Copula-based joint probability function for PGA and CAV: a case study from Taiwan

Yun Xu1,2, Xiao-Song Tang1,2,*†, J. P. Wang2 and H. Kuo-Chen3

1 State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, China
2 Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
3 Department of Earth Sciences, National Central University, Taiwan

SUMMARY

This study aims to develop a joint probability function of peak ground acceleration (PGA) and cumulative absolute velocity (CAV) for the strong ground motion data from Taiwan. First, a total of 40,385 earthquake time histories are collected from the Taiwan Strong Motion Instrumentation Program. Then, the copula approach is introduced and applied to model the joint probability distribution of PGA and CAV. Finally, the correlation results using the PGA-CAV empirical data and the normalized residuals are compared. The results indicate that there exists a strong positive correlation between PGA and CAV. For both the PGA and CAV empirical data and the normalized residuals, the multivariate lognormal distribution composed of two lognormal marginal distributions and the Gaussian copula provides adequate characterization of the PGA-CAV joint distribution observed in Taiwan. This finding demonstrates the validity of the conventional two-step approach for developing empirical ground motion prediction equations (GMPEs) of multiple ground motion parameters from the copula viewpoint. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS: correlation; joint probability function; PGA; CAV; copulas

1. INTRODUCTION

Given that a single intensity measure (IM) is often insufficient to represent the severity of a ground motion, the use of multiple IMs that capture different features of a ground motion is desired for earthquake engineering (e.g., [1]). Therefore, attempts have been made to study the correlations between various IMs during the past few years. For example, Baker and Cornell [2], Baker and Jayaram [3], Goda and Hong [4], and Goda and Atkinson [5] studied the correlations between spectral acceleration (SA) values at different vibration periods. Baker [6] examined the correlations between SA, peak ground acceleration (PGA), and Arias intensity (Ia). Bradley [7] developed empirical correlation equations between PGA, SA, spectrum intensity (SI), and acceleration spectrum intensity (ASI). Bradley [8] further developed empirical equations for predicting the correlation coefficients between peak ground velocity (PGV) and several IMs, namely PGA, SA, SI, and ASI.

To the best of our knowledge, PGA-based (peak ground acceleration) pseudo-static analysis is still playing a predominant role in earthquake-resistant design, as it is prescribed by several technical references like Eurocode 8 [9] and ASCE 7-10 [10]. However, the limitation of the PGA-based...
design has been long discussed (e.g., [11]). That is, structural instability and response to an earthquake motion is of high complexity and uncertainty, and in addition to PGA, the parameters such as the duration of the ground shaking should also be critical to the onset of structural damage (e.g., [12]).

In order to consider the influence of ground motion duration, parameters incorporating the cumulative shaking effects were proposed, and cumulative absolute velocity (CAV) is one of those parameters to characterize the intensity of an earthquake ground motion. Specifically, the mathematical expression of CAV was defined as follows (e.g., [13]):

$$\text{CAV} = \int_0^{t_{\text{max}}} |a(t)| dt$$  \hspace{1cm} (1)

where $t_{\text{max}}$ is the total duration of the ground shaking; $|a(t)|$ is the absolute acceleration of the earthquake time history. For a better explanation to CAV, we made a graph in Figure 1 to illustrate the CAV of a hypothetical ground motion, with the motion’s CAV equal to the sum of the solid black areas.

After the development of CAV, some other studies focusing on its relationship with and its influence on structural safety were reported. For example, from several major earthquake events, the study of EPRI [13] pointed out that CAV has a stronger correlation with structural damage than PGA. Similarly, with the data from Italy and Greece, strong correlations between CAV and structural damage were found in recent earthquakes of the regions (e.g., [12]). Furthermore, Campbell and Bozorgnia [14] developed a new empirical ground motion prediction equation (GMPE) for the horizontal component of CAV using the PEER-NGA strong motion database. Campbell and Bozorgnia [15] further developed a relationship between the standardized version of CAV and the Japan Meteorological Agency (JMA) and modified Mercalli (MMI) instrumental seismic intensities based on the PEER-NGA strong motion database.

On the other hand, new engineering assessments were also developed using this energy-related parameter. For example, Kramer and Mitchell [16] proposed a new framework for soil liquefaction assessment based on the CAV of an earthquake ground motion, as Fahjan et al. [17] proposed an algorithm for earthquake early warning based on the CAV of the early ground motions detected. In addition to the academic works, some recent technical references have prescribed a CAV threshold as the safe-shutdown earthquake for critical infrastructures (e.g., [18, 19]).

Given both PGA and CAV are important indicators to structure’s seismic safety, a logical design framework is to consider both parameters in earthquake-resistant analysis. However, prior to the development of PGA-CAV design methodology, it is essential to study the statistical model for the two parameters, such as their joint probability function. Therefore, the key scope/novelty of this

![Image of acceleration time history](image-url)

Figure 1. Definition of cumulative absolute velocity (CAV) from an acceleration time history.
The study is to model the PGA-CAV joint probability function with the copula approach. Specifically, we used the strong ground motion data in Taiwan as a case study, and successfully calibrated a copula model for their joint probability distribution. The copula approach has been used in geotechnical and structural engineering (e.g., [20, 21]). In earthquake engineering, Goda and Atkinson [22] applied the copula approach to study the interperiod dependence of ground motion parameters and validated the conventional two-step approach for developing empirical GMPEs.

The paper in the following is organized with the sections: The data mining on the PGA and CAV data for this statistical study is given in Section Two; the introduction of the copula theory and its applications are given in Section Three; the calibration for the PGA-CAV joint probability function using empirical data is detailed in Section Four; the PGA-CAV joint probability function subject to given magnitude and distance is further developed in Section Five; followed by some discussion on the adequacy of the multivariate lognormal distribution for modeling multiple ground motion parameters given in Section Six.

2. THE DATA MINING

2.1. Taiwan strong motion instrumentation program

The region around Taiwan is known for high seismicity. On average, around 2,000 earthquakes above $M_L$ 3.0 occur in the region every year, with a catastrophic event like the $M_L$ 7.3 Chi-Chi Earthquake in 1999 that could reoccur in decades [23]. Since the 1990s, the Taiwan Strong Motion Instrumentation Program (TSMIP) has been launched for collecting high-quality earthquake data for the study region. As shown in Figure 2, around 700 earthquake stations of TSMIP are in operation as of now, forming a ‘dense’ instrumentation network to ‘watch’ the local seismicity more ‘closely.’

Since the operation of TSMIP, the database has been essential to a variety of earthquake studies. For example, Lin et al. [24] developed several GMPEs with the data from TSMIP, which are essential to recent seismic hazard studies for Taiwan using the regional models (e.g., [23, 25]). In addition, the

![Figure 2. Locations of 708 earthquake stations of the Taiwan Strong Motion Instrumentation Program.](image-url)
database is critical to the investigation of the basin effects on site response in northern Taiwan [26], as well as to the development of earthquake early warning systems in Taiwan [27].

2.2. The PGA and CAV data used in this study
In this study, 40,385 earthquake time histories associated with 312 earthquakes above $M_L 5.0$ were first collected from TSMIP. Note that all the 312 earthquakes are main shocks, in order to make this statistical study more consistent in terms of data mining. After some calculations on each time history, 40,385 pairs of PGA and CAV were then collected for this statistical study.

Table I summarizes the statistics of the 40,385 PGA and CAV samples, including mean, standard deviation (SD), etc. Figure 3(a) shows the PGA-CAV joint histogram. It shows that the joint distribution is highly skewed, or the distribution is very asymmetrical against the mean values of PGA or CAV. Moreover, it also shows that PGA and CAV are ‘positively’ correlated, with a Pearson correlation coefficient of $\rho=0.70$ or a Kendall correlation coefficient of $\tau=0.66$.

3. COPULA APPROACH FOR MODELING A JOINT PROBABILITY FUNCTION
The word *copula*, from the Latin, means ‘link’ or ‘tie’ of different things. Therefore, in probability and statistics, a copula refers to a function that can link marginal probability distributions to a joint probability function. According to Sklar’s theorem [28], the bivariate cumulative distribution function (CDF) for $X_1$ and $X_2$, denoted as $F_{X_1,X_2}(x_1, x_2)$, can be expressed as follows (e.g., [29]):

$$F_{X_1,X_2}(x_1, x_2) = \Pr(X_1 \leq x_1, X_2 \leq x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2); \theta) = C(u_1, u_2; \theta) \quad (2)$$

where $C(u_1, u_2; \theta)$ is a bivariate copula function, $\theta$ is the copula parameter describing the dependency

<table>
<thead>
<tr>
<th>Table I. Descriptive statistics for PGA-CAV empirical data and normalized residuals.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Pearson correlation ($\rho$)</td>
</tr>
<tr>
<td>Kendall correlation ($\tau$)</td>
</tr>
</tbody>
</table>

Peak ground acceleration (PGA); cumulative absolute velocity (CAV).

Figure 3. Histograms of observed and simulated peak ground acceleration-cumulative absolute velocity data.
between $X_1$ and $X_2$, and $F_{X_1}(x_1)$, and $F_{X_2}(x_2)$ are the marginal CDFs for $X_1$ and $X_2$, respectively. As shown in Equation (2), usually $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are denoted as $u_1$ and $u_2$ ranging from 0 to 1 in the formula.

Here, we would like to provide an example to help explain the copula algorithms. For the case with $X_1$ and $X_2$ both following normal distributions, we can first write their marginal CDFs, $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$, as follows:

\[ u_1 = F_{X_1}(x_1) = \Pr(X_1 \leq x_1) = \Phi \left( \frac{x_1 - p_1}{q_1} \right) \]  

(3)

and

\[ u_2 = F_{X_2}(x_2) = \Pr(X_2 \leq x_2) = \Phi \left( \frac{x_2 - p_2}{q_2} \right) \]  

(4)

where $p_1$ (i.e., mean value of $X_1$) and $q_1$ (i.e., standard deviation of $X_1$) are the model parameters for $X_1$, $p_2$, and $q_2$ are those for $X_2$, and $\Phi$ denotes the CDF of a standard normal distribution (mean=0 and standard deviation=1).

Next, we use the Clayton copula in this example to develop their joint probability function. By definition, the mathematical expression of the Clayton copula is as follows (e.g., [29]):

\[ C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\theta} \]  

(5)

As a result, by combining Equations (3), (4), and (5), the joint CDF for $X_1$ and $X_2$ can be expressed with Equation (6) from the copula theory:

\[ F_{X_1, X_2}(x_1, x_2) = \left\{ \left[ \Phi \left( \frac{x_1 - p_1}{q_1} \right) \right]^{-\theta} + \left[ \Phi \left( \frac{x_2 - p_2}{q_2} \right) \right]^{-\theta} - 1 \right\}^{-1/\theta} \]  

(6)

where $p_1$, $q_1$, $p_2$, $q_2$, and $\theta$ are the model parameters, which will be calibrated from the data.

Therefore, the key tasks of using the copula approach to model the joint probability function of bivariate statistics can be summarized as follows: (i) searching for the best-fit marginal distributions for $X_1$ and $X_2$, and (ii) searching for the best-fit copula that can capture their dependence structure. As far as this study is concerned, the two corresponding tasks are to calibrate the marginal distributions for PGA and CAV, and to search for a copula that can link the two marginals to form a joint probability function for the given PGA-CAV data. In Section 4, we will elaborate the development for the PGA-CAV joint probability function, along with the key algorithms used in the calibration based on the copula theory.

The copula approach has been widely applied to engineering probabilistic analyses, especially for bivariate model developments (e.g., [20, 21, 30–37]). For example, Goda [21] employed the copula approach to model the joint probability function of structure’s maximum displacement and permanent displacement under seismic loading. Li et al. [30, 31] and Huffman and Stuedlein [36] used copulas to model the bivariate distribution of the two hyperbolic curve-fitting parameters underlying load-settlement curves of piles. Li and Tang [20] and Zhang et al. [37] constructed the bivariate distribution of soil’s cohesion and friction angle using copulas, and concluded that the joint probability function would improve geotechnical reliability analysis with a better simulation of their correlation structures.

4. PGA-CAV JOINT PROBABILITY FUNCTION USING EMPIRICAL DATA

4.1. Calibration of marginal distributions

As other univariate studies, the calibrations of the marginal probability functions for PGA and CAV in this study are to search for a best-fit probability model for them, among the six commonly-used models, namely the normal distribution, normal truncated below zero (TruncNormal), Gumbel distribution,
Gumbel truncated below zero (TruncGumbel), lognormal distribution, and Weibull distribution. As summarized in Table II, the six models were used altogether for the calibration. Figure 4(a) shows the comparisons between the observed frequency and theoretical curves for the given 40,385 PGA data from the TSMP database, indicating that the lognormal distribution outperforms the others for the modeling. Similarly, Figure 4(b) shows the graph for the CAV data, also indicating the lognormal distribution could model the observations the best among the six probability functions.

It must be noted that the determination of the best-fit model is not based on our own ‘feeling’ by looking at the graphs, but based on some quantitative measure of the curves. As recommended by other copula studies (e.g., [20]), we used the Akaike Information Criterion (AIC) [38] to quantify the level of goodness-of-fit between a theoretical model and the given samples, and when a smaller AIC score is obtained, it means the selected model shows a better agreement with the observations. The AIC scores for the six selected marginal distributions are summarized in Table III. As mentioned previously, the lognormal distribution with the lowest AIC scores among the six distributions was then considered the best marginal probability function for both PGA and CAV, in contrast to the normal distribution with the highest AIC scores that shows the least agreement with the observations. Note that the AIC scores were negative, making the one with the largest number followed by a ‘−’ sign becoming the smallest.

4.2. Calibration of copula

With the two marginal distributions determined, the next step is to search for a copula that can capture the dependence structure of PGA and CAV appearing in the 40,385 pairs of observations. It should be

Table II. Cumulative distribution functions associated with the selected six distributions.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>F(x, p, q)</th>
<th>μ = p, σ² = q²</th>
<th>Range of p</th>
<th>Range of q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Φ(x/p, q)</td>
<td>p = p, q² = q²</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
<tr>
<td>TruncNormal</td>
<td>Φ(x−p)/1−Φ(p/q)</td>
<td>p = p, σ² = q²</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
<tr>
<td>Gumbel</td>
<td>exp[−exp[−q(x−p)]]</td>
<td>μ = p + 0.5772/q, σ² = π²/(6q²)</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
<tr>
<td>TruncGumbel</td>
<td>exp[−exp[−q(x−p)]−exp[−exp(pq)]]/1−exp[−exp(pq)]</td>
<td>μ = p + 0.5772/q, σ² = π²/(6q²)</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>Φ[ln(x/p)/q]</td>
<td>μ = exp(p + 0.5q²), σ² = [exp(q²) − 1]exp(2p + q²)</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
<tr>
<td>Weibull</td>
<td>1 − exp[−(x/p)²]</td>
<td>μ = pΓ(1 + 1/q), σ² = p²[Γ(1 + 2/q) − Γ²(1 + 1/q)]</td>
<td>(−∞, ∞)</td>
<td>(0, ∞)</td>
</tr>
</tbody>
</table>

The Φ denotes the standard normal distribution function and Γ is the gamma function.

Figure 4. Frequency curves of fitted marginal distributions for peak ground acceleration (PGA) and cumulative absolute velocity (CAV).
noted that a dependence structure is represented by a copula function in standard uniform space (e.g., [20]), because bivariate copula functions are defined in the domain of \([0, 1]^2\). Therefore, to view the dependence structure, the PGA and CAV data are first transformed into standard uniform data \((U_1, U_2)\). As most copula studies (e.g., [20, 21]), we adopted the empirical distributions of PGA and CAV for the task, calculating their empirical distributions as follows:

\[
\begin{align*}
U_{1i} & = \frac{\text{rank}(\text{PGA}_i)}{N + 1} \\
U_{2i} & = \frac{\text{rank}(\text{CAV}_i)}{N + 1} \quad i = 1, 2, \ldots, N
\end{align*}
\] (7)

in which \(\text{rank}(\text{PGA}_i)\) [or \(\text{rank}(\text{CAV}_i)\)] denotes the rank of \(\text{PGA}_i\) (or \(\text{CAV}_i\)) among the list \(\{\text{PGA}_1, \ldots, \text{PGA}_N\}\) (or \(\{\text{CAV}_1, \ldots, \text{CAV}_N\}\)) in an ascending order.

Figure 5(a) shows the \(U_1–U_2\) joint histogram for the 40,385 pairs of PGA-CAV data from the TSMIP database. It is found that the dependence structure of PGA and CAV is nearly symmetrical against the 45° diagonal plane in the standard uniform space. As a result, based on previous studies (e.g., [20, 29]), we selected the Gaussian copula, Plackett copula, Frank copula, Clayton copula, Gumbel copula and Independent copula, which could as a group provide adequate diversity, for the model calibration. Table IV is a summary of the six copulas, including their mathematical expressions.

The copula parameter \(\theta\) can be determined through the Pearson correlation coefficient \(\rho\) or the Kendall correlation coefficient \(\tau\) as illustrated in Li and Tang [20]. Note that the determination of copula parameters using the Kendall correlation coefficient depends on the correlation coefficient only. It is independent of the underlying marginal distributions. Thus, the Kendall correlation coefficient \(\tau\) is adopted to determine the copula parameter \(\theta\). The relationships between \(\tau\) and \(\theta\) for the selected copulas are shown in Table IV, and the copula parameters for each copula are shown in Table V. For example, given the Kendall correlation coefficient \(\tau = 0.66\) from the observations, the copula parameter \(\theta\) is equal to 0.86 for the Gaussian copula.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PGA</th>
<th>CAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>-137934</td>
<td>-12655</td>
</tr>
<tr>
<td>TruncNormal</td>
<td>-166190</td>
<td>-39720</td>
</tr>
<tr>
<td>Gumbel</td>
<td>-164459</td>
<td>-41691</td>
</tr>
<tr>
<td>TruncGumbel</td>
<td>-196368</td>
<td>-71902</td>
</tr>
<tr>
<td>Lognormal</td>
<td>-229233</td>
<td>-105062</td>
</tr>
<tr>
<td>Weibull</td>
<td>-204994</td>
<td>-83610</td>
</tr>
</tbody>
</table>

The Akaike information criterion (AIC) score is underlined if the corresponding distribution is the best-fit distribution; peak ground acceleration (PGA); cumulative absolute velocity (CAV).

Figure 5. Observed and simulated \(U_1–U_2\) data for peak ground acceleration and cumulative absolute velocity.
Table IV. Summary of the adopted six copula functions.

<table>
<thead>
<tr>
<th>Copula</th>
<th>$C(u_1, u_2; \theta)$</th>
<th>$\tau$</th>
<th>Generator function, $\phi_0(t)$</th>
<th>Range of $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>$u_1 u_2$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp \left[ -\frac{r_1^2 - 2\theta r_1 r_2 + r_2^2}{2(1-\theta^2)} \right] , dr_1 , dr_2$</td>
<td>$\frac{2\sqrt{\pi \cdot \theta}}{\pi}$</td>
<td>—</td>
<td>$[-1, 1]$</td>
</tr>
<tr>
<td>Plackett</td>
<td>$S = \sqrt{S^2 - 4u_1 u_2 (\theta - 1)} / 2(\theta - 1)$, $S = 1 + (\theta - 1)(u_1 + u_2)$</td>
<td>$4\int_0^1 \int_0^1 C(u_1, u_2; \theta) , dC(u_1, u_2; \theta) - 1$</td>
<td>—</td>
<td>$(0, \infty)$</td>
</tr>
<tr>
<td>Frank</td>
<td>$-\frac{1}{\theta} \ln \left[ 1 + \frac{(e^{-u_1 \theta} - 1)(e^{-u_2 \theta} - 1)}{e^{-u_1 \theta} - 1} \right]$</td>
<td>$1 + \frac{4}{\theta} \left[ \int_0^\theta t \frac{\exp(t) - 1}{\exp(t) - 1} , dt - 1 \right]$</td>
<td>$-\ln \frac{e^{-u_1 \theta} - 1}{e^{-u_1 \theta} - 1}$</td>
<td>$(-\infty, \infty)$</td>
</tr>
<tr>
<td>Clayton</td>
<td>$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$</td>
<td>$\frac{0}{\theta + 2}$</td>
<td>$\frac{1}{\theta} (t^{-\theta} - 1)$</td>
<td>$(0, \infty)$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\exp \left[ -{(\ln u_1)^0 + (\ln u_2)^0}^{\theta/\theta} \right]$</td>
<td>$\frac{\theta - 1}{\theta}$</td>
<td>$(-\ln t)^0$</td>
<td>$[1, \infty)$</td>
</tr>
</tbody>
</table>

The symbol "—" denotes that the mentioned item is not available; $\Phi^{-1}$ is the inverse standard normal distribution function.
Like the calibration of marginal distributions, the calibration of copula is also a ‘semi’ trial-and-error process: Among the six candidates, we fitted each of them to the observed dependence structure, then calculating the level of goodness-of-fit in terms of AIC. A summary of the calibration for the six copulas is also given in Table V. Accordingly, the Gaussian copula, with the lowest AIC score, is the best among the six for coupling the two marginal distributions (i.e., lognormal distributions) to form a joint probability function for PGA and CAV. By contrast, the independent copula is the least capable of modeling the PAG-CAV bivariate statistics from the TSMIP database. The earlier conclusions can also be drawn from the simulated samples (sample size N=40,385 as the observations) from the Gaussian and Independent copulas, showing that the best-fit Gaussian copula in Figure 5(b) can effectively reproduce the observed dependence structure in Figure 5(a). On the contrary, the Independent copula in Figure 5(c) is incapable of simulating the observed dependence structure.

4.3. Joint probability distribution for PGA and CAV in Taiwan

With the calibrations indicating the combination of the lognormal distributions and the Gaussian copula is the best-estimate model for the PGA-CAV data in the study, their joint probability function can be expressed as follows:

\[
F_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp \left[ -\frac{r_1^2 - 2\theta r_1 r_2 + r_2^2}{2(1-\theta^2)} \right] \, dr_1 \, dr_2
\]

where \(X_1\) and \(X_2\) represent PGA and CAV, respectively; for the model parameters, they are as follows: \(p_1 = -4.44, q_1 = 1.22, p_2 = -2.84, q_2 = 1.19, \) and \(\theta = 0.86\). Note that as mentioned previously, the model parameters were calculated from the 40,385 PGA and CAV samples collected from the TSMIP database in Taiwan.

With the copula model, Figure 3(b) shows the model forecast for the joint PGA-CAV frequency in terms of a sample size in 40,385 as of the observation. The results show that the empirical model (Figure 3(b)) and the observations (Figure 3(a)) are in good agreement, demonstrating the capability of the copula model for the PGA-CAV modeling. For comparison, Figure 3(c) shows the model forecast based on the Independent copula along with two lognormal distributions (i.e., best-fit marginal distributions for PGA and CAV). It can be seen that the change from the best-fit Gaussian copula makes the joint distribution deviate from the observations significantly, as a result of the model on the basis of the Independent copula ignoring the observed correlation between PGA and CAV.

5. PGA-CAV JOINT PROBABILITY FUNCTION SUBJECT TO GIVEN MAGNITUDE AND DISTANCE

Understandably, the analyses and model presented above focused on the PGA-CAV empirical data from the region around Taiwan associated with a wide range of magnitude and distance. As a result,
the model can provide the information about the PGA-CAV joint probability in general, while unable to estimate PGA-CAV joint probability subject to given magnitude and distance. However, in seismic hazard analysis, ground motion parameters are usually estimated using GMPEs based on explanatory variables of a ground motion, such as earthquake magnitude, source-to-site distance (e.g., [22]). Therefore, the PGA-CAV joint probability function subject to given magnitude and distance is essential to seismic hazard assessment. In the following, we adopt the copula approach to further develop the PGA-CAV joint probability function subject to given magnitude and distance based on our data used in this study.

5.1. PGA- and CAV-based GMPEs for Taiwan

The PGA- and CAV-based GMPEs for Taiwan are developed with the 40,385 earthquake data (PGA, CAV, magnitude, distance all available). Specifically, the GMPEs are developed with the functional form recommended in the literature [12]:

$$\ln(\text{IM}) = aM_L + b \ln \sqrt{D^2 + h^2} + c + \omega$$ (9)

where \(\ln\) is the natural logarithm; \(\text{IM}\) represents a specific intensity measure (i.e., PGA or CAV in this study); \(M_L\) is local earthquake magnitude; \(D\) is epicentral distance; \(\omega\) is model error; \(a\), \(b\), \(c\), and \(h\) are model parameters, with \(h\) referred to as fictitious focal depth [12]. Note that the four model parameters are obtained through the least-square algorithm to minimize the sum of squared error (SSE):

$$\text{SSE} = \sum_{i=1}^{N} [\ln(\text{IM}_i) - \ln(\text{IM}_i)]^2$$ (10)

where \(\text{IM}_i\) and \(\text{IM}_i\) denote the \(i\)-th observation and prediction of \(\text{IM}\), respectively; \(N=40,385\) is the sample size of the strong ground motion data.

On the basis of the functional form and the 40,385 strong ground motion data from TSMIP, the PGA- and CAV-based GMPEs are developed as follows:

$$\ln(\text{PGA}) = 0.926M_L - 0.787\ln \sqrt{D^2 + 26.783^2} - 6.098 + \omega_{\ln(\text{PGA})}$$ (11)

$$\ln(\text{CAV}) = 1.338M_L - 0.599\ln \sqrt{D^2 + 26.446^2} - 7.821 + \omega_{\ln(\text{CAV})}$$ (12)

where \(\omega_{\ln(\text{PGA})}\) and \(\omega_{\ln(\text{CAV})}\) are model errors following normal distributions with a mean of 0 and standard deviations of \(\sigma_{\ln(\text{PGA})} = 0.775\) and \(\sigma_{\ln(\text{CAV})} = 0.683\), respectively.

5.2. Normalized residuals for PGA and CAV

The predictions of \(\ln(\text{PGA})\) and \(\ln(\text{CAV})\) (denoted as \(\ln(\text{PGA})\) and \(\ln(\text{CAV})\)) subject to given magnitude and distance can be obtained through Equations (11) and (12). Then, together with the observations of \(\ln(\text{PGA})\) and \(\ln(\text{CAV})\) and the standard deviations (denoted as \(\sigma_{\ln(\text{PGA})}\) and \(\sigma_{\ln(\text{CAV})}\)) of the error terms in the two GMPEs, the normalized residuals of PGA and CAV can be calculated with the following formulae (e.g., [7, 8, 22, 39]):

$$\varepsilon_{\ln(\text{PGA})} = \frac{\ln(\text{PGA}_i) - \ln(\text{PGA})}{\sigma_{\ln(\text{PGA})}}$$ (13)

$$\varepsilon_{\ln(\text{CAV})} = \frac{\ln(\text{CAV}_i) - \ln(\text{CAV})}{\sigma_{\ln(\text{CAV})}}$$ (14)

where \(\varepsilon_{\ln(\text{PGA})}\) and \(\varepsilon_{\ln(\text{CAV})}\) denote the normalized residuals corresponding to the \(i\)-th ground motion record for PGA and CAV, respectively.
Figure 6(a) shows the joint histogram of the 40,385 samples of $\varepsilon_{\ln(\text{PGA})}$ and $\varepsilon_{\ln(\text{CAV})}$, and their statistics are summarized in Table I. In comparison with the PGA-CAV empirical data in Figure 3 (a), the normalized residuals have a more symmetrical and concentrated distribution against the mean values. Moreover, there is an evident correlation between the normalized residuals. Specifically, a Pearson correlation coefficient of $\rho = 0.84$ and a Kendall correlation coefficient of $\tau = 0.64$ are obtained from the samples. Note that the Pearson correlation coefficient varies significantly (i.e., from $\rho = 0.70$ to $\rho = 0.84$) while the Kendall correlation coefficient remains almost unchanged (i.e., from $\tau = 0.66$ to $\tau = 0.64$) when transforming the PGA and CAV data into the normalized residuals. The reason is that the Kendall correlation coefficient is based on the ranks of the samples, in contrast to the Pearson correlation coefficient relying on the values of the samples that could vary substantially after data transformation (e.g., [21, 22]). Note that the Pearson correlation coefficient between $\varepsilon_{\ln(\text{PGA})}$ and $\varepsilon_{\ln(\text{CAV})}$ is used as the Pearson correlation coefficient between $\ln(\text{PGA})$ and $\ln(\text{CAV})$ in conventional correlation analyses (e.g., [7, 8, 39]), which will introduce errors in quantifying the correlation between PGA and CAV. Therefore, in multivariate modeling of ground motion parameters, the Kendall correlation coefficient (based on rank) should be used over the Pearson correlation coefficient (based on value).

5.3. Marginal distributions and copula subject to given magnitude and distance

It is understood that $\ln(\text{PGA})$ and $\ln(\text{CAV})$ can be predicted using the regression models in Equations (11) and (12) given magnitude $M_L$ and distance $D$. According to the fundamentals of regression analysis, $\ln(\text{PGA})$ and $\ln(\text{CAV})$ follow normal distributions (PGA and CAV follow lognormal distributions) because the model errors $\varepsilon_{\ln(\text{PGA})}$ and $\varepsilon_{\ln(\text{CAV})}$ in the two regression models are defined to follow normal distributions. The means of $\ln(\text{PGA})$ and $\ln(\text{CAV})$ are the predictions by the regression models (i.e., $0.926M_L - 0.787\ln\sqrt{D^2 + 26.783^2} - 6.098$ and $1.338M_L - 0.599\ln\sqrt{D^2 + 26.446^2} - 7.821$, respectively). The standard deviations of $\ln(\text{PGA})$ and $\ln(\text{CAV})$ are the standard deviations of the model errors (i.e., 0.775 and 0.683, respectively). Therefore, the normalized residuals defined in Equations (13) and (14) will have a standard normal distribution, which can be demonstrated by the statistics in Table I and Figure 7 showing the empirical distributions of the normalized residuals are in good agreement with the standard normal distribution.

Besides, we also performed copula analysis on the PGA and CAV residuals, aiming to characterizing a best-fit model to capture their dependence structure. The AIC scores for the six copulas in Table IV are listed in Table V. The copula parameters are obtained from the Kendall correlation coefficient of $\tau = 0.64$, which are also summarized in Table V. It is clear that the best-fit copula for the normalized residuals is also the Gaussian copula with the lowest AIC score. It is not surprising that the normalized residuals have the same dependence structure or copula as the PGA and CAV data. Like the Kendall correlation coefficient, the copula function is also rank-dependent and is invariant against strictly monotonic linear and nonlinear transformations of random variables (e.g., [21, 22]). Therefore, PGA and CAV will have a dependence structure of the Gaussian copula regardless of the GMPEs considered. Figure 8 shows the similar plots with Figure 5. Like the PGA and CAV empirical data, the best-fit Gaussian copula in Figure 8(b) can effectively reproduce the
observed dependence structure between the normalized residuals in Figure 8(a). On the contrary, the Independent copula in Figure 8(c) is incapable of simulating the observed dependence structure.

5.4. The resulting PGA-CAV joint probability distribution

Similarly, the combination of the lognormal distributions and the Gaussian copula is the best-estimate model for the PGA-CAV data subject to given magnitude and distance. As a result, the PGA-CAV joint probability function subject to given magnitude $M_L$ and distance $D$ can be expressed as follows:

$$F_{X_1,X_2}(x_1,x_2|M_L,D) = \int_{-\infty}^{\ln x_1/p_1} \int_{-\infty}^{\ln x_2/p_2} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp \left[ -\frac{r_1^2 - 2\theta r_1 r_2 + r_2^2}{2(1-\theta^2)} \right] dr_1 dr_2$$

(15)

where $X_1$ and $X_2$ denote PGA and CAV, respectively; The model parameters are as follows: $p_1 = 0.926 M_L - 0.787 \ln \sqrt{D^2 + 26.783^2 - 6.098}$, $p_2 = 1.338 M_L - 0.599 \ln \sqrt{D^2 + 26.446^2 - 7.821}$, $q_1 = 0.775$, $q_2 = 0.683$, and $\theta = 0.84$.

Note that although the PGA-CAV joint probability function subject to given magnitude and distance (i.e., Equation (15)) has the same functional form as the PGA-CAV joint probability function irrespective of magnitude and distance (i.e., Equation (8)), they have rather different model parameters. The model parameters in Equation (15) are functions of the given magnitude and distance and the joint probability function varies with changing magnitude and distance. This feature is especially useful for seismic hazard analysis. This is because seismic hazard analysis needs to calculate the exceedance probability for one parameter (e.g., PGA) conditioned on the other parameter (e.g., CAV) given magnitude and distance of possible earthquake ruptures (e.g., [1, 7, 8, 39]). Figures 6(b) and 6(c) show the model forecasts for the joint frequency of PGA-CAV normalized residuals using the Gaussian copula and the Independent copula, respectively. The same conclusion as that for the

Figure 7. Peak ground acceleration (PGA)-cumulative absolute velocity (CAV) normalized residuals and the standard normal distribution.

Figure 8. Observed and simulated $U_1$-$U_2$ data for peak ground acceleration-cumulative absolute velocity normalized residuals.
PGA-CAV empirical data in Figure 3 can be drawn: the Gaussian copula outperforms the Independence copula in the PGA-CAV modeling.

6. DISCUSSION: ADEQUACY OF THE MULTIVARIATE LOGNORMAL DISTRIBUTION

It can be concluded from the previous sections that PGA and CAV can be marginally modeled as a lognormal variate, and their dependence structure can be captured by the Gaussian copula either the empirical data or the normalized residuals are considered. The combination of lognormal marginals and the Gaussian copula is essentially the multivariate lognormal distribution. Therefore, the multivariate lognormal distribution (e.g., [1]) provides adequate characterization of the joint probability distribution of multiple ground motion parameters in Taiwan. This finding also validates the conventional two-step approach [22] for developing empirical GMPEs of multiple ground motion parameters, which involves regression analysis of the individual parameters and subsequently characterizing their dependence by evaluating the Pearson correlation coefficient of the logarithm of the ground motion parameters corrected by using a GMPE.

7. CONCLUSIONS

This study successfully used the copula approach to calibrate the PGA-CAV joint probability function for the strong ground motion data from Taiwan. Both the PGA-CAV joint probability function subject to given magnitude and distance and the PGA-CAV joint probability function irrespective of magnitude and distance were developed. The model development was based on a total of 40,385 earthquake time histories (associated with 312 earthquakes above $M_L$ 5.0) from the TSMIP. Empirical GMPEs for predicting PGA and CAV were developed, and the correlation results using the empirical data and the normalized residuals were compared. It was found that a strong correlation existed between PGA and CAV for both the empirical data and the normalized residuals. The multivariate lognormal distribution composed of two lognormal marginal distributions and the Gaussian copula provides adequate characterization of the PGA-CAV joint distribution observed in Taiwan. This finding validates the conventional two-step approach for developing empirical GMPEs of multiple ground motion parameters from the copula viewpoint.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Project No. 51509188), the Natural Science Foundation of Hubei Province of China (Project No. 2015CFBS09), and China Postdoctoral Science Foundation (Project No. 2015M572196). The constructive comments from the two anonymous reviewers are greatly appreciated.

REFERENCES