Soil Dynamics and Earthquake Engineering 65 (2014) 126–130

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

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

Epistemic uncertainty in on-site earthquake early warning on the use of PGV–PD3 empirical models

J.P. Wang, Yih-Min Wu *

[a] Department of Civil & Environmental Engineering, the Hong Kong University of Science and Technology, Kowloon, Hong Kong, China
[b] Department of Geosciences, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 27 January 2013
Received in revised form
20 March 2014
Accepted 7 June 2014

Keywords:
Earthquake early warning
Epistemic uncertainty
PGV–PD3 empirical model

ABSTRACT

From the literature, we found that PGV–PD3 regressions for on-site earthquake early warning (EEW) can be quite different depending on the presumption whether or not PGV–PD3 data from different regions should be "mixable" in regression analyses. As a result, this becomes a source of epistemic uncertainty in the selection of a PGV–PD3 empirical relationship for on-site EEW. This study is aimed at examining the influence of this epistemic uncertainty on EEW decision-making, and demonstrating it with an example on the use of PGV–PD3 models developed with data from Taiwan, Japan, and Southern California. The analysis shows that using the "global" PGV–PD3 relationship for Southern California would accompany a more conservative EEW decision-making (i.e., early warning is activated more frequently) than using the local empirical model developed with the PGV–PD3 data from Southern California only. However, the influence of this epistemic uncertainty on EEW is not that obvious for the cases of Taiwan and Japan.

1. Introduction

Earthquake prediction has been proved a controversial subject given those recent catastrophic earthquakes unpredicted [1]. Under the circumstances, alternatives such as seismic hazard analysis and earthquake early warning are accepted as a more practical approach for earthquake risk mitigation [1–8]. In short, seismic hazard analysis is to best estimate an earthquake ground motion at the site with earthquake data such as seismicity and fault locations; by contrast, earthquake early warning is a real-time approach sending out warning messages and taking timely responses before the arrival of destructive motions.

Uncertainty plays a key role in an analysis and decision-making. Nowadays, two types of uncertainty are generally categorized [9]. The first is the aleatory uncertainty resulting from natural randomness, such as random earthquake sizes and locations. In contrast, the other type is the epistemic uncertainty owing to our imperfect knowledge, and a common instance about epistemic uncertainty in earthquake studies is in the selections of suitable ground motion models for a specific application given so many empirical models are available [10–13].

Some assessments have considered both uncertainties during an earthquake analysis, and one of the examples is Probabilistic Seismic Hazard Analysis (PSHA) [14,15]. In a PSHA study, the aleatory uncertainty in earthquake sizes and locations is considered and governed by the best-estimate probability density functions developed. On the other hand, the epistemic uncertainty such as selecting ground motion models is taken into account by using a logic-tree analysis to integrate multiple models in the analysis.

On the use of an empirical relationship, on-site earthquake early warning (EEW) is to utilize initial motions that have been detected to predict the peak motion arriving a few seconds later at the site. Understandably, when the expected peak motion is relatively large, earthquake early warning will be activated for the site. However, since the empirical model is developed with earthquake data subject to some natural randomness, inevitably the empirical model is associated with some error (i.e., aleatory uncertainty). In a recent study, this aleatory uncertainty in on-site EEW was specifically discussed, as well as its influence on EEW decision-making. Accordingly, the study suggested a probability-based decision-making framework for on-site EEW, considering the error of the empirical model owing to the aleatory uncertainty in earthquake data [16].

By contrast, this paper is aimed at discussing the epistemic uncertainty in on-site earthquake early warning. Like the previous study discussing the aleatory uncertainty, we also used the PD3–PGV empirical model (model details were given in the following section) as an example to discuss a source of epistemic uncertainty in on-site EEW (i.e., whether or not PGV–PD3 earthquake data are "mixable" in an analysis), as well as its influence on EEW decision-making for three regions (i.e., Southern California, Japan, Taiwan). The analysis shows that when earthquake data is considered "mixable", the empirical model developed would lead to a more

* Corresponding author.
E-mail address: drymwu@ntu.edu.tw (Y.-M. Wu).

http://dx.doi.org/10.1016/j.soildyn.2014.06.003
0267-7261/

"conservative" decision-making scheme for on-site EEW systems in Southern California, than using the model developed with local earthquake data only, considering PGV–PD3 data from different regions should not be mixable in developing an EEW empirical relationship for a specific region.

2. Overview of the PGV–PD3 empirical model

As mentioned previously, on-site earthquake early warning is on the basis of using an empirical model to predict the magnitude of peak motions arriving a few seconds later given early motions that have been detected. Based on data on 780 earthquakes from Taiwan, Japan, and Southern California, Wu and Kanamori (2008) proposed an empirical relationship between first-three second ground displacement (PD3) and peak ground velocity (PGV) for on-site earthquake early warning. In their study, although the PGV–PD3 samples of Southern California are quite different from those in Taiwan and Japan in terms of their magnitudes, the data from the three regions were considered "mixable." Therefore, as shown in Fig. 1, an empirical PD3–PGV model based on the mixed data from the three regions was then developed as Eq. 1, and this "global" model would be employed for the development of an on-site EEW for any region in the world, including Southern California:

$$\log PGV = 1.52 + 0.81 \times \log PD3 \pm 0.32 \quad (1)$$

where PGV and PD3 are in cm/s and cm, respectively; the term ± 0.32 is the model error or the standard deviation of the error term ε in a regression model.

3. An epistemic uncertainty in developing the PGV–PD3 relationship for Southern California

However, from Fig. 2 we can see more clearly that the PGV–PD3 samples in Southern California would present a different group of data from those in Taiwan and Japan. As a result, for developing an on-site EEW for the regions around Southern California, one could consider the data should not be "mixable" given the difference, and suggest that an on-site EEW for Southern California should be present on the basis of using a PGV–PD3 empirical model developed with those local data only, like the following equation:

$$\log PGV = 1.02 + 0.59 \times \log PD3 \pm 0.34 \quad (2)$$

As a result, the two models (i.e., Eqs. 1 and 2) from different perspectives are causing an epistemic uncertainty as for the selection of a PD3–PGV relationship for developing on-site EEW

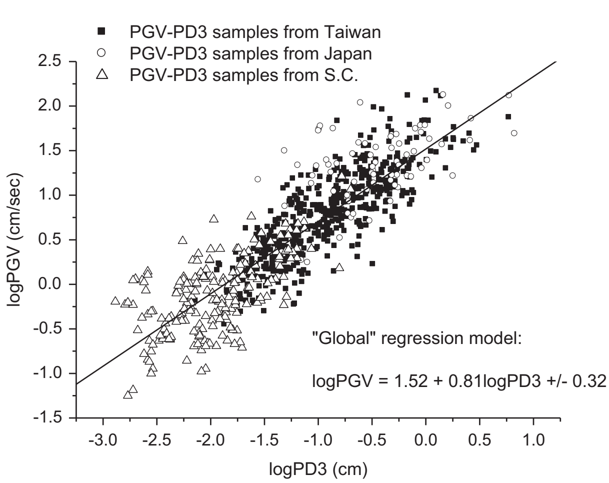

Fig. 1. The PGV–PD3 data from three regions, and the "global" regression model developed with the mixed data.

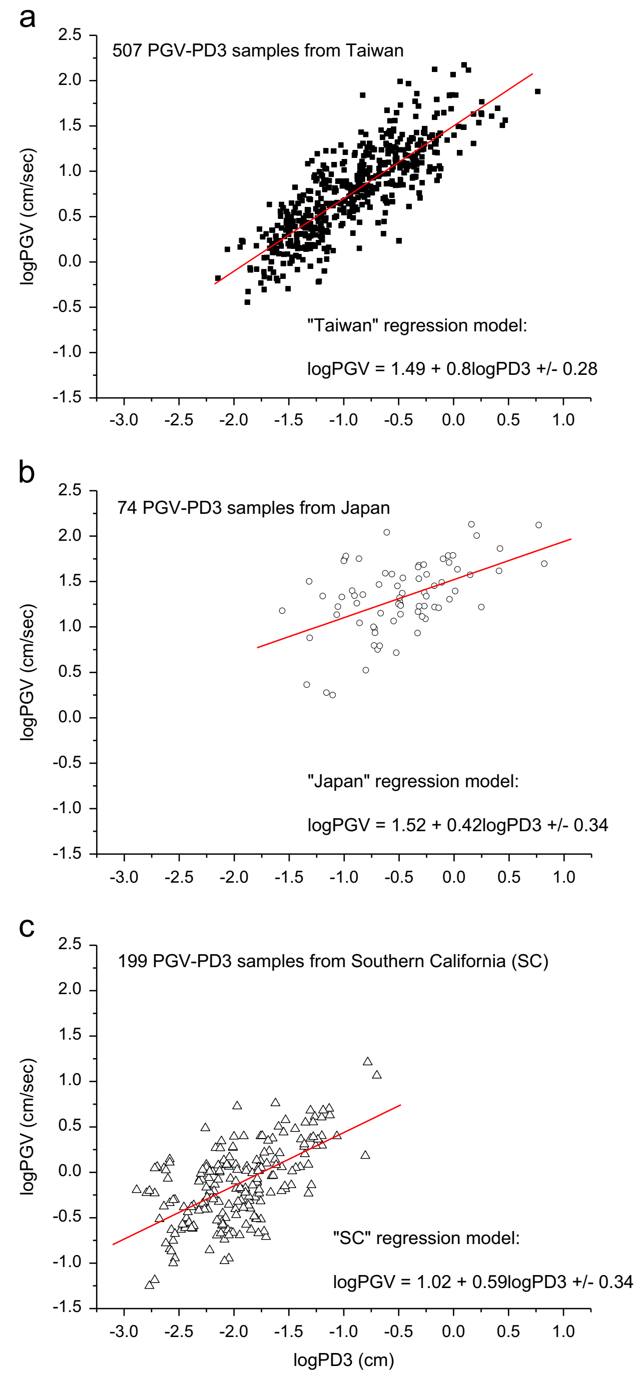

Fig. 2. The PGV–PD3 data from the three regions, and the three respective "local" regression models.

for a target area (Southern California in this example), given our current understandings could not verify whether or not PD3–PGV datasets from different regions should be "mixable" in an analysis such as regression. Nevertheless, this paper is not aimed at verifying this question. Instead, the following sections would present and discuss the influence of this epistemic uncertainty on on-site EEW, which is the scope and novelty of this paper.

4. The probability distribution of *Y* given a new observation *x** on the use of a regression model

On the use of a regression model $Y = f(x) \pm \varepsilon$ developed with n samples, the forecast of Y given a new observation x^* is a random

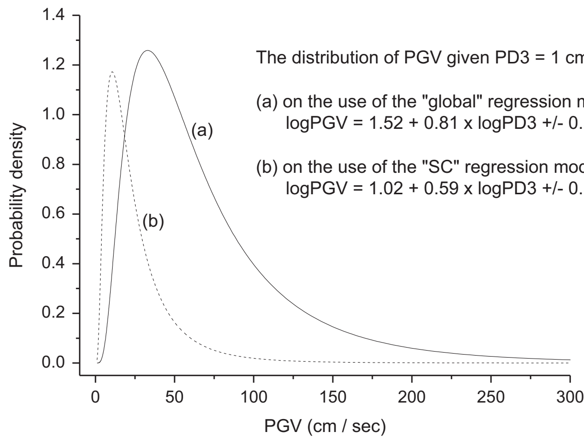

Fig. 3. The probability distributions of PGV given PD3 = 1 cm on the use of the "global" PGV–PD3 relationship and the "SC" model developed with the data from Southern California.

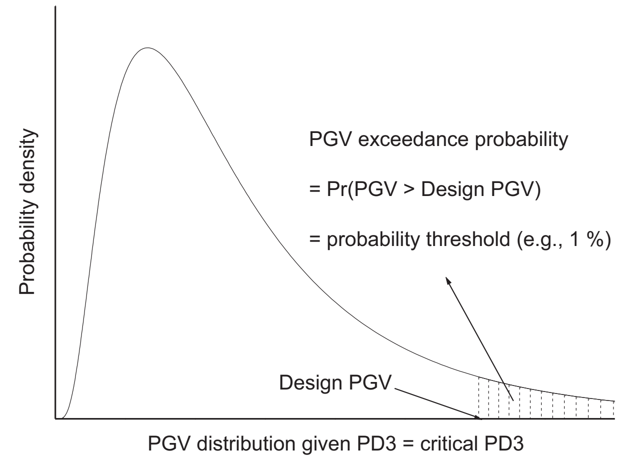

Fig. 4. A schematic diagram illustrating the probability-based decision-making for activating on-site earthquake early warning.

variable following the *t*-distribution governed by sample size *n*, model error ε, etc. (the complete algorithm is given in the Appendix A). As a result, for this EEW study the two probability distributions of PGV given PD3 = 1 cm are shown in Fig. 3, on the use of the "global" regression model developed with the data from three regions (i.e., Eq. (1)), and on the use of the "SC" model (i.e., Eq. (2)) developed with the data in Southern California only.

From this example, we can see that the distributions of PGV given PD3 = 1 cm are quite different given the "global" model or the "SC" model was used. As a result, selecting a PGV–PD3 relationship among the two models for developing on-site EEW for Southern California is an epistemic uncertainty, and its influence on the PGV forecast could be substantial like the example shown in Fig. 3.

5. Probability-based decision-making framework for on-site EEW

To incorporate the model error of a PD3–PGV relationship into on-site EEW decision-making, the study discussing the aleatory uncertainty in on-site EEW suggested a probability-based decision-making framework, on the basis of a small exceedance probability (say 1%) that the design PGV would be exceeded by the actual PGV arriving later [16]. In other words, if the PGV exceedance probability given a PD3 value that has been detected is greater than the probability threshold, earthquake early warning is needed; otherwise it is not needed. Furthermore, the so-called critical PD3 value can be back calculated, and when the PD3 detected is greater than the critical PD3, earthquake early warning is needed; otherwise it is not needed.

Fig. 4 is a schematic diagram illustrating such a probability-based decision-making framework for on-site EEW. At PD3 equal to the critical PD3 value, the probability distribution of PGV shown in Fig. 4 can be computed, and on this condition the PGV exceedance probability is equal to a small probability threshold prescribed. Understandably, when a PD3 value detected is greater than the critical PD3, the PGV distribution will shift to the right, and the PGV exceedance probability will exceed the threshold prescribed, and earthquake early warning is then activated.

6. The critical PD3 values for Southern California

Given PGV exceedance probability threshold as 1% and a structure's design PGVs equal to 100, 150, 200, and 250 cm/s, this section presents the calculations of critical PD3 values on the use of the "global" and "SC" empirical relationships. As shown in Fig. 5, the critical PD3 is 0.48 cm given design PGV = 100 cm/s on the use of the "global" model (i.e., logPGV = 1.52 + 0.81 × logPD3 ± 0.32), which is lower than the critical PD3 of 2.06 cm calculated with the "SC" model (logPGV = 1.02 + 0.59 × logPD3 ± 0.34). As for the other three PGV design values, the critical PD3 values on the basis of the "global" model are also lower than those using the "SC" model. As a result, on this condition (i.e., exceedance probability = 1%) the on-site earthquake early warning should be activated more frequently when the "global" model is adopted, given a lower critical PD3 value that is attained. In other words, using the "global" PD3–PGV relationship for developing on-site EEW for Southern California is considered more "conservative" than using the empirical relationship developed with local earthquake data only.

Besides, we found that at a given probability threshold = 1%, the differences in the critical PD3 could be quite substantial especially when the design PGV is large. For example, given a structure's design PGV = 250 cm/s, the on-site EEW using the "global" PGV–PD3 model will be activated as long as the PD3 detected is higher than the critical value of 1.5 cm, but the EEW using the "SC" model will be activated when PD3 is greater than 9.8 cm, a much higher threshold than 1.5 cm in the other option.

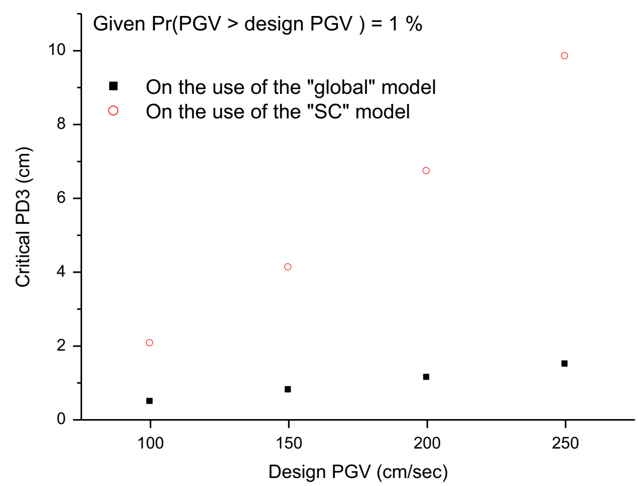

Fig. 5. Critical PD3 calculations given PGV exceedance probability as 1% and the four design PGVs, on the use of the "global" and "SC" PGV–PD3 relationships.

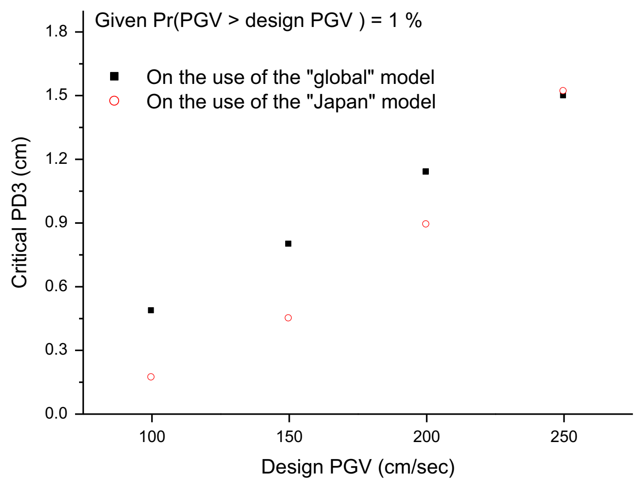

Fig. 6. Critical PD3 calculations given PGV exceedance probability as 1% and the four design PGVs, on the use of the "global" and "Japan" PGV–PD3 relationships.

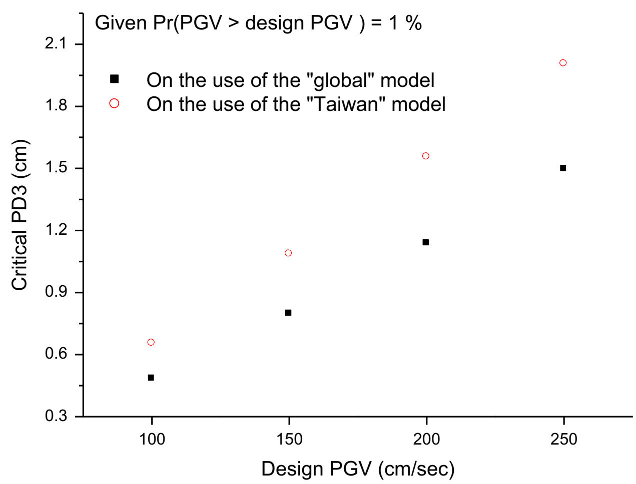

Fig. 7. Critical PD3 calculations given PGV exceedance probability as 1% and the four design PGVs, on the use of the "global" and "Taiwan" PGV–PD3 relationships.

7. The critical PD3 values for Japan and Taiwan

Repeating the calculations, Fig. 6 shows the critical PD3 on the use of a local PGV–PD3 empirical model (i.e., $\log PGV = 1.52 + 0.41 \times \log PD3 \pm 0.34$) developed with the data from Japan shown in Fig. 2b. With the empirical relationship, the critical PD3 was found equal to 0.17 cm given design PGV = 100 cm/s and exceedance probability threshold = 1%, which is lower than the critical PD3 value of 0.47 cm calculated with the "global" model. As a result, for developing an on-site EEW for Japan, using the "Japan" PGV–PD3 relationship is considered more conservative than using the "global" model, based on the critical PD3 values of the two options shown in Fig. 6.

Similarly, Fig. 7 shows the critical PD3 on the use of the PGV–PD3 empirical model (i.e., $\log PGV = 1.52 + 0.41 \times \log PD3 \pm 0.34$) developed with the data from Taiwan shown in Fig. 2a. Similar to the "California" case shown in Fig. 5, using the "global" model for Taiwan is more conservative (i.e., EEW is activated more frequently) than using the "Taiwan" PGV–PD3 relationship developed with local data only. However, it is worth noting that the differences in critical PD3 calculations between the two options are not significant in this case. In other words, the influence of the epistemic uncertainty in the selections of the global or local PGV–PD3 models on on-site EEW in Taiwan is not as substantial as that for Southern California.

8. Summary and conclusions

In the developing of a PGV–PD3 empirical model for on-site earthquake early warning, whether or not PGV–PD3 data from different regions could be mixable in an analysis is a source of epistemic uncertainty, not to mention from the PGV–PD3 datasets from Southern California are quite different from those in Taiwan and Japan. As a result, for developing on-site EEW for a target region, using a local PGV–PD3 model or a global relationship are equally convincing, depending on the presumption whether PGV–PD3 data from different regions are mixable or not, a source of epistemic uncertainty in on-site earthquake early warning.

On the basis of the PGV–PD3 data from Taiwan, Japan, and Southern California in the literature, we found that the influence of the epistemic uncertainty on EEW activations in Southern California is more significant than in Japan and Taiwan. To be more specific, using the "global" PGV–PD3 model developed with mixed data from the three regions would accompany a lower critical PD3 than using the local empirical relationship developed with the data in Southern California only. In other words, the on-site EEW will be activated more frequently given the "global" model is used, which is considered a more conservative decision-making scheme, in a comparison to the use of the PGV–PD3 relationship developed with the local data from Southern California.

On the contrary, the analysis shows that for the regions around Japan, using the local PGV–PD3 relationship will accompany a more conservative decision-making for EEW activations, in a comparison to the use of the "global" model. However, the differences in critical PD3 values in this case are less significant than that in Southern California. In other words, the influence of the epistemic uncertainty in choosing the local or global model on earthquake early warning in Japan is less obvious than its influence on the case of Southern California. Similarly, the influence of this epistemic uncertainty on on-site EEW in Taiwan is not that obvious either, given a small difference in critical PD3 values calculated with the local or "global" PGV–PD3 relationship.

Acknowledgments

We appreciate the valuable comments from the anonymous reviewers for making our work much improved on this submission.

Appendix A. The distribution of forecast *Y* given a new observation *x**

Given n samples (x_1, y_1) to (x_n, y_n), a simple linear regression model between X and Y can be developed as $Y = \beta_0 + \beta_1 x^* \pm \varepsilon$, where ε is the model standard deviation. According to probability and statistics, the variance of Y given a new observation x^* is as follows [17]:

$$V[Y_{x^*}] = \varepsilon^2 \left[1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}} \right] \quad (A.1)$$

where $\bar{x}$ is the average of $x_1 \ldots x_n$, and S_{xx} is equal to

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (A.2)$$

Besides, a standardized variable T can be expressed as follows:

$$T = \frac{Y_{x^*} - (\beta_0 + \beta_1 x^*)}{\varepsilon \sqrt{1 + (1/n) + ((x^* - \bar{x})^2 / S_{xx})}} \quad (A.3)$$

where T follows a t-distribution with degree of freedom of $n-2$. As a result, the distribution of Y_{x*} (the forecast Y given a new observation x^*) and the exceedance probability $\Pr(Y_{x^*} > y^*)$ can be computed accordingly.

References

[1] Geller RJ, Jackson DD, Kagan YY, Mulargia F. Earthquake cannot be predicted. Science 1997;275:1616.
[2] Kanamori H, Hauksson E, Heaton T. Real-time seismology and earthquake hazard mitigation. Nature 1997;390:461–4.
[3] Allen RM, Kanamori H. The potential for earthquake early warning in Southern California. Science 2003;300:786–9.
[4] Teng TL, Wu YM, Shin TC, Tsai YB, Lee WHK. One minute after: strong-motion map, effective epicenter, and effective magnitude. Bull Seismol Soc Am 1997;87:1209–19.
[5] Wu YM, Kanamori H. Development of an earthquake early warning system using real-time strong motion signals. Sensors 2008;8:1–9.
[6] Wang JP, Huang D, Yang Z. The deterministic seismic hazard map for Taiwan developed using an in-house Excel-based program. Comput Geosci 2012;48:111–6.
[7] Wang JP, Yun X, Wu YM. A FOSM calculation for seismic hazard assessment with the consideration of uncertain size conversion. Nat Hazards Earth Syst Sci 2013;13:2649–57.
[8] Wu YM, Shin TC, Chang CH. Near real-time mapping of peak ground acceleration and peak ground velocity following a strong earthquake. Bull Seismol Soc Am 2001;91:1218–28.
[9] Ang A, Tang W. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering. New Jersey: John Wiley & Sons; 2007; 1–2.
[10] Cheng CT, Chiou SJ, Lee CT, Tsai YB. Study on probabilistic seismic hazard maps of Taiwan after Chi-Chi earthquake. J GeoEng 2007;2:19–28.
[11] Wang JP, Huang D, Cheng CT, Shao KS, Wu YC, Chang CW. Seismic hazard analysis for Taipei City including deaggregation, design spectra, and time history with Excel applications. Comput Geosci 2013;52:146–54.
[12] Sabetta F, Lucantoni A, Bungum H, Bommer J. Sensitivity of PSHA results to ground motion prediction relations and logic-tree weights. Soil Dyn Earthq Eng 2005;25:317–29.
[13] Barani S, Spallarossa D, Bazzurro P, Eva C. Sensitivity analysis of seismic hazard for Western Liguria (North Western Italy): a first attempt towards the understanding and quantification of hazard uncertainty. Tectonophysics 2007;435:13–35.
[14] Cornell CA. Engineering seismic risk analysis. Bull Seismol Soc Am 1968;58:1583–606.
[15] McGuire RK. Probabilistic seismic hazard analysis and design earthquakes: closing the loop. Bull Seismol Soc Am 1995;85:1275–84.
[16] J.P. Wang, Y.M. Wu, T.L. Lin, L. Brant, The uncertainty of a Pd3–PGV onsite earthquake early warning system, Soil Dyn Earthq Eng 36 (2012) 32-37.
[17] Devore JL. Probability and Statistics for Engineering and the Sciences. Pacific Grove, CA: Duxbury Press; 2008.