applied sciences

MDPI

Article

Stochastic Generation of Peak Ground Accelerations Based on Single Seismic Event Data for Safety Assessment of Structures

Jihoon Seok and Jeeho Lee *

Department of Civil and Environmental Engineering, Dongguk University, Seoul 04620, Republic of Korea; te06088@dgu.edu

* Correspondence: jeeholee@dgu.edu; Tel.: +82-2-2260-3352

Abstract: The Korean Peninsula, characterized by low-to-moderate seismicity, faces a shortage of strong ground motion records, posing challenges for the seismic safety assessment of critical infrastructures. Given the rarity of large-magnitude earthquakes, generating a variety of earthquakes with rational values of Peak Ground Acceleration (PGA) is essential for robust seismic fragility and risk analysis. To address this, a new stochastic approach is proposed to simulate artificial earthquakes at multiple source-to-site distances and derive the probability distribution of PGA based on recorded data from a single seismic event. Two key source parameters, seismic moment and corner frequency, are treated as random variables with a negative correlation, reflecting their uncertainties and dependence on source-to-site distance. The Monte Carlo simulation with copula sampling of the key source parameters generates Fourier spectra for artificial earthquakes, which are transformed into the time domain to yield PGA distributions at various distances. A comparison with recorded data shows that the proposed method effectively simulates ground motion intensities, with no statistically significant differences between the simulated results and recorded data ($p > 0.05$). The present method of determining PGA distributions provides a robust framework to enhance seismic risk analysis for the safety assessment of structures.

Keywords: stochastic approach; peak ground acceleration; monte carlo simulation; copula sampling; seismic safety assessment; artificial earthquake generation; response spectrum

check for updates

Citation: Seok, J.; Lee, J. Stochastic Generation of Peak Ground Accelerations Based on Single Seismic Event Data for Safety Assessment of Structures. *Appl. Sci.* **2024**, *14*, 10031. https://doi.org/10.3390/app142110031

Academic Editors: Heui Yung and Yu Ping Yuen

Received: 7 October 2024
Revised: 29 October 2024
Accepted: 30 October 2024
Published: 3 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

On 15 November 2017, Pohang, South Korea, was struck by an earthquake measuring M_w 5.4 according to the Korea Meteorological Administration (KMA) or M_w 5.5 by the U.S. Geological Survey (USGS), marking it the second largest earthquake event in the country over the past century [1,2]. This event called attention to the seismic risk on the Korean Peninsula, a region generally characterized by low-to-moderate seismicity, and highlighted the necessity for robust seismic safety assessments, especially for critical infrastructures such as high-rise buildings, large dams, and nuclear power plants near the epicenter [3,4]. Given the limited availability of large-scale actual data, it is essential to generate artificial earthquakes that account for the stochastic nature of Peak Ground Acceleration (PGA), which ensures a diverse range of rational input ground motions for comprehensive seismic performance assessments.

Artificial earthquakes can be generated by means of: (1) scaling historical earthquakes, (2) matching a target spectrum through random vibration theory [5], or (3) simulating ground motions through stochastic approaches [6–9]. While scaling historical records offers simplicity and availability, they represent only a limited range of earthquake event scenarios [10–12]. Since a recorded PGA only reflects the strongest ground motion and does not account for surface characteristics or the complexity of seismic events, amplifying historical records to match a target PGA leads to inadequate generation results. The 2017 Pohang Earthquake, with a focal depth of 7 km (KMA) and a rare combination

of reverse and strike-slip faulting [1], exemplifies the limitations of relying solely on historical data for earthquake prediction. Meanwhile, the second method, generating earthquakes to match a design spectrum [5], which is conventionally constructed using a single predetermined PGA and simplified envelope curves, often fails to fully capture the complex dynamic characteristics of actual earthquakes [13]. This is because design spectra frequently underestimate or overestimate the spectral acceleration at certain periods [14,15].

For a stochastic approach to ground motion generation, a Ground Motion Model (GMM) is employed. These models are categorized into two types: empirical and physics-based [6–9]. Empirical GMMs, using simplified representations of seismic and geological properties [8,9], lead to less accurate predictions in capturing the complex interactions of seismic activity [16]. In contrast, physics-based GMMs provide higher accuracy by integrating seismic sources, wave propagation, and site characteristics, making them more suitable for regions like the Korean Peninsula, where historical records of large-magnitude earthquakes are limited [16,17]. Due to the significant computational costs required for 3D crustal modeling [18], this study employs a stochastic point-source model, a simplified version of physics-based GMMs [19].

In regions with low-to-moderate seismic activity, the limited availability of records from similar epicentral distances, despite data collection from various stations, hinders the development of probability distributions of PGA for specific source-to-site distances. PGA's variability is influenced by factors beyond distance and magnitude, including seismic event uncertainties, site conditions, and complex wave propagation. Simple interpolation can lead to inaccuracies, since PGAs fluctuate randomly even at the same distance. Moreover, the limited proximity of stations to epicenters and restricted access to seismic data further complicate sufficient data acquisition. To address these challenges, it is crucial to determine a rational probability distribution for PGA, which can facilitate a well-established description of the target response spectrum for fragility and risk analysis of critical infrastructures subjected to extreme seismic events.

This study proposes a new method for deriving the probability distribution of PGA using seismic data recorded from a single event, with the aim of enhancing structural safety assessments. First, artificial earthquakes are generated stochastically, from which a distribution of PGA is subsequently determined. This resulting probability distribution can be utilized to select an appropriate group of artificial earthquakes for the seismic safety assessment of critical infrastructures, particularly in regions where strong seismic records are extremely rare.

To simulate ground motions using actual seismic records from a single seismic event, the present study employs a point-source stochastic model characterized by two key source parameters: seismic moment (M_0) and corner frequency (f_c) [20]. Due to their inherent uncertainties, these parameters are treated as random variables, which can result in variations in the predicted PGAs [21,22]. Their independence from source-to-site distance further justifies this assumption. Probability distributions are fitted to observed values of both parameters, f_c and M_0, and copula sampling is applied to account for their negative correlation. The optimal copula function, selected from five copula functions based on the highest log-likelihood, is used to model the dependency between f_c and M_0. The Monte Carlo Simulation (MCS) with random sampling of the parameters generates Fourier spectra for artificial earthquakes at various distances, which are then transformed into the time domain to produce their time histories and corresponding PGA distributions. The resulting probability distributions of PGA for these simulated artificial earthquakes at various distances are then compared and validated against recorded earthquake data.

2. Methodology

2.1. Ground Motion Database

The seismic records were collected from 130 stations in the KMA's National Earthquake Comprehensive Information System (NECIS) when the M_w 5.4 earthquake struck Pohang, located in the southeast region of the Korean Peninsula, on 15 November 2017. Out of

these stations, data from 111 stations with complete accelerograms were used to calculate parameters for the stochastic point-source model in the present study. Table 1 shows basic information about the 2017 Pohang Earthquake.

Table 1. Detailed information on the 2017 Pohang Earthquake.

Information	Details
UTC time	15 November 2017 05:29:31
KST time (local time)	15 November 2017 14:29:31
Magnitude	M_L 5.4, M_w 5.4
Depth (km)	7 km
PGA (g)	0.58 g ($R = 7.46$ km)

NECIS provides raw data in the miniSEED format, a simplified version of the Standard for the Exchange of Earthquake Data (SEED) to facilitate the digital exchange of seismic data [23]. This format includes observational data in digital counts, which are numerical values recorded by seismometers before any preprocessing is performed to convert them into physical quantities such as ground motion acceleration, velocity, and displacement [24]. Accurate physical measurements require that seismographs be properly calibrated. Observational data from all stations were sampled at a rate of 100 samples per second, corresponding to an interval of 0.01 s. Calibration sheets for each station are used to assess the sensitivity of the seismometers, ensuring accurate conversion of seismograph units into measurements of gravitational acceleration (g).

2.2. Stochastic Point-Source Model

2.2.1. Parameters of Stochastic Point-Source Model

The Fourier Amplitude Spectrum (FAS) of ground motion can be determined using the stochastic point-source model [6]:

$$A(M_0, R, f) = E(M_0, f)P(R, f)G(f)I(f) \tag{1}$$

where M_0 is the seismic moment, R is the source-to-site distance, and f is the frequency. This equation consists of four components: the source effect, $E(M_0, f)$, which characterizes the earthquake source mechanism; the path effect, $P(R, f)$, which describes the attenuation of seismic waves along their propagation path; the site effect, $G(f)$, which accounts for local site amplification and attenuation; and the type of motion, $I(f)$, which facilitates expressing the spectrum as acceleration $(= -2\pi f)$.

The site effect, $G(f)$, is divided into two distinct factors: the site amplification factor, denoted by $A^*(f)$, and the diminution factor, denoted by $D(f)$:

$$G(f) = A^*(f)D(f) \tag{2}$$

The expression for $D(f)$ involves the high-frequency spectral decay factor, κ_0.

$$D(f) = \exp\{-\pi f \kappa_0\} \tag{3}$$

To determine κ_0, the slope of the Fourier amplitude is calculated for each station that recorded seismic data during the 2017 Pohang Earthquake [25]. For Fourier spectra with a negative slope, the decrease in amplitude beyond a specific frequency is estimated [26]. In this study, a κ_0 value of 0.014 s is employed, where it represents the zero-distance intercept of the values obtained from each station, as originally defined in [25]. The site amplification factor is determined using the horizontal-to-vertical spectral ratio, with a damping ratio of 5% applied at each station [27].

$$A^*(f) = \frac{RSA_H(f)}{RSA_V(f)} \tag{4}$$

where $RSA_H(f)$ and $RSA_V(f)$ denote the response spectra for the horizontal and vertical components of ground motion, respectively.

The path effect, $P(R, f)$, consists of geometrical spreading, R^n, and the residual term, which is the anelastic attenuation function:

$$P(R, f) = R^n \exp\left\{\frac{-\pi f R}{\beta Q(f)}\right\} \tag{5}$$

where the geometrical spreading function determined for low-to-moderate seismicity regions of Eastern North America (ENA) is uniformly applied in this study as follows [28]:

$$R^n = \begin{cases} R^{-1.3} & (R < 70 \text{ km}) \\ 70^{-1.3}(R/70)^{0.2} & (70 \text{ km} \le R < 130 \text{ km}) \\ 70^{-1.3}(130/70)^{0.2}(R/70)^{-0.5} & (130 \text{ km} \le R) \end{cases} \tag{6}$$

In Equation (5), β, representing the shear wave velocity near the source, is 3.5 km/s according to [29]. The quality factor, $Q(f)$, is derived by separating observed earthquake ground motions into unknown terms associated with the source, path, and site effects. Using this methodology, the quality factor for frequencies above 1 Hz is obtained through regression analysis, resulting in the formula: $Q(f) = 163f^{0.745}$.

The source effect, $E(M_0, f)$ in Equation (1), is expressed as:

$$E(M_0, f) = CM_0S(f) = \frac{CM_0}{1 + (f/f_c)^2} \tag{7}$$

Here, f_c is the corner frequency and the constant C is given by:

$$C = \frac{R_pVF}{4\pi\rho\beta^3R_0} \tag{8}$$

where R_p (the wave radiation pattern), V (the partition of total shear-wave energy into horizontal components), and F (a parameter indicating the effect of the free surface) are assumed to be 0.55, 0.707, and 2, respectively. The value of 0.55 is widely accepted as the average of body-wave radiation patterns over various azimuths and take-off angles [30,31]. The parameter $F = 2$ is a standard value applied in most ground motion simulations [6]. The density near the source, ρ is 2.7 g/cm^3, based on the seismic velocity structure of the Korean Peninsula [32], while the reference distance R_0 is conventionally 1 km [16]. In the present study, f_c and M_0, are set as key source parameters and given by [20,33,34]:

$$f_c = \left[\frac{J}{2\pi^3\Omega_0^2}\right]^{\frac{1}{3}} \tag{9}$$

$$M_0 = \frac{4\pi\rho\beta^3R_0\Omega_0}{R_pVF} \tag{10}$$

where Ω_0 is the long-term flat-level function as:

$$\Omega_0 = 2\left[\frac{K^3}{J}\right]^{\frac{1}{4}} \tag{11}$$

The constants K and J are calculated by integrating the displacement amplitude spectrum of the observed ground motion and multiplying angular frequency followed by integration, as used in [33,34].

2.2.2. Transformation into Time Domain

In this study, generated Fourier spectra is transformed into the time domain to determine the PGAs of the corresponding artificial earthquakes. The following steps based on the point-source GMM are employed [6,35]. The procedure involves calculations in both time and frequency domains. Steps (1), (2), and (5) are performed in the time domain, while steps (3) and (4) are conducted in the frequency domain:

(1) Generate Gaussian or uniform white noise for the desired time duration.
(2) Apply a shaping window function to the white noise by multiplying them.
(3) Transform the windowed noise into the frequency domain and normalize it by the square root of the mean square (RMS) amplitude spectrum.
(4) Multiply the normalized noise with the modeled FAS as defined in Equation (1).
(5) Apply Inverse Fast Fourier Transformation to obtain the predicted ground motion time histories.

The shaping window function for time t, and scaled time t_η for different window functions is defined as follows [36]:

$$w(t;\epsilon,\eta,t_\eta) = a(t/t_\eta)^b \exp\{-c(t/t_\eta)\} \tag{12}$$

where b, a, and c are:

$$b = -\frac{\epsilon \ln \eta}{1+\epsilon(\ln\epsilon - 1)} \tag{13}$$

$$a = (\exp\{1\}/\epsilon)^b \tag{14}$$

$$c = b/\epsilon \tag{15}$$

Following previous research [36], the values of the parameters ϵ and η are set to 0.2 and 0.05, respectively. The scaled time t_η is defined as:

$$t_\eta = f_{Tgm} T_{gm} \tag{16}$$

where parameter f_{Tgm} has a value of 1 for the box window and 2 for the exponential window [37]. This study follows the recommendation to use $f_{Tgm} = 2$ to ensure accurate comparisons between response spectra computed using both window types [6]. For the duration model of ground motions in low-to-moderate seismicity regions such as ENA, the following expression is used [38]:

$$T_{gm} = \frac{1}{2f_c} + hR \tag{17}$$

where h is the slope over a given source-to-site distance. For ENA, the values of h are 0.16 for 10 km $\leq R <$ 70 km, -0.03 for 70 km $\leq R <$ 130 km, and 0.04 for 130 km $\leq R <$ 1000 km [38].

2.3. Seismic Source Characterization and PGA Distribution Modeling

2.3.1. Statistical Modeling of Source Parameters for the 2017 Pohang Earthquake

Two critical parameters affecting the source effect, f_c and M_0, were analyzed for the 2017 Pohang Earthquake. These parameters were obtained from 111 out of 130 stations, all located within a 500 km radius of the epicenter, with the source-to-site distances ranging from 11.58 to 413.18 km. According to Equation (1), the Fourier amplitude is influenced by the source-to-site distance, R, and assumes stochastic variations in f_c and M_0. The randomness of these parameters introduces uncertainty into the Fourier amplitude, leading to variability in the predicted PGAs for artificial earthquakes.

Spearman correlation analysis explored relationships between source parameters and source-to-site distance, and among the source parameters themselves. Despite appearing random within a valid range of distance, as depicted in Figure 1a,b, a dependency exists

among the source parameters, as illustrated in Figure 1c. The correlation coefficient between source-to-site distance and f_c was 0.0537, indicating virtually no correlation. Similarly, the correlation coefficient between source-to-site distance and M_0 was 0.0349, suggesting a negligible correlation. In contrast, the correlation between f_c and M_0 was estimated at –0.4570, indicating a weak negative correlation. This finding underscores the importance of generating sampling points that account for this correlation.

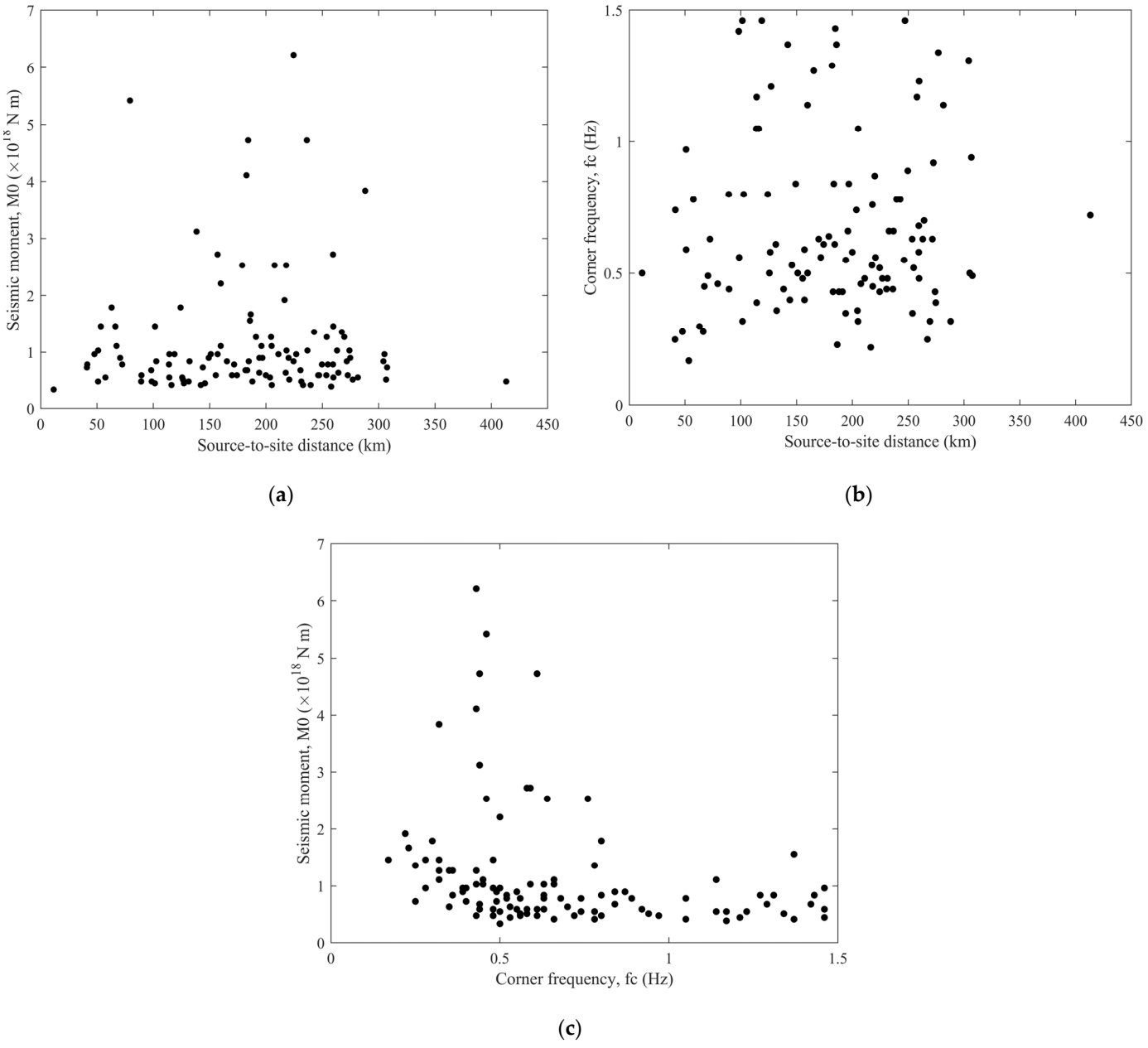


Figure 1. Relationship between (**a**) the source-to-site distance and seismic moment, (**b**) the source-to-site distance and corner frequency, and (**c**) corner frequency and seismic moment.

Copula functions are widely used to estimate joint probability density functions because they effectively capture the dependence structure between two variables while allowing for the coupling and separation of each probability distribution [39–41]. The application of copulas is also well-established in earthquake research [42,43]. A copula is a multivariate cumulative distribution function that uses the complex dependency structures between marginal probability distributions and random variables. According to Sklar's

theorem, the joint cumulative distribution function of the source parameters y_1 and y_2 is defined as follows [44]:

$$F(y_1, y_2) = C(F(y_1), F(y_2)|\theta) \tag{18}$$

where $C(F(y_1), F(y_2)|\theta)$ represents the copula, $F(y_1)$ and $F(y_2)$ are the marginal cumulative distribution functions of the source parameters, and θ is a parameter indicating the dependency.

In the comparative evaluation of five copula functions (Clayton, Frank, Gaussian, Gumbel, and t), the Frank copula was selected as the superior option for calculating the joint probability distribution and generating sampling points for f_c and M_0 supported by its highest log-likelihood value as detailed in Table 2.

Table 2. Log-likelihood estimates and Spearman correlation coefficients for various copula functions.

Copula Function	Log-Likelihood Estimate	Spearman Correlation Coefficient
Clayton	–0.000049	–0.00489
Frank	11.88	–0.457
Gaussian	11.15	–0.414
Gumbel	–0.000089	–0.0138
t	11.21	–0.445

This selection was further reinforced by the Frank copula's precise alignment with the target Spearman correlation coefficient of –0.4570, derived from the observed source parameters. The Frank copula function is expressed as follows:

$$C(u, v) = -\frac{1}{\theta}\log\left[1 + \frac{(\exp\{-\theta u\} - 1)(\exp\{-\theta v\} - 1) - 1}{\exp\{-\theta\} - 1}\right] \tag{19}$$

where u and v are the uniformly distributed random variables and θ is a parameter that characterizes the dependence structure between these variables. It is observed that recorded values of f_c and M_0 from the 111 data points are more suitably represented by distinct probability distributions. The f_c values follow a lognormal distribution, as expressed in:

$$f_1(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{\frac{-(\ln x - \mu)^2}{2\sigma^2}\right\} \tag{20}$$

where the mean $\mu = -0.507$, and the variance $\sigma^2 = 0.228$. Meanwhile, M_0 is better modeled by a Burr distribution written as:

$$f_2(x|\alpha, d, k) = \frac{\frac{kd}{\alpha}\left(\frac{x}{\alpha}\right)^{d-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^d\right)^{k+1}} \tag{21}$$

where $\alpha = 0.485$ is the scale parameter, and $d = 8.971$ and $k = 0.168$ are the shape parameters. Figure 2 shows 10,000 pairs of sampled f_c and M_0 from the five copula functions. The sampled pairs in Figure 2b exhibited a Spearman correlation of –0.4566, closely matching the Spearman correlation coefficient of –0.4570 in the observed source parameters. It is confirmed that the sampled data points effectively reflect the dependence structure of the original data. Moreover, the sampling points generated using the Frank copula function preserve the probability distribution of the recorded source parameters as shown in Figure 3. It not only underscores the accuracy of the Frank copula in capturing the underlying data structure but also emphasizes its capability to maintain the integrity of statistical relationships.

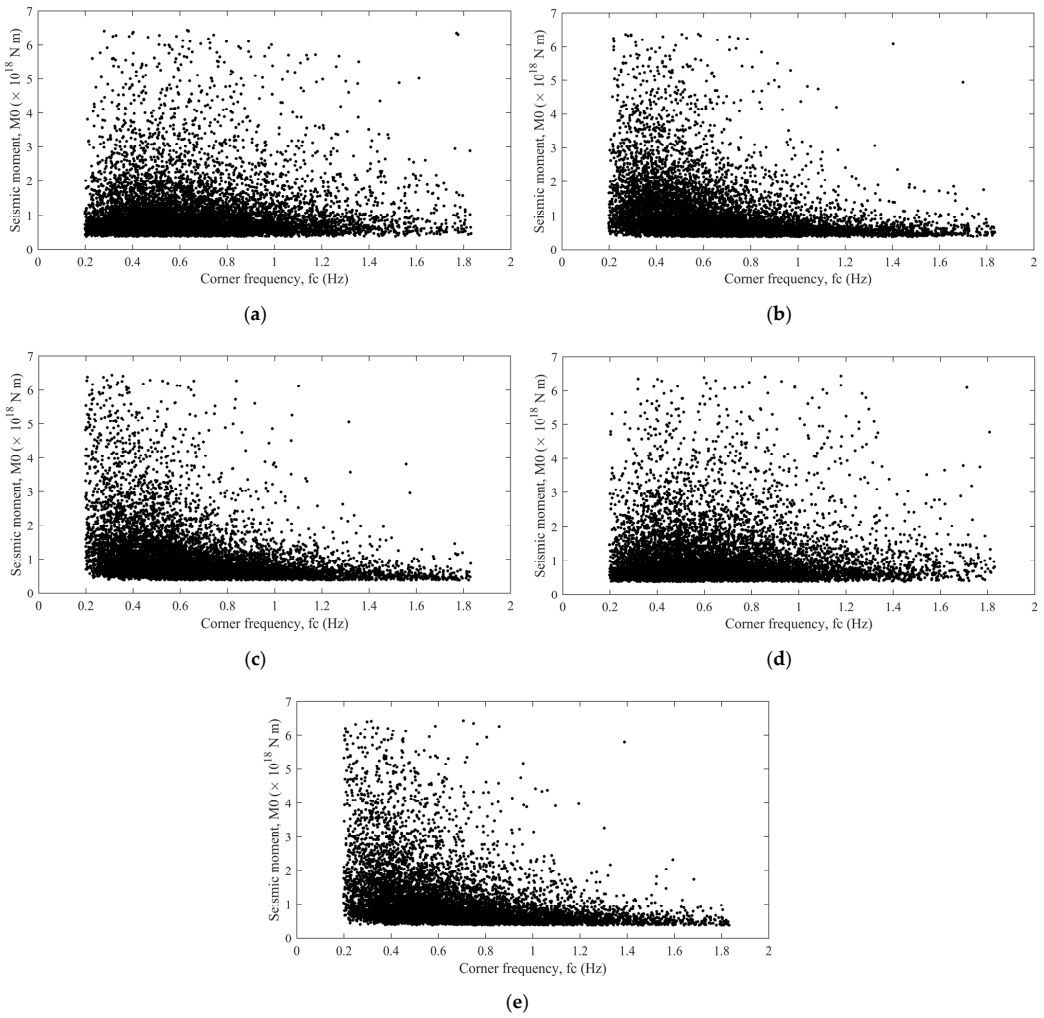


Figure 2. Sampled corner frequency and seismic moment using five different copula functions: (**a**) Clayton, (**b**) Frank, (**c**) Gaussian, (**d**) Gumbel, and (**e**) t.

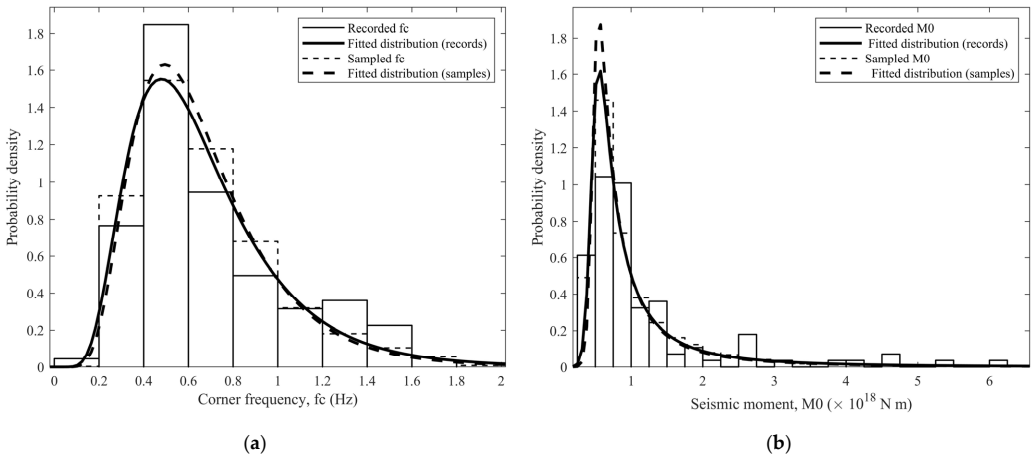


Figure 3. Probability density histograms and probability distributions of (**a**) corner frequency, modeled with a lognormal distribution, and (**b**) seismic moment, modeled with a Burr distribution (left panel) The lines represent the histogram and distribution curve fitted to records (solid) and samples (dashed).

2.3.2. Computation of PGA Probability Distributions from Artificial Earthquakes

Using Frank copula sampling, 10,000 sets of f_c and M_0 were extracted. Equation (1), with the parameters κ_0 and $Q(f)$ from Section 2.2.1, was then applied in the Monte Carlo Simulation (MCS) to generate artificial earthquakes as follows:

$$A(M_{0i}, R, f) = (-2\pi f)\left(\frac{CM_{0i}}{1+(f/f_{ci})^2}\right)R^n \exp\left\{\frac{-\pi f R}{\beta 163 f^{0.745}}\right\}A^*(f)\exp\{-0.014\pi f\} \quad (22)$$

where M_{0i} denotes the i-th sample of the seismic moment, and f_{ci} represents the i-th sample of the corner frequency, as both M_{0i} and f_{ci} are paired values obtained through copula sampling. The MCS allows the exploration of a broad range of scenarios by incorporating variability in the source parameters. Statistical analysis was then conducted on the PGAs of the 10,000 generated artificial earthquakes.

The procedure of the proposed methodology to estimate probability distributions of PGA is as follows:

(1) Obtain recorded source parameters from the earthquake source in the vicinity of the site of interest.
(2) Determine the Probability Density Function (PDF) of source parameters using Equations (20) and (21).
(3) Generate sample pairs of f_c and M_0 by Frank copula sampling for use in MCS.
(4) Define the seismological and geological parameters needed to compute the FAS.
(5) Use the MCS to generate artificial earthquakes with Equation (22) and transform them into the time domain as described in Section 2.2.2.
(6) Determine PGAs and calculate their probability distribution.

3. Results

This section presents a numerical simulation example demonstrating the effectiveness of the proposed method. In this example, different source-to-site distances (10, 25, 50, and 75 km) were selected to estimate the PGA probability distributions, while obtaining corresponding time histories and response spectra. At each distance, 10,000 samples of source parameters were used in the MCS to generate artificial earthquakes with a magnitude of M_w 5.4, identical to the 2017 Pohang Earthquake.

The proposed methodology highlights the importance of detailed seismic analysis to ensure the safety and preparedness of critical infrastructure in proximity to the focus. For example, the Hangsa Dam, currently under construction and set to reach a height of 53 m, is located approximately 25 km from the epicenter. Additionally, the Wolseong Nuclear Power Plant is situated about 45 km from the focus. Another critical facility potentially at risk is the Andong Dam, which stands 83 m tall and is located 75 km from the epicenter.

Artificial earthquakes were generated as outlined in Section 2.3.2, resulting in the derivation of 10,000 PGAs for source-to-site distances of 10, 25, 50, and 75 km. A Kolmogorov–Smirnov test was then performed on the artificially generated PGAs at each distance. The test yielded p-values of 0.87, 0.30, 0.15, and 0.30, respectively, all of which exceed the 5% significance level, supporting the null hypothesis that the log-transformed PGAs follow a normal distribution. This confirms the assumption of normality for further statistical analysis. The probability distribution of PGAs at each distance was determined by fitting the data to a lognormal distribution, as in Equation (20), where it is also used for the probability distribution of f_c. The parameters of this lognormal distribution for PGA, μ (mean) and σ (standard deviation) are shown in Table 3.

Table 3. Parameters of PGA probability distributions for different source-to-site distances.

	10 km	**25 km**	**50 km**	**75 km**
μ	−1.437	−3.233	−4.452	−5.236
σ	0.934	0.852	0.827	0.823

The time history corresponding to the median PGA from 10,000 artificial earthquakes at source-to-site distances of 10, 25, 50, and 75 km is presented in Figure 4, with the median PGAs of 0.24, 0.039, 0.012, and 0.0053 g, respectively. Figure 5 illustrates the response spectra for 100 artificial earthquakes across four different distances. These 100 spectra were selected from a set of 10,000 simulations, based on the smallest deviation between their individual PGAs and the median PGA for each distance. The PGAs from all generated earthquakes, created using the methodology in Section 2.3.2, are compared with recorded data from a total of 20 stations, as shown in Figure 6. Additionally, the PDFs for PGAs derived from artificial earthquakes closely conform to a lognormal distribution, as demonstrated in Figure 7. The parameters characterizing the lognormal distribution for each distance are detailed in Table 3.

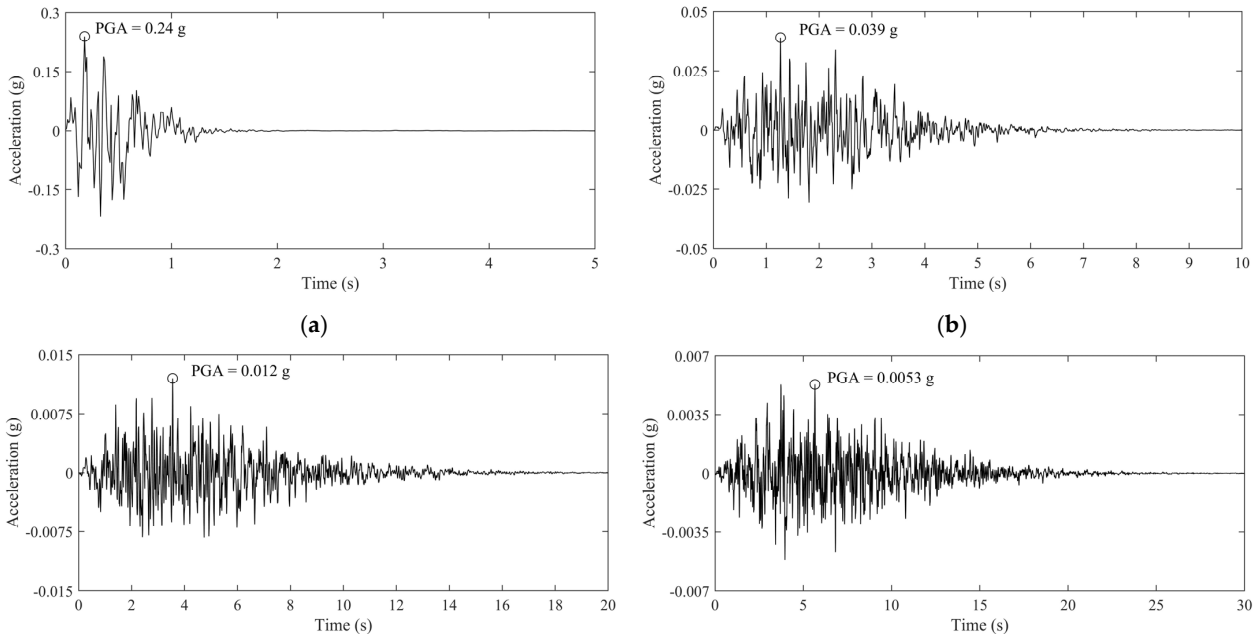


Figure 4. Time histories of artificial earthquakes with median PGA at source-to-site distances of (**a**) 10 km, (**b**) 25 km, (**c**) 50 km, and (**d**) 75 km. Circles indicate the PGAs.

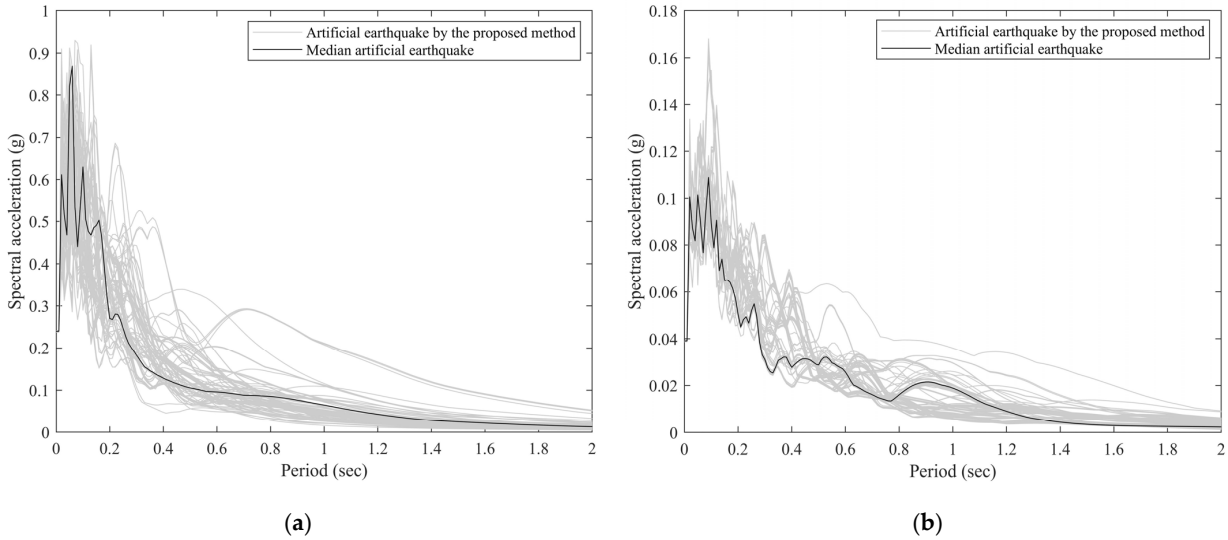
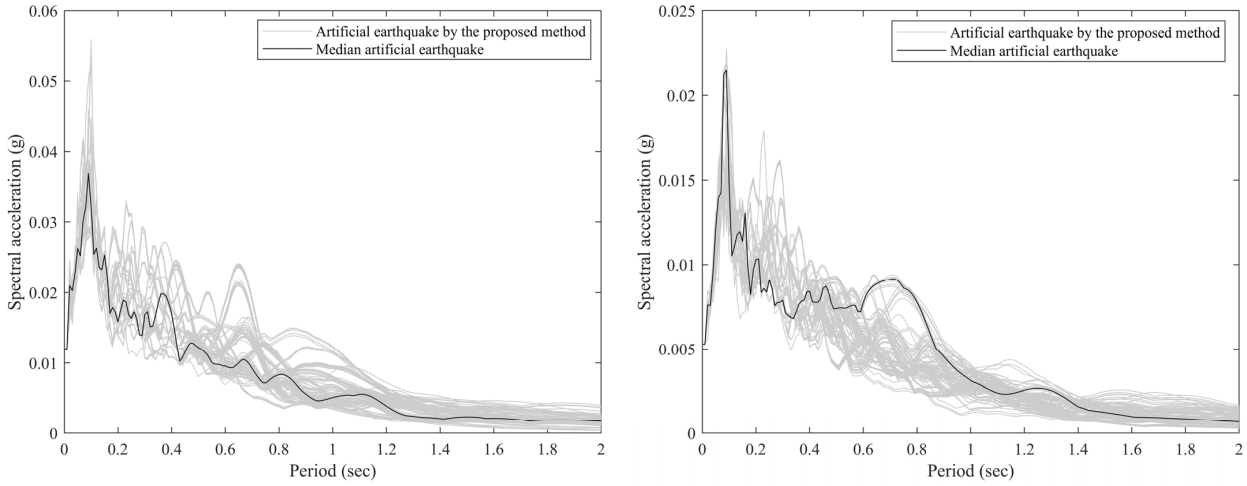


Figure 5. *Cont.*



(**c**) (**d**)

Figure 5. Response spectra for 100 out of 10,000 generated artificial earthquakes at source-to-site distances of (**a**) 10 km, (**b**) 25 km, (**c**) 50 km, and (**d**) 75 km. The black solid line represents the response spectrum for the median PGA of artificial earthquakes. The 100 spectra were chosen based on the smallest deviation between their individual PGAs and the median PGA.

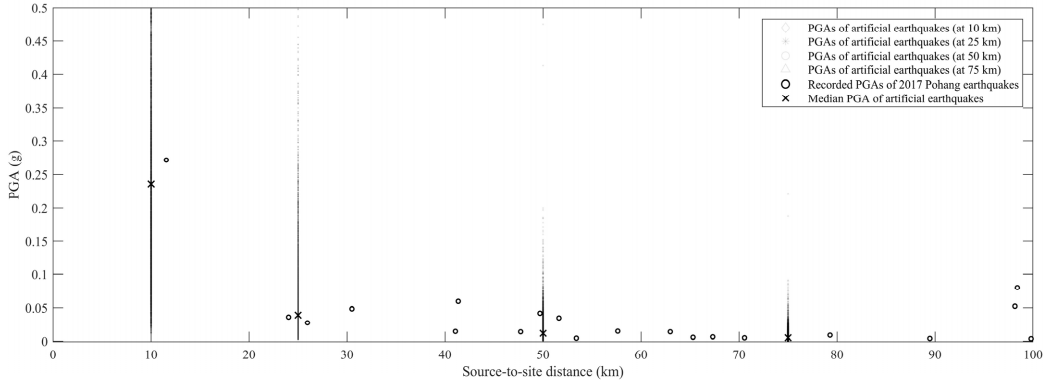
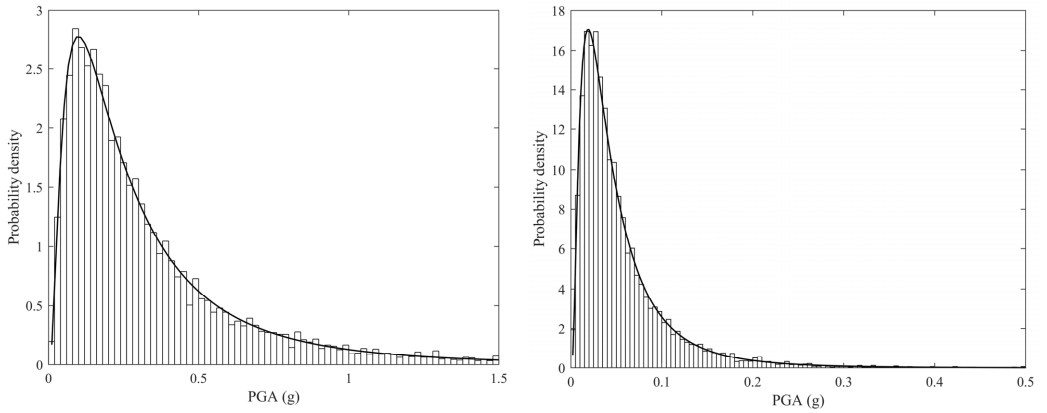
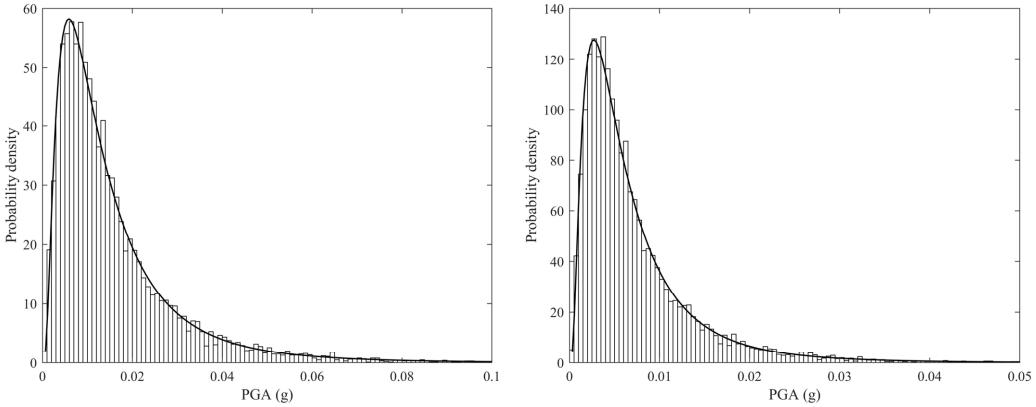


Figure 6. Distribution of PGAs from artificial earthquakes at four different source-to-site distances, along with the PGAs from actual earthquakes at 20 stations. Circles represent individual PGAs from stations, while x marks indicate the median PGAs of artificial earthquakes. It is noted that generated PGAs for all source-to-site distances are truncated at 0.5 g for data visibility.



(**a**) (**b**)

Figure 7. *Cont.*



(**c**) (**d**)

Figure 7. Probability density histograms and corresponding probability distribution curves for PGAs at a source-to-site distance of (**a**) 10 km, (**b**) 25 km, (**c**) 50 km, and (**d**) 75 km. Each distribution curve, depicted by a solid line, was fitted with a lognormal distribution.

4. Discussion

The PGAs recorded from 130 stations during the 2017 Pohang Earthquake are presented in Figure 8. Of these records, the PGAs at 20 stations within a source-to-site distance range of 11.56 to 99.8 km are used to validate the artificially generated earthquakes proposed in this study. While sufficient empirical data would enable more accurate PGA estimations, regions like the Korean Peninsula continue to face data limitations. This scarcity complicates direct comparisons between the PGAs of empirical records from actual earthquakes and those of artificial earthquakes.

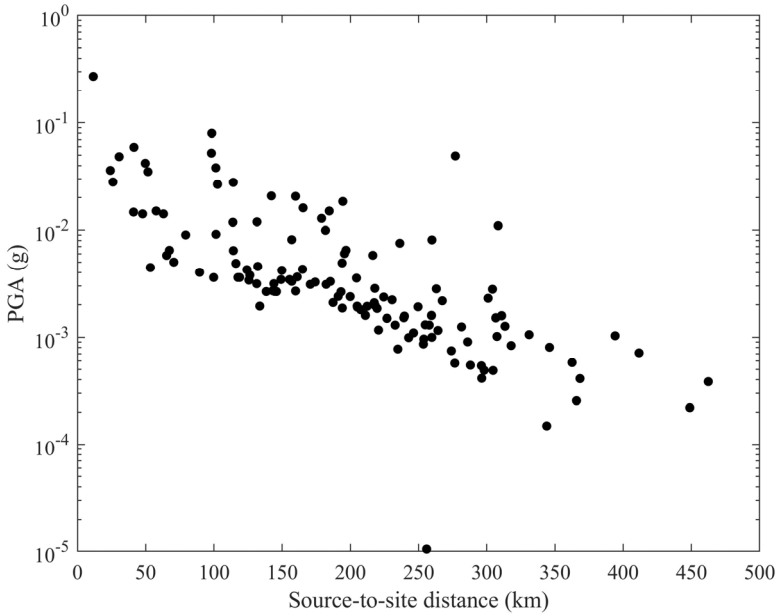


Figure 8. Source-to-site distance versus observed PGAs for the 2017 Pohang Earthquake recorded at 130 stations.

The close alignment between the median PGAs of the artificial and the recorded earthquake data, as shown in Figure 6, is attributed to the effective sampling of the randomly paired f_c and M_0 sets. This representative sampling ensures consistency between the simulated and empirical medians, even with the limited number of available empirical records. By using consistent geological and seismological parameters across all distances and ensuring that the randomly sampled pairs reflect realistic physical characteristics, the

model provides reliable predictions of actual ground motions. Moreover, according to Table 4, the percentage differences between the median PGAs of the empirical data and the artificial earthquakes are 11.76%, 8%, 22.22%, and 20.34% for the source-to-site distance ranges of 10, 25, 50, and 75 km, respectively, with an overall average difference of 15.58%.

Table 4. Comparison of empirical and simulated PGAs by source-to-site distances.

Case	Station	Source-to-Site Distance (km)	PGA (g)	Median PGA of Records (g)	Median PGA of Simulations (g)
1	PHA2	11.56	0.27	0.27	0.24
2	HAK	24.04	0.036	0.036	0.039
	CHS	25.96	0.028		
	DKJ	30.50	0.048		
3	YOCB	41.06	0.015	0.015	0.012
	CSO	41.35	0.059		
	YODB	47.72	0.014		
	ADO2	49.69	0.042		
	USN2	51.63	0.035		
	JEJB	53.40	0.0045		
	DAG2	57.63	0.015		
4	YEYB	62.99	0.014	0.0065	0.0053
	ULJ2	65.31	0.0058		
	EUSB	67.32	0.0065		
	GUWB	70.56	0.0050		
	ADOA	79.29	0.0090		
	CIGB	89.47	0.0040		
	GUM	98.15	0.052		
	UJN	98.40	0.080		
	BUS2	99.80	0.0036		

Although the dataset expands as the source-to-site distance increases, the percentage difference between the empirical records and simulations does not necessarily decrease. However, a Mann–Whitney U test, which compares the distribution of PGAs rather than just their medians, yielded p-values of 0.84, 0.84, 0.09, and 0.10 for each case. These results, showing no significant statistical difference between the empirical and simulated PGA distributions ($p > 0.05$), indicate that the overall distributions and median PGAs of both the empirical records and the artificial earthquakes exhibit a high degree of similarity. This similarity reinforces the predictive accuracy of the present stochastic model, as more empirical data become available, enhancing its reliability as a consistent estimator. This correspondence with actual ground motion distributions provides a basis for further validating the proposed method.

Furthermore, the stochastic approach utilized in this study effectively models the probability distribution of PGA as a lognormal distribution, in agreement with findings from [45], which suggest that peak values of strong motion parameters are likely to follow a lognormal distribution. Figure 7 demonstrates that the PGAs from the artificial earthquakes generated in this study also exhibit this lognormal behavior. It supports the accuracy of the lognormal fittings applied and further illustrates the effectiveness of the stochastic approach for predicting PGA behavior in seismic analysis. By validating the model's ability to generate reliable PGA probability distributions, the evidence underscores the robustness of the proposed method in accurately capturing realistic seismic behavior patterns.

5. Conclusions

In this study, we propose a new stochastic approach for generating probability distribution of Peak Ground Acceleration (PGA) based on recorded data from a single seismic event in regions with limited strong ground motion data. Conventional ground motion prediction techniques often fall short in these areas because they rely on historical data

from seismically active regions. To address these limitations, the present study incorporates local site characteristics, source characteristics, and seismic wave propagation from a single seismic event into a stochastic point-source model, enabling the simulation of multiple artificial earthquakes.

In the proposed approach, it is assumed that the source parameters, f_c and M_0, which are recorded data from a single seismic event, are random variables. These parameters are demonstrated to be independent of the source-to-site distance, thereby incorporating the inherent uncertainty associated with them. It is also shown that the optimal copula function for capturing the relationship between f_c and M_0, is the Frank copula function. This function was selected from five copula functions based on the highest log-likelihood. By using the Frank copula function for sampling points, the reliability of the generated artificial earthquakes is significantly enhanced. It is concluded that the present approach, utilizing Monte Carlo Simulation (MCS) with random sampling of the two key parameters, successfully generates sets of artificial ground motions and their corresponding PGA distributions. The method's effectiveness is validated by comparing the median PGAs of artificial earthquakes, generated from a single seismic event, across four different source-to-site distances with recorded data.

The derived probability distributions of PGA provide a rational method for selecting an appropriate set of artificial earthquakes for seismic safety assessments for high-rise buildings, large dams, nuclear power plants, and other critical infrastructures in regions like the Korean Peninsula, where strong seismic records are extremely rare. Given the scarcity of large-amplitude recorded ground motions in this region, the ability to generate and select reasonable artificial earthquakes from a single seismic event is particularly valuable. Consequently, the proposed stochastic approach is expected to significantly enhance the reliability of seismic safety assessments and provide a solid foundation for future research on earthquake ground motion prediction in regions with limited strong motion records. To further enhance the method's reliability and broader applicability, future work should include validation against multiple seismic events of varying magnitudes. This comprehensive validation approach would better establish the robustness of the proposed method across diverse seismic scenarios.

Author Contributions: Conceptualization, J.L.; Data curation, J.S.; Formal analysis, J.S.; Funding acquisition, J.L.; Investigation, J.S.; Methodology, J.S. and J.L.; Project administration, J.L.; Resources, J.L.; Software, J.S.; Validation, J.S. and J.L.; Writing—original draft, J.S. and J.L.; Writing—review and editing, J.S. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was partly funded by the Dongguk University Graduate School.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This work was supported by the Dongguk University Research Fund of 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. NECIS-National Earthquake Comprehensive Information System. Available online: https://necis.kma.go.kr/necis-dbf/usermain/new/common/mainMapForm.do (accessed on 10 April 2024).
2. Search Earthquake Catalog. Available online: https://earthquake.usgs.gov/earthquakes/search/ (accessed on 10 April 2024).
3. Nuclear Power in South Korea. Available online: https://world-nuclear.org/information-library/country-profiles/countries-o-s/south-korea.aspx (accessed on 10 April 2024).
4. Kwater. Available online: https://www.kwater.or.kr/eng/busi/resoPage.do?s_mid%5Cu003d1179= (accessed on 10 April 2024).
5. Gasparini, D.; Vanmarcke, E.H. *SIMQKE: A Program for Artificial Motion Generation: User's Manual and Documentation*; Department of Civil Engineering, Massachusetts Institute of Technology: Massachusetts, MA, USA, 1976.
6. Boore, D.M. Simulation of ground motion using the stochastic method. *Pure. Appl. Geophys.* **2003**, *160*, 635–676. [CrossRef]

7. Campbell, K.W. Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America. *Bull. Seismol. Soc. Am.* **2003**, *93*, 1012–1033. [CrossRef]
8. Rezaeian, S.; Der Kiureghian, A. A stochastic ground motion model with separable temporal and spectral nonstationarities. Earthq. *Eng. Struct. Dyn.* **2008**, *37*, 1565–1584. [CrossRef]
9. Rezaeian, S.; Der Kiureghian, A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq. Eng. Struct. Dyn.* **2010**, *39*, 1155–1180. [CrossRef]
10. Nau, J.M.; Hall, W.J. Scaling methods for earthquake response spectra. *J. Struct. Eng.* **1984**, *110*, 1533–1548. [CrossRef]
11. Vidic, T.; Fajfar, P.; Fischinger, M. Consistent inelastic design spectra: Strength and displacement. *Earthq. Eng. Struct. Dyn.* **1994**, *23*, 507–521. [CrossRef]
12. Khansoltani, E.; Shakeri, K. Scaling of ground motion records for nonlinear time history analysis of structures. In Proceedings of the 7th International Conference on Seismology and Earthquake Engineering, Teheran, Iran, 18–21 May 2015.
13. Vanmarcke, E.H.; Gasparini, D.A. Simulated earthquake ground motions. In Proceedings of the 4th International Conference on Structural Mechanics in Reactor Technology, San Francisco, CA, USA, 15–19 August 1977.
14. Tolis, S.V.; Faccioli, E. Displacement design spectra. *J. Earthq. Eng.* **1999**, *3*, 107–125. [CrossRef]
15. Erduran, E.; Lindholm, C. A Critical Look at the Use of Design Spectrum Shape for Seismic Risk Assessment. *Earthq. Spectra* **2012**, *28*, 1711–1721. [CrossRef]
16. Baker, J.; Bradley, B.; Stafford, P. Physics-Based Ground-Motion Characterization. In *Seismic Hazard and Risk Analysis*; Cambridge University Press: Cambridge, UK, 2021; pp. 196–233.
17. Bora, S.S.; Scherbaum, F.; Kuehn, N.; Stafford, P. Fourier spectral-and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions. *Bull. Earthq. Eng.* **2014**, *12*, 467–493. [CrossRef]
18. Shearer, P.M. The Seismic Wave Equation. In *Introduction to Seismology*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2019; pp. 39–60.
19. Sun, X.; Hartzell, S.; Rezaeian, S. Ground-motion simulation for the 23 August 2011, Mineral, Virginia, earthquake using physics-based and stochastic broadband methods. *Bull. Seismol. Soc. Am.* **2015**, *105*, 2641–2661. [CrossRef]
20. Brune, J.N. Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res.* **1970**, *75*, 4997–5009. [CrossRef]
21. Shi, J.; Kim, W.Y.; Richards, P.G. The corner frequencies and stress drops of intraplate earthquakes in the northeastern United States. *Bull. Seismol. Soc. Am.* **1998**, *88*, 531–542. [CrossRef]
22. Jee, H.W.; Han, S.W. Regional ground motion prediction equation developed for the Korean Peninsula using recorded and simulated ground motions. *J. Earthq. Eng.* **2022**, *26*, 5384–5406. [CrossRef]
23. Beyreuther, M.; Barsch, R.; Krischer, L.; Megies, T.; Behr, Y.; Wassermann, J. ObsPy: A Python toolbox for seismology. *Seismol. Res. Lett.* **2010**, *81*, 530–533. [CrossRef]
24. Kwon, J.; Heo, T.; Kim, J.K.; Oh, H.S. A New P-Wave Detector via Moving Empirical Cumulative Distribution Function. *Bull. Seismol. Soc. Am.* **2018**, *108*, 2080–2089. [CrossRef]
25. Anderson, J.G.; Hough, S.E. A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bull. Seismol. Soc. Am.* **1984**, *74*, 1969–1993.
26. Ktenidou, O.J.; Gélis, C.; Bonilla, L.F. A study on the variability of kappa (κ) in a borehole: Implications of the computation process. *Bull. Seismol. Soc. Am.* **2013**, *103*, 1048–1068. [CrossRef]
27. Zhao, J.X.; Irikura, K.; Zhang, J.; Fukushima, Y.; Somerville, P.G.; Asano, A.; Ohno, Y.; Oouchi, T.; Takahashi, T.; Ogawa, H. An empirical site-classification method for strong-motion stations in Japan using H/V response spectral ratio. *Bull. Seismol. Soc. Am.* **2006**, *96*, 914–925. [CrossRef]
28. Atkinson, G.M. Empirical attenuation of ground-motion spectral amplitudes in southeastern Canada and the northeastern United States. *Bull. Seismol. Soc. Am.* **2004**, *94*, 1079–1095. [CrossRef]
29. Atkinson, G.M.; Silva, W. Stochastic modeling of California ground motions. *Bull. Seismol. Soc. Am.* **2000**, *90*, 255–274. [CrossRef]
30. Boore, D.M.; Boatwright, J. Average body-wave radiation coefficients. *Bull. Seismol. Soc. Am.* **1984**, *74*, 1615–1621. [CrossRef]
31. Atkinson, G.M.; Boore, D.M. Ground-motion relations for eastern North America. *Bull. Seismol. Soc. Am.* **1995**, *85*, 17–30. [CrossRef]
32. Kim, S.K. A study on the crustal structure of the Korean Peninsula. *J. Geol. Soc. Korea.* **1995**, *31*, 393–403. (In Korean)
33. Andrews, D.J. Objective determination of source parameters and similarity of earthquakes of different size. In *Earthquake Source Mechanics*; Das, S., Boatwright, J., Scholz, C.H., Eds.; American Geophysical Union: Washington, DC, USA, 1986; Volume 37, pp. 259–267.
34. Snoke, J.A. Stable determination of (Brune) stress drops. *Bull. Seismol. Soc. Am.* **1987**, *77*, 530–538.
35. Lam, N.; Wilson, J.; Hutchinson, G. Generation of synthetic earthquake accelerograms using seismological modelling: A review. *J. Earthq. Eng.* **2000**, *4*, 321–354. [CrossRef]
36. Saragoni, G.R.; Hart, G.C. Simulation of artificial earthquakes. *Earthq. Eng. Struct. Dyn.* **1973**, *2*, 249–267. [CrossRef]
37. Boore, D.M. Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bull. Seismol. Soc. Am.* **1983**, *73*, 1865–1894.
38. Boatwright, J.; Choy, G.L. Acceleration source spectra anticipated for large earthquakes in northeastern North America. *Bull. Seismol. Soc. Am.* **1992**, *82*, 660–682.

39. De Michele, C.; Salvadori, G.; Canossi, M.; Petaccia, A.; Rosso, R. Bivariate Statistical Approach to Check Adequacy of Dam Spillway. *J. Hydrol. Eng.* **2005**, *10*, 50–57. [CrossRef]
40. 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. [CrossRef]
41. Genest, C.; Rémillard, B.; Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.* **2009**, *44*, 199–213. [CrossRef]
42. Goda, K.; Atkinson, G.M. Interperiod Dependence of Ground-Motion Prediction Equations: A Copula Perspective. *Bull. Seismol. Soc. Am.* **2009**, *99*, 922–927. [CrossRef]
43. Xu, Y.; Tang, X.S.; Wang, J.P.; Kuo-Chen, H. Copula-based joint probability function for PGA and CAV: A case study from Taiwan: Joint Probability Function for PGA and CAV. *Earthq. Eng. Struct. Dyn.* **2016**, *45*, 2123–2136. [CrossRef]
44. Sklar, M. *Fonctions de Répartition à N Dimensions et Leurs Marges*; Annales de l'ISUP: Paris, France, 1959; pp. 229–231.
45. Kramer, S.L. Estimation of Ground Motion Parameters. In *Geotechnical Earthquake Engineering*; Pearson Education India: Chennai, India, 1996; p. 87.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.