indicate that the impact of correlations between the MPs on uncertainty quantification is insignificant, and it is conservative to neglect those. These valuable findings are based on the study of one school building. Additional studies on other buildings and structural systems can be conducted to further validate these findings. With the framework introduced here, alternative ML tools, possibly fused with physics-based constraints, can be considered for the prediction of the structural response if they are proven to be more accurate than the statistical surrogate model employed here.

DECLARATIONS

Authors' contributions
Conceptualization, implementation, numerical analysis, model assessment, visualization: Bose S
Conceptualization, methodology, supervision, model assessment, funding acquisition: Stavridis A
Conceptualization, model assessment: Anastasopoulos PC
Conceptualization, supervision: Sett K

Availability of data and materials
The data and materials pertinent to this paper can be provided upon request.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1. 1994 Northridge earthquake reconnaissance report, Volume 11, Issues S1 and S2. 1996. Available from: https://journals.sagepub.com/toc/eqsa/11/1_suppl [Last accessed on 20 Sep 2023].
2. Brando G, Rapone D, Spacone E, et al. Damage reconnaissance of unreinforced masonry bearing wall buildings after the 2015 Gorkha,

Nepal, Earthquake. *Earthq Spectra* 2017;33:243-73. DOI

3. Mehrabi AB, Shing PB, Schuller MP, Noland JL. Performance of masonry-infilled RC frames under in-plane lateral loads, Report CU/SR-94-6; Boulder, CO: University of Colorado, 1994. Available from: https://nehrpsearch.nist.gov/static/files/NSF/PB95174025.pdf [Last accessed on 20 Sep 2023].

4. Stavridis A, Martin-Tempestti J, Bose S. Updating the ASCE 41 provisions for infilled RC frames, 2017 SEAOC Convention, San Diego, CA, USA. 2017. Available from: https://www.researchgate.net/publication/322759831_Updating_the_ASCE_41_Provisions_for_Infilled_RC_Frames [Last accessed on 20 Sep 2023].

5. Crisafulli FJ, Carr AJ. Proposed macro-model for the analysis of infilled frame structures. *Bull New Z Soc Earthq Eng* 2007;40:69-77. DOI

6. Burton H, Deierlein G. Simulation of seismic collapse in nonductile reinforced concrete frame buildings with masonry infills. *J Struct Eng* 2014;140:A4014016. DOI

7. Sattar S, Liel AB. Seismic performance of nonductile reinforced concrete frames with masonry infill walls - I: development of a strut model enhanced by finite element models. *Earthq Spectra* 2016;32:795-818. DOI

8. Allouzi R, Irfanoglu A. Development of new nonlinear dynamic response model of reinforced concrete frames with infill walls. *Adv Struct Eng* 2018;21:2154-68. DOI

9. Al-Chaar G, Mehrabi AB, Manzouri T. Finite element interface modeling and experimental verification of masonry-infilled R/C frames. *J Masonry Soc* 2008;261:47-65. Available from: https://www.researchgate.net/publication/284415010_Finite_element_interface_modeling_and_experimental_verification_of_masonry [Last accessed on 20 Sep 2023]

10. Stavridis A, Shing PB. Finite-element modeling of nonlinear behavior of masonry-infilled RC frames. *J Struct Eng* 2010;136:285-96. DOI

11. Narinder, Goswami K, Bose S, Banerjee A, Chakraborty S. ML-based surrogate model for RC bridge pier collapse fragility prediction. In: Shrikhande M, Agarwal P, Kumar PCA, editors. In Proceedings of 17th Symposium on Earthquake Engineering. Singapore: Springer Nature; 2023. Volume 2, pp. 661-71. Available from: https://link.springer.com/chapter/10.1007/978-981-99-1604-7_49 [Last accessed on 20 Sep 2023].

12. Parida SS, Bose S, Butcher M, Apostolakis G, Shekhar P. SVD enabled data augmentation for machine learning based surrogate modeling of non-linear structures. *Eng Struct* 2023;280:115600. DOI

13. Giovanis DG, Fragiadakis M, Papadopoulos V. Epistemic uncertainty assessment using incremental dynamic Analysis and neural networks. *Bull Earthq Eng* 2016;14:529-47. DOI

14. Dabiri H, Faramarzi A, Dall'asta A, Tondi E, Micozzi F. A machine learning-based analysis for predicting fragility curve parameters of buildings. *J Build Eng* 2022;62:105367. DOI

15. Rosti A, Del Gaudio C, Rota M, et al. Empirical fragility curves for Italian residential RC buildings. *Bull Earthq Eng* 2021;19:3165-83. DOI

16. Liel AB, Haselton CB, Deierlein GG, Baker JW. Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Struct Saf* 2009;31:197-211. DOI

17. Esteghamati M, Flint MM. Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design. *Eng Struct* 2021;245:112971. DOI

18. Hwang S, Mangalathu S, Shin J, Jeon J. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. *J Build Eng* 2021;34:101905. DOI

19. Micheli L, Alipour A, Laflamme S. Multiple-surrogate models for probabilistic performance assessment of wind-excited tall buildings under uncertainties. *ASCE-ASME J Risk Uncertainty Eng Syst Part A Civ Eng* 2020;6:04020042. DOI

20. Haselton CB, Deierlein GG. Assessing seismic collapse safety of modern reinforced concrete moment-frame buildings, PEER report 2007-08. Ph.D. Dissertation, Stanford University, USA. 2006. Available from: https://peer.berkeley.edu/publications/2007-08 [Last accessed on 20 Sep 2023].

21. Bose S, Martin J, Stavridis A. Simulation framework for infilled RC frames subjected to seismic loads. *Earthq Spectra* 2019;35:1739-62. DOI

22. Bose S, Nozari A, Mohammadi ME, et al. Structural assessment of a school building in Sankhu, Nepal damaged due to torsional response during the 2015 Gorkha earthquake. In: Pakzad S, Juan C, editors. Dynamics of civil structures, Volume 2. Cham: Springer International Publishing; 2016. pp. 31-41. DOI

23. Tempestti JM, Stavridis A. Simplified method to assess lateral resistance of infilled reinforced concrete frames. In Proceedings of 16th World Conference on Earthquake Engineering, Santiago, Chile. 2017. Available from: https://www.researchgate.net/publication/312939259_SIMPLIFIED_METHOD_TO_ASSESS_LATERAL_RESISTANCE_OF_INFILLED_REINFORCED_CONCRETE_FRAMES [Last accessed on 20 Sep 2023].

24. ASCE/SEI 41-13. Seismic evaluation and retrofit of existing buildings. Reston, VA, USA: American Society of Civil Engineers; 2017. Available from: https://ascelibrary.org/doi/book/10.1061/9780784412855 [Last accessed on 20 Sep 2023].

25. Yousefianmoghadam S, Behmanesh I, Stavridis A, Moaveni B, Nozari A, Sacco A. System identification and modeling of a dynamically tested and gradually damaged 10-story reinforced concrete building. *Earthq Eng Struct Dyn* 2018;47:25-47. DOI

26. McKenna F, Fenves GL, Scott MH, Jeremic B. Open system for earthquake engineering Simulation (OpenSEES), Berkeley, California, USA: University of California; 2000. Available from: https://opensees.berkeley.edu/ [Last accessed on 20 Sep 2023].

27. Spacone E, Filippou FC, Taucer FF. Fibre beam-column model for non-linear analysis of r/c frames: part I. formulation. *Earthq Eng*

Struct Dyn 1996;25:711-25. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-9845%28199607%2925%3A7%3C711%3A%3AAID-EQE576%3E3.0.CO%3B2-9 [Last accessed on 20 Sep 2023]

28. Menegotto M, Pinto PE. Method of analysis for cyclically loaded reinforced concrete plane frames including changes in geometry and nonelastic behavior of elements under combined normal force and bending, IASBE Symposium on Resistance and Ultimate Deformability of structures Acted on by Well-Defined Repeated Loads, Final Report, Lisbon. 1973. Available from: https://xueshu.baidu.com/usercenter/paper/show?paperid=672527f0ca29d742b9e32522a8210769 [Last accessed on 20 Sep 2023].

29. Filippou FC, Popov EP, Bertero VV. Effects of bond deterioration on hysteretic behavior of reinforced concrete joints. Report EERC 83-19. Berkeley: University of California; 1983. Available from: https://nehrpsearch.nist.gov/static/files/NSF/PB84192020.pdf [Last accessed on 20 Sep 2023].

30. Yassin M, Hisham M. Nonlinear analysis of prestressed concrete structures under monotonic and cyclic loads. Ph.D. Dissertation, University of California, Berkeley, USA. 1994. Available from: https://www.proquest.com/openview/e0565f57989505eb9b0cbcdfba94e500/1?pq-origsite=gscholar&cbl=18750&diss=y [Last accessed on 20 Sep 2023].

31. Chaulagain H, Rodrigues H, Spacone E, Varum H. Seismic assessment of RC structures with infill masonry panels in Nepal - Sensitivity analysis. Second European Conference on Earthquake Engineering and Seismology, Istanbul, Turkey. 2014. Available from: http://hdl.handle.net/10773/12676 [Last accessed on 20 Sep 2023].

32. Pradhan PM. Composite action of partial infill wall in reinforced concrete frame under lateral load. Ph.D. Dissertation, Tribhuvan University, Nepal. 2009. Available from: https://www.researchgate.net/publication/268387175_Composite_Action_of_Partial_Infill_Wall_in_Reinforced_Concrete_Frame_Under_Lateral_Load [Last accessed on 20 Sep 2023].

33. Takai N, Shigefuji M, Rajaure S, et al. Strong ground motion in the Kathmandu Valley during the 2015 Gorkha, Nepal, earthquake. *Earth Planets Space* 2016:68. DOI

34. United State Geological Survey. 2015. Available from: http://earthquake.usgs.gov/earthquakes/eventpage/us20002926#scientific_waveforms [Last accessed on 20 Sep 2023].

35. ATC. FEMA-P695. Quantification of building seismic performance factors. Federal Emergency Management Agency. 2009. Available from: https://store.atcouncil.org/index.php?dispatch=products.view&product_id=210 [Last accessed on 20 Sep 2023].

36. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn* 2002;31:491-514. DOI

37. Hancock J, Bommer JJ. The effective number of cycles of earthquake ground motion. *Earthq Eng Struct Dyn* 2005;34:637-64. DOI

38. Garini E, Gazetas G. Damage potential of near-fault records: sliding displacement against conventional "Intensity Measures". *Bull Earthq Eng* 2013;11:455-80. DOI

39. Biasio MD. Ground motion intensity measures for seismic probabilistic risk analysis. Universite de Grenoble, France. 2014. Available from: https://www.researchgate.net/publication/290182432_Ground_motion_intensity_measures_for_seismic_probabilistic_risk_analysis [Last accessed on 20 Sep 2023].

40. Anastasopoulos PC, Mannering FL. Analysis of pavement overlay and replacement performance using random parameters hazard-based duration models. *J Infrastruct Syst* 2015;21:04014024. DOI

41. Housner GW. Intensity of ground motion during strong earthquakes. California Institute of Technology, USA. 1952. Available from: https://authors.library.caltech.edu/records/s16tc-ah892#:~:text=A%20measure%20of%20the%20surface,Intensities%20for%20the%20same%20earthquakes [Last accessed on 20 Sep 2023].

42. Zhang Y, Shen R, Sun B, Liu T, Shi Y, Li L. Selection of ground motion intensity measures in fragility analysis of a mega-scale steel frame structure at separate limit states: a case study. *Buildings* 2022;12:1530. DOI

43. Trifunac MD, Brady AG. A study on the duration of strong earthquake ground motion. *Bull Seismol Soc Am* 1975;65:581-626. Available from: https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/65/3/581/101795/A-study-on-the-duration-of-strong-earthquake?redirectedFrom=fulltext [Last accessed on 20 Sep 2023].

44. Sarma SK, Yang KS. An evaluation of strong motion records and a new parameter A_{95}. *Earthq Eng Struct Dyn* 1987;15:119-32. DOI

45. Bommer JJ, Martinez-Pereira A. The prediction of strong-motion duration for engineering design. In Proceedings of 11th World Conference on Earthquake Engineering, London, UK. 1996. Available from: https://www.iitk.ac.in/nicee/wcee/article/11_84.PDF [Last accessed on 20 Sep 2023].

46. Pejovic J, Jankovic S. Selection of ground motion intensity measure for reinforced concrete structure. *Procedia Eng* 2015;117:588-95. DOI

47. Du W. An empirical model for the mean period (Tm) of ground motions using the NGA-West2 database. *Bull Earthq Eng* 2017;15:2673-93. DOI

48. Buratti N. A comparison of the performances of various ground motion intensity measures. In Proceedings of 15th World Conference on Earthquake Engineering, Lisbon, Portugal. 2012. Available from: https://www.semanticscholar.org/paper/A-comparison-of-the-performances-of-various-ground-Buratti/328f6876a90872715501321c05e24538f65652d2 [Last accessed on 20 Sep 2023].

49. Adam C, Kampenhuber D, Ibarra LF. Optimal intensity measure based on spectral acceleration for P-delta vulnerable deteriorating frame structures in the collapse limit state. *Bull Earthq Eng* 2017;15:4349-73. DOI

50. Campbell KW, Bozorgnia Y. Cumulative absolute velocity (CAV) and seismic intensity based on the PEER-NGA database. *Earthq Spectra* 2012;28:457-85. DOI

51. Washington SP, Karlaftis MG, Mannering FL. Statistical and econometric methods for transportation data analysis. 2nd edition. Chapman & Hall/CRC. 2011. Available from: https://academic.oup.com/jrsssa/article/185/2/731/7068967 [Last accessed on 20 Sep 2023].

52.   Zellner A, Theil H. Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica* 1962;30:54-78. DOI

53.   Thomas A. Seismic damage assessment of representative motorway bridges: a three-stage least squares approach. Master's Dissertation. 2016. Available from: https://www.proquest.com/openview/26c1ce1526a8c2de795412ff3fa59679/1?pq-origsite=gscholar&cbl=18750 [Last accessed on 20 Sep 2023].

54.   Yeh I. Modeling of strength of high-performance concrete using artificial neural networks. *Cem Concr Res* 1998;28:1797-808. DOI

55.   Bose S, Rai DC. Behavior of AAC infilled RC frame under lateral loading. Master's Thesis, Indian Institute of Technology, Kanpur, India. 2013. Available from: https://www.researchgate.net/publication/288099033_Behavior_of_AAC_infilled_RC_frame_under_lateral_loading [Last accessed on 20 Sep 2023].

56.   Yousefianmoghadam S, Song M, Stavridis A, Moaveni B. System identification of a two-story infilled RC building in different damage states. In Proceedings of 2nd ATC-SEI Conference on Improving the Seismic Performance of Existing Buildings and other Structures, San Francisco, CA, USA. 2015. DOI

57.   Yousefianmoghadam S, Song M, Mohammadi ME, et al. Nonlinear dynamic tests of a reinforced concrete frame building at different damage levels. *Earthq Engng Struct Dyn* 2020;49:924-45. DOI

58.   Gao X. Experimental and numerical assessment of the seismic performance of an innovative retrofit scheme for infilled RC frames. 2021. DOI

59.   Melchers RE, Beck AT. Structural reliability analysis and prediction. Wiley: New York; 1999. DOI

60.   Panagiotakos TB, Fardis MN. Deformations of reinforced concrete members at yielding and ultimate. *ACI J Struct Eng* 2001;98:135-48. DOI

61.   Bose S. Analytical and numerical framework for the seismic assessment of infilled RC frames. PhD thesis, University at Buffalo, USA. 2019. Available from: https://www.proquest.com/openview/ba3539943af5c626206276c25d6be72c/1?pq-origsite=gscholar&cbl=18750&diss=y [Last accessed on 20 Sep 2023].

62.   Porter KA, Beck JL, Shaikhutdinov RV. Sensitivity of building loss estimates to major uncertain variables. *Earthq Spectra* 2002;18:719-43. DOI

63.   Grünthal G. European macroseismic scale 1998 (EMS-98). Luxembourg: Centre Europèen de Géodynamique et de Séismologie; 1998. Available from: https://www.gfz-potsdam.de/en/EMS-98%20%20-%20%20European%20Macroseismic%20Scale%201998 [Last accessed on 20 Sep 2023].

64.   National Society for Earthquake Technology-Nepal (NSET). Seismic vulnerability evaluation guideline for private and public buildings, Part II: post disaster damage assessment, Kathmandu, Nepal. 2009. Available from: https://www.preventionweb.net/files/2605_NSET20Profile202007.pdf [Last accessed on 20 Sep 2023].

65.   Brzev S, Pandey B, Maharjan DK, Ventura C. Seismic vulnerability assessment of low-rise reinforced concrete buildings affected by the 2015 Gorkha, Nepal, Earthquake. *Earthq Spectra* 2017;33:275-98. DOI

66.   Bose S, Rai DC. Lateral load behavior of an open-ground-story RC building with AAC infills in upper stories. *Earthq Spectra* 2016;32:1653-74. DOI

67.   Hilly R, Bose S, Hu Y, Stavridis A. Distribution of seismic demand during the 2015 Gorkha earthquake. In proceedings of the 17th Symposium of Earthquake Engineering, Roorkee, India. 2022. Available from: https://www.researchgate.net/publication/365347010_Distribution_of_Seismic_Demand_During_the_2015_Gorkha_Earthquake [Last accessed on 20 Sep 2023].

68.   Porter K, Kennedy R, Bachman R. Creating fragility functions for performance-based earthquake engineering. *Earthq Spectra* 2007;23:471-89. DOI

69.   Baker JW. Efficient analytical fragility function fitting using dynamic structural analysis. *Earthq Spectra* 2015;31:579-99. DOI

70.   Cristofaro MT, D'Ambrisi A, De Stefano M, Tanganelli M. Concrete compressive strength extracted from existing buildings. In The New Boundaries of Structural Concrete: Session D - Concrete Quality Control on Site California, ACIIC - Ancona 2011. Available from: https://www.ingenio-web.it/files/d2.pdf [Last accessed on 27 Sep 2023].

71.   Schueremans L, Van Gemert DA. Probability density functions for masonry material parameters - a way to go? *Struct Anal Hist Construct* 2006;2:921-8. Available from: https://www.researchgate.net/publication/312457877_Probability_density_functions_for_masonry_material_parameters_-_A_way_to_go [Last accessed on 27 Sep 2023]

72.   Shimizu Y, Hirosawa M, Zhou J. Statistical analysis of concrete strength in existing reinforced concrete buildings in Japan. In Proceedings of the 12th World Conference on Earthquake Engineering, Aukland, New Zealand, 1999. Available from: https://api.semanticscholar.org/CorpusID:34788826 [Last accessed on 27 Sep 2023].