International Journal of Steel Structures (2022) 22(4):1002–1024
https://doi.org/10.1007/s13296-022-00619-w

Online ISSN 2093-6311
Print ISSN 1598-2351

Evaluation of the Probabilistic Distribution of Statistical Data Used in the Process of Developing Fragility Curves

Vahid Mohsenian[1] · Nima Gharaei-Moghaddam[2] · Alireza Arabshahi[2]

Received: 28 June 2021 / Accepted: 23 May 2022 / Published online: 15 June 2022

Abstract
In most cases, the normal or log-normal distributions are assumed for the recorded statistical data in the process of fragility curve construction. This study aims to evaluate the validity of these assumptions and their influence on the accuracy of the results. For this purpose, considering the intensity corresponding to different damage levels as the statistical data, the analytical method is used for the development of fragility curves, taking advantage of incremental dynamic analysis for six multi-story moment-resisting steel frame structures with 3, 5, 7, 10, 12 and 15 stories. Comparison of the fragility curves attained by the assumptions of normal and log-normal distribution shows close agreement between the results, such that the maximum difference for different performance levels in the frame structures is determined to be about 13%. According to the outcomes of numerical tests such as Shapiro–Wilk and Kolmogorov–Smirnov and the graphical and descriptive tests performed on the attained statistical results, the assumption of the normal distribution is not incorrect for all of the performance levels. However, the assumption of the log-normal distribution is a more reliable hypothesis. Accordingly, it is proposed to utilize this assumption for the development of fragility curves in the reliability evaluation of structures subjected to seismic loading.

Keywords Seismic reliability · Fragility analysis · Statistical distribution · Shapiro–Wilk test · Kolmogorov–Smirnov test · Descriptive test

1 Introduction

Due to the wide range of uncertainties in the process of seismic analysis of structures, the best and most reasonable method seems to be the probabilistic approach (Ang & Tang, 2007). In this regard, reliability and fragility analysis are the most effective approaches. The development of fragility curves requires a statistic and probabilistic analysis that can be performed by different methodologies (including analytical, experimental, and combined methods and even methods based on engineering judgment) depending on the desired accuracy (Baharvand & Ranjbaran, 2020; Zuo et al., 2019). However, the analytical techniques are frequently used in the case of accurate computer models due to acceptable accuracy and ease in controlling data and the attained statistical sets.

Considering the special features of Incremental Dynamic Analysis (IDA) in dealing with the inherent uncertainties of ground motions records and providing an appropriate statistical population, this approach is usually used in the development of fragility curves (Vamvatsikos & Cornell, 2002, 2004).

Assuming that parameter r represents the structural response, R stands for the limit state corresponding to a predefined damage level, IM designates the earthquake intensity measure, and im is a given excitation intensity, then a fragility curve in the form of Eq. 1 determines the probability that the response exceeds the limit states for the given intensity:

$$\text{Fragility} = P[r \geq R|\text{IM} = \text{im}] \tag{1}$$

Fragility curves are the cumulative distribution function of damage (Hao, 2011). There are generally two approaches of constant damage level (IM-based; IM stands for Intensity Measure) and constant hazard level (EDP-based; EDP stands for Engineering Demand Parameter) methods for fragility analysis (Mohsenian et al., 2021c; Zareian et al., 2010). In

✉ Nima Gharaei-Moghaddam
Nima.Gharaeimoghaddam@mail.um.ac.ir

1 Department of Civil Engineering, University of Science and Culture, Tehran, Iran

2 Department of Civil Engineering, School of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

the first approach (IM-based), the probability of exceeding from limit responses corresponding to a predefined performance level is determined for various earthquake intensity levels. In this method, which is more common, the intensity measure corresponding to the considered performance level is used for establishing a statistical population. On the other hand, in the EDP-based method, which is deemed more appropriate for seismic retrofitting purposes, the probability of exceeding different performance levels for a given intensity of the applied excitation is determined. In this approach, the structural response is considered as statistical data. In some cases, the fragility curves may be established using both methods. For this purpose, the exceedance probabilities for different intensities are derived discretely using the EDP-based approach. Then, using the IM-based method, the fragility curve is described as the best curve fitted to the extracted points in the previous step.

The development of fragility curves provides a powerful means for evaluating the influence of different parameters on the seismic responses of structures. Accordingly, fragility analysis has been performed by various investigators for different purposes. For instance, some researchers utilized the fragility curves for seismic performance evaluation of different structural systems under different types of excitations (Ghowsi & Sahoo, 2013; Kim & Leon, 2013; Lee et al., 2018; Mohsenian, Filizadeh, et al., 2021). Fragility curves are also powerful tools for assessing the effectiveness of different retrofitting and strengthening methods on the seismic response of damaged or weak structures (Mohsenian et al., 2020, 2021b). For instance, Ahmadi and Ebadi Jamkhaneh (2021) utilized fragility analysis to evaluate the effectiveness of the energy dissipation devices on improving the seismic performance of structures with a soft story. Shafaei and Naderpour (2020) utilized fragility analysis to evaluate the seismic performance of reinforced concrete frame structures retrofitted by FRP and subjected to main shock-after shock sequence. Montazeri et al. (2021) performed fragility analysis to assess the seismic performance of retrofitted conventional bridges. Kima and Shinozuka (2004) developed fragility curves for bridges retrofitted by steel jacketing.

A review of previous studies shows the extensive applications of fragility analysis. Evaluations showed that in most of the past investigations, the fragility curves were developed assuming a probabilistic distribution for the statistical data. However, the log-normal distribution is a more common assumption, but the normal distribution has also been used in previous studies for fragility curve development (Mohsenian et al., 2021c). Although the accuracy of the fragility analysis directly depends on such assumption, its validity is not verified and even in some cases, investigators attempted to propose alternative methods to prevent making such assumptions (Sudret et al., 2014). Needless to say, in a project depending on the size and importance, the accuracy of the results can considerably affect the safety and economical aspects. According to the authors' best knowledge, the only available study that evaluates the validity of assumptions regarding the distribution of statistical data used in fragility curve development is the research performed by Shinozuka et al. (2000). In this study, the authors tested the goodness of fit of the fragility curves developed by assuming two-parameter log-normal distribution and estimated the confidence intervals of the two parameters (median and log-standard deviation) of the distribution. However, this study was performed on a bridge structure.

The present study tends to evaluate the accuracy of the assumptions of normal and log-normal distribution for the data used in fragility curve development in building structures, and also determine the sensitivity of the results of fragility analysis to these assumptions. What makes this study distinctive from the previous similar research works and the major novelties of the present paper are its focus on the building structures, the utilized performance-based viewpoint, and a clear and applicable methodology. Moreover, assessment of the sensitivity of fragility curves to different assumed distributions (normal or log-normal) is another novelty of the present study. For this purpose, six multi-story moment-resisting steel frames with 3, 5, 7, 10, 12, and 15 stories are designed. Considering different performance and hazard levels, fragility curves are developed assuming both normal and log-normal distributions for the data derived from Incremental Dynamic Analysis (IDA). Different numerical tests such as Shapiro–Wilk and Kolmogorov–Smirnov, as well as the graphical and descriptive tests, are performed on the utilized data sets for fragility curve development to assess the validity of the assumed distributions, and according to the outcomes of the performed statistical tests, the assumption of the log-normal distribution is more reliable, although the normal assumption is also not incorrect for all of the performance levels.

This study is organized into six sections. Sections 2 and 3 present the details of the studied models and the adopted assumptions for nonlinear modeling of the structures. The hierarchy of incremental dynamic analysis and fragility analysis of the studied frame structures using both normal and log-normal assumptions are discussed in Sect. 4. The performed tests on the attained data for fragility curve development to determine the appropriate statistical distribution and the attained results are presented in Sect. 5. Finally, Sect. 6 concludes this study.

2 Characteristics of the Studied Models

In this study, 2-dimensional intermediate moment-resisting steel frame structures, depicted in Fig. 1, are used. The gravitational dead (Q_D) and live (Q_L) loads in the stories are 31.5

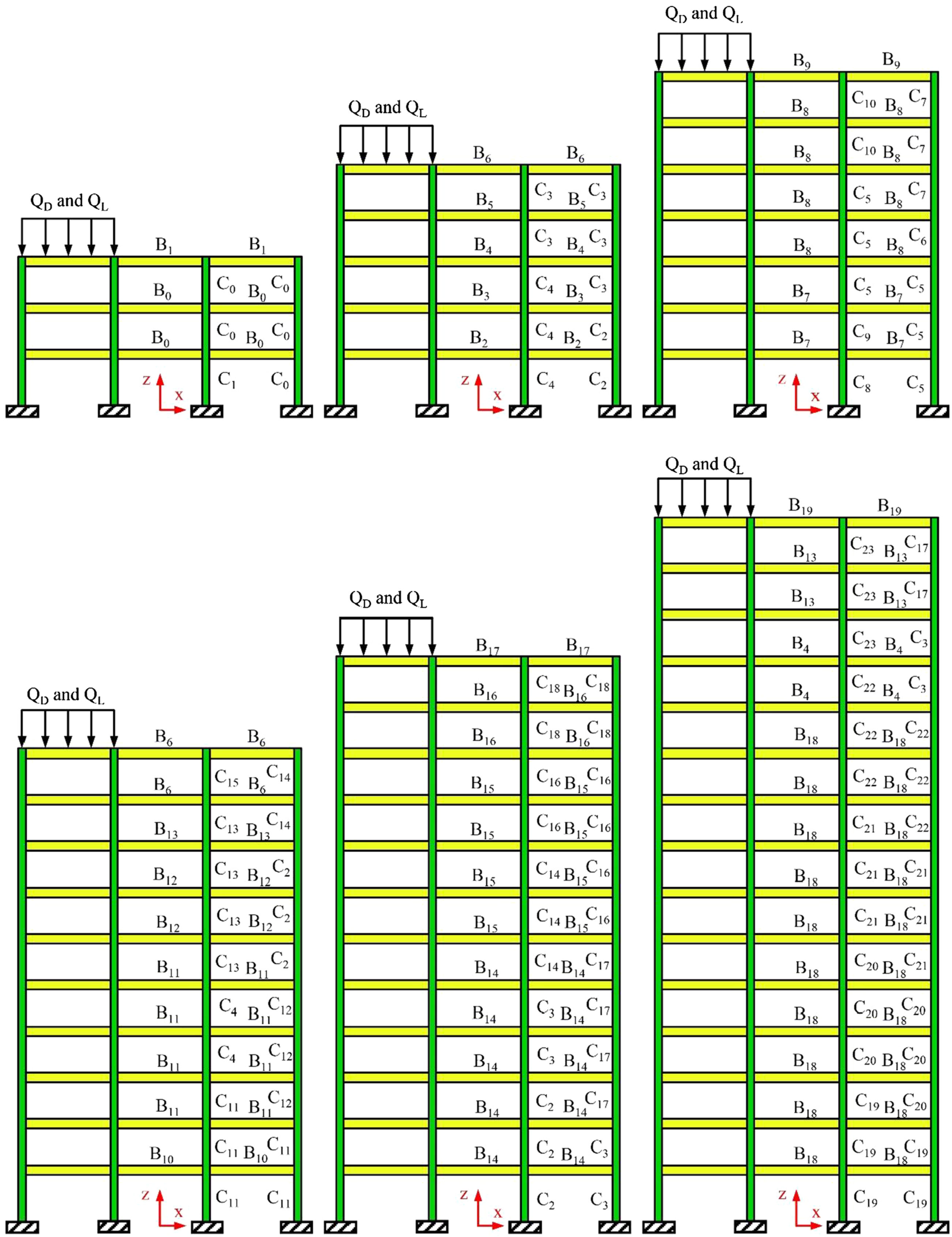

Fig. 1 Geometrical properties and loading details of the studied structures

and 10 kN/m^2, respectively. The roof live load (Q_L) is 7.5 kN/m^2. The span length and story heights are identical for all the structures and equal to 5 and 3.2 m, respectively. To investigate the effect of structural height, 3, 5, 7, 10, 12, and 15-story structures are designed. The selected heights for the modeled structures are in the range of allowable height range for the moment-resisting steel frame system (a maximum of 50 m from the base level).

It is assumed in the design phase that the structures belong to the category of ordinary buildings and are located in a site with high seismicity (PGA = 0.35 g). The site soil is considered to be type C according to ASCE7 (ASCE, 2010) categorization (stiff soil with the shear wave velocity between 375 to 750 m/s). The frame structures are designed according to AISC360 (AISC, 2010) using ETABS software (CSI, 2015).

For the beams and columns, I-shaped and box sections are used, respectively. The properties of the beam and columns sections, which are specified by B_i and C_i in Fig. 1, are presented in Table 1.

It is noteworthy that the geometry, member sections, and loading of the frame structures are symmetrical relative to the z-axis (see Fig. 1). Rigid diaphragms are also considered at each story level. A360 steel grade with the yield stress, Poisson's ratio, and modulus of elasticity equal to 250 MPa, 0.26, and 210 GPa is considered for the structural components of the designed buildings (ASTM, 2019).

3 Modeling Nonlinear Behavior of Structures

PERFORM-3D software (CSI, 2017) is used for 2-dimensional nonlinear modeling and analysis of the structure. The gravitational loading assumptions for the nonlinear model are the same as the linear model. It should be noted that in the combination of gravitational and lateral loads, the effects of the gravitational loads (Q_G) is considered according to Eq. 2, in which Q_D and Q_L stand for the dead and live loads, respectively (ASCE, 2017):

$$Q_G = Q_D + 0.25Q_L \tag{2}$$

The generalized load-deformation curve depicted in Fig. 2 is used for nonlinear modeling of beams and columns of the frame structures. The parameters a, b, and c in this figure are extracted from the table of acceptance criteria of steel members according to the yielding mode and compactness of the

Table 1 Geometrical properties of the beam and column section of the designed structures (dimensions are in mm)

Columns		Beams	
ID	Section (width × thikness)	ID	Section (width × thikness)
C_0	Box(180 × 8)	B_0	Web(350 × 6)-Flanges(150 × 8)
C_1	Box(240 × 8)	B_1	Web(270 × 6)-Flanges(120 × 8)
C_2	Box(350 × 15)	B_2	Web(350 × 10)-Flanges(180 × 20)
C_3	Box(300 × 15)	B_3	Web(350 × 10)-Flanges(200 × 20)
C_4	Box(400 × 20)	B_4	Web(300 × 10)-Flanges(200 × 20)
C_5	Box(240 × 10)	B_5	Web(270 × 10)-Flanges(150 × 20)
C_6	Box(200 × 8)	B_6	Web(240 × 10)-Flanges(150 × 20)
C_7	Box(180 × 8)	B_7	Web(400 × 6)-Flanges(150 × 10)
C_8	Box(300 × 10)	B_8	Web(300 × 6)-Flanges(150 × 10)
C_9	Box(270 × 10)	B_9	Web(270 × 6)-Flanges(120 × 10)
C_{10}	Box(150 × 8)	B_{10}	Web(300 × 15)-Flanges(180 × 15)
C_{11}	Box(500 × 20)	B_{11}	Web(350 × 10)-Flanges(250 × 25)
C_{12}	Box(400 × 15)	B_{12}	Web(350 × 10)-Flanges(200 × 25)
C_{13}	Box(300 × 20)	B_{13}	Web(270 × 10)-Flanges(180 × 20)
C_{14}	Box(250 × 15)	B_{14}	Web(300 × 8)-Flanges(180 × 15)
C_{15}	Box(200 × 20)	B_{15}	Web(270 × 8)-Flanges(180 × 15)
C_{16}	Box(200 × 15)	B_{16}	Web(240 × 8)-Flanges(150 × 15)
C_{17}	Box(240 × 15)	B_{17}	Web(200 × 8)-Flanges(150 × 15)
C_{18}	Box(180 × 15)	B_{18}	Web(350 × 10)-Flanges(300 × 20)
C_{19}	Box(600 × 25)	B_{19}	Web(200 × 10)-Flanges(150 × 20)
C_{20}	Box(550 × 25)	–	–
C_{21}	Box(450 × 20)	–	–
C_{22}	Box(350 × 20)	–	–
C_{23}	Box(270 × 15)	–	–

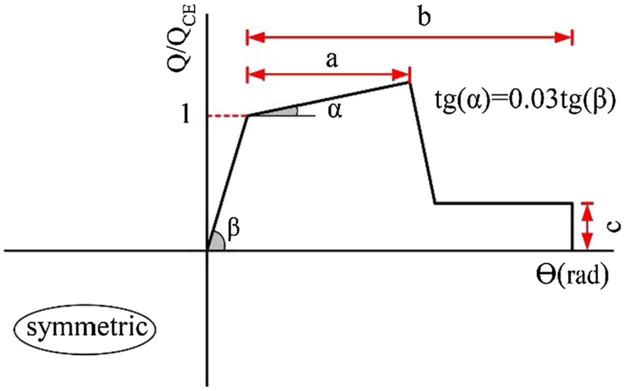

Fig. 2 The generalized force–deformation curve of steel structural elements (ASCE, 2017)

structural elements (ASCE, 2017). According to Fig. 2, the slope of the initial hardening stage of steel, tg(α), is set to be the 3% of the slope of the elastic branch, tg(β). (ASCE, 2017). In this figure, θ represents the plastic hinge rotation.

The maximum expected strength, Q_{CE}, of the beams is derived from Eq. 6, while for the columns Eqs. 4 and 5 are used:

$$Q_{CE} = ZF_{ye} \tag{3}$$

$$Q_{CE} = ZF_{ye}\left(1 - \frac{|P|}{2P_{ye}}\right)(for)\left(\frac{|P|}{P_{ye}} < 0.2\right) \tag{4}$$

$$Q_{CE} = ZF_{ye}\frac{9}{8}\left(1 - \frac{|P|}{P_{ye}}\right)(for)\left(\frac{|P|}{P_{ye}} \geq 0.2\right) \tag{5}$$

In these equations, Z and F_{ye} are the plastic section modulus and the expected yield stress of materials, respectively. P stands for the axial force of the member at the beginning of the dynamic analysis, P_{ye} is the axial load corresponding to the axial yielding of the column, which is derived by multiplying the cross-section of the element by the expected yield stress of materials ($P_{ye} = AF_{ye}$).

It is also noteworthy that concentrated flexural-axial hinges are considered for the beam and column elements at the critical locations (both ends).

4 Developing Fragility Curves Using Incremental Dynamic Analysis

First, the selected ground motions are applied to the structure and the modeled structures are analyzed using the Incremental Dynamic Analysis (IDA). 30 pairs of ground motion records corresponding to the site condition (the shear wave velocity between 375 to 750 m/s) are extracted from the PEER database (PEER, http://peer.berkeley.edu/peer-ground-motion-database). It is obvious that the number of utilized records is much greater than the minimum required number of ground motion records for IDA analysis (Shome, 1999). It should be noted that the minimum required number of records should also comply with the limitations of normality tests as well (Ghasemi & Zahediasl, 2012). This issue is completely discussed in Sect. 5. However, given the significant reduction in the inherent uncertainties due to the number of records used in this study, the results of IDA are expected to be sufficiently reliable.

The selected records and properties of their main components are given in Table 2. The selected records are classified as far-fault ground motions. Between the horizontal components of each ground motion, the one with the maximum highest spectral acceleration in the vibration frequency range of frame structures is selected as the main component and used in IDA. According to Fig. 3, these records are selected such that their average spectrums have a good agreement with the site design spectrum. As it is evident in this figure, the difference between the design spectrum and the average spectrum of the records in the governing mode of each frame is negligible.

For IDA, the Peak Ground Acceleration (PGA(g)) and the maximum inter-story drifts are opted as the Intensity (IM) and Demand Measures (DM), respectively. The IDA results in a graphical relationship between DM and IM, which is called the IDA curve. In order to improve accuracy of the analysis, the increment of IM measure in the analysis is selected equal to 0.05 g. Accordingly, the peak ground acceleration of the record in nth step (PGA_n) is derived from Eq. (6). Given PGA_0 as the initial peak ground acceleration of ground motion records (see Table 2), the scale factor in nth step (SF_n) is derived from Eq. (7) (Mohsenian & Mortezaei, 2019)

$$PGA_n(g) = 0.05\mathrm{n} \tag{6}$$

$$SF_n = PGA_n/PGA_0 \tag{7}$$

The IDA curves for the studied structures are demonstrated in Fig. 4. The limit states corresponding to different performance levels of the Immediate Occupancy (IO), Life Safety (LS), and Collapse Prevention (CP) are also depicted in this figure (ASCE, 2017). However, since this study aims to compare the results of two different assumptions for the statistical distribution of data used in fragility curve development, other arbitrary limit states can also be used.

Having the results of IDA in hand, the subsequent steps are followed to develop the fragility curves:

i. According to Fig. 5, the intensity measure corresponding to a given performance level of the system (in

Table 2 Properties of the main component of the selected ground motion records for IDA

Records	Earthquake and Year	R[a](km)	Component	M_w	PGA_0(g)
R_1	Cape Mendocino (USA), 1992	41.97	90	7.1	0.18
R_2	Cape Mendocino (USA), 1992	19.95	0	7.1	0.12
R_3	Cape Mendocino (USA), 1992	25.91	270	7.1	0.26
R_4	Cape Mendocino (USA), 1992	26.51	0	7.1	0.23
R_5	Chi-Chi (Taiwan), 1999	26.31	E	7.6	0.25
R_6	Chi-Chi (Taiwan), 1999	19.0	E	7.6	0.25
R_7	Chi-Chi (Taiwan), 1999	15.0	E	7.6	0.16
R_8	Chi-Chi (Taiwan), 1999	24.1	W	7.6	0.19
R_9	Chi-Chi (Taiwan), 1999	19.8	N	7.6	0.64
R_{10}	Chi-Chi (Taiwan), 1999	20.0	W	7.6	0.23
R_{11}	Chi-Chi (Taiwan), 1999	15.0	N	7.6	0.30
R_{12}	Chi-Chi (Taiwan), 1999	28.79	N	7.6	0.12
R_{13}	Chi-Chi (Taiwan), 1999	43.17	N	7.6	0.14
R_{14}	Chi-Chi (Taiwan), 1999	42.87	E	7.6	0.11
R_{15}	Chuetsu-oki (Japan), 2007	17.93	NS	6.8	0.32
R_{16}	Darfield (New Zealand), 2010	24.5	E	7.0	0.63
R_{17}	Hector Mine (USA), 1999	41.81	90	7.1	0.18
R_{18}	Hector Mine (USA), 1999	31.06	360	7.1	0.19
R_{19}	Iwate (Japan), 2008	28.9	NS	6.9	0.28
R_{20}	Iwate (Japan), 2008	25.56	NS	6.9	0.24
R_{21}	Kern County (USA), 1952	38.42	111	7.3	0.18
R_{22}	Kocaeli (Turkey), 1999	30.73	90	7.5	0.12
R_{23}	Landers (USA), 1992	34.86	90	7.4	0.13
R_{24}	Landers (USA), 1992	45.34	210	7.4	0.11
R_{25}	Landers (USA), 1992	25.02	45	7.4	0.21
R_{26}	Loma Prieta (USA), 1989	20.34	285	6.9	0.48
R_{27}	Northridge (USA), 1994	23.07	180	6.7	0.25
R_{28}	Northridge (USA), 1994	31.69	90	6.7	0.10
R_{29}	Northridge (USA), 1994	19.74	352	6.7	0.24
R_{30}	San Fernando (USA), 1971	25.47	90	6.6	0.11

[a]Closest Distance to Fault Rupture

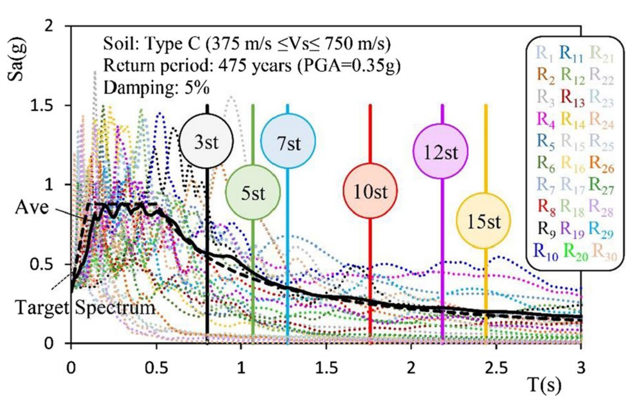

Fig. 3 Comparison of the average spectrum of the selected ground motion records with the site design spectrum

this study the IO, LS, and CP performance levels) is extracted from the IDA curves. At this step, for each performance level, a statistical community containing 30 members will be available.

ii. Assuming normal or log-normal distribution for the collected data sets in the previous step, after calculating the mean value (μ) and standard deviation (σ), the density probability functions are established using Eqs. 8 and 9 (Nowak & Collins, 2012):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}EXP\left(\frac{(x-\mu)^2}{-2\sigma^2}\right) \tag{8}$$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}EXP\left(\frac{((\ln(x)-\mu)^2}{-2\sigma^2}\right) \tag{9}$$

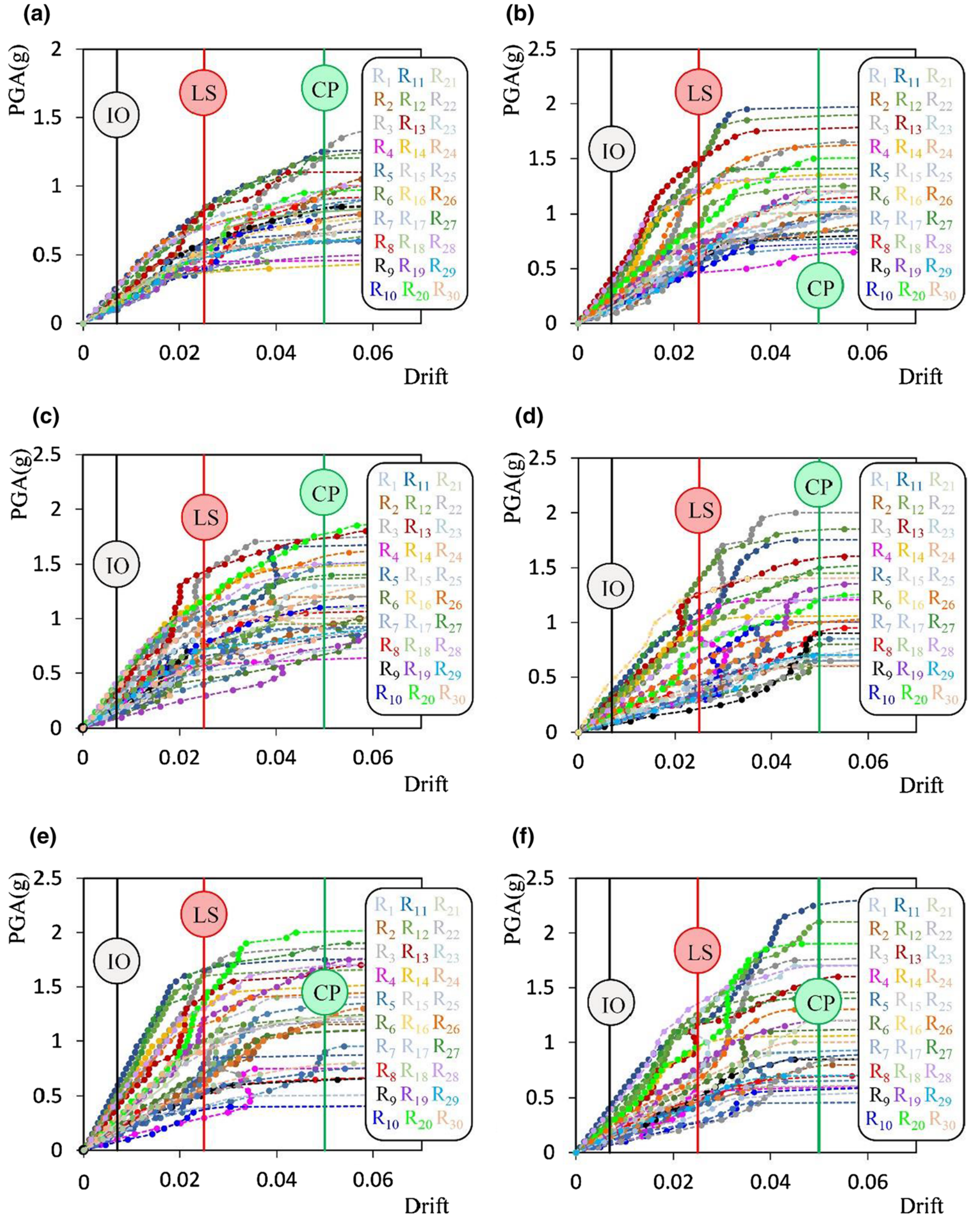

Fig. 4 IDA curves and the limit states corresponding to different performance levels of **a** 3- **b** 5- **c** 7- **d** 10- **e** 12- and **f** 15-story frames

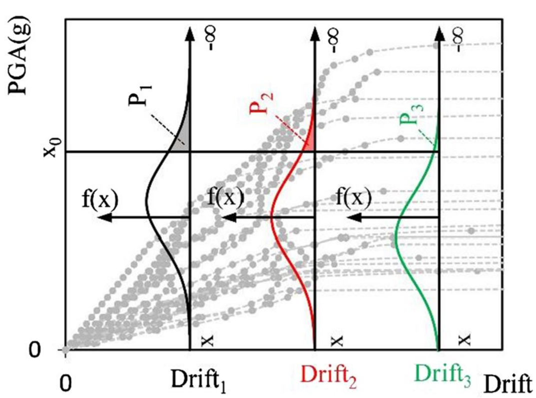

Fig. 5 Calculation of the exceedance probability for a certain performance level under a given intensity (x_0) using IDA results

iii. According to Fig. 5, taking x_0 as a specific intensity, the integral of the probability density function (the area under the curve) from $-\infty$ to x_0 determines the exceedance probability (P) for the considered damage level. This means that at this specific intensity, there is a probability of P that the structural response exceeds the response corresponding to the considered damage (performance) level.

iv. Subtracting P from 1 gives the reliability (P_0) of the system for the considered damage (performance) level, and this means that at a certain intensity, there is a probability of P_0 that the structure does not experience the considered performance level (Mohsenian, Filizadeh et al., 2021).

The fragility curves are derived for different performance levels of the studied structures according to the described methodology, assuming normal and log-normal distributions. The attained fragility curves are demonstrated in Fig. 6.

As evident in Fig. 6, however, there are differences between the curves derived from different distribution assumptions, but there is no clear trend for these differences. In most cases, for higher seismic intensities, the normal distribution assumptions result in lower exceedance probabilities. This is more evident for taller frame structures and higher performance levels. Vice versa, under lower seismic intensities, the log-normal assumption gives lower exceedance probabilities.

In the following, the maximum differences between the fragility curves for each performance level (D_{IO}, D_{LS}, and D_{CP}) of the frame structures are extracted up to the peak ground acceleration of 1.0 g (PGA = 1.0 g). The attained results are presented in Fig. 7 ($Difference = (Fragility_{Normal} - Fragility_{Log-Normal}) \times 100$).

For the IO performance level, the maximum difference between the curves is about 10%. According to Fig. 7, most of the differences occur around the intensity of 0.35 g, which indicates the design basis earthquake according to many of the design codes. For the LS and CP performance levels, the maximum differences between the curves are 10 and 13.5%, respectively. These maximums for the mentioned performance levels have occurred around the ground motion intensities of 0.95 and 0.8 g, which are higher than

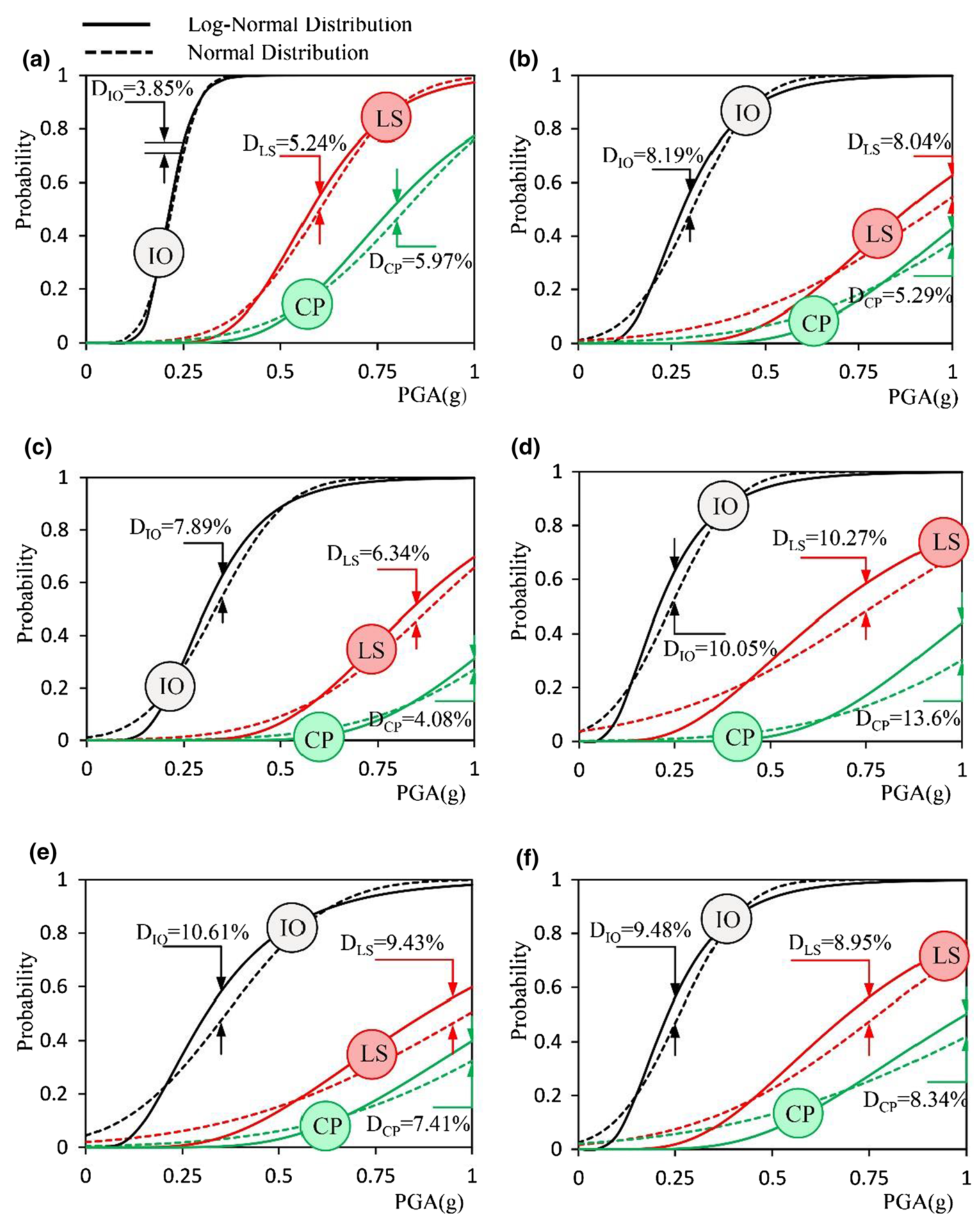

Fig. 6 The extracted fragility curves for different performance levels using IDA results **a** 3- **b** 5- **c** 7- **d** 10- **e** 12- and **f** 15-story frames

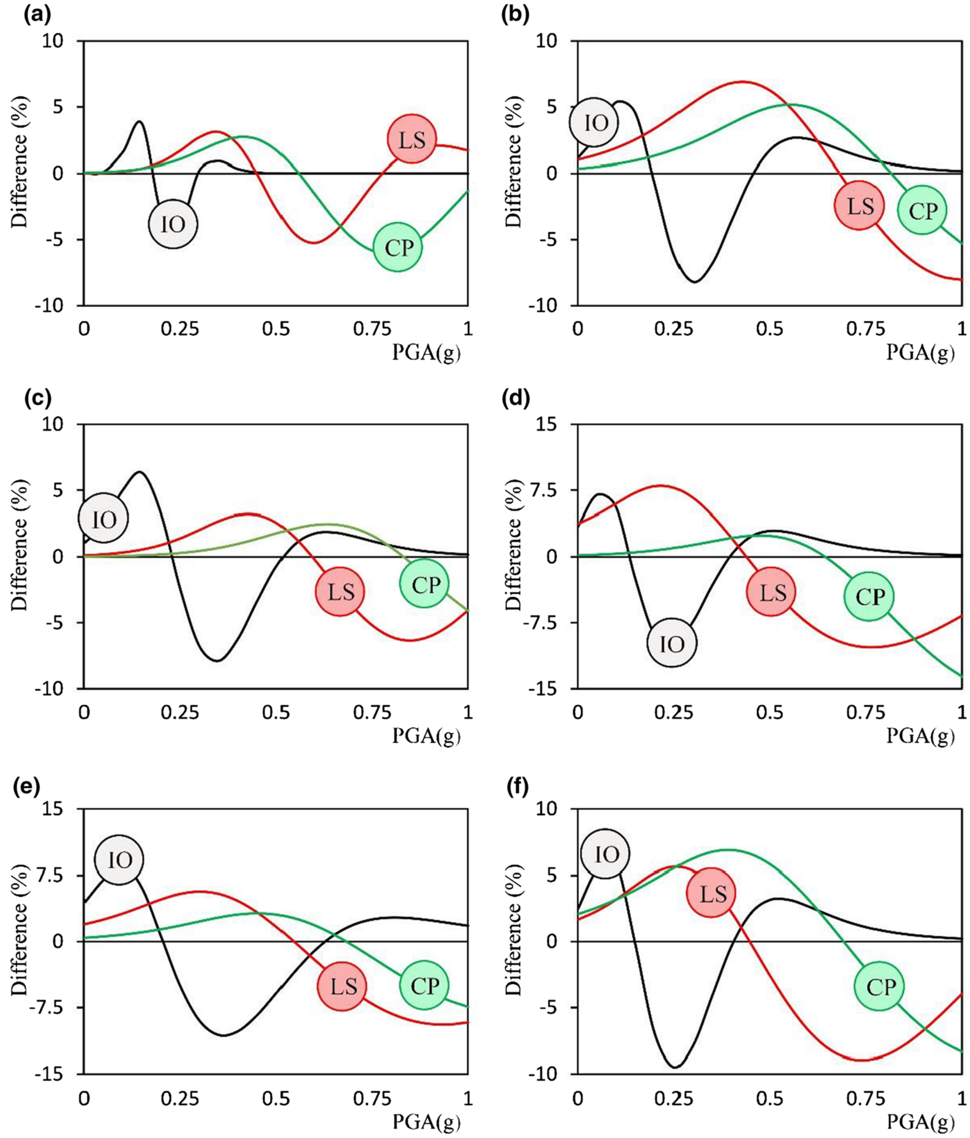

Fig. 7 The curves of difference percentage between the developed fragility curves using normal and log-normal distribution assumptions **a** 3- **b** 5- **c** 7- **d** 10- **e** 12- and **f** 15-story frames

the intensity corresponding to the maximum considered earthquake (0.55 g) (see Fig. 7).

5 Evaluation of the Assumed Statistical Distributions

Controlling the dispersion and central tendencies of parts of the data (sample variables) and consequently providing a suitable distribution function are among the valid statistical methods that are often used for the probabilistic evaluation of larger communities. When a statistical population has a normal distribution, the normality of data is evaluated using different methods that fall into three broad categories: numerical (significance tests), descriptive, and graphical (Mishra et al., 2019). In this section, the mentioned methods are first briefly described and defined. Then, using these methods, the accuracy of the normal and log-normal assumptions of the data derived from IDA analysis will be evaluated. SPSS Statistics software (Statistics, 2013) is used for this purpose.

5.1 Numerical Normality Test Methods

The numerical normality test methods usually use well-known statistics such as Kolmogorov–Smirnov (K-S), Lilliefors corrected (K-S), Shapiro–Wilk, and Anderson–Darling (Barton, 2005; Öztuna et al., 2006; Shapiro & Wilk, 1965).

Although the estimation accuracy of all statistics depends on the sample size (small sample size leads to estimation error), studies have shown that for all possible distributions and sample sizes, the Shapiro–Wilk statistic has the highest accuracy in the estimation process, and

Kolmogorov–Smirnov statistics is in the second place (Razali & Wah, 2011). Thus, for the small sample size, the Shapiro–Wilk statistic is usually recommended. The high computational volume of IDA, which is a time-consuming process, encourages the authors to utilize the minimum possible number of ground motion records (Han & Chopra, 2006; Vamvatsikos & Allin Cornell, 2006). Accordingly, the utilized sample size in this study is small (all samples consist of 30 data points which guarantee the minimum required number of data points for the utilized tests (Ghasemi & Zahediasl, 2012)). According to the provided explanations, in the present study, only the Shapiro–Wilk and Kolmogorov–Smirnov tests were used. For each frame, the results of the mentioned tests were for both normal and log-normal distribution assumptions are presented in Tables 3 and 4. In these tables, column df stands for the "degrees of freedom" which is equal to the sample size. The two other columns are used to check whether the normality (or log-normality) assumption is correct or not. Sig. represents the significance, and the significance values lower than 0.05 mean that the data set do not follow normal (or log-normal) distribution. However, for the tests with significance values higher than 0.05, there is a higher probability of normal (log-normal) distribution provided that the value of the statistics (first column) is closer to 1.

As mentioned, achieving the significance values (Sig.) Less than 0.05 in the statistics (which is the acceptable limit in statistical analysis) means rejecting the normality (log-normality) assumption of the distribution function governing the statistical population, and higher significance values (closer to 1) means more reliable assumptions. As evident in Table 3, the assumption of a normal distribution is ruled out in many cases (see the yellow cells) and is also a weak assumption for other cases based on the statistics values. In comparison, the log-normal distribution assumption for the data is much stronger (see Table 4). As can be seen, both statistics agree on the accuracy of the log-normal distribution assumption for the data. Given the explanations provided, the log-normal distribution is preferable.

Table 3 The results of Shapiro–Wilk and Kolmogorov–Smirnov tests for controlling the assumption of the normal distribution of data used in fragility curve development

	Kolmogorov-Smirnov[a]			Shapiro–Wilk		
	Statistic	df	Sig	Statistic	df	Sig
3-Storey						
IO	0.114	30	0.200*	0.960	30	0.235
LS	0.144	30	0.065	0.945	30	0.078
CP	0.105	30	0.200*	0.966	30	0.336
5-Storey						
IO	0.124	30	0.189	0.951	30	0.121
LS	0.153	30	**0.036**	0.949	30	0.103
CP	0.143	30	0.069	0.933	30	**0.035**
7-Storey						
IO	0.119	30	0.200*	0.937	30	**0.047**
LS	0.110	30	0.200*	0.963	30	0.273
CP	0.106	30	0.200*	0.965	30	0.312
10-Storey						
IO	0.129	30	0.150	0.930	30	**0.028**
LS	0.157	30	**0.029**	0.930	30	**0.029**
CP	0.137	30	0.094	0.938	30	**0.048**
12-Storey						
IO	0.148	30	0.051	0.949	30	0.105
LS	0.124	30	0.190	0.950	30	0.115
CP	0.124	30	0.196	0.954	30	0.154
15-Storey						
IO	0.089	30	0.200*	0.962	30	0.257
LS	0.145	30	0.061	0.952	30	0.131
CP	0.135	30	0.107	0.937	30	**0.045**

Bold values designate the cases that the significance values are below 0.05, which reject the normality assumption

*This is a lower bound of true significance

[a]Lilliefors Significance Correction

Table 4 The results of Shapiro–Wilk and Kolmogorov–Smirnov tests for controlling the assumption of the log-normal distribution of data used in fragility curve development

	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig	Statistic	df	Sig
3-Storey						
IO	0.094	30	0.200*	0.987	30	0.951
LS	0.093	30	0.200*	0.976	30	0.621
CP	0.112	30	0.200	0.960	30	0.225
5-Storey						
IO	0.079	30	0.200*	0.964	30	0.307
LS	0.101	30	0.200*	0.976	30	0.626
CP	0.106	30	0.200*	0.970	30	0.452
7-Storey						
IO	0.057	30	0.200*	0.983	30	0.861
LS	0.095	30	0.200*	0.971	30	0.474
CP	0.087	30	0.200*	0.975	30	0.586
10-Storey						
IO	0.105	30	0.200*	0.961	30	0.250
LS	0.066	30	0.200*	0.973	30	0.540
CP	0.072	30	0.200*	0.977	30	0.657
12-Storey						
IO	0.118	30	0.200*	0.972	30	0.513
LS	0.101	30	0.200*	0.957	30	0.180
CP	0.105	30	0.200*	0.954	30	0.147
15-Storey						
IO	0.106	30	0.200*	0.957	30	0.187
LS	0.107	30	0.200*	0.964	30	0.299
CP	0.117	30	0.200*	0.969	30	0.426

5.2 Descriptive Normality Test Method

The descriptive method is based on evaluating the frequency, mean (μ), and standard deviation (σ) of the data. The normal distribution has a symmetrical bell-shaped curve, and for the normal distribution in a statistical population, 68.2, 99.7 and 95.4% of the observations would be between $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$, respectively (Altman & Bland, 1995). Skewness and kurtosis are the important parameters that describe asymmetry. Since the values of these two parameters in a normal distribution are zero, a significant deviation of them from zero will undermine the normality assumption (Thode, 2002). Converting these parameters to a Z score, and providing a tolerance interval would be a good measure of normality. In the latter case, the results obtained between + 1.96 and 1.96 indicate the correctness of the normality assumption for the statistical population (Ghasemi & Zahediasl, 2012).

The results of the descriptive test for the studied frames at different performance levels for both normal and log-normal distributions are as presented in Table 5. It should be noted that in the case of log-normal assumption, the tests are performed on the logarithm of the data derived from IDA analysis. If the normal assumption is verified for those values, the main data has a log-normal distribution. According to the results, both assumptions for the distribution of data are in the significance intervals, but in comparison, the assumption of the log-normal distribution of data is certainly more reliable, given the lower values of the score statistic.

5.3 Graphical Normality Test Methods

The graphical methods are the approximate approaches for examining the hypothesis of normality distribution. Due to the low reliability of this method, it is used only as an auxiliary tool along with other methods (Öztuna et al., 2006) In these approaches, histograms, stem-and-leaf plots, boxplots, and quantile–quantile (Q-Q) plots are used to evaluate the hypothesis. As mentioned earlier, for a normal distribution, the histogram is bell-shaped and symmetrical related to the mean (Ghasemi & Zahediasl, 2012).

The stem and leaf diagrams are similar to the histograms and are used to illustrate the probability distribution shape of quantitative data (Das & Imon, 2016). Since the use of this diagram requires the availability of large-size samples, this method is not very common. Accordingly, it has not been used in the present study. For the studied frames, histogram diagrams are for both normal and log-normal (normality of the logarithms of the attained data from IDA) assumptions are shown in Figs. 8, 9, 10, 11, 12, 13.

As mentioned, histograms are a way to show the shape of the distribution of experimental data. The closer the histogram shape is to the Gaussian or bell-shaped distribution, the more the data fit the normal distribution. As evident from Figs. 8, 9, 10, 11, 12, 13, the histograms attained for the logarithms of the data sample are closer to the Gaussian distribution shape. Therefore, it is concluded that the log-normal distribution is a stronger assumption for all the studied structures and different considered performance levels. This finding is in agreement with the results numerical and descriptive tests results.

The Q-Q plots show the observed and expected values. In a normal distribution, the observed values are almost equal to the expected values. Deviation from this correspondence will reduce the validity of the normal distribution. Figures 14, 15, 16, 17, 18, 19 depict the attained Q-Q plots for the normal and log-normal assumptions. If the data belongs to the normal distribution, the points should be around a straight line, otherwise, this shows a null hypothesis, which means the data will not follow the normal distribution. According to Figs. 14, 15, 16, 17, 18, 19, the log-normal assumption seems a more valid hypothesis.

In a box diagram, the mean of the statistical population is drawn as a line inside the box, and the range between the first and third quartiles of frequency is considered as the length of the box (Altman & Bland, 1995). If the box is symmetric relative to the mean line, the assumption of a normal distribution for the data is supported. For the studied frames, considering the different performance levels and different hypotheses for the normal and log-normal distribution of the data, the mentioned diagrams are extracted for the IDA results and their logarithm values. The attained results are depicted in Fig. 20. Considering the box plots of normal distribution assumptions, it seems there is a low probability that the statistical data has a normal distribution, but by for the log-normal assumption (box diagrams on the right) the logarithm of the IDA results follow a normal distribution with a high probability, i.e., it is assumed that the statistical data follow the log-normal distribution. This finding is in agreement with the results of the graphical tests, as well as numerical and descriptive methods.

6 Conclusion

In this study, the reliability of the normal and log-normal probability distribution assumptions for the fragility curve development has been investigated considering different performance levels in the structural system. For this purpose, three numerical (significance), descriptive and graphical test methods have been utilized. To evaluate the effects of the statistical distribution on the accuracy of the analysis, the attained fragility curves using both assumptions have been compared. Based on the adopted assumptions, the following conclusions can be made:

Table 5 The results of the descriptive tests for controlling the normal and log-normal distribution assumptions of data used in developing fragility curves of different performance levels

	Normal Distribution			Log-Normal Distribution		
	IO	LS	CP	IO	LS	CP
3-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.772	0.612	0.179	− 0.048	0.007	− 0.343
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	1.363	− 0.291	− 0.695	0.267	− 0.423	− 0.698
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	1.941	1.538	0.451	− 0.119	0.017	− 0.861
Z KURTOSIS	1.752	− 0.374	− 0.894	0.344	− 0.544	− 0.898
5-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.548	0.575	0.661	− 0.374	− 0.141	0.135
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	− 0.442	− 0.448	− 0.550	− 0.387	− 0.569	− 0.897
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	1.377	1.446	1.663	− 0.940	− 0.355	0.339
Z KURTOSIS	− 0.568	− 0.575	− 0.707	− 0.497	− 0.731	− 1.153
7-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.912	0.412	0.388	− 0.170	− 0.227	− 0.136
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	0.603	− 0.632	− 0.680	− 0.277	− 0.763	− 0.743
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	2.294	1.035	0.975	− 0.427	− 0.570	− 0.343
Z KURTOSIS	0.774	− 0.812	− 0.875	− 0.356	− 0.981	− 0.955
10-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.595	0.730	0.733	− 0.073	− 0.255	− 0.015
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	− 0.704	− 0.359	− 0.047	− 1.138	− 0.564	− 0.427
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	1.495	1.835	1.844	− 0.184	− 0.642	− 0.037
Z KURTOSIS	− 0.905	− 0.462	− 0.060	− 1.464	− 0.725	− 0.549
12-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.613	0.356	0.156	− 0.252	− 0.292	− 0.484
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	− 0.279	− 0.953	− 1.157	− 0.653	− 0.948	− 0.475
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	1.541	0.895	0.392	− 0.633	− 0.735	− 1.216
Z KURTOSIS	− 0.358	− 1.225	− 1.487	− 0.839	− 1.218	− 0.610
15-Storey						
Valid no	30	30	30	30	30	30
Missing	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.502	0.544	0.625	− 0.437	− 0.312	0.005

Table 5 (continued)

	Normal Distribution			Log-Normal Distribution		
	IO	LS	CP	IO	LS	CP
Std. Error of Skewness	0.398	0.398	0.398	0.398	0.398	0.398
Kurtosis	0.125	− 0.314	− 0.496	− 0.560	− 0.192	− 1.003
Std. Error of Kurtosis	0.778	0.778	0.778	0.778	0.778	0.778
Z SKEWNESS	1.264	1.367	1.570	− 1.098	− 0.784	0.013
Z KURTOSIS	0.161	− 0.404	− 0.638	− 0.719	− 0.247	− 1.289

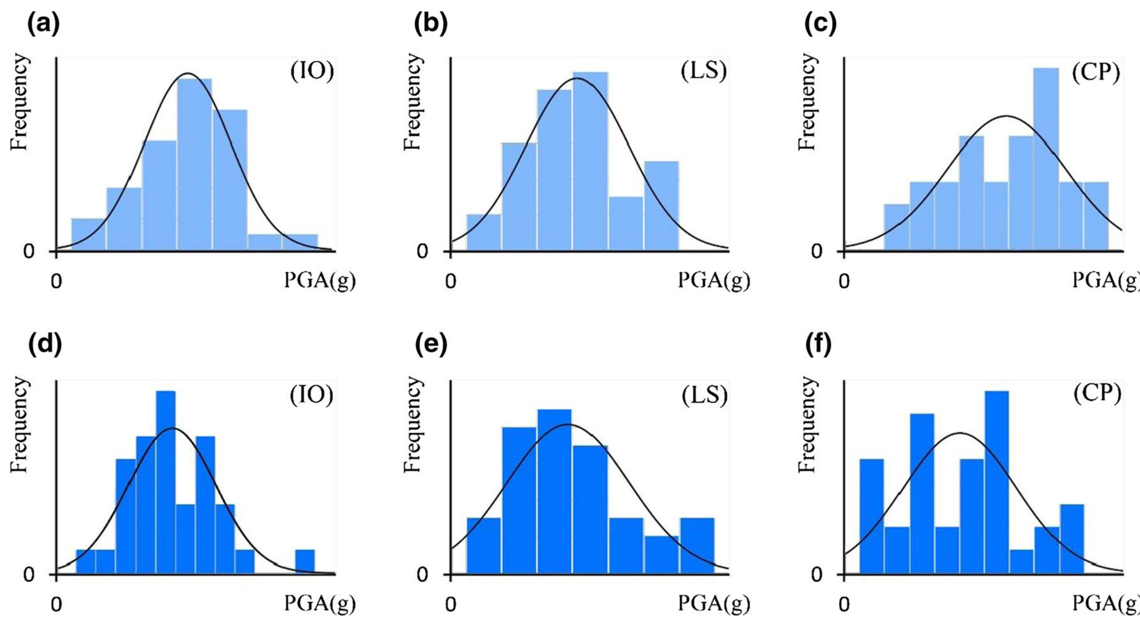

Fig. 8 The histogram diagrams at different performance levels for 3-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

1. Although the fragility curves derived from both normal and log-normal distribution assumptions are similar, for different earthquakes intensities, up to 13% difference is observed between them. Studies have shown that the differences in the distribution of probability values in both assumptions do not follow a specific trend.
2. Based on the results of numerical tests (significance tests) and descriptive methods, the assumption of normal distribution for the data is not false, but it is not a strong hypothesis. Because the results of the numerical test oppose this assumption in some cases. Moreover, in some other cases, the results of numerical and descriptive methods for this assumption are not in agreement. Therefore, it is concluded that the findings do not support the assumption of normal distribution for the data used in fragility curve development.
3. The results of both numerical tests, i.e., Shapiro–Wilk and Kolmogorov–Smirnov confirm the accuracy of the log-normal distribution assumption for statistical data with a high probability. In this regard, the results of descriptive tests also confirm the accuracy of this assumption. In addition, there is consistency between the findings of both numerical and descriptive and graphical tests. Accordingly, the log-normal distribution assumption for statistical data used in the process fragility curve development is verified.

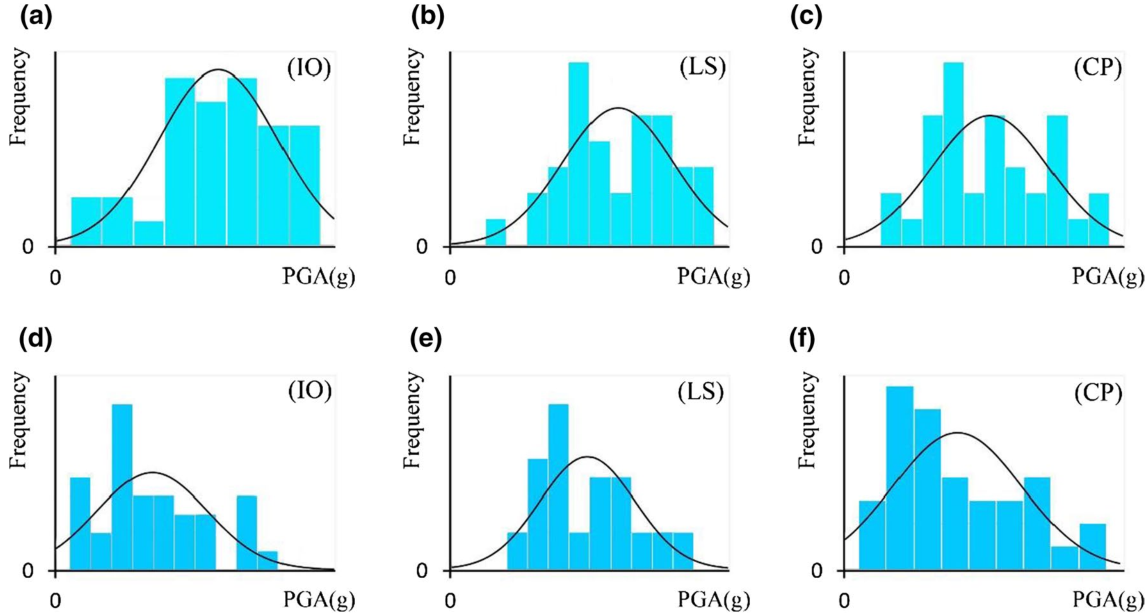

Fig. 9 The histogram diagrams at different performance levels for 5-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

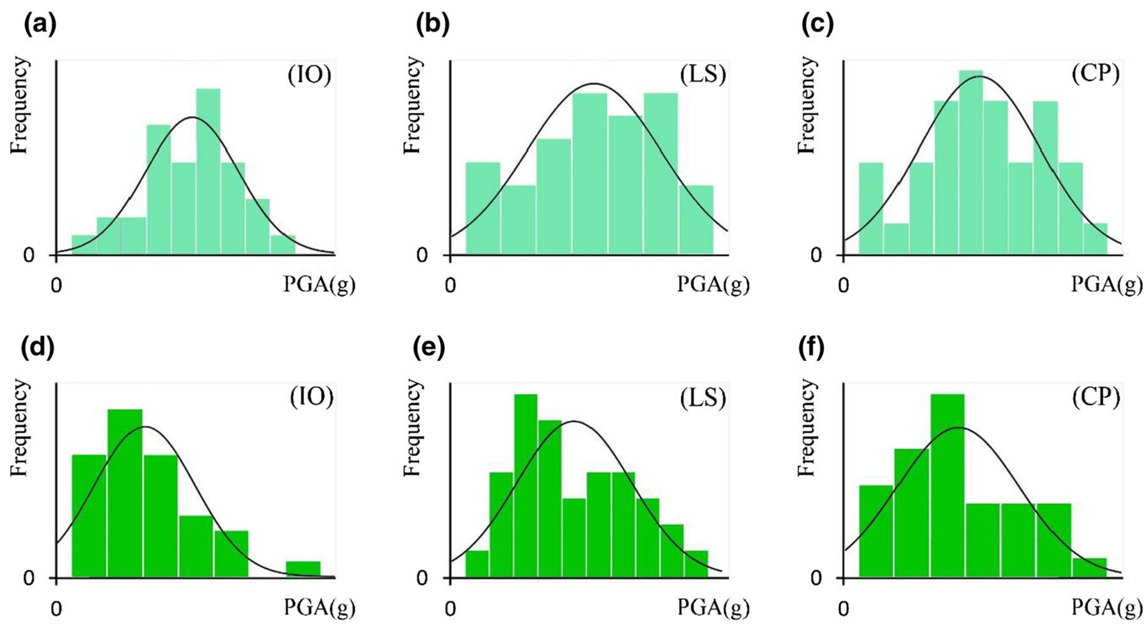

Fig. 10 The histogram diagrams at different performance levels for 7-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

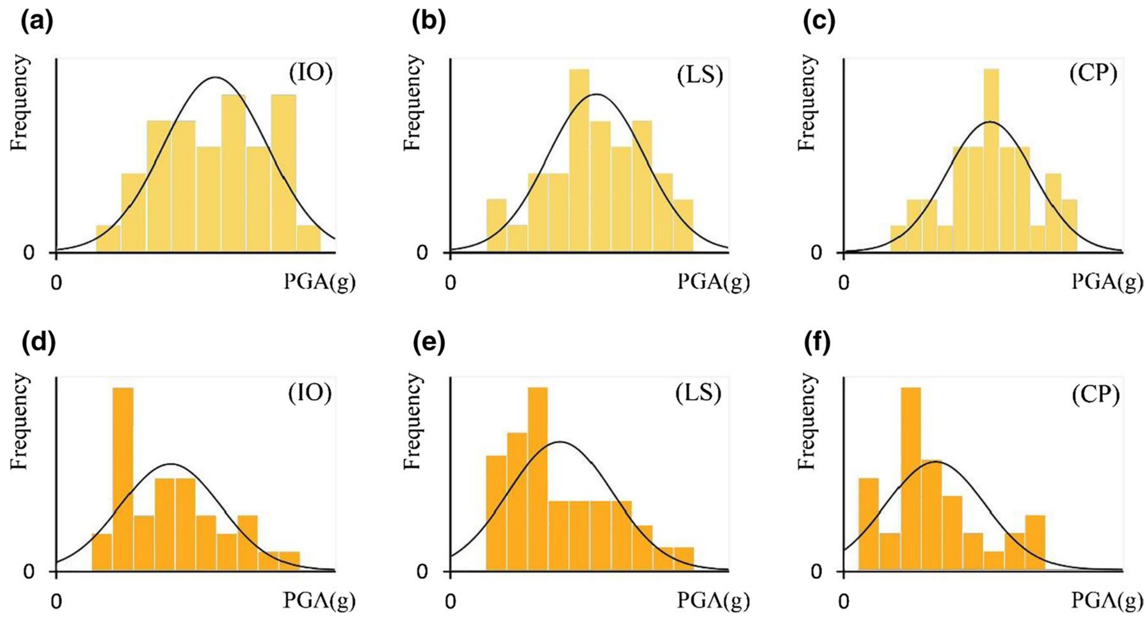

Fig. 11 The histogram diagrams at different performance levels for 10-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

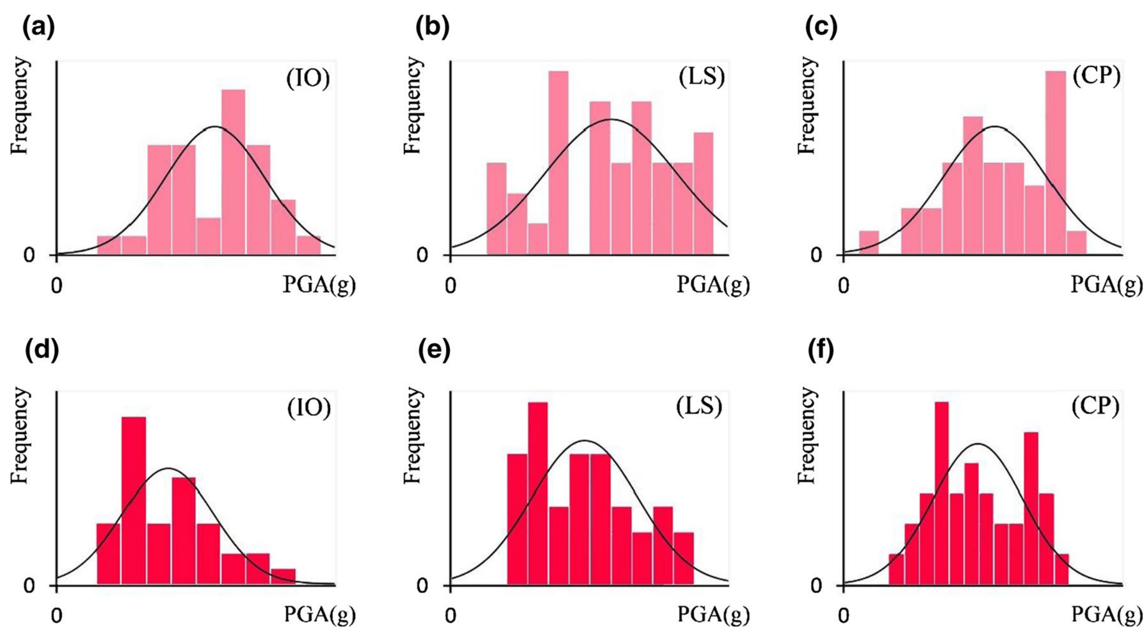

Fig. 12 The histogram diagrams at different performance levels for 12-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

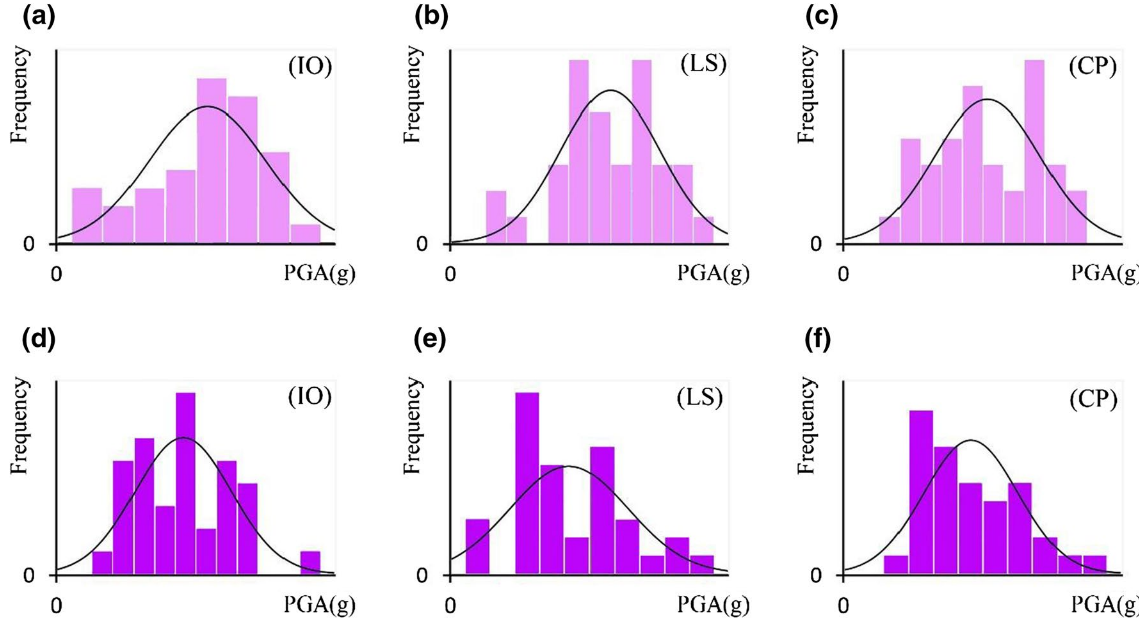

Fig. 13 The histogram diagrams at different performance levels for 15-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

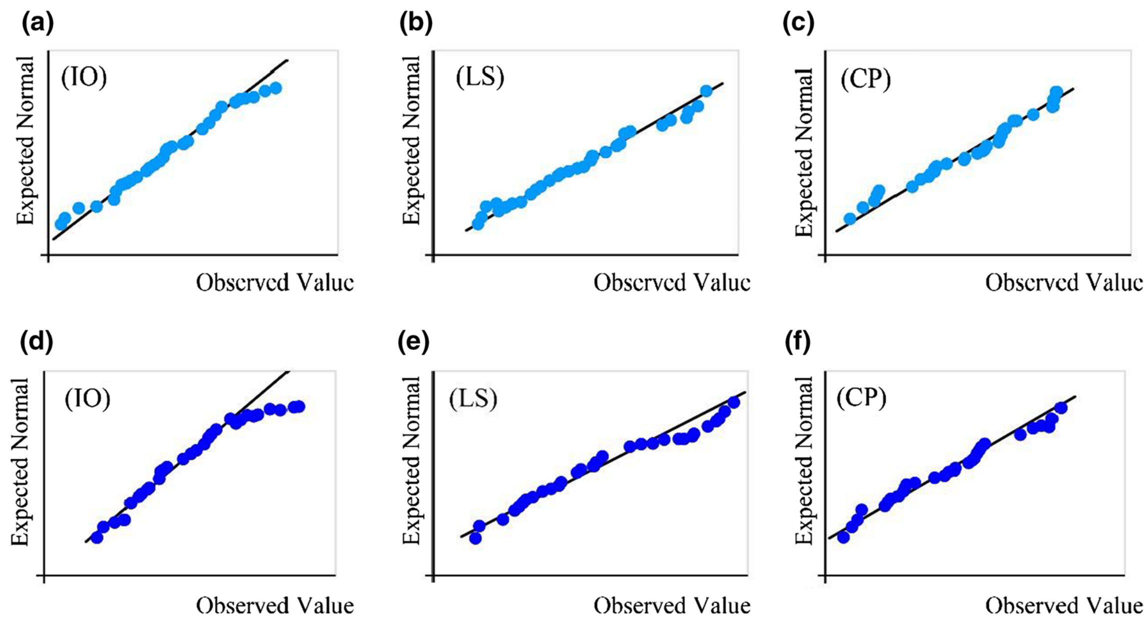

Fig. 14 The Q-Q plots for 3-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

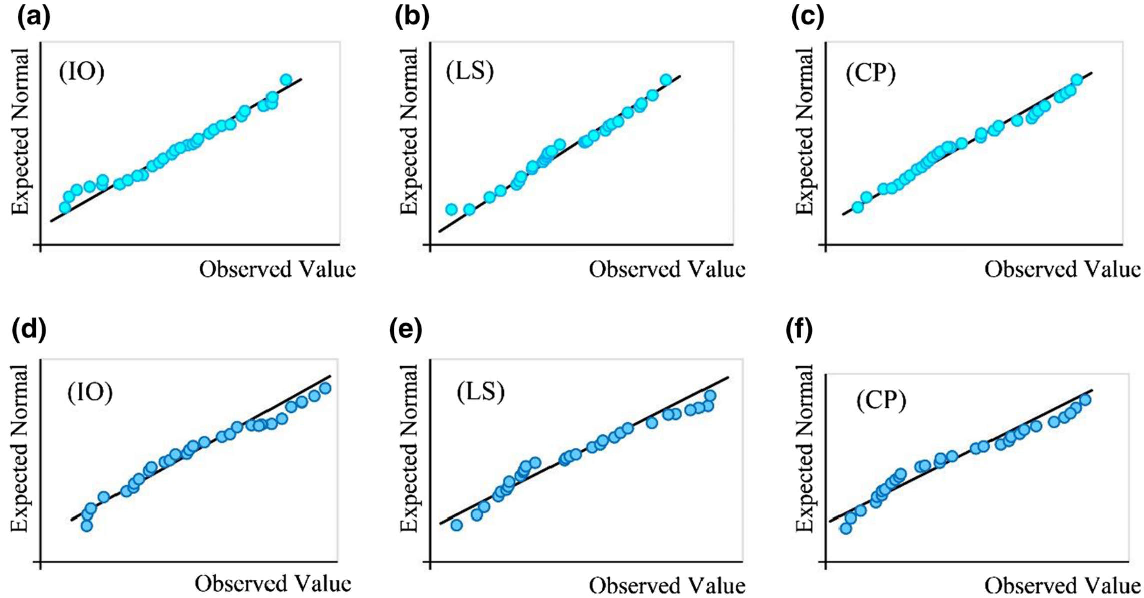

Fig. 15 The Q-Q plots for 5-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

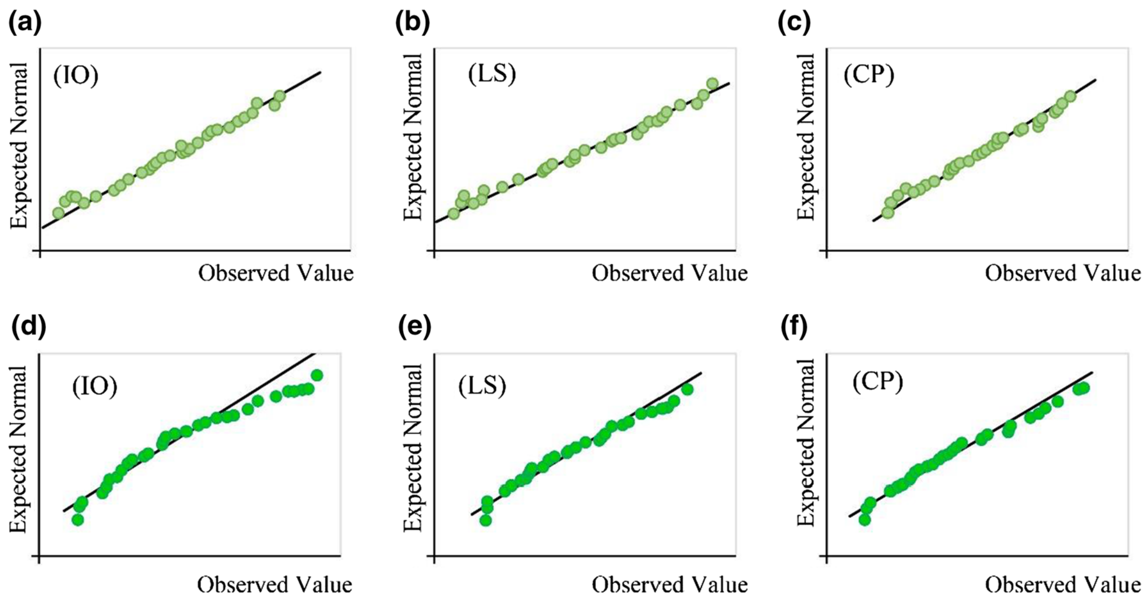

Fig. 16 The Q-Q plots for 7-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

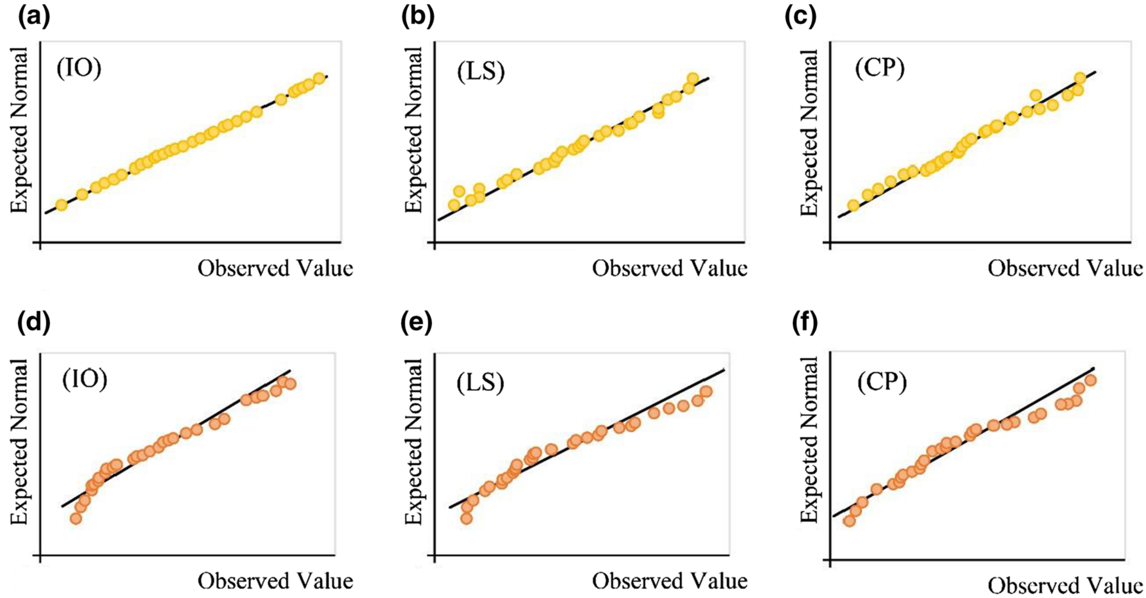

Fig. 17 The Q-Q plots for 10-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

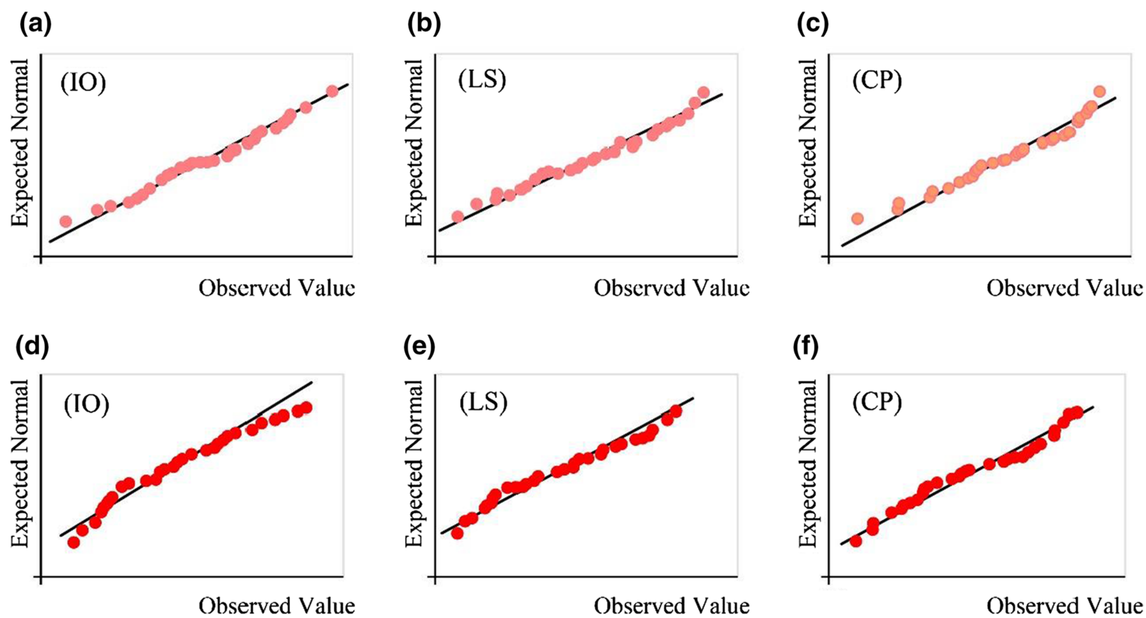

Fig. 18 The Q-Q plots for 12-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

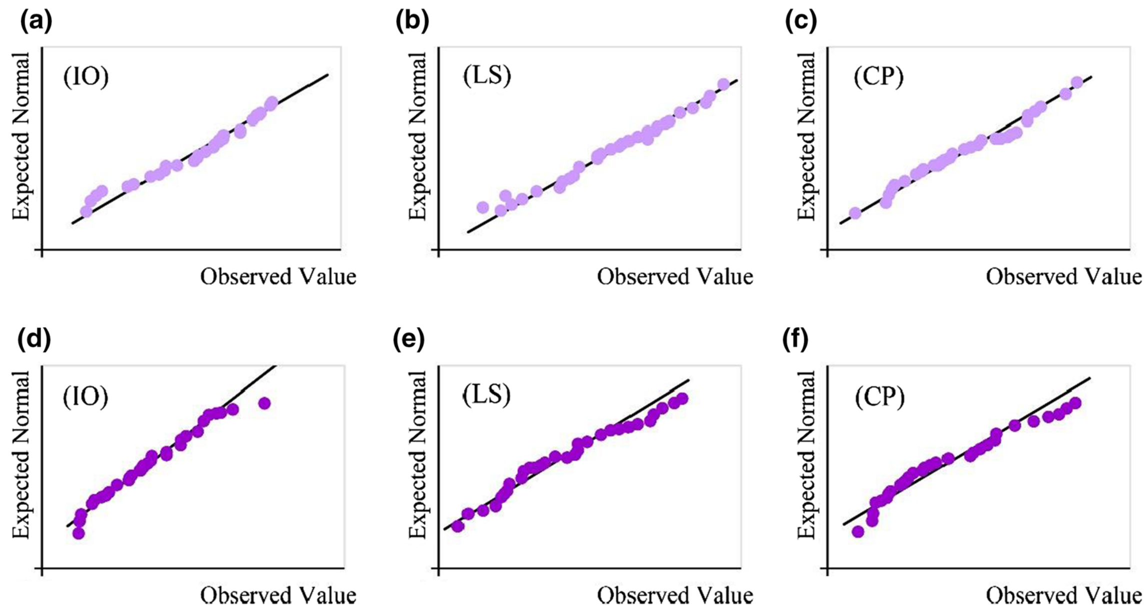

Fig. 19 The Q-Q plots for 15-story frame (**a**, **b** and **c**) log-normal distribution (**d**, **e** and **f**) normal distribution

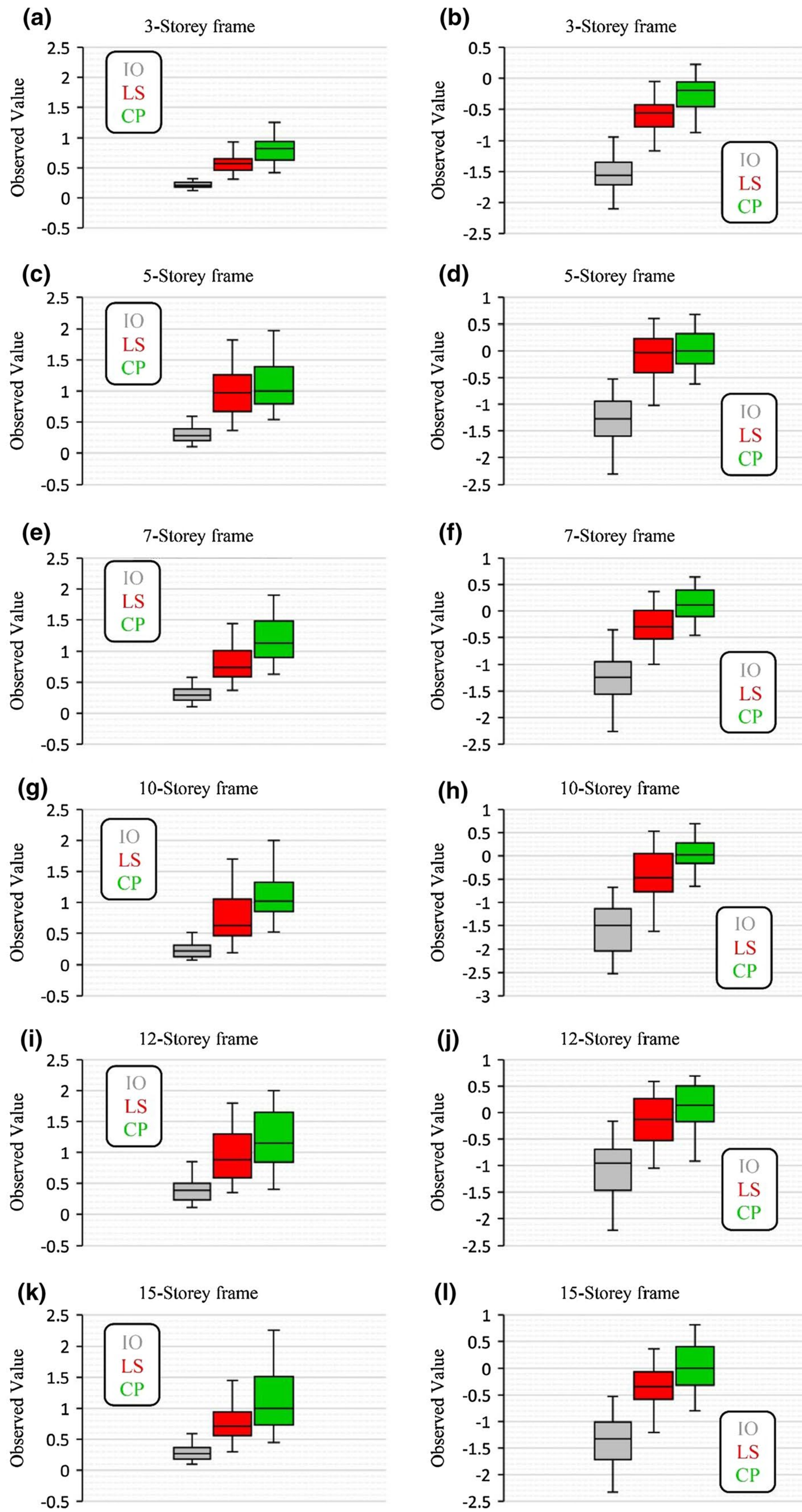

Fig. 20 The box plots at different performance levels of the studied structures (**a**, **c**, **e**, **g**, **i** and **k**) normal distribution (**b**, **d**, **f**, **h**, **j** and **l**) log-normal distribution

Authors Contribution Credit roles for the paper: "Evaluation of the Probabilistic Distribution of Statistical Data Used in the Process of Developing Fragility Curves". Conceptualization: Vahid Mohsenian; Data Curation: Vahid Mohsenian; Formal Analysis: Vahid Mohsenian, Alireza Arabshahi; Funding Acquisition: -; Investigation: Vahid Mohsenian, Alireza Arabshahi, Nima Gharaei-Moghaddam; Methodology: Vahid Mohsenian; Project administration: Vahid Mohsenian, Nima Gharaei-Moghaddam; Resources: Vahid Mohsenian, Alireza Arabshahi, Nima Gharaei-Moghaddam; Software: Alireza Arabshahi, Vahid Mohsenian; Supervision: Nima Gharaei-Moghaddam; Validation: Vahid Mohsenian, Alireza Arabshahi; Visualization: Vahid Mohsenian, Alireza Arabshahi; Writing—original draft: Vahid Mohsenian, Nima Gharaei-Moghaddam; Writing – review & editing: Nima Gharaei-Moghaddam;

Funding This study was not funded by any company.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmadi, M., & Ebadi Jamkhaneh, M. (2021). Numerical investigation of energy dissipation device to improve seismic response of existing steel buildings with Soft-First-Story. *International Journal of Steel Structures, 21*, 691–702.

AISC (2010). Specification for Structural Steel Buildings, ANSI/AISC 360-10. American Institute of Steel Construction.

Altman, D. G., & Bland, J. M. (1995). Statistics notes: The normal distribution. *Bmj, 310*, 298.

Ang, A. H., & Tang, W. H. (2007). *Probability concepts in engineering planning: Emphasis on applications to civil and environmental engineering*. John Wiley and Sons.

ASCE. (2010). *Minimum Design Loads and Associated Criteria for Buildings and Other Structures, ASCE/SEI 7–10*. American Society of Civil Engineers.

ASCE. (2017). *Seismic Evaluation and Retrofit of Existing Buildings ASCE/SEI 41–17*. American Society of Civil Engineers.

ASTM (2019). *Standard Specification for Carbon Structural Steel—ASTM A36/A36M-19*. ASTM International

Baharvand, A., & Ranjbaran, A. (2020). A new method for developing seismic collapse fragility curves grounded on state-based philosophy. *International Journal of Steel Structures, 20*, 583–599.

Barton, P. (2005). A guide to data analysis and critical appraisal. *Review International Journal of Endocrinology and Metabolism,* 486–489.

CSI (2015). Structural and Earthquake Engineering Software, ETABS, Extended three dimensional analysis of building systems nonlinear. 15.2.2 ed., Computers and Structures Inc.

CSI (2017). Structural and earthquake engineering software, PERFORM-3D nonlinear analysis and performance assessment for 3D structures. 7.0.0 ed., Computers and Structures Inc.

Das, K. R., & Imon, A. (2016). A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics, 5*, 5–12.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*, 486.

Ghowsi, A. F., & Sahoo, D. R. (2013). Seismic performance of buckling-restrained braced frames with varying beam-column connections. *International Journal of Steel Structures, 13*, 607–621.

Han, S. W., & Chopra, A. K. (2006). Approximate incremental dynamic analysis using the modal pushover analysis procedure. *Earthquake Engineering and Structural Dynamics, 35*, 1853–1873.

Hao, R. N. (2011). Critical compassionate pedagogy and the teacher's role in first-generation student success. *New Directions for Teaching and Learning, 2011*, 91–98.

Statistics, I. (2013). IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. IBM Corp. *Google Search*.

Kim, D.-H., & Leon, R. T. (2013). Fragility analyses of mid-rise T-stub PR frames in the Mid-America earthquake region. *International Journal of Steel Structures, 13*, 81–91.

Kim, S.-H., & Shinozuka, M. (2004). Development of fragility curves of bridges retrofitted by column jacketing. *Probabilistic Engineering Mechanics, 19*, 105–112.

Lee, J., Kong, J., & Kim, J. (2018). Seismic performance evaluation of steel diagrid buildings. *International Journal of Steel Structures, 18*, 1035–1047.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia, 22*, 67.

Mohsenian, V., Filizadeh, R., Ozdemir, Z. & Hajirasouliha, I. (2020). Seismic performance evaluation of deficient steel moment-resisting frames retrofitted by vertical link elements. Structures, Elsevier, 724–736.

Mohsenian, V., Hajirasouliha, I. & Filizadeh, R. (2021c). Seismic reliability assessment of steel moment-resisting frames using Bayes estimators. *Proceedings of the Institution of Civil Engineers-Structures and Buildings*, 1–15.

Mohsenian, V., Filizadeh, R., Hajirasouliha, I., & Garcia, R. (2021). Seismic performance assessment of eccentrically braced steel frames with energy-absorbing links under sequential earthquakes. *Journal of Building Engineering, 33*, 101576.

Mohsenian, V., Hajirasouliha, I., & Filizadeh, R. (2021b). Seismic reliability analysis of steel moment-resisting frames retrofitted by vertical link elements using combined series–parallel system approach. *Bulletin of Earthquake Engineering, 19*, 831–862.

Mohsenian, V., & Mortezaei, A. (2019). New proposed drift limit states for box-type structural systems considering local and global damage indices. *Advances in Structural Engineering, 22*, 3352–3366.

Montazeri, M., Ghodrati Amiri, G., & Namiranian, P. (2021). Seismic fragility and cost-benefit analysis of a conventional bridge with retrofit implements. *Soil Dynamics and Earthquake Engineering, 141*, 106456.

Nowak, A. S., & Collins, K. R. (2012). *Reliability of structures*. CRC Press.

Öztuna, D., Elhan, A. H., & Tüccar, E. (2006). Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. *Turkish Journal of Medical Sciences, 36*, 171–176.

PEER. *PEER ground motion database* [Online]. Available: http://peer.berkeley.edu/peer-ground-motion-database [Accessed].

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*, 21–33.

Shafaei, H., & Naderpour, H. (2020). Seismic fragility evaluation of FRP-retrofitted RC frames subjected to mainshock-aftershock records. *Structures, 27*, 950–961.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591–611.

Shinozuka, M., Feng, M. Q., Lee, J., & Naganuma, T. (2000). Statistical analysis of fragility curves. *Journal of Engineering Mechanics, 126*, 1224–1231.

Shome, N. (1999). *Probabilistic seismic demand analysis of nonlinear structures*, Stanford University.

Sudret, B., Mai, C. & Konakli, K. (2014). Computing seismic fragility curves using non-parametric representations. *arXiv preprint* arXiv:1403.5481.

Thode, H. C. (2002). *Testing for normality*. CRC Press.

Vamvatsikos, D., & Allin Cornell, C. (2006). Direct estimation of the seismic demand and capacity of oscillators with multi-linear static pushovers through IDA. *Earthquake Engineering and Structural Dynamics, 35*, 1097–1117.

Vamvatsikos, D., & Cornell, C. A. (2002). Incremental dynamic analysis. *Earthquake Engineering and Structural Dynamics, 31*, 491–514.

Vamvatsikos, D., & Cornell, C. A. (2004). Applied incremental dynamic analysis. *Earthquake Spectra, 20*, 523–553.

Zareian, F., Krawinkler, H., Ibarra, L., & Lignos, D. (2010). Basic concepts and performance measures in prediction of collapse of buildings under earthquake ground motions. *The Structural Design of Tall and Special Buildings, 19*, 167–181.

Zuo, Y., Li, W., & Li, M. (2019). Seismic fragility analysis of steel frame structures containing initial flaws in beam-column connections. *International Journal of Steel Structures, 19*, 504–516.