SCIENTIFIC Publishing Limited

Prevention and Treatment of Natural Disasters

https://ojs.ukscip.com/index.php/ptnd

Research Article

Modeling Dependence of Peak Floor Acceleration and Maximum Interstory Drift Ratios with Gaussian Copulas

Huda Munjy * and Stephanie Botros

Department of Civil and Environmental Engineering, California State University, Fullerton, CA 92831, USA

* Correspondence: hmunjy@fullerton.edu; Tel.: +1-559-907-7956

Received: 16 January 2025; **Revised:** 3 February 2025; **Accepted:** 16 February 2025; **Published:** 18 February 2025

Abstract: This study introduces a multivariate demand model for Engineering Demand Parameters (EDPs) in Performance Based Seismic Design (PBSD), utilizing Gaussian copulas to characterize the dependence structure of the demand vector. The effectiveness of this approach is assessed by comparing EDPs generated using Gaussian copulas against those assumed under a joint lognormal distribution. This validation study is further carried forward to values of economic loss for the four special steel moment frames obtained via the three sets of EDPs. The Performance Assessment Calculation Tool (PACT) developed by the Federal Emergency Management Agency (FEMA) P-58 (2015) is used for loss estimation. Results indicate that using copulas to represent the dependence structure of EDPs better captures the characteristics of the population of EDPs rather than assuming a joint lognormal distribution. Distributions of economic loss generated using copulas match the loss generated from the true observations of EDPs better than loss generated assuming a joint lognormal distribution. The sample size of the selected and scaled ground motions required for the generation of realizations of building response via nonlinear dynamic analysis is also investigated, which proves to yield more accurate values of response but, at the expense of using a larger number of initial observations.

Keywords: Earthquake Engineering; Seismic Performance Assessment; Loss Estimation; Statistical Modeling; Engineering Demand Parameters

1. Introduction

Accurately assessing earthquake-related losses is essential for mitigating the risks associated with seismic events. Performance-Based Seismic Design (PBSD) has emerged as a critical framework for achieving this goal, focusing on designing structures that meet specific seismic performance objectives while addressing key factors such as financial impact, fatalities, and operational downtime. PBSD requires integrating conditional probabilities at multiple stages, including site hazard, structural response, damage assessment, and loss estimation, while managing uncertainties across these stages. Core components of this process include estimating ground motion hazards, predicting structural responses, and evaluating the resulting damages and losses. One of the most crucial elements in PBSD is the quantification of economic losses due to seismic damage, as this directly informs the effectiveness of the design approach.

Since its inception in 2001, organizations such as FEMA, PEER, and ATC have recognized the potential of PBSD in improving earthquake resilience. Collaborating with FEMA, ATC developed the Next-Generation Performance-Based Seismic Design Guidelines [1], which laid the foundation for more comprehensive seismic performance assessments. In particular, FEMA introduced FEMA P-58, a set of guidelines that provide a framework for performance assessment and a tool known as the Performance Assessment Calculation Tool (PACT), designed to practically implement these methodologies [2]. However, several studies that have evaluated FEMA P-58 using real-world data, such as from the Canterbury earthquake sequence, have highlighted discrepancies between predicted and actual structural losses [3, 4].

FEMA P-58 models the dependence structure of Engineering Demand Parameters (EDPs)—critical variables for quantifying earthquake-induced damage and losses—using a joint lognormal distribution. Previous research has generally assumed that individual EDPs follow lognormal distributions but has not sufficiently tested whether the collective distribution, or EDP vector, truly fits this model. Esmaili et al. (2016), addressing sample size limitations, applied Bayesian statistics while maintaining the assumption of a joint lognormal distribution for EDP dependencies [5]. Moreover, peak floor acceleration (PFA) and maximum interstory drift ratio (MaxIDR), two pivotal EDPs, have been examined for their influence on structural damage. Studies have found that PFA's fit to the lognormal distribution is varied, especially in cases involving collapse scenarios [6]. Accurate modeling of EDPs is thus essential for precisely quantifying earthquake-induced damage and estimating economic losses. Building upon this prior research, Goda (2010) and Goda and Tesfamariam (2015) explored alternative methods for modeling EDP dependencies, such as copulas, which enable joint probabilistic modeling of multiple EDPs. These studies connected such probabilistic models to economic loss scenarios, improving our understanding of earthquake impacts. This study extends this line of inquiry by comparing the use of copulas versus joint lognormal distributions for generating simulations of PFA and MaxIDR, two key EDPs in PBSD. This comparison is critical for enhancing the accuracy of seismic response assessments in high-risk regions like Los Angeles, where the annual economic risk due to earthquakes can reach hundreds of millions of dollars [7, 8].

While seismic performance assessment is heavily reliant on the accurate statistical representation of EDPs, current design guidelines often assume a uniform statistical dependence structure—typically a joint lognormal distribution—for all EDPs. This assumption, although widely used, may lead to inefficiencies in estimating seismic losses. Therefore, this study investigates the differences in estimated earthquake losses by comparing simulations of PFA and MaxIDR generated using copulas versus those using a joint lognormal distribution. PFA and MaxIDR were selected due to their direct impacts on structural acceleration and drift, providing a comprehensive view of structural response. In addition, a Kolmogorov-Smirnov test is employed to determine the optimal number of initial ground motions required for accurate seismic response assessments when EDPs are modeled using copulas versus a joint lognormal distribution. Copulas, widely used in fields like financial risk management for modeling multivariate dependencies [9, 10], offer a robust alternative that better captures the complex relationships between multiple datasets. The methodologies are applied to special steel moment-resisting frame buildings in Los Angeles, a region characterized by high seismic risk and significant annual economic exposure [11].

This research makes two key contributions: (1) it evaluates the use of copulas and joint lognormal distributions for modeling the dependencies between EDPs and estimating economic losses from earthquakes in buildings, and (2) it offers recommendations on the optimal number of ground motions required for precise seismic response assessments in building evaluations.

2. Materials and Methods

2.1. Modeling Dependence Structure of EDPs Using Lognormal Distribution Assumption

Performance-Based Earthquake Engineering (PBEE) methodology is utilized to probabilistically assess the resilience of structural systems during seismic events. It enables various stakeholders—engineering designers, developers, and building owners—to collaboratively determine the desired level of building performance while ensuring compliance with building code standards. The PBEE process typically involves hazard analysis, structural analysis, and damage-loss assessment. In structural analysis, multiple scenarios of demand parameters are simulated based on a limited set of initial inputs. Mathematically, PBEE methods are often represented using a triple integral formulation derived from the total probability theorem [12]:

$$v\,(DV) = \int\int\int G\langle DV|DM\rangle\,|dG\langle DM|EDP\rangle|\;dG\langle EDP|IM\rangle\,|d\lambda(\,IM) \tag{1}$$

IM represents the intensity measure derived from hazard analysis, EDP denotes the engineering demand parameter obtained through structural analysis, DM corresponds to the damage measure from damage analysis, DV signifies the decision variable from loss analysis, and λ(IM) denotes the mean annual rate of exceedance from probabilistic seismic hazard analysis procedures. These parameters can be scalar or vectors. This mathematical representation is intricate and encompasses multiple layers of variability. This study specifically explores the variability emerging immediately after structural analysis, focusing on the methods used to generate realizations of EDPs, and the variability in estimated losses stemming from these methods. Spectral acceleration at the fundamental period of vibration serves as a single-variable IM, while peak floor acceleration (PFA) and maximum interstory drift ratio (MaxIDR) are considered as multivariate EDPs.

According to FEMA P-58-1 Appendix G (2015), simulated sets of demands are determined using Monte Carlo simulation, generating numerous demands from a small number of analyses under the assumption that EDPs follow a joint lognormal distribution. Following nonlinear time history analysis (NLTHA), results from structural analysis are organized into an m×n matrix, where m represents the number of rows (one per analysis), and n represents the number of columns (one per peak EDP, e.g., PFA, MaxIDR). This EDP matrix is assumed to follow a joint lognormal distribution. Post obtaining the demand matrix, the typical procedure involves determining median values for each parameter and the covariance matrix, then mathematically simulating a large number of demand vectors using a random number selection process based on the median and covariance matrices. The number of required analyses varies depending on several factors. Appendix G of FEMA P-58-1 notes that datasets with non-full rank covariance matrices (n > m, where the number of variables exceeds the number of initial inputs) may indicate insufficient data for generating the covariance matrix. Nonetheless, PACT is designed to handle such non-full rank EDP covariance matrices. Yang developed the algorithm employed for generating simulated demands [13, 14]. Equations (2) and (3) are used for simulating EDPs, where Z represents the vector of natural log demand parameters; U stands for a matrix of uncorrelated standard normal variables with a mean of 0 and an identity covariance matrix; and L and D denote the lower-triangular decomposition of the correlation matrix and the matrix of standard deviations obtained from the covariance matrix of EDPs, respectively.

$$Z = \lambda U + \mu \tag{2}$$

$$\lambda = LD \tag{3}$$

2.2. Modeling Dependence Structure of EDPs Using Copulas

Copulas are mathematical tools that can capture the dependence structure of multivariate data. They have several applications and have been widely applied in the fields of finance, insurance, reliability theory, and more recently, in the field of hydrology (Genst and Favre, 2007). Consider two EDPs, X and Y, whose distribution functions are $F\left(x\right) = P\left[X \leq x\right]$ and, $G\left(x\right) = P\left[Y \leq y\right]$, respectively. The joint distribution of X and Y is therefore $H\left(x, y\right) = \left[X < x,\ Y < y\right]$. For all EDPs (x, y) we now have three other numbers that can be associated: $F\left(x\right)$, $G\left(y\right)$, and $H(x, y)$, with all of these numbers lying in the interval $[0, 1]$. Therefore, each ordered pair $\left(F\left(x\right), G\left(y\right)\right)$ corresponds to a number for the joint distribution $H(x, y)$ that lies within $[0, 1]$ and that this link between the ordered pair and the joint distribution is called a copula [15], and according to Sklar's theorem, this copula is represented by C in Equation (4) [16].

$$H\left(x, y\right) = C\left(F\left(x\right), G\left(y\right)\right) \tag{4}$$

Using Sklar's theorem, the joint distribution can be modeled using the marginal distributions of the EDPs and the copula, without having any information regarding the joint distribution of the EDPs. This allows for greater flexibility when modeling the dependence structure of random multivariate data, like that of EDPs. For the dependence structure of X and Y, an empirical copula is used, represented by Equation (5) where R_i and S_i are the rank of X_i and Y_i in ascending order, respectively and u and v represent $F\left(x\right)$ and $F\left(y\right)$, respectively [10].

$$C_n\left(u, v\right) = \frac{1}{n}\sum\nolimits_{i=1}^{n} 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \tag{5}$$

For modeling dependencies using copulas, commonly used linear correlation coefficients, such as Pearson's ρ, are not suitable as these linear measures of correlation are not preserved by the copula. This means that two pairs of variables that have the same copula can have differing linear correlation coefficients. Therefore, Kendall's tau, denoted by τ, is used as it is a constant of the copula, which means that any correlated variables with the same copula will have the τ that corresponds to that copula. Therefore, when the marginal distributions are continuous, Kendall's τ only depends on the copula and not on the marginal distributions of the variables. Kendall's τ is calculated as $P\,[concordance] - P\,[discordance]$, where two pairs of observations of EDPs, (X_i, Y_i) and (X_j, Y_j), are concordant if $X_i < X_j$ and $Y_i < Y_j$ or if $X_i > X_j$ and $Y_i > Y_j$ and are discordant if $X_i < X_j$ and $Y_i > Y_j$ or if $X_i > X_j$ and $Y_i < Y_j$. Kendall's τ can be written in the form of Equation (6) where, for random variables X_1 and X_2, $(\tilde{X}_1, \tilde{X}_2)$ is the expectation of (X_1, X_2) [17]. If both probabilities are equal, then $\tau\,(X_1, X_2) = 0$.

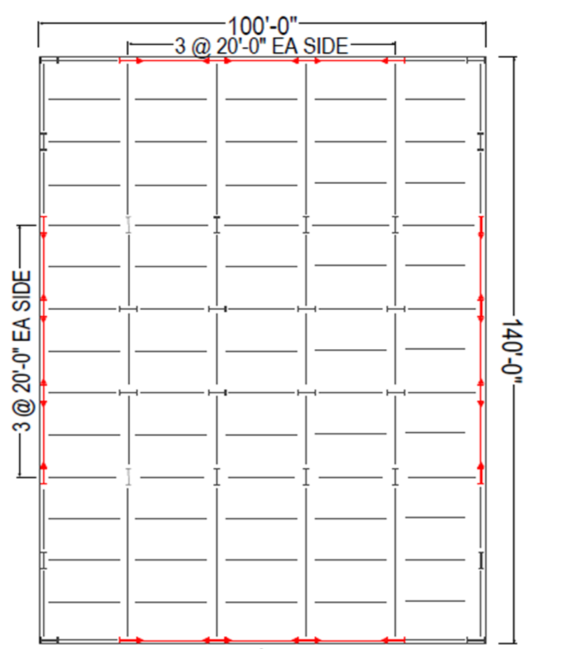
$$\tau\,(X_1, X_2) = P\left[\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right) > 0\right] - P\left[\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right) < 0\right] \tag{6}$$

One of the primary challenges in utilizing copulas to model the dependence structure of variables lies in selecting the appropriate parametric family or distribution of the copula that effectively captures the relationship within the data. In this study, a Gaussian copula is chosen due to its widespread application in handling multivariate data. Furthermore, previous research has demonstrated that a Gaussian copula can accurately represent the characteristics of the dependence structure between maximum interstory drift ratio and peak floor acceleration [8]. The objective of this study is to showcase the application of Gaussian copulas in accurately modeling the dependence of Engineering Demand Parameters (EDPs) to yield precise estimates of seismic loss, rather than comparing EDPs modeled under different copula families. Hence, Gaussian copulas exclusively serve to model the dependence structure of EDPs in this research.

3. Results

3.1. Structural Modeling and Ground Motion Selection and Modification for Case Study SMRFs

Four special steel moment resisting frames (SMRFs) are studied as case study structures. These include buildings of 2, 4, 8, and 12 stories, each having first-mode building periods of 0.92, 1.61, 2.28, and 3.10 seconds, respectively. The design of these buildings follows ASCE/SEI 7-02 [16] and ANSI/AISC 341-05 [17] standards for a site located in downtown Los Angeles, California. Detailed plan and elevation views of the buildings can be seen in **Figure 1**.



SMRF Beam and Column Schedule			
Beam		Column	
Label	Size	Label	Size
SMF BM1	W21x57	SMF C1	W24x31
SMF BM2	W21x73	SMF C2	W24x62
SMF BM3	W16x31	SMF C3	W24x33
SMF BM4	W30x132	SMF C4	W24x103
SMF BM5	W21x68	SMF C8	W24x84
SMF BM6	W24x84	SMF C9	W24x146
SMF BM7	W27x94	SMF C10	W24x162
SMF BM8	W30x116	SMF C11	W24x207
SMF BM9	W30x108		

Figure 1. *Cont.*

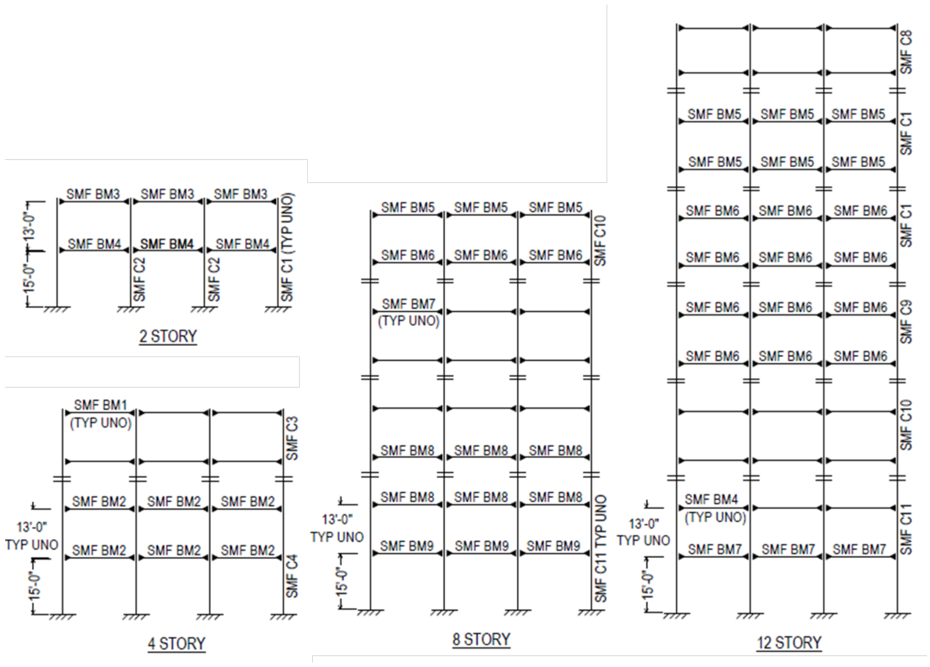


Figure 1. Plan and elevation views of case study structures.

Static pushover curves, illustrating the general load-deflection relationship, are presented in **Figure 2.** For a comprehensive understanding of the modeling approach for these SMRFs, please refer to the National Institute of Standards and Technology GCR 10-917-8 report (2010).

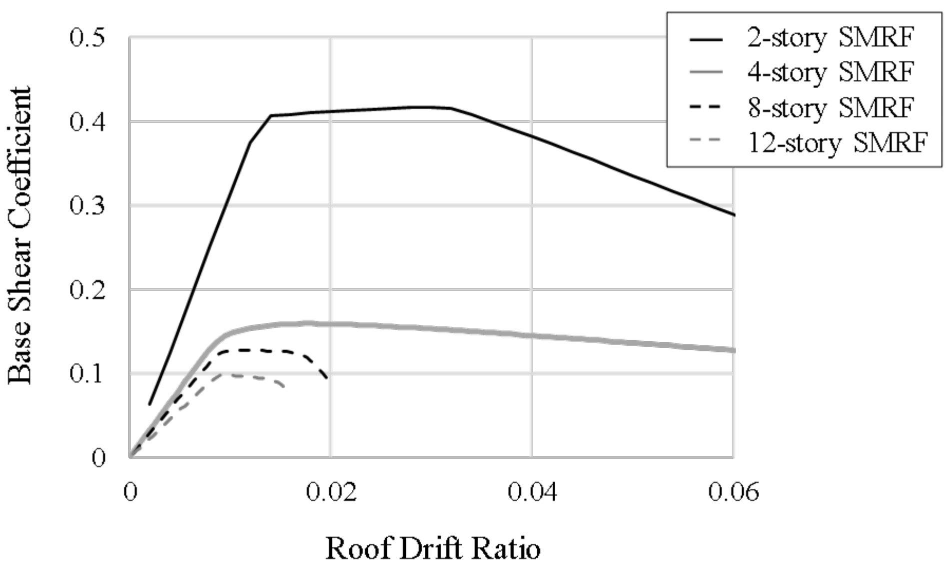


Figure 2. Static pushover curve for 2-, 4-, 8-, and 12-story case study SMRFs.

Ground motion selection and scaling (GMSM) is done for a total of 200 ground motions per case study building based on the Conditional Mean Spectrum (CMS) method [3]. Disaggregation for all buildings is accomplished using Open Source Seismic Hazard Analysis [18]. GMSM is performed for 10% probability of exceedance in 50 years seismic hazard (i.e., equivalent to the design basis earthquake, DBE), and a frequent event with 50% probability of exceedance in 50 years seismic hazard (Vs_{30} = 760 m/s is used). These are cases at which seismic loss and damages are highly sensitive to EDPs with not much contribution to collapse cases for well-engineered structures. The ground motions are selected from the Next Generation of Attenuation Relationships (PEER-NGA) database, which does not include records of event foreshocks or aftershocks. 100 ground motions per hazard level (100 for 10% in 50 years and 100 for 50% in 50 years) per building are selected and scaled to match the target conditional mean spectra for the location of the buildings (all buildings are modeled at the same location).

Nonlinear time history analyses (NLTHA) are conducted to derive engineering demand parameters (EDPs) such as peak floor acceleration (PFA) and maximum inter-story drift ratio (IDR) at all story levels. The resulting EDPs from 100 selected and scaled ground motions per hazard level per case study structure are referred to as population EDPs. These EDPs are directly generated through NLTHA without assuming any dependence between them. They serve as the baseline for comparing against EDPs generated under assumptions of copula and joint lognormal distributions.

3.2. Sample EDPs Generated Using Gaussian Copulas

To evaluate the accuracy of engineering demand parameters (EDPs) generated with varying initial data points, random samples of 5, 11, 15, 25, 50, and 100 points are selected from a collection of 100 EDPs (referred to as population EDPs) obtained from nonlinear response history analyses. For each sample size, 100 sets of EDPs are simulated using Gaussian copulas (referred to as copula EDPs) and joint lognormal distributions (referred to as lognormal EDPs). This process is repeated across different case study SMRFs (2-, 4-, 8-, 12-story) and hazard levels (10% and 50% in 50 years). For each sample size (5, 11, 15, 25, 50, 100), subsets of EDPs are randomly chosen from the population dataset to generate 100 sets of copula EDPs and 100 sets of lognormal EDPs. This procedure is iterated multiple times per initial sample size, encompassing each case study SMRF and hazard level. To assess how well the generated EDPs match their respective populations, a Kolmogorov-Smirnov (KS) goodness of fit test is utilized. The KS test is preferred for its nonparametric nature, enabling evaluation of distributional conformity without assuming adherence to a specific distributional form, unlike methods such as Student's t-test.

The KS statistic measures the greatest distance between two empirical distribution functions and is calculated by Equation (7) where D_{mn} is the KS test statistic, $\underset{EDP_i}{sup}$, denotes the supremum, $F_m\,(EDP_i)$ is the cumulative distribution function of each population EDP_i, and $F_n\,(EDP_i)$ is the cumulative distribution function for each sample EDP_i.

$$D_{mn} = \underset{EDP_i}{sup}\ |F_m\,(EDP_i) - F_n\,(EDP_i)| \tag{7}$$

The null hypothesis, H_0, posits that the sample data follows the same distribution as the population, while the alternative hypothesis, H_a, suggests that the sample data does not follow the same distribution as the population. A total of 60 random samples (6 different initial sample sizes) are drawn from the population EDPs for each hazard level and case study SMRF. Each random sample from the population EDPs is used to generate 100 sets of copula EDPs and 100 sets of lognormal EDPs. **Figures 3–6** depict the results of the KS tests, where each number in the table represents the percentage of failures among the 10 random samples for each sample size considered.

2-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	10	20	10	0	10
PFA1	20	10	10	10	0	10
PFA2	30	20	10	10	20	0
IDR1	10	20	10	10	10	0
IDR2	10	10	20	10	0	0

2-story SMRF Copula Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	20	0	0	10	0
PFA1	10	20	10	10	0	0
PFA2	20	20	20	0	0	0
IDR1	20	10	0	0	10	10
IDR2	10	10	0	20	0	0

Figure 3. *Cont.*

2-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	20	10	10	10	10	0
PFA1	20	10	0	20	10	10
PFA2	30	20	20	20	20	10
IDR1	10	20	20	10	10	0
IDR2	10	30	30	20	10	0

2-story SMRF Copula Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	10	0	0	0	0
PFA1	20	10	0	0	0	10
PFA2	20	10	10	20	0	0
IDR1	20	20	0	20	0	0
IDR2	10	10	0	10	0	0

Figure 3. Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 2-story SMRF and 10% and 50% in 50 years hazard with the greyscale indicating levels of failure.

4-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	20	20	40	40	20	10
PFA1	50	40	30	30	10	10
PFA2	50	30	50	50	40	30
PFA3	30	20	20	20	30	30
PFA4	30	30	30	40	10	30
IDR1	40	30	30	10	10	30
IDR2	50	30	30	30	20	30
IDR3	20	20	20	40	20	30
IDR4	40	30	30	40	20	30

4-story SMRF Copula Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	40	30	10	0	0
PFA1	50	30	10	30	0	0
PFA2	10	20	30	10	10	10
PFA3	20	40	20	30	0	0
PFA4	20	20	30	40	20	0
IDR1	50	30	20	10	10	10
IDR2	40	40	30	10	20	10
IDR3	50	20	40	20	30	30
IDR4	40	40	20	20	10	30

4-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	40	40	40	30	30	0
PFA1	50	30	40	30	30	20
PFA2	50	40	30	30	10	20
PFA3	40	30	30	30	20	30
PFA4	30	30	30	20	10	20
IDR1	30	30	40	20	10	30
IDR2	40	20	40	20	20	10
IDR3	20	30	40	20	20	20
IDR4	30	20	20	30	30	20

4-story SMRF Copula Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	20	10	40	10	10
PFA1	20	20	10	20	10	0
PFA2	50	30	20	20	10	0
PFA3	40	50	30	20	20	10
PFA4	30	40	20	10	0	0
IDR1	60	20	30	20	0	30
IDR2	60	20	30	10	10	20
IDR3	20	20	30	10	20	10
IDR4	30	30	40	10	20	20

Figure 4. Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 4-story SMRF and 10% and 50% in 50 years hazard with the greyscale indicating levels of failure.

Darker shades of grey indicate higher frequencies of failures for that specific sample size and EDP type. As the number of stories increases, the minimum required sample size to predominantly match the population distributions also increases. For instance, in the case of the 2-story SMRF, the highest observed failure percentage across all cases is 30%, with some instances showing all samples matching the population distribution. Conversely, for the 8-story building, there are cases where failure rates reach 60% with smaller sample sizes, particularly notable with lognormal EDPs even when using a full rank covariance matrix. Furthermore, with a sample size of 15 for the 4-story SMRF—exceeding the number of EDPs (9 EDPs for the 4-story SMRF)—instances of 40% failure are observed at both hazard levels when employing a joint lognormal distribution. Even with an increased sample size of 25 at the 10% hazard level, several cases exhibit 40% or more failure rates in matching the population distribution. However, in lower-rise buildings (specifically 2-story and 4-story systems) with smaller sample sizes, lower rates of failure are observed for lognormal EDPs compared to copula EDPs. For instance, the 4-story copula EDPs exhibit two instances of 60% failure under the 50% in 50 years hazard level, which is higher than observed with lognormal EDPs. This disparity can be attributed to the fundamental characteristics of each method: a joint lognormal distribution imposes a strict model on the variables, resulting in more consistent failure rates that do not fluctuate significantly with changes in sample size. In contrast, copulas provide greater flexibility and require less stringent assumptions about the joint distribution of variables, leading to reduced failure rates as sample size increases. In summary, the recommendations regarding initial sample size are as follows: for achieving higher accuracy in simulated demand sets, albeit with reduced efficiency, utilize copulas with sufficient initial observations to establish a full rank covariance matrix. Conversely, for optimizing efficiency at the cost of accuracy in simulated demand sets, employ the joint lognormal assumption with a smaller number of initial observations.

8-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	70	40	50	30	30	20
PFA1	70	50	50	30	30	20
PFA2	60	40	40	20	20	20
PFA3	30	40	30	30	40	20
PFA4	30	50	20	30	30	10
PFA5	30	40	20	10	30	20
PFA6	50	30	40	20	30	10
PFA7	50	40	50	20	40	30
PFA8	50	40	50	30	30	30
IDR1	60	40	50	10	20	30
IDR2	60	50	50	40	30	30
IDR3	60	50	50	40	30	30
IDR4	40	50	40	30	30	30
IDR5	40	50	50	40	30	10
IDR6	60	70	60	50	40	40
IDR7	60	40	40	30	30	30
IDR8	60	30	50	50	40	30

8-story SMRF Copula samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	40	30	20	10	0
PFA1	50	50	30	20	10	0
PFA2	50	40	30	10	10	0
PFA3	60	20	50	20	10	10
PFA4	40	20	30	30	10	0
PFA5	40	20	30	20	20	20
PFA6	40	30	40	20	20	10
PFA7	40	40	30	20	30	10
PFA8	40	20	30	20	10	20
IDR1	40	30	20	30	20	10
IDR2	60	40	40	20	20	20
IDR3	40	40	40	10	10	30
IDR4	40	30	30	10	10	30
IDR5	40	40	30	10	20	20
IDR6	40	60	20	20	30	10
IDR7	40	40	20	30	20	20
IDR8	50	40	20	20	10	20

8-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	30	40	20	20	10	20
PFA1	30	40	40	20	10	20
PFA2	40	30	30	20	20	30
PFA3	30	30	40	30	40	10
PFA4	50	40	40	10	30	10
PFA5	60	50	50	40	20	0
PFA6	60	40	40	40	20	20
PFA7	50	50	40	20	10	30
PFA8	60	30	30	30	40	30
IDR1	60	30	30	40	20	20
IDR2	60	30	40	30	20	20
IDR3	50	30	40	30	20	10
IDR4	50	40	50	30	10	20
IDR5	60	40	40	40	20	20
IDR6	60	40	30	30	20	10
IDR7	40	40	40	20	30	10
IDR8	50	50	40	20	30	20

8-story SMRF Copula samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	60	30	30	10	10	10
PFA1	50	50	30	20	10	10
PFA2	50	40	30	20	10	10
PFA3	60	40	30	40	10	30
PFA4	60	40	20	20	20	30
PFA5	40	40	40	30	30	20
PFA6	40	30	40	30	20	30
PFA7	70	40	40	30	20	0
PFA8	60	40	30	40	10	20
IDR1	40	50	40	30	20	20
IDR2	40	40	40	20	30	10
IDR3	40	60	30	20	30	10
IDR4	50	30	20	10	50	10
IDR5	60	20	30	30	20	10
IDR6	60	20	20	30	20	0
IDR7	50	40	20	10	10	10
IDR8	60	40	20	20	20	10

Figure 5. Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 8-story SMRF and 10% and 50% in 50 years hazard with the greyscale indicating levels of failure.

12-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	40	60	50	60	30	30
PFA1	40	60	40	60	40	30
PFA2	50	60	40	60	40	30
PFA3	60	50	50	50	50	20
PFA4	60	50	40	60	20	20
PFA5	60	60	50	50	20	10
PFA6	40	60	50	30	10	20
PFA7	60	70	60	50	20	20
PFA8	70	50	40	50	30	20
PFA9	60	50	30	60	30	30
PFA10	70	40	40	30	10	30
PFA11	70	40	30	30	30	10
PFA12	70	40	30	40	20	20
IDR1	50	50	40	40	50	10
IDR2	60	60	30	30	20	20
IDR3	60	40	50	30	20	20
IDR4	40	40	50	30	30	20
IDR5	40	40	40	50	40	30
IDR6	50	30	40	30	40	20
IDR7	40	40	50	40	10	10
IDR8	50	40	30	40	10	10
IDR9	60	50	30	30	30	20
IDR10	70	60	40	40	20	20
IDR11	70	50	30	40	20	10
IDR12	70	60	30	50	20	30

12-story SMRF Copula samples 10% in 10 years

Sample Size	5	11	15	25	50	100
PGA	60	50	20	50	40	10
PFA1	70	50	40	20	30	10
PFA2	60	50	50	10	30	0
PFA3	50	50	40	50	20	30
PFA4	60	60	30	40	40	20
PFA5	50	50	40	30	10	20
PFA6	30	50	40	30	20	10
PFA7	40	40	40	30	10	20
PFA8	40	40	30	30	30	20
PFA9	30	40	20	30	30	0
PFA10	50	50	40	40	30	10
PFA11	40	40	30	20	30	30
PFA12	40	50	50	40	20	10
IDR1	40	50	40	50	40	20
IDR2	50	50	60	20	40	30
IDR3	60	60	50	20	30	30
IDR4	30	50	60	10	30	10
IDR5	40	40	40	20	0	0
IDR6	50	30	40	50	10	10
IDR7	60	50	20	30	20	20
IDR8	50	50	60	10	10	20
IDR9	40	40	40	50	30	0
IDR10	50	40	40	40	30	10
IDR11	60	50	50	20	20	10
IDR12	60	40	40	50	20	30

12-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	80	50	20	50	30	20
PFA1	60	40	50	30	40	10
PFA2	60	50	50	40	40	20
PFA3	30	60	60	50	30	20
PFA4	40	60	40	40	20	20
PFA5	30	70	50	20	30	30
PFA6	50	50	60	60	30	20
PFA7	50	60	50	50	0	10
PFA8	50	70	30	40	30	10
PFA9	40	70	40	50	20	20
PFA10	70	70	20	50	10	10
PFA11	60	40	60	30	40	20
PFA12	80	50	40	40	20	30
IDR1	60	60	40	40	20	30
IDR2	50	50	30	30	30	30
IDR3	60	50	50	30	40	20
IDR4	40	60	40	40	40	30
IDR5	40	40	30	50	20	10
IDR6	50	30	40	20	50	30
IDR7	50	60	50	40	40	20
IDR8	50	60	60	40	20	10
IDR9	60	70	30	30	10	10
IDR10	60	50	30	30	20	30
IDR11	50	40	40	50	20	20
IDR12	60	50	50	60	30	20

12-story SMRF Copula samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	60	70	40	30	10	0
PFA1	60	40	50	50	30	30
PFA2	70	60	60	20	40	20
PFA3	60	60	50	20	20	20
PFA4	50	50	50	40	20	10
PFA5	60	50	60	30	0	30
PFA6	70	70	40	10	10	20
PFA7	40	50	30	30	40	30
PFA8	80	40	30	20	20	0
PFA9	50	30	30	30	30	10
PFA10	60	50	40	10	0	20
PFA11	70	60	50	20	30	20
PFA12	40	60	40	30	20	10
IDR1	60	70	40	30	20	0
IDR2	50	50	60	40	10	10
IDR3	70	70	40	60	30	10
IDR4	50	40	40	30	30	10
IDR5	60	50	40	30	40	20
IDR6	50	60	40	40	20	30
IDR7	40	60	50	30	20	30
IDR8	50	50	30	40	10	30
IDR9	70	50	60	20	20	30
IDR10	50	70	20	20	20	20
IDR11	70	60	50	10	10	10
IDR12	70	40	20	40	10	20

Figure 6. Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 12-story SMRF and 10% and 50% in 50 years hazard with the greyscale indicating levels of failure.

3.3. Estimation of Economic Loss

The Performance Assessment Calculation Tool (PACT), developed by the Applied Technology Council (ATC), is a computer-based calculation tool that includes a repository of fragility and consequence data to perform probabilistic calculations and accumulation of losses according to the methodology described in FEMA P-58-1 (2015). A PACT model is developed for each case study SMRF, and several tests are run to generate the economic loss associated with the input demand matrices. Collapse cases are not considered for these analyses as the aim of this study is to assess the effect of different methods for generating EDPs on estimated loss, not on whether or not the buildings collapse. Moreover, the probability of collapse of the buildings used in this study is insignificant at the studied hazard levels and chances of attaining a collapse damage state are highly unlikely. Therefore, realizations indicating a collapse limit state are limited to a couple, if any, which would only minutely affect one tail of the loss distribution. For each case study SMRF, the *population* EDPs for the 10% and 50% hazard levels are input into PACT to generate losses (herein referred to as *population* loss).

Along with *population* loss, the economic loss associated with 100 joint lognormal (herein referred to as *lognormal* loss) and 100 *copula* EDPs (herein referred to as *copula* loss) is also generated. Based on the results from the previous section, for each building, enough initial realizations are used to create a full rank covariance matrix. Therefore, for the 2- and 4-story buildings, 11 initial *population* EDPs are used to generate 100 *copula* and *lognormal* EDPs and for the 8- and 12-story buildings, 25 and 50 initial *population* EDPs are used to generate 100 *copula* and *lognormal* EDPs, respectively. All of the structural and nonstructural components in the case study SMRFs are defined by fragility and loss functions based on the normative values provided in Appendix F of FEMA P-58-1 (2015) for the 'Research' occupancy category. All values of monetary loss resulting from PACT analyses were normalized by average maximum loss of damageable components, which was calculated per building by pushing the buildings to maximum capacity, forcing them to the largest possible damage state. The following are the monetary values of average maximum loss, in millions of dollars, for the 2-, 4-, 8- and 12- story buildings, respectively: 7.46, 14.92, 29.84, 44.76.

The results of these analyses are shown in **Figures 7–10**, where the black asterisk curves represent the CDFs of *population* loss, and the red dotted curves represent the CDFs of the *copula* and *lognormal* EDPs, respectively.

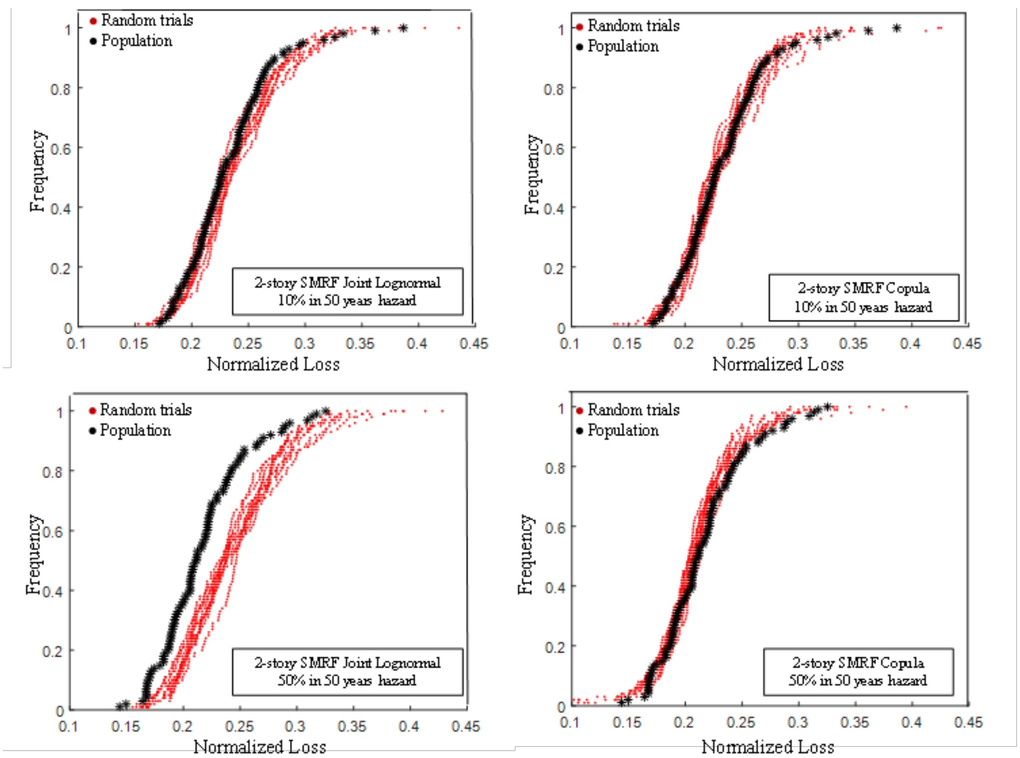


Figure 7. Empirical cumulative distribution functions of loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 2-story SMRF considering 10% and 50% in 50 years hazard.

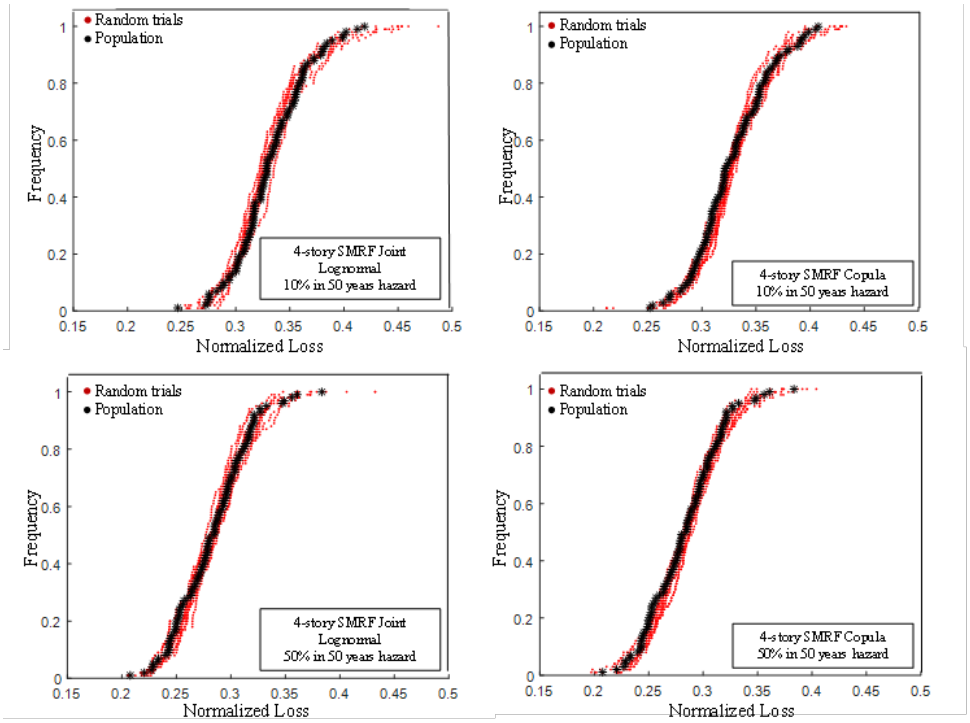


Figure 8. Empirical cumulative distribution functions of loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 4-story SMRF considering 10% and 50% in 50 years hazard.

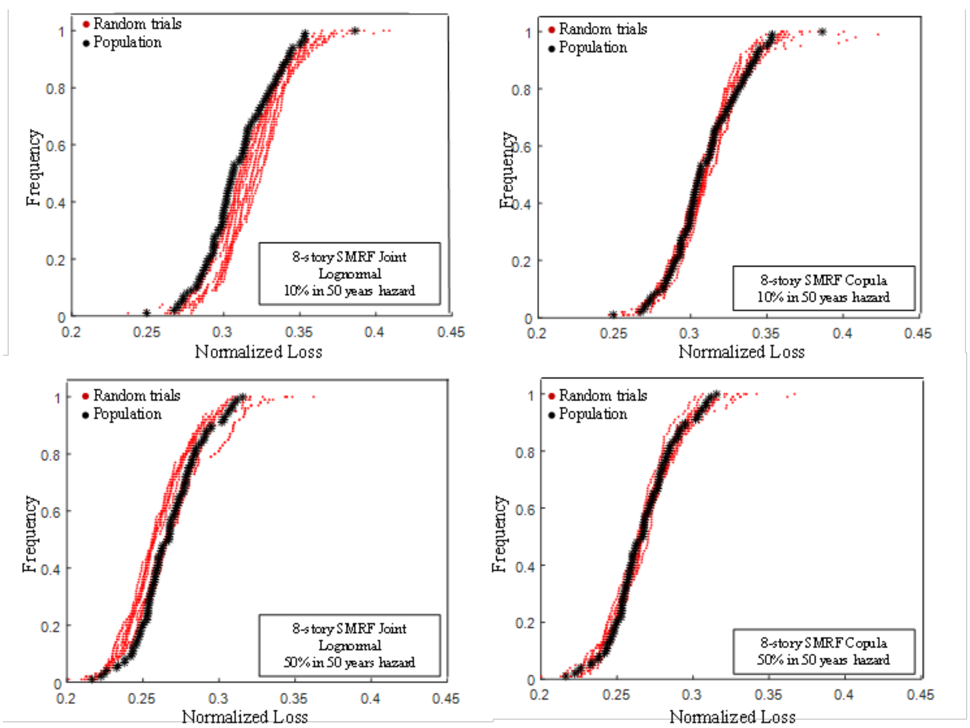


Figure 9. Empirical cumulative distribution functions of loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 8-story SMRF considering 10% and 50% in 50 years hazard.

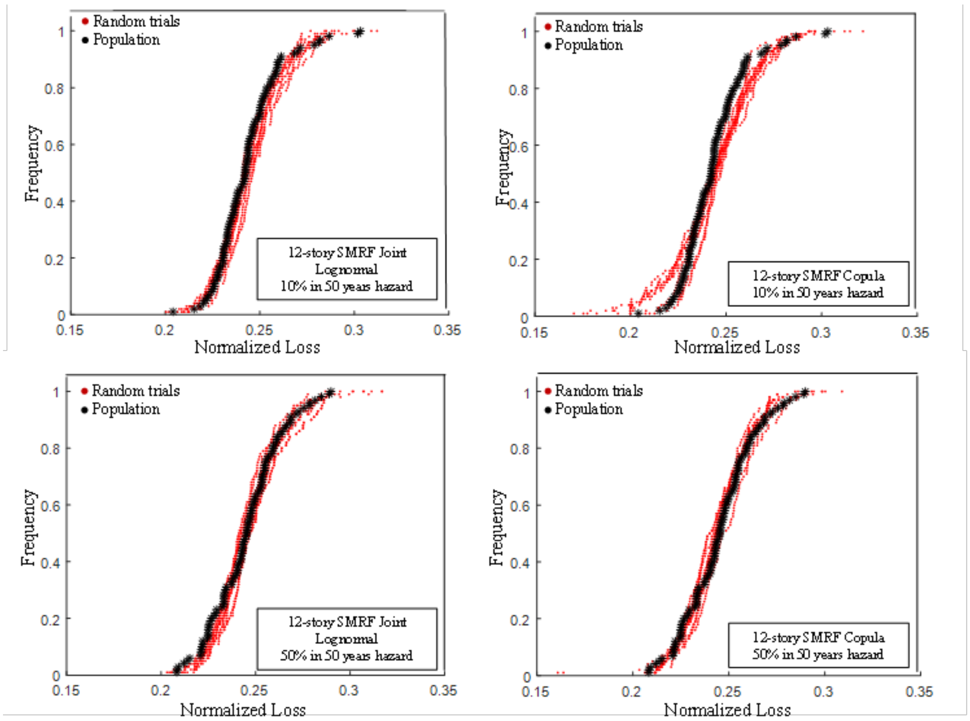


Figure 10. Empirical cumulative distribution functions of loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 12-story SMRF considering 10% and 50% in 50 years hazard.

For the 2-story SMRF, both *copula* and *lognormal* EDPs perform well in estimating loss considering 10% in 50 years hazard, but for the 50% in 50 years hazard level, *copula* EDPs are capable of replicating losses generated by the *population* EDPs. While, in general, both methods are acceptable in matching loss generated from *population* EDPs, for the majority of cases, *copula* EDPs are superior to *lognormal* EDPs except for the 4-story building and 8-story at 10% in 50 years hazard. In the latter, *lognormal* EDPs generate loss values that are better at matching the *population* loss. It is worth noting that as story height increases, a smaller difference in loss is noticed between the two ground motion hazard levels, which is especially noticeable for the 12-story building. The demand for the higher rise systems was relatively small which caused most of the damage to fall within the same damage state for both hazard levels, which then corresponds to similar values of loss, regardless of ground motion intensity. To quantify the difference between the distribution of losses obtained from *copula* and *lognormal* EDPs, with *population* loss, respectively, Kullback-Leibler (KL) divergence is used. KL divergence measures the distance between one probability distribution and another reference distribution, which is calculated. A KL divergence of 0 corresponds to two identical distributions. Let P and Q be discrete probability distributions and D_{KL} be the KL divergence between P and Q, represented by Equation (8).

$$D_{KL}(P||Q) = \sum\nolimits_{x \in X} P(x)\log(\frac{Q(x)}{P(x)}) \quad (8)$$

The KL divergence of each *copula* and *lognormal* loss CDF and *population* loss CDF are calculated. For each hazard level and SMRF, the average KL divergence is calculated and presented in **Table 1**. For example, the average KL divergence of all 10 generated *copula* loss samples with *population* loss for the 2-story building is 0.042. In all but one case (12-story 10% in 50 years hazard) *copula* loss has a smaller KL divergence compared with *lognormal* loss. These findings are in line with previous results which showed that *copula* EDPs can better match *population* EDPs at smaller sample sizes in most cases, especially when a full rank covariance matrix is achieved.

Table 1. Average KL Divergence values between population and simulation CDFs for 10% and 50% probability of exceedance in 50 years.

	Lognormal (10% in 50 years)	Copula (10% in 50 years)	Lognormal (50% in 50 years)	Copula (50% in 50 years)
2-story	0.0196	0.0186	0.0617	0.0237
4-story	0.025	0.0161	0.0145	0.0132
8-story	0.0475	0.0301	0.0142	0.0125
12-story	0.0118	0.0123	0.0183	0.0116

4. Discussion

This study builds on previous research in seismic performance assessment, particularly the work done by FEMA and ATC with the development of FEMA P-58 and the associated Performance Assessment Calculation Tool (PACT) [1, 2]. While these guidelines have been widely adopted in earthquake engineering, studies evaluating their application, such as those using data from the Canterbury earthquake sequence [3, 4], have identified discrepancies between predicted and actual structural losses, underscoring the limitations of assuming a joint lognormal distribution for EDPs. The research conducted by Aslani (2010) and Goda and Tesfamariam (2015) [6, 7] demonstrated the potential of copula-based probabilistic models to better capture the dependencies between multiple EDPs, an area this study expands upon. In particular, our findings align with Goda and Tesfamariam's (2015) conclusions, as the use of Gaussian copulas in this study led to more accurate EDP simulations and loss estimations compared to the joint lognormal distribution. Moreover, while Esmaili et al. (2016) applied Bayesian statistics to address sample size limitations while maintaining the joint lognormal assumption [5], this study's results reinforce that Gaussian copulas, when paired with sufficient ground motion data, outperform the lognormal approach in both accuracy and reliability, especially for higher-rise buildings. In this context, our study contributes to the ongoing evolution of performance-based seismic design methodologies by providing a more accurate and efficient means of modeling EDP dependencies and enhancing the precision of seismic loss predictions.

This study introduces an approach employing Gaussian copulas to model the dependence among engineering demand parameters (EDPs) used in assessing the seismic performance of buildings. Current guidelines, as outlined in FEMAP-58 (2015), adopt a methodology assuming EDPs follow a joint lognormal distribution. By using Gaussian copulas, this study aims to generate a comprehensive set of EDPs without relying on potentially inaccurate assumptions about EDP dependencies. Peak floor acceleration and maximum interstory drift ratios from selected and scaled ground motion records are obtained for four special steel moment resisting frame buildings. These values serve as the basis for generating EDPs that exhibit the dependence structure modeled by Gaussian copulas, as well as conforming to the joint lognormal distribution prescribed by FEMAP-58 (2015). Findings indicate that for lower-rise buildings (e.g., 2- and 4-story systems), assuming a joint lognormal distribution can achieve approximately 30–40% accuracy in simulated demand sets even with small initial sample sizes. However, Gaussian copulas demonstrate instances where simulated demands achieve 90–100% accuracy, contingent upon using a larger number of initial observations to achieve this precision. Conversely, even with increased initial observations, EDPs generated under a joint lognormal distribution do not consistently attain the same level of accuracy as those generated using Gaussian copulas. In higher-rise buildings (e.g., 8- and 12-story systems), both Gaussian copulas and joint lognormal distributions yield inaccurate results when using significantly smaller sample sizes relative to the number of variables. However, as the sample size increases, Gaussian copulas show a clear trend toward higher accuracy in EDP simulations, whereas this trend is less pronounced for joint lognormal distributions. In summary, employing copulas alongside sufficient initial observations to establish a full rank covariance matrix results in more accurate simulated demand sets. In contrast, assuming a joint lognormal distribution may offer greater efficiency but generally results in lower overall accuracy, particularly for lower-rise buildings. For higher-rise buildings, Gaussian copulas generally provide more precise representations of EDPs.

In terms of calculating and assessing losses, the findings indicate that in most instances, losses computed using engineering demand parameters (EDPs) generated with Gaussian copulas closely align with losses derived from observed or population EDPs, more so than losses calculated from EDPs following a joint lognormal distribution. These results underscore the efficacy of statistical tools, such as Gaussian copulas, which require fewer assumptions and can be effectively employed to generate EDP realization vectors. Moreover, the study illuminates the sources of variability inherent in performance-based assessments of economic loss at the EDP level, thereby influencing damage and loss estimation. By proposing a methodology that minimizes the inaccuracies associated with simulating suites of demand sets from a limited number of initial analyses, this research addresses the challenge of reducing reliance on potentially erroneous assumptions in EDP-level performance assessment. Ultimately, this study contributes to advancing a more comprehensive and accurate methodology, enhancing the reliability of engineered structures in seismic risk assessment and mitigation strategies.

The main advantage of copulas is their ability to more accurately represent the relationships and dependencies between EDPs. Joint lognormal distributions assume that the EDPs are individually lognormally distributed and that their dependencies follow a pre-specified correlation structure. However, this assumption often oversimplifies the complexity of real-world data, where EDPs may exhibit non-linear dependencies or dependencies that vary across different conditions, which joint lognormal distributions cannot capture as effectively. Gaussian copulas, on the other hand, provide a more flexible framework that allows for modeling complex, non-linear dependencies between multiple variables without assuming specific marginal distributions for the EDPs. This results in a more accurate representation of the underlying physical relationships between structural responses, as seen in our study, where copulas generated EDP sets with significantly higher accuracy than the joint lognormal distributions, particularly when using a sufficient sample size.

As for why the increased accuracy is more pronounced for Gaussian copulas in high-rise buildings, this can be attributed to the higher complexity in the relationships between EDPs for taller structures. High-rise buildings typically exhibit more complex interdependencies between the structural parameters due to factors such as increased floor interaction, wind forces, and greater sensitivity to seismic motion at various levels of the building. When using smaller sample sizes for high-rise buildings, both methods—the Gaussian copulas and joint lognormal distributions—tend to yield inaccurate results due to the difficulty of capturing the intricate dependencies with limited data. However, as the sample size increases, Gaussian copulas are better able to capture the nuanced dependencies between the multiple EDPs at play, which is critical for accurate simulations. This improvement is less pronounced in joint lognormal distributions because of their inherent limitations in representing complex, non-linear relationships. In other words, the copula method's flexibility in modeling higher-dimensional dependence structures becomes increasingly valuable for high-rise buildings as the sample size grows, leading to more accurate simulations and, consequently, more reliable loss estimations. This trend highlights the superior adaptability of Gaussian copulas for complex structural systems, especially in tall buildings where the interrelationships among EDPs are more intricate and varied.

5. Conclusions

This study demonstrates the advantages of using Gaussian copulas for modeling the dependence structure of Engineering Demand Parameters (EDPs) in Performance-Based Seismic Design (PBSD) over the traditionally assumed joint lognormal distribution. The findings highlight that Gaussian copulas, when paired with an appropriate number of ground motions, lead to more accurate and reliable simulations of EDPs, particularly for higher-rise buildings. In contrast, assuming a joint lognormal distribution, while efficient, can result in lower accuracy, especially for lower-rise buildings where demand simulations often deviate significantly from observed data. This research builds upon previous work in seismic performance assessment, specifically the methodologies outlined by FEMA P-58 and the contributions of other researchers [1, 2, 5–7]. By improving the representation of multivariate dependencies between EDPs, this study provides a more precise approach for assessing economic losses and structural performance in seismic risk evaluations. Furthermore, the application of copulas offers a promising alternative to the joint lognormal assumption, addressing the inefficiencies in current design guidelines and enhancing the accuracy of seismic loss predictions. In conclusion, the integration of copulas into PBSD frameworks offers substantial improvements in the reliability of seismic performance assessments and provides a more robust foundation for earthquake loss mitigation strategies, particularly in high-risk regions.

However, there are some limitations to this study. First, while the use of copulas provided more accurate simulations, it requires a larger sample size of initial ground motions, which may not always be feasible in practice. Additionally, the study focused on a specific class of structural systems—special steel moment resisting frame buildings—limiting the generalizability of the findings to other types of structures or regions with different seismic characteristics. Another limitation is the potential computational complexity of the copula-based approach, which may be more resource-intensive compared to traditional methods, especially for large-scale, real-time assessments. Future work should explore the applicability of this methodology to a broader range of structural systems, investigate ways to optimize sample sizes for ground motion selection, and further assess the computational feasibility for large-scale applications.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

All data used in this research is available upon request. Please contact corresponding author to request this information.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Applied Technology Council (ATC). *FEMA P - 58: Next - Generation Seismic Performance Assessment for Buildings, Volume 2 – Implementation Guide*. Federal Emergency Management Agency: Washington, DC, USA, 2015. Available from: https://www.fema.gov/media-library/assets/documents/106118
2. FEMA. *Next - Generation Methodology for Seismic Performance Assessment of Buildings*. Report No. FEMA P - 58. The Applied Technology Council for the Federal Emergency Management Agency: Washington, DC, USA, 2015.
3. Baker, J.W. Conditional mean spectrum: Tool for ground - motion selection. *J. Struct. Eng.* **2010**, *137*, 322–331.
4. Baker, J.W.; Cremen, G.; Giovinazzi, S.; et al. Benchmarking FEMA P - 58 performance predictions against observed earthquake data–A preliminary evaluation for the Canterbury earthquake sequence. In Proceedings of the 2016 Annual Conference of the New Zealand Society for Earthquake Engineering, Christchurch, New Zealand
5. Esmaili, O.; Grant Ludwig, L.; Zareian, F. Improved performance - based seismic assessment of buildings by utilizing Bayesian statistics. *Earthquake Engi. Struct. Dyn.* **2016**, *45*, 581–597.
6. Aslani, H. Probabilistic Earthquake Loss Estimation and Loss Disaggregation in Buildings [Ph.D. dissertation]. Dept. of Civil and Environmental Engineering, Stanford Univ.: Stanford, CA, USA, 2005.
7. Goda, K.; Tesfamariam, S. Multi - variate seismic demand modelling using copulas: Application to non - ductile reinforced concrete frame in Victoria, Canada. *Struct. Saf.* **2015**, *56*, 39–51.
8. Habiboellah, F. Copulas, Modeling dependencies in Financial Risk Management [Master Thesis]. Vrije Universiteit: Amsterdam, The Netherlands.
9. Genest, C.; Favre, A.C. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.* **2007**, *12*, 347–368.
10. Olshansky, R.B.; Wu, Y. Earthquake risk analysis for Los Angeles County under present and planned land uses. *Environ. Plann. B Plann. Des.* **2001**, *28*, 419–432.
11. Moehle, J.; Deierlein, G.G. A framework methodology for performance - based earthquake engineering. In

Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004.

12. Yang, T.Y.; Moehle, J.; Stojadinovic; B.; et al. *An Application of PEER Performance - Based Earthquake Engineering Methodology*. Research Digest No. 2005 - 1. Pacific Earthquake Engineering Research Center: Richmond, CA, USA.
13. Yang, T.Y., Moehle, J., Stojadinovic, B.; et al. Seismic performance evaluation of facilities: methodology and implementation. *J. Struct. Eng.* **2009**, *135*, 1146–1154
14. Nelsen, R.B. *An Introduction to Copulas.* Springer: New York, NY, USA, 1994; pp. 1–4.
15. Sklar, A. Random variables, joint distribution functions, and copulas. *Kybernetika* **1973**, *9*, 449–460.
16. ASCE. *Minimum Design Loads for Buildings and Other Structures*. ASCE/SEI 7 - 10. American Society of Civil Engineers: Reston, VA, USA, 2010.
17. ANSI, A. *AISC 358–05 Prequalified Connections for Special and Intermediate Steel Moment Frames for Seismic Applications*. American Institute of Steel Construction Inc.: Chicago, IL, USA, 2005.
18. Field, E.H.; Gupta, N.; Gupta, V.; et al. Hazard calculations for the WGCEP - 2002 earthquake forecast using OpenSHA and distributed object technologies. *Seismol. Res. Lett.* **2005**, *76*, 161–167.