Soil Dynamics and Earthquake Engineering 115 (2018) 324–330

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

Copula-based earthquake early warning decision-making strategy

J.P. Wang[b], Xiao-Song Tang[a,*], Yih-Min Wu[c], Dian-Qing Li[a]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan University, Wuhan 430072, China

[b] Department of Civil Engineering, National Central University, Taiwan

[c] Department of Geosciences, National Taiwan University, Taiwan

ARTICLE INFO

Keywords:
Earthquake early warning
Copula approach
Joint distribution
Reliability

ABSTRACT

This paper presents a new copula-based earthquake early warning (EEW) decision-making strategy, aiming to characterize missed-alarm and false-alarm probabilities for an on-site EEW and determine an optimum threshold at which EEW should be set off with the lowest missed-alarm and false-alarm probabilities combined. On the basis of an existing PD_3-PGV (PD_3: peak ground displacement within the first three seconds after P-wave arrives; PGV: peak ground velocity) on-site EEW, the analysis shows that a copula model consisting of the Lognormal distribution, Weibull distribution, and Frank copula can satisfactorily model the PD_3-PGV joint probability distribution. Accordingly, the optimum PD_3 triggering thresholds for different PGV warning thresholds from 5 to 35 cm/s are presented for future references in the use of the PD_3-PGV on-site EEW with maximum reliability.

1. Introduction

From recent events, it is a fact that we cannot predict earthquakes for effective hazard mitigation. The main reason for not being able to predict earthquakes is that we cannot monitor the stress and strain conditions of rock a few kilometers below the ground surface where the focal points are located [1]. As a result, several methods were developed for seismic hazard mitigation, such as seismic hazard assessment based on geological and seismological data. For example, Wang et al. [2] conducted a probabilistic seismic hazard analysis (PSHA) for Taipei city, estimating the levels of PGA (peak ground acceleration) that the sites could encounter considering different periods of time. Other PSHA case studies include those by Ayele [3], Silacheva et al. [4], and Nakajima et al. [5], all aiming to best estimate earthquake-induced PGA (or SA: spectral acceleration) in different return periods for preparing a corresponding earthquake-resistant design for the target sites.

Unlike seismic hazard analysis that develops site-specific earthquake-resistant designs considering earthquake uncertainties, earthquake early warning (EEW) aims to detect and interpret an occurring earthquake as early as possible, then disseminate warning messages to the public before the arrival of peak earthquake motions (mainly utilizing the nature that radio waves can travel much faster than seismic waves). Nowadays, EEW has been implemented in several regions around the world, including Taiwan, Japan, California, Mexico, Romania and Turkey [6–19]. Understandably, EEW aims to send out warning messages tens of seconds before the arrival of peak ground shaking. Within the lead time, immediate actions can be taken for hazard mitigation, including shutting down the operation of critical facilities (e.g., nuclear power plants), decelerating high-speed trains, etc. Generally speaking, there are two types of EEW, referred to as regional and on-site systems. The former is to interpret data that have been detected by multiple seismic stations close to the earthquake, then sending out warning messages to farther sites. By contrast, an on-site system is based on the less strong ground motions that have been detected at a site, from which the magnitude of peak shaking (arriving later at the same site) can be estimated with an empirical relationship; and as the estimated peak shaking is strong, on-site EEW will be set off [16–19].

As a result, the essential of an on-site EEW is the relationship between early motions and peak motions (the larger the early motions, the larger the peak motions). For instance, Wu and Kanamori [18] collected 780 strong-motion data from Taiwan, Japan, and Southern California, and discovered a moderate correlation between peak ground displacement within the first three seconds after P-wave arrives (PD_3) and peak ground velocity (PGV), which has become the backbone of the on-site EEW system implemented in Taiwan [19].

Like warning systems for landslides, debris flows, etc., missed and false alarms in EEW are inevitable due to imperfect prediction models. Therefore, in addition to EEW methodology, many studies on EEW reliability have also been reported. For instance, Iervolino et al. [20]

* Corresponding author.
E-mail address: xstang@whu.edu.cn (X.-S. Tang).

https://doi.org/10.1016/j.soildyn.2018.08.031
Received 26 December 2017; Received in revised form 19 July 2018; Accepted 25 August 2018

considered ground motion prediction equations play a more critical role in regional EEW than the models for determining the real-time forecasts of earthquake magnitude and location. Wang et al. [16] proposed that using multiple precursors should be able to improve EEW reliability, based on a case study using the PEER (Pacific Earthquake Engineering Research Center) ground motion database. Note that such a suggestion is in agreement with Böse et al. [21] who found the frequency of false alarms can be substantially reduced with multiple precursors added to the equation. Other studies concerning EEW reliability include the adoption of risk-based and performance-based decision-making for EEW [22,23], selection of a proper empirical model as the core in EEW computation [24], and reliability evaluation for the on-site EEW implemented in Taiwan [19].

In this paper, a new application of the copula approach to estimate the probabilities of missed alarm and false alarm for an on-site EEW was presented. On the basis of an existing PD_3-PGV on-site EEW, we found that a copula model consisting of the Lognormal distribution, Weibull distribution, and Frank copula can satisfactorily model the PD_3-PGV joint probability distribution. More importantly, based on the joint distribution model for PD_3 and PGV, the probability of missed alarm (the probability of PD_3 less than a triggering threshold while PGV greater than a warning threshold) and the probability of false alarm (the probability of PD_3 greater than a triggering threshold while PGV less than a warning threshold) were estimated. Finally, the optimum PD_3 triggering thresholds for different PGV warning thresholds were presented for future references in the use of the PD_3-PGV on-site EEW with maximum reliability.

2. The copula approach

2.1. Methodology

The copula approach is to develop a joint probability distribution for a given multivariate dataset, i.e., to calibrate a joint distribution function based on the best-fit marginal distributions for each variable, and the best-fit copula function for modeling the dependence structure between variables. As a result, the calibration process is referred to as a de-coupled calibration, which can be conducted more easily than the direct calibration on a multivariate model (e.g., a multivariate Normal distribution). Since the copula approach can provide a number of joint distribution models with different combinations of marginal distributions and copula functions, the chance of capturing the trend of a multivariate dataset can be significantly increased [25].

To further illustrate the copula approach, a bivariate example is presented in the following. Given both X and Y follow the Uniform distribution from a to b (a and b are constants; $a < b$), their marginal cumulative distribution functions (CDFs), $F_X(x)$ and $F_Y(y)$, can be respectively expressed as:

$$F_X(x) = \Pr(X \le x) = \frac{x-a}{b-a} \tag{1}$$

and

$$F_Y(y) = \Pr(Y \le y) = \frac{y-a}{b-a} \tag{2}$$

Then considering their dependence structure can be modeled by the Clayton copula, its copula function, $C(u_X, u_Y; \theta)$, is given by [25]:

$$C(u_X, u_Y; \theta) = \{u_X^{-\theta} + u_Y^{-\theta} - 1\}^{-1/\theta} = \{[F_X(x)]^{-\theta} + [F_Y(y)]^{-\theta} - 1\}^{-1/\theta} \tag{3}$$

where θ is the copula parameter for the Clayton copula; $u_X = F_X(x)$ and $u_Y = F_Y(y)$ are the marginal CDFs for X and Y, respectively. Note that for a copula $C(u_X, u_Y; \theta)$, its parameter θ can be determined from the Kendall rank correlation coefficient, τ, between X and Y using the following relationship [26]:

$$\tau = 4\int_0^1 \int_0^1 C(u_X, u_Y; \theta)\mathrm{d}C(u_X, u_Y; \theta) - 1 \tag{4}$$

With the three sub-models calibrated, the joint CDF for X and Y, $F_{X,Y}(x, y)$, can be developed by substituting Eq. (1) and Eq. (2) into Eq. (3):

$$F_{X,Y}(x, y) = Pr(X \le x \mathbf{AND} Y \le y) = \left\{\left[\frac{x-a}{b-a}\right]^{-\theta} + \left[\frac{y-a}{b-a}\right]^{-\theta} - 1\right\}^{-1/\theta} \tag{5}$$

2.2. Applications

The copula approach has been increasingly utilized in different probabilistic studies, including slope stability assessment [27–30], flood frequency evaluation [31], traffic data analysis [32], financial engineering [33], as well as earthquake engineering and applied seismology [34–41]. For instance, in order to develop the joint distribution model for a structure's peak and residual displacements under earthquake condition, Goda [34] developed a copula model consisting of the Frechet distribution, generalized Pareto distribution, and asymmetrical Gumbel copula that can satisfactorily model the target joint probability. Goda and Tesfamariam [35] proposed a model consisting of two Frechet distributions, one generalized Pareto distribution, one asymmetrical Gumbel copula, and one Gaussian copula for capturing the joint distribution of the three variables: namely maximum inter-story drift ratio, residual inter-story drift ratio, and peak floor acceleration. Xu et al. [38] discovered a model consisting of two Lognormal distributions and one Gaussian copula can satisfactorily capture the joint distribution function for PGA (peak ground acceleration) and CAV (cumulative acceleration velocity), two very important ground motion intensity measures in earthquake engineering.

3. PD_3-PGV data and joint distribution

As mentioned previously, Wu and Kanamori [18] collected 780 pairs of PD_3 and PGV data for on-site EEW development. As shown in Fig. 1, such an early precursor (PD_3) and peak motion (PGV) has a Kendall rank correlation coefficient of $\tau = 0.70$. It is clear that the two variables' joint probability function cannot be delineated directly from Fig. 1. As a result, we reprocessed the data and plotted their observed joint histogram as Fig. 2, which will be modeled with the copula approach.

3.1. Calibration of marginal distributions

The first step of the model development is to calibrate the marginal CDFs for PD_3 and PGV individually. Four common univariate

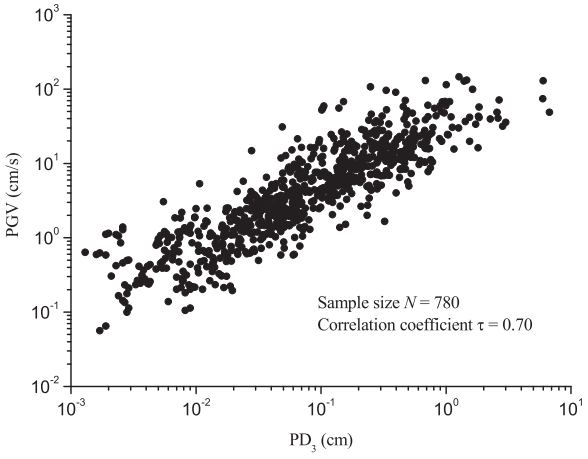

Fig. 1. Scatter plot of the 780 paired PD_3-PGV observations [18].

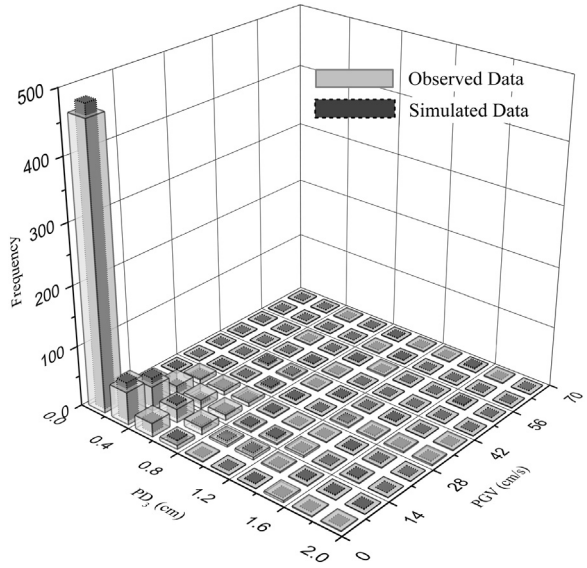

Fig. 2. The observed and predicted joint histogram for PD_3 and PGV.

distributions were selected, namely the Normal distribution truncated below zero (referred to as TruncNomal), Gumbel distribution truncated below zero (referred to as TruncGumbel), Lognormal distribution and Weibull distribution. Note that all the four distributions are defined on positive real numbers, thus satisfying the requirements of positive values for both PD_3 and PGV. In this study, Akaike Information Criterion (AIC) [26] is adopted to identify the best-fit marginal distribution. The formula of AIC for identifying the best-fit marginal distribution is shown in Appendix. It is clear that AIC is related to the log-likelihood for a specified distribution and the number of unknown parameters to be calibrated in this distribution. Therefore, AIC favors a distribution that has the maximum log-likelihood and minimum number of unknown parameters. That is, a distribution producing the lowest AIC value among the set of candidate distributions is considered as the best-fit distribution.

Table 1 summarizes the AIC values for the four distributions. Based on the data for PD_3 in Fig. 1, the AIC values for the TruncNormal, TruncGumbel, Lognormal and Weibull distributions are calculated as 560, −128, −1062 and −974, respectively. It is clear that the Lognormal distribution with the lowest AIC value of −1062 is the best-fit distribution for PD_3. Similarly, with the data for PGV in Fig. 1, the AIC values are obtained as 6190, 5648, 5142 and 5058 for the TruncNormal, TruncGumbel, Lognormal and Weibull distributions, respectively. Therefore, the Weibull distribution with the lowest AIC value of 5058 is the best-fit distribution for PGV. To verify the goodness-of-fit of the identified best-fit distributions, Figs. 3 and 4 show the observations (sample size $N = 780$) along with the four model predictions for PD_3 and PGV, respectively. From the plots, it can also be seen that the Lognormal distribution provides the best fit to the PD_3 observations, whereas the Weibull distribution is the best-fit distribution to model the PGV observations.

Table 1
Summary of the calibration for the best-fit marginal distributions.

Marginal distribution	AIC value	
	PD_3	PGV
TruncNormal	560	6190
TruncGumbel	−128	5648
Lognormal	**−1062** (best-fit)	5142
Weibull	−974	**5058** (best-fit)

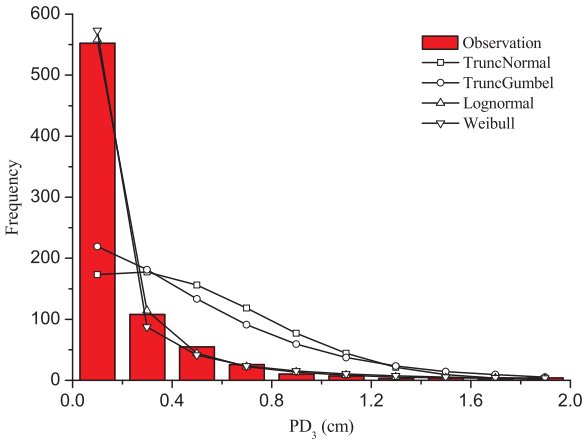

Fig. 3. Marginal distribution for PD_3: observation and four model predictions.

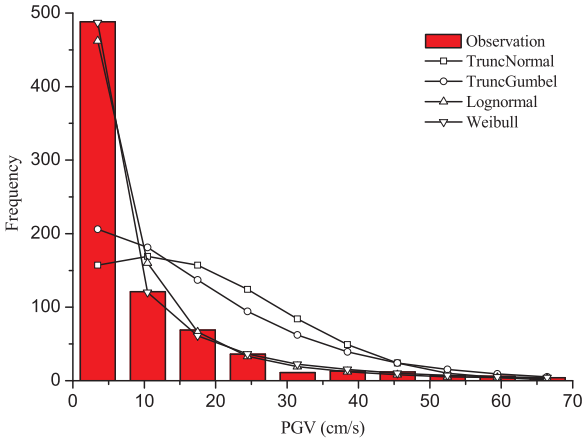

Fig. 4. Marginal distribution for PGV: observation and four model predictions.

3.2. Calibration of dependence structure with copula

Next, we calibrated the dependence structure between PD_3 and PGV. Like the marginal CDF calibration, six commonly used copulas were selected, namely the Gaussian copula, Plackett copula, Frank copula, Clayton copula, Gumbel copula and Independent copula. The copula functions for the six copulas are summarized in Table 2. According to Xu et al. [38], this group of copulas provides adequate diversity that should be able to model a given bivariate dataset. Similarly, AIC is used in this study to identify the best-fit copula. The formula of AIC for identifying the best-fit copula is also shown in Appendix. It can be seen that AIC is related to the log-likelihood for a specified copula and the number of unknown copula parameters to be calibrated in this copula. As a result, AIC favors a copula that has the maximum log-likelihood and minimum number of unknown copula parameters. That is, a copula with the lowest AIC value among the set of candidate copulas is taken as the best-fit copula. Based on the data in Fig. 1, the AIC values for the Gaussian, Plackett, Frank, Clayton, Gumbel and Independent copulas are obtained as −1020, −1044, −1127, −465, −888 and 2, respectively. It is clear that the Frank copula with the lowest AIC value of −1127 is the best-fit copula for describing the dependence structure between PD_3 and PGV.

3.3. Joint distribution model for PD_3 and PGV

Given the marginal distributions for PD_3 and PGV are respectively calibrated as the Lognormal distribution and the Weibull distribution, and the dependence structure between PD_3 and PGV is calibrated as the Frank copula, the joint CDF for PD_3 and PGV, $F_{PD_3,PGV}(PD_3^*, PGV^*)$, can

Table 2
Summary of the calibration for the best-fit copula.

Copula	Functional form, $C(u_X, u_Y; \theta)$	AIC value
Independent	$u_X u_Y$	2
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u_X)} \int_{-\infty}^{\Phi^{-1}(u_Y)} \frac{1}{2\pi\sqrt{1-\theta^2}} exp\left[-\frac{r_X^2 - 2\theta r_X r_Y + r_Y^2}{2(1-\theta^2)}\right] dr_X dr_Y,\ \theta \in \left[-1,\ 1\right]$	−1020
Plackett	$\frac{S-\sqrt{S^2-4u_X u_Y \theta(\theta-1)}}{2(\theta-1)},\ S = 1 + (\theta-1)(u_X + u_Y),\ \theta \in (0,\ \infty)/\{1\}$	−1044
Frank	$-\frac{1}{\theta} ln\left[1 + \frac{(e^{-\theta u_X}-1)(e^{-\theta u_Y}-1)}{e^{-\theta}-1}\right],\ \theta \in (-\infty,\ \infty)/\{0\}$	**−1127** (best-fit)
Clayton	$(u_X^{-\theta} + u_Y^{-\theta} - 1)^{-1/\theta},\ \theta \in (0,\ \infty)$	−465
Gumbel	$exp\{-[(-ln\ u_X)^\theta + (-ln\ u_Y)^\theta]^{1/\theta}\},\ \theta \in (1,\ \infty)$	−888

be expressed as Eq. (6) by substituting the two marginal distributions into the Frank copula:

$$F_{PD_3,PGV}(PD_3^*,\ PGV^*) = Pr(PD_3 \le PD_3^* \textbf{AND} PGV \le PGV^*) = -\frac{1}{\theta} ln\left[1 + \frac{(e^{-\theta u_{PD_3}}-1)(e^{-\theta u_{PGV}}-1)}{e^{-\theta}-1}\right] \quad (6)$$

where θ is the copula parameter for the Frank copula, and is estimated as 11.48 by solving Eq. (4) using $\tau = 0.70$; PD_3^* and PGV* are the arguments for PD_3 and PGV, respectively; u_{PD_3} and u_{PGV} are the marginal CDFs for PD_3 and PGV, respectively. For PD_3 following the Lognormal distribution, u_{PD_3} is as follows:

$$u_{PD_3} = F_{PD_3}(PD_3^*) = \Pr(PD_3 \le PD_3^*) = \Phi\left(\frac{\ln(PD_3^*) - p_1}{q_1}\right) \quad (7)$$

where Φ denotes the CDF of the standard Normal distribution; p_1 and q_1 are the two distribution parameters governing the marginal CDF for PD_3. Based on the mean and standard deviation of PD_3, p_1 and q_1 are respectively estimated as −2.37 and 1.34 using the method of moments. For PGV following the Weibull distribution, u_{PGV} is:

$$u_{PGV} = F_{PGV}(PGV^*) = \Pr(PGV \le PGV^*) = 1 - \exp\left[-\left(\frac{PGV^*}{p_2}\right)^{q_2}\right] \quad (8)$$

where p_2 and q_2 are the two distribution parameters governing the marginal CDF for PGV. Specifically, p_2 and q_2 are the so-called scale and shape parameters for the Weibull distribution. Based on the mean and standard deviation of PGV, p_2 and q_2 are respectively estimated as 7.26 and 0.62 using the method of moments.

Fig. 2 also shows the predicted joint histogram by the joint distribution model for PD_3 and PGV in Eq. (6). Note that the predicted joint histogram is produced using 780 pairs of simulated PD_3 and PGV data from the fitted joint distribution model. The choice of a sample size of $N = 780$ for the simulated data is to facilitate comparison with the observed joint histogram which is also produced using 780 pairs of observed PD_3 and PGV data. It can be seen from Fig. 2 that the predicted joint histogram is in good agreement with the observed joint histogram, indicating that the fitted joint distribution model for PD_3 and PGV in Eq. (6) can satisfactorily model the given PD_3-PGV bivariate dataset.

4. Model application to on-site EEW reliability assessment

4.1. Missed-alarm and false-alarm probabilities

As mentioned previously, missed alarm and false alarm in EEW are inevitable owing to natural randomness. For the PD_3-PGV on-site EEW, missed alarm refers to the situation when the actual PGV is greater than a pre-determined warning threshold, WT_{PGV}, while EEW is not set off (i.e., the detected PD_3 is less than a pre-determined triggering threshold, TT_{PD_3}). On the other hand, false alarm refers to the situation when the actual PGV is less than WT_{PGV}, while EEW is set off (i.e., the detected PD_3 is greater than TT_{PD_3}). Therefore, the probability of missed alarm is the probability of PD_3 less than TT_{PD_3} while PGV greater than WT_{PGV}, whereas the probability of false alarm is the probability of PD_3 greater than TT_{PD_3} while PGV less than WT_{PGV}. Mathematically, the probabilities of missed alarm and false alarm can be respectively expressed as:

$$Pr(Missed\ alarm) = Pr(PD_3 \le TT_{DD_3} \textbf{AND} PGV > WT_{PGV}) \quad (9)$$

and

$$Pr(False\ alarm) = Pr(PD_3 > TT_{PD_3} \textbf{AND} PGV \le WT_{PGV}) \quad (10)$$

According to probability theory, Eqs. (9) and (10) can be respectively rewritten as:

$$Pr(Missed\ alarm) = F_{PD_3}(TT_{PD_3}) - F_{PD_3,PGV}(TT_{PD_3},\ WT_{PGV}) \quad (11)$$

and

$$Pr(False\ alarm) = F_{PGV}(WT_{PGV}) - F_{PD_3,\mathrm{PGV}GV}(TT_{PD_3},\ WT_{PGV}) \quad (12)$$

It is clear that the probabilities of missed alarm and false alarm are functions of TT_{PD_3} and WT_{PGV}. Therefore, given TT_{PD_3} and WT_{PGV}, the probabilities of missed alarm and false alarm can be readily obtained by using Eqs. (11) and (12) based on the calibrated marginal and joint CDFs for PD_3 and PGV. For instance, given $TT_{PD_3} = 1$ cm and $WT_{PGV} = 10$ cm/s, the PD_3-PGV on-site EEW should have a missed-alarm probability of 25.85% and a false-alarm probability of 0.16%. The surface plots of the probabilities of missed alarm and false alarm are shown in Figs. 5 and 6, respectively. It can be seen that the increase of TT_{PD_3} or decrease of WT_{PGV} increases the probability of missed alarm

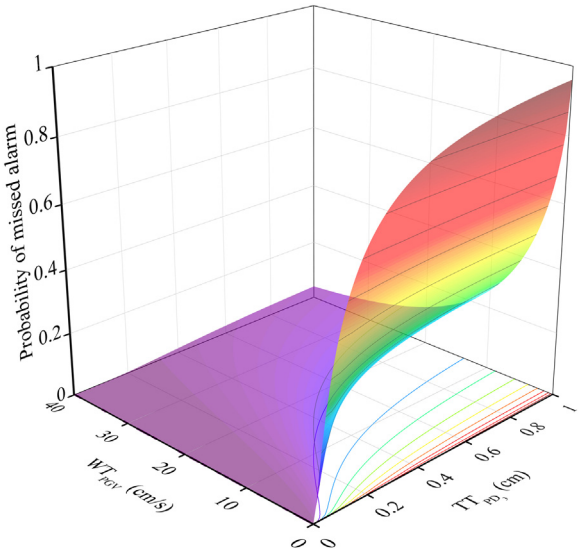

Fig. 5. Surface plot of probability of missed alarm against TT_{PD_3} and WT_{PGV}.

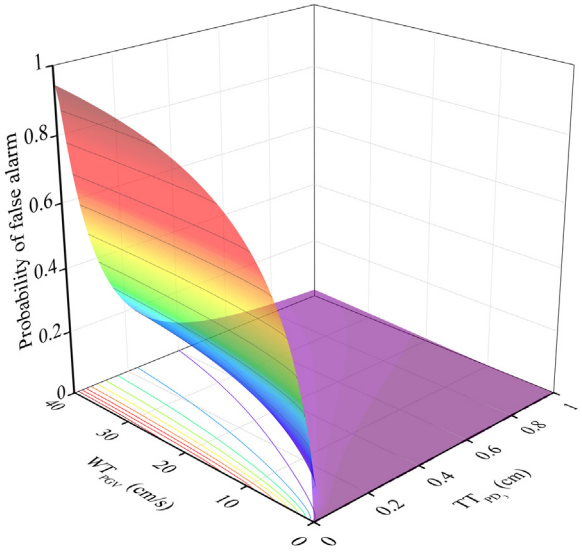

Fig. 6. Surface plot of probability of false alarm against $TT_{PD_3}$and WT_{PGV}.

but decreases the probability of false alarm.

4.2. Method for determining optimum triggering threshold

For the PD_3-PGV on-site EEW, WT_{PGV} is generally pre-determined by engineers based on the characteristics of a structure, such as design resistance, the consequence of structural failure, etc. That is, if a structure is designed and constructed with high capacity against PGV, a large value of WT_{PGV} is adopted; otherwise a lower one is used. Then, TT_{PD_3} is determined using a PD_3-PGV empirical equation based on WT_{PGV}. Note that this method for determining TT_{PD_3} ignores the uncertainties in the PD_3-PGV regression model. In this study, a feasible and more objective method for determining TT_{PD_3} based on WT_{PGV} is proposed, which involves an optimization of the probability of missed + false alarm with respect to TT_{PD_3}. The optimum TT_{PD_3} is the one that minimizes the probability of missed + false alarm, or equivalently maximizes the reliability for the PD_3-PGV on-site EEW.

Fig. 7 shows the surface plot of the probability of missed + false alarm. Like the probabilities of missed alarm and false alarm, the probability of missed + false alarm is also a function of TT_{PD_3} and

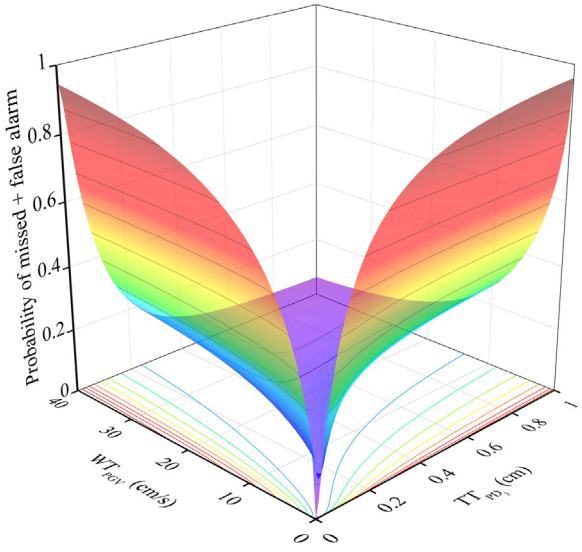

Fig. 7. Surface plot of probability of missed + false alarm against $TT_{PD_3}$and WT_{PGV}.

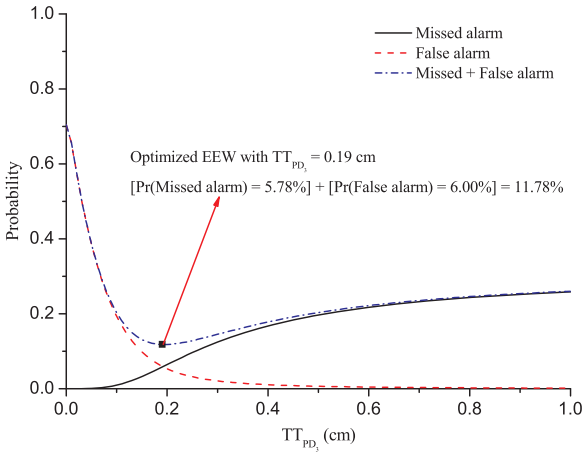

Fig. 8. Probabilities of missed alarm, false alarm, and missed + false alarm for various $TT_{PD_3}$when WT_{PGV} = 10 cm/s.

WT_{PGV}. However, for a given or fixed WT_{PGV}, the probabilities of missed alarm, false alarm, and missed + false alarm are all a function of TT_{PD_3} only. Under this condition, the probabilities of missed alarm, false alarm, and missed + false alarm are reduced from a surface to a curve. As an example, Fig. 8 shows the curves for the probabilities of missed alarm, false alarm, and missed + false alarm when WT_{PGV} = 10 cm/s. It can be seen that the probability of missed alarm increases with increasing TT_{PD_3}, whereas the probability of false alarm decreases with increasing TT_{PD_3}. On the other hand, the probability of missed + false alarm firstly decreases, and then increases when TT_{PD_3} ranges from 0 to 1 cm. As a result, an optimum TT_{PD_3} of 0.19 cm is obtained from the optimization, which produces the minimum probability of missed + false alarm of 11.78%, or the maximum reliability of 88.22% for the PD_3-PGV on-site EEW.

4.3. Optimum triggering thresholds given WT_{PGV} from 5 to 35 cm/s

Fig. 9 shows the optimum TT_{PD_3} and the resulting minimum probabilities of missed + false alarm for WT_{PGV} ranging from 5 to 35 cm/s. These results are obtained by applying the proposed method in Section 4.2 and changing the given WT_{PGV}. Note that the range of WT_{PGV} from 5 to 35 cm/s is broad enough to cover all kinds of scenarios. It can be seen that the optimum TT_{PD_3} increases from 0.11 cm to 1.51 cm when WT_{PGV} increases from 5 cm/s to 35 cm/s. On the other hand, the minimum probability of missed + false alarm decreases from 12.01% to 6.93% when WT_{PGV} ranges from 5 cm/s to 35 cm/s. The derived optimum

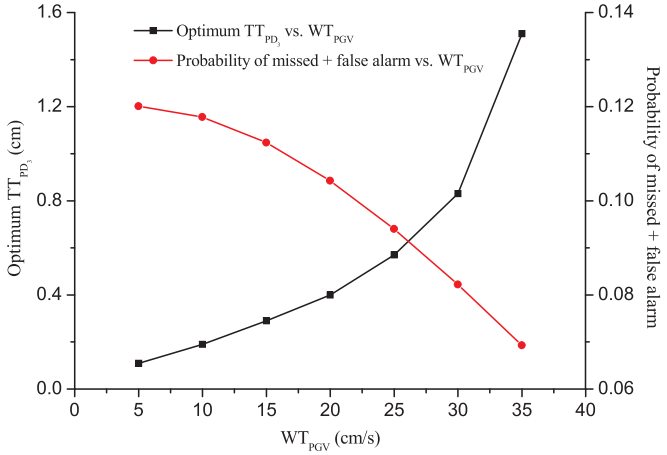

Fig. 9. Optimum $TT_{PD_3}$and resulting probabilities of missed + false alarm for WT_{PGV} from 5 to 35 cm/s.

TT_{PD_3} for different given WT_{PGV} in Fig. 9 can serve as future references in the use of the PD3-PGV on-site EEW with maximum reliability.

5. Summary and conclusion

The paper presents a new application of the copula approach to on-site EEW reliability assessment. On the basis of an existing PD_3-PGV on-site EEW, we found that a copula model consisting of the Lognormal distribution, Weibull distribution, and the Frank copula can satisfactorily model the PD_3-PGV joint probability distribution. Based on the calibrated PD_3-PGV joint probability distribution, the probabilities of missed alarm and false alarm can be easily calculated.

Another highlight of the study is the new method for determining the optimum PD_3 triggering threshold subject to a pre-determined PGV warning threshold that depends on the characteristics of structures. The proposed method is based on optimization in searching for the optimum value in corresponding to the minimum probability of missed + false alarm. Notably, the calculation can be achieved with the calibrated joint probability distribution present herein for calculating missed-alarm and false-alarm probabilities, a new application of the copula approach to propose decision-making strategies for on-site earthquake early warning.

Acknowledgements

The authors appreciate the review and comments from the Editor and Reviewers. The financial support on the study from National Central University Taiwan, the National Natural Science Foundation of China (Project No. 51509188, 51579190, 51779189) and the Fundamental Research Funds for the Central Universities (Project No. 2042017kf0247, 2042018kf0243) is also appreciated.

Appendix. : AIC

The AIC value for identifying the best-fit marginal distribution is defined as

$$AIC = -2\sum_{i=1}^{N} ln\, f\,(PD_{3i}\ or\ PGV_i;\ p,\ q) + 2k_1 \tag{A.1}$$

where $\sum_{i=1}^{N} ln\, f\,(PD_{3i}\ or\ PGV_i;\ p,\ q)$ is the log-likelihood for a specified distribution in which $f\,(PD_3\ or\ PGV;\ p,\ q)$ is the probability density function (PDF) for PD_3 or PGV; PD_{3i} or PGV_i is the i-th observation for PD_3 or PGV; $N = 780$ is the sample size of the observed PD_3 and PGV data; $k_1 = 2$ is the number of distribution parameters for all the four selected distributions.

Similarly, to identify the best-fit copula, the AIC value is given by

$$AIC = -2\sum_{i=1}^{N} ln\, c\left(u_{PD_{3i}}, u_{PGV_i}; \theta\right) + 2k_2 \tag{A.2}$$

where $\sum_{i=1}^{N} ln\, c\,(u_{PD_{3i}}, u_{PGV_i};\ \theta)$ is the log-likelihood for a specified copula in which $c\,(u_{PD_3},\ u_{PGV};\ \theta) = \partial^2 C\,(u_{PD_3},\ u_{PGV};\ \theta)/\partial u_{PD_3}\partial u_{PGV}$ is the copula density function; $k_2 = 1$ is the number of copula parameters for the Gaussian, Plackett, Frank, Clayton and Gumbel copulas and $k_2 = 0$ is the number of copula parameters for the Independent copula; $\{(u_{PD_{3i}},\ u_{PGV_i}),\ i = 1,\ 2,\ ...,\ N\}$ are the empirical distribution values of observed PD_3 and PGV data, which are given by

$$\begin{cases} u_{PD_{3i}} = \frac{rank\,(PD_{3i})}{N+1} \\ u_{PGV_i} = \frac{rank\,(PGV_i)}{N+1} \end{cases} i = 1,\ 2,\ ...,N \tag{A.3}$$

in which $rank(PD_{3i})$ [or $rank(PGV_i)$] denotes the rank of PD_{3i} (or PGV_i) among the list $\{PD_{31}, ..., PD_{3N}\}$ (or $\{PGV_1, ..., PGV_N\}$) in an ascending order.

References

[1] Geller RJ, Jackson DD, Kagan YY, Mulargia F. Earthquakes cannot be predicted. Science 1997;275(5306):1616.
[2] Wang JP, Huang D, Cheng CT, Shao KS, Wu YC, Chang CW. Seismic hazard analyses for Taipei city including deaggregation, design spectra, and time history with excel applications. Comput Geosci 2013;52:146–54.
[3] Ayele A. Probabilistic seismic hazard analysis (PSHA) for Ethiopia and the neighboring region. J Afr Earth Sci 2017;134:257–64.
[4] Silacheva NV, Kulbayeva UK, Kravchenko NA. Probabilistic seismic hazard assessment of Kazakhstan and Almaty city in peak ground accelerations. Geod Geodyn 2018;9(2):131–41.
[5] Nakajima M, Choi IK, Ohtori Y, Choun YS. Evaluation of seismic hazard curves and scenario earthquakes for Korean sites based on probabilistic seismic hazard analysis. Nucl Eng Des 2007;237(3):277–88.
[6] Wu YM, Kanamori H. Experiment on an onsite early warning method for the Taiwan early warning system. Bull Seismol Soc Am 2005;95(1):347–53.
[7] Allen RM, Gasparini P, Kamigaichi O, Böse M. The status of earthquake early warning around the world: an introductory overview. Seismol Res Lett 2009;80(5):682–93.
[8] Erdik M, Fahjan Y, Ozel O, Alcik H, Mert A, Gul M. Istanbul earthquake rapid response and the early warning system. Bull Earthq Eng 2003;1(1):157–63.
[9] Wu YM, Teng TL. A virtual subnetwork approach to earthquake early warning. Bull Seismol Soc Am 2002;92(5):2008–18.
[10] Hoshiba M, Kamigaichi O, Saito M, Tsukada S, Hamada N. Earthquake early warning starts nationwide in Japan. EOS 2008;89(8):73–4.
[11] Nakamura Y. Development of earthquake early-warning system for the hinkansen: some recent earthquake engineering research and practice in Japan. Jpn Natl Comm Int Assoc Earthq Eng 1984:224–38.
[12] Alcik H, Ozel O, Apaydin N, Erdik M. A study on warning algorithms for Istanbul earthquake early warning system. Geophys Res Lett 2009;36(5). [L00B05].
[13] Espinosa-Aranda JM, Jiménez A, Ibarrola G, Alcantar F, Aguilar A, Inostroza M, Maldonado S. Mexico city seismic alert system. Seismol Res Lett 1995;66(6):42–53.
[14] Hsiao NC, Wu YM, Shin TC, Teng TL. Development of earthquake early warning system in Taiwan. Geophys Res Lett 2009;36(5). [L00B02].
[15] Zollo A, Iannaccone G, Lancieri M, Cantore L, Convertito V, Emolo A, Festa G, Gallovič F, Vassallo M, Martino C, Satriano C, Gasparini P. Earthquake early warning system in southern Italy: methodologies and performance evaluation. Geophys Res Lett 2009;36(5). [L00B07].
[16] Wang JP, Huang D, Chang SC, Brant L. On-site earthquake early warning with multiple regression analysis: featuring two user-friendly applications for Excel. Comput Geosci 2013;58:1–7.
[17] Wang JP, Wu YM, Lin TL, Brant L. The uncertainties of a Pd3-PGV onsite earthquake early warning system. Soil Dyn Earthq Eng 2012;36:32–7.
[18] Wu YM, Kanamori H. Development of an earthquake early warning system using real-time strong motion signals. Sensors 2008;8(1):1–9.
[19] Xu Y, Wang JP, Wu YM, Kuo-Chen H. Reliability assessment on earthquake early warning: a case study from Taiwan. Soil Dyn Earthq Eng 2017;92:397–407.
[20] Iervolino I, Giorgio M, Galasso C, Manfredi G. Uncertainty in early warning predictions of engineering ground motion parameters: what really matters? Geophys Res Lett 2009;36(5). [L00B06].
[21] Böse M, Hauksson E, Solanki K, Kanamori H, Wu YM, Heaton TH. A new trigger criterion for improved real-time performance of onsite earthquake early warning in Southern California. Bull Seismol Soc Am 2009;99(2A):897–905.
[22] Iervolino I, Giorgio M, Manfredi G. Expected loss-based alarm threshold set for earthquake early warning systems. Earthq Eng Struct Dyn 2007;36(9):1151–68.
[23] Iervolino I. Performance-based earthquake early warning. Soil Dyn Earthq Eng 2011;31(2):209–22.
[24] Wang JP, Wu YM. Epistemic uncertainty in on-site earthquake early warning on the use of PGV-PD3 empirical models. Soil Dyn Earthq Eng 2014;65:126–30.
[25] Nelsen RB. An introduction to copulas. 2nd ed. New York: Springer; 2006.

[26] Li DQ, Tang XS. Modeling and simulation of bivariate distribution of shear strength parameters using copulas. In: Phoon Kok-Kwang, Ching Jianye, editors. Risk and Reliability in Geotechnical Engineering. Boca Raton: CRC Press; 2014. p. 77–128.
[27] Tang XS, Li DQ, Rong G, Phoon KK, Zhou CB. Impact of copula selection on geotechnical reliability under incomplete probability information. Comput Geotech 2013;49:264–78.
[28] Tang XS, Li DQ, Zhou CB, Phoon KK. Copula-based approaches for evaluating slope reliability under incomplete probability information. Struct Saf 2015;52:90–9.
[29] Tang XS, Li DQ, Cao ZJ, Phoon KK. Impact of sample size on geotechnical probabilistic model identification. Comput Geotech 2017;87:229–40.
[30] Zhang L, Li DQ, Tang XS, Cao ZJ, Phoon KK. Bayesian model comparison and characterization of bivariate distribution for shear strength parameters of soil. Comput Geotech 2018;95:110–8.
[31] Ozga-Zielinski B, Ciupak M, Adamowski J, Khalil B, Malard J. Snow-melt flood frequency analysis by means of copula based 2D probability distributions for the Narew River in Poland. J Hydrol Reg Stud 2016;6:26–51.
[32] Zou Y, Zhang YL. A copula-based approach to accommodate the dependence among microscopic traffic variables. Transp Res Part C Emerg Technol 2016;70:53–68.
[33] Kleinow J, Moreira F. Systemic risk among European banks: a copula approach. J Int Financ Mark Inst Money 2016;42:27–42.
[34] Goda K. Statistical modeling of joint probability distribution using copula: application to peak and permanent displacement seismic demands. Struct Saf 2010;32(2):112–23.
[35] 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.
[36] Goda K, Ren JD. Assessment of seismic loss dependence using copula. Risk Anal 2010;30(7):1076–91.
[37] Goda K, Atkinson GM. Interperiod dependence of ground-motion prediction equations: a copula perspective. Bull Seismol Soc Am 2009;99(2A):922–7.
[38] Xu Y, Tang XS, Wang JP, Kuo-Chen H. Copula-based joint probability function for PGA and CAV: a case study from Taiwan. Earthq Eng Struct Dyn 2016;45(13):2123–36.
[39] Nikoloulopoulos AK, Karlis D. Fitting copulas to bivariate earthquake data: the seismic gap hypothesis revisited. Environmetrics 2008;19(3):251–69.
[40] Grigoriu M. Do seismic intensity measures (IMs) measure up? Probabilistic Eng Mech 2016;46:80–93.
[41] Zentner I. A general framework for the estimation of analytical fragility functions based on multivariate probability distributions. Struct Saf 2017;64:54–61.