MODELLING SEISMIC RECORD AND SOIL TEST RESULT BY NEURAL COMPUTING WITH GENETIC ALGORITHM

Tienfuan KERH
Professor, Department of Civil Engineering
National Pingtung University of Science and Technology Pingtung 91201, Taiwan
tfkerh@mail.npust.edu.tw

Yu-Hsiang SU
Graduate student, Department of Civil Engineering
National Pingtung University of Science and Technology Pingtung 91201, Taiwan
appleatein@gmail.com

Ayman MOSALLAM
Professor, Department of Civil and Environmental Engineering University of California, Irvine, California 92697, USA
mosallam@uci.edu

Keywords: Neural Network, Genetic Algorithm, Peak Ground Acceleration, Unchecked Site, Prediction

ABSTRACT

Earthquakes are natural hazards which occur quite often worldwide every year, and the recorded data sets can be used to analyse the characteristics of seismic response in a specified region. In this study, a genetic algorithm based neural network model is developed to improve the reliability of predicting peak ground acceleration, the key element to evaluate earthquake response and to setup seismic design standard. Three seismic parameters including local magnitude, epicentre distance, and epicenter depth, are taken in the input layer for developing the fundamental estimation model. Then, two geological conditions including standard penetration test value and shear wave velocity, are added for developing a new model to reflect the site response more adequately. Based on the earthquake records and soil test data from 86 checking stations within 24 seismic subdivision zones in Taiwan area, the computational results show that the combination of using neural network and genetic algorithm can achieve a better performance than that of using neural network model solely. This preferred model can be extended to predict peak ground acceleration at unchecked sites, and can be applied to check the design standard in building code. This study may provide a new approach to solve this type of earthquake related nonlinear problem.

1. INTRODUCTION

For an event of severe strong ground motion, it may cause a large scale of structural damages and result in tremendous casualties and property losses directly and indirectly. Accumulated results have shown that earthquakes accounting for nearly 60% of all disaster-related mortality in the past decade (Bartels and VanRooyen, 2012). To reduce various negative impacts from this natural disaster, a wide range of relating research topics, such as earthquake mechanism and potency investigation, prediction and warning system development, instrumental measurement and data analysis, have been extensively reported (Bailey et al., 2009; Rhoades and Evision, 2004; Zobin et al., 2014) in the field of applied geophysics as well as in the community of earthquake engineering.

Regarding seismic data analysis, the sources are in general come from checking stations in certain regions. Except of earthquake occurring time and location, the historical record data may also include some of important seismic parameters such as local magnitude (Mg), epicenter distance (Di), epicenter depth (De), and peak ground acceleration (PGA) in vertical (V), north-south (NS), and east-west (EW) directions, respectively. These data can be used to describe the characteristics of strong ground motion in a specified area, and can be taken as a basis for setting up anti-earthquake design standard in building code. Therefore, there also exist many papers to deal with this important practical problem. For instance, the use of conventionally statistical method with linear and nonlinear models applied in estimating PGA (Yuen and Mu, 2011). The recently developed method of using artificial neural network (NN) to model the above mentioned seismic parameters can also be found frequently (García et al., 2007; Kerh et al., 2008). These reports did provide useful information for the cases studied and relevant engineering applications.

In neural network approach, the commonly supervised learning paradigm and the steepest gradient descent based search method in neural computing may converge to a local minimum of the error function if the initial connection weights are randomly selected. To overcome this problem, the connection weights could be selected using genetic algorithm (GA) that increases the probability of finding the global minimum of the error function (Vonk et al., 1997). Therefore, this study focuses on developing a genetic algorithm based neural network model (NN+GA) to improve the reliability of predicting model.

The application of using this combination model in predicting PGA is very limited, and the research only focused on investigating seismic basic parameters (Kerh et al., 2010). It may still lack some of crucial factors such as geological difference to reflect the truly site effect. Therefore, in addition to seismic parameters (Mg, Di, De), this study takes two seismic related soil test results, that is, shear wave velocity (Vs) and standard penetration test value (SPT-N) to the input layer of the combination model. The new prediction model developed in the present study is expected to have a better reliability and represent site response more adequately.

2. INFORMATION IN RESEARCH AREA

The island of Taiwan is located within the “ring of fire”, the primary trigger force to create earthquake is come from the intrusion of Eurasia sea plate and Philippine sea plate. There are 33 major active faults distributed in the whole island (CGS, 2013), so strong ground motions are frequently occurred in this region. Figure 1 (left) shows a yearly earthquake records in recent (11/2011-10/2012, Chen and Chang, 2012), it can be seen that thousands of strong ground motions occurred in the neighborhood of Taiwan island, where it no lacks of destructive earthquakes, where the most significant one is the so-called 921 earthquake (Mg=7.3, De=8km) occurred at the central part of Taiwan in 1999, resulted in tremendous casualties and structural damages, which equivalent to about 10 billion USD property losses.

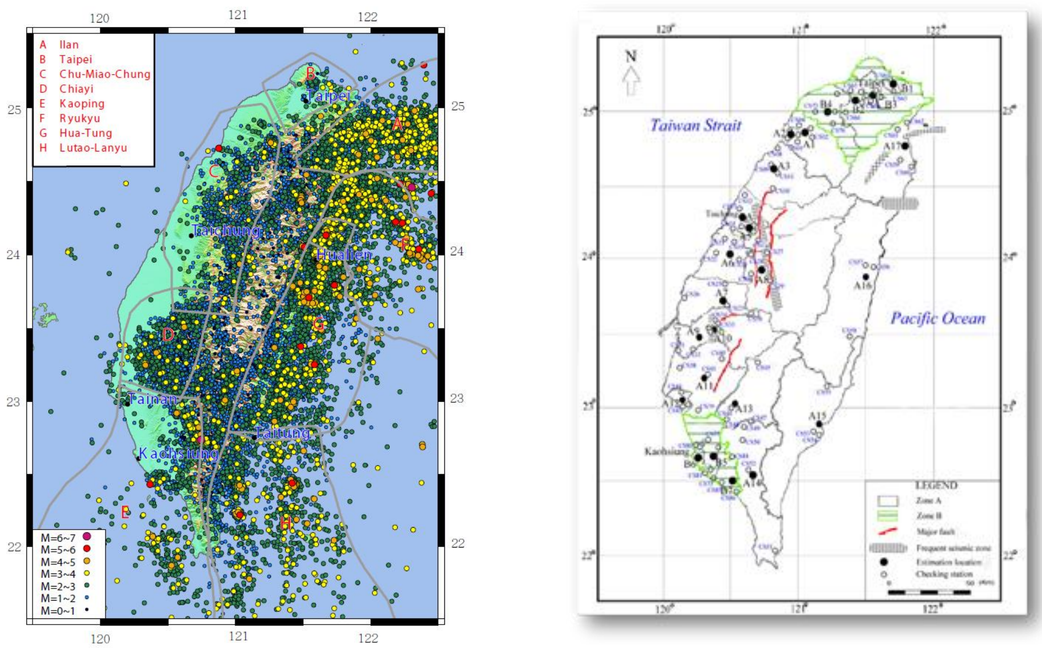

Figure 1. Distribution of earthquakes (left) and sketch of research area (right).

After several revisions and to respond to the actual seismic conditions, the currently building code in Taiwan divided the entire island into two division zones (Zone A and Zone B), and there are 24 subdivision zones based on boundary of cities and counties for more precisely. The design earthquake area coefficients of horizontal acceleration for Zone A and Zone B, are 0.33g and 0.23g, respectively (CPA, 2006). Note that 1g = 981gal (cm/s^2), it can be used for calculating earthquake force in structural design.

From the sketch of the present study area in Figure 1 (right), it can be seen that the white color region represents Zone A, which has a total of 17 subdivision zones (A1-A17). The strip line with green color area represents Zone B, which has a total of 7 subdivision zones (B1-B7). Additionally, the 86 white dot signs denote checking stations to provide historical seismic records from the year of 1994 to the year of 2013. The dark dot signs are the location of unmeasured sites to stand for the 24 subdivision zones, and seismic data sets from 2 to 4 checking stations are taken for analysis in these subdivision zones.

Basically, the geological conditions in the island of Taiwan may consist of six major regions, The soil condition in the western side of Taiwan is in general softer than the other regions because of its geologically loose structure. Hence, ground motion in this region may be more sensitive to site effects and should be considered more carefully in engineering design. The difference of geological conditions may cause to generate different degree of amplification effect during an earthquake, and so the acceleration response spectrum may vary under different geological environments.

The site effect is an important factor to revise nonlinear acceleration response spectrum, which is used for anti-earthquake design in currently building code of Taiwan (Wun et al., 2004). In general, the averaged soil layer at 30m underground of the construction site is used for determining the characteristics of soil condition. The definitions of the standard penetration test value (SPT-N) and the shear wave velocity (V_s) may be written as:

$$\text{SPT-N} = \sum_{i=1}^{n} t_i \Big/ \sum_{i=1}^{n} (t_i / N_i) \qquad (1)$$

$$V_s = \sum_{i=1}^{n} t_i \Big/ \sum_{i=1}^{n} (t_i / V_{si}) \qquad (2)$$

where t_i is the thickness (*m*) of soil layer; V_{si} is the averaged shear wave velocity (*m/s*) at the *i* soil layer; and N_i denotes standard penetration test value for each soil layer.

From the principle of soil mechanics and seismic wave theory (Wikipedia, 2014), it is understand that the SPT-N value may provide an indication of the relative density of granular deposits, and may be used to reflect the resistant of liquefaction due to earthquake. For a seismic body wave, the P-wave (sometimes referred to as the pressure wave) propagates very quickly and only lasts for a short time. Thus, it causes relatively insignificant structural damage. Whereas, the S-wave (or secondary wave), propagates more slowly than the P-wave, and it may cause greater structural damage. Therefore in this study, both soil parameters of SPT-N value and S-wave velocity are considered to represent the site effect during strong ground motions.

3. COMBINING NEURAL NETWORK MODEL WITH GENETIC ALGORITHM

In the field of artificial intelligence, artificial neural network has been well developed in recent years. There exist several types of neural networks but in general can be categorized into two types: supervised learning networks, and unsupervised learning networks (Chang and Chang, 2010). The frequently used back-propagation neural network is a multi-layered feed-forward network, which uses the supervised learning process to treat nonlinear mapping relationship between inputs and outputs. This neural network model has various engineering applications due to its simplicity and effectiveness. The detailed principle, operational logic routine, and transfer function of this multi-layered neural network can be found in many of the related literatures (Shanga, 2005; Mandic and Chambers, 2001). For simplification, the basic equations for the neural network model can be written as:

$$Y_j = F(\sum W_{ij} X_i - \theta_j) \tag{3}$$

where Y_j is the output of neuron *j*, W_{ij} represents the connection weight from neuron *i* to neuron *j*, X_i is the input signal generated for neuron *i*, θ_j is the bias term associated with neuron *j*, and $F(x)$ is the nonlinear activation function.

The performance of a neural network model can generally be evaluated by using the coefficient of correlation (R) and the root mean square error (RMSE), defined as follows, respectively:

$$\mathrm{R} = \sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\sum_{n}^{N}(T_n - Y_n)^2 / N} \tag{5}$$

where x_i and $\bar{x}$ are the recorded value and its average value, respectively, y_i and $\bar{y}$ are the estimated value and its average value, respectively, and m denotes the number of data points in the analysis. In addition, *N* is the number of learning cases, T_n is the target value for case *n*, and Y_n is the output value for case *n*.

This study employs the neural network toolbox in MATLAB to perform neural network calculations (Wu et al., 2003). The main steps for developing the model are: 1) generating a vector file for the seismic records and soil test data (Mg, Di, De, PGA, SPT-N, Vs) in Excel; 2) importing the data set into the neural network toolbox; 3) choosing the training function, learning function, performance function, and transfer function to create a new network; 4) training the network and adapting the weight and bias values for completing the model development; and 5) simulating to estimate PGA in the three directions for each of the seismic subdivision zones.

It is well-known that the typical genetic algorithm requires a genetic representation of the solution domain, and a fitness function to evaluate the performance of the solutions in the domain (Wikipedia, 2014). This particular class of evolutionary computation does not use the original parameters but applies operators to a coded representation of the parameters. The fundamental structure of genetic algorithm including five major operators: 1) encoding and decoding, 2) fitness function, 3) selection strategy, 4) crossover, and 5) mutation. With these computational operators, the use of genetic algorithm is not only restricted to searching the solution space, but also for obtaining the global optimum solution or near global optimum solutions.

For more details, the basic procedures for incorporating a genetic algorithm into neural network model are: 1) input weight and bias values associated with each layer from the neural network model; 2) use binary coding for weight and bias to be represented as a continuous chromosome; 3) develop a fitness function by using the root mean square error between estimations and records; 4) set up population parameters and evolution parameters; 5) start the search for obtaining a new set of weight and bias; 6) insert the new set values to the neural network model for getting a new estimation; 7) repeat the procedure until it converge to get the optimal solution.

In this study, the software package GeneHunter (Ward, 2004) is taken to perform the calculations required to obtain the best weight and bias set for the neural network model. At first, the relating parameters of neural network model such as inputs, targets, layers, neurons, weight, and bias are generated in Excel interface. Then, by loading software tool into Excel and set up various parameters, and start to search the best solution finally. With this approach, a more reliable PGA estimation can be expected and this will provide an improved basis for analysing the seismic problem studied herein.

4. MODEL PERFORMANCE AND COMPARISON

In the present neural network model, the hyperbolic-tangent is used as the transfer function, which values ranged between -1 to +1. To scale the input parameters and to prevent the effect of extreme values, the input data sets need to be normalized by using the following equation:

$$V_{new} = \frac{2(V_{old} - V_{\min})}{V_{\max} - V_{\min}} - 1 \tag{6}$$

where V_{new} is the value after normalization; V_{old} is the original data; $V_{\max}$ and $V_{\min}$ are the original maximum and minimum data, respectively (Yeh, 2004). For the totally random data sets, three parts of data, i.e. 70%, 20%, and 10%, are used for train, adapt, and simulate, respectively in the neural network model calculation stages. Note that there is no particular rule to arrange the data set, the performance of each calculation stage can be evaluated from the indices of correlation coefficient and root mean square error as defined previously.

The use of three seismic parameters (Mg, Di, De) in the input layer of neural network model has been proved to have an acceptable performance for PGA predictions. Whereas, the addition of using the two geological parameters (SPT-N, Vs) in the neural network model may require a further check to verify the sensitivity of these parameters. As shown in Table 1 is the comparison of performance including correlation coefficient and root mean square error between NN model and NN+GA model with different input parameters. It can be observed that the NN+GA model do perform better than that of NN model. Also, the inputs to include both seismic and geological parameters do have a better performance than that of using seismic parameter solely. These results prove the use of NN+GA model is more reliable, and the geological conditions do have a positive influence on the PGA prediction. Therefore, this preferred model will be used for predicting PGA at unmeasured sited to be discussed in the following subsection.

Table 1. Comparison and performance of NN model and NN+GA model.

Model		NN		NN+GA	
Input parameter / Performance		trained	simulate	trained	simulate
Seismic (Mg, Di, De)	R^2	0.65614	0.46806	0.74626	0.53744
	RMSE	0.26482	0.35815	0.18427	0.18659
Seismic + Geological (Mg, Di, De, SPT-N, V_s)	R^2	0.64505	0.47678	0.83365	0.55943
	RMSE	0.29995	0.41748	0.16640	0.20245

For the 24 seismic subdivision zones in Taiwan area, the averaged coordinate of checking stations within each of the subdivision zone is taken to represent the estimation location, and the PGA at this unmeasured site will be evaluated by using the developed NN+GA model. Initially, both seismic records and soil test results for the total 86 checking stations studied herein are trained and developed by NN+GA model for each station individually. Then, by taking all estimation results with target values, the relationships are shown in Figure 2 for V, NS, EW directions, and for all data sets, respectively. It can be found that the correlation coefficient in the range of $0.7 \le |R| < 1$ for all cases, and it can reach up to $R^2 = 0.83708$ for all data sets. That is, the estimated PGA has a high relationship with seismic record. Meanwhile, the root mean square error is as small as 0.01747, which shows a sufficient accuracy of the prediction results by NN+GA model.

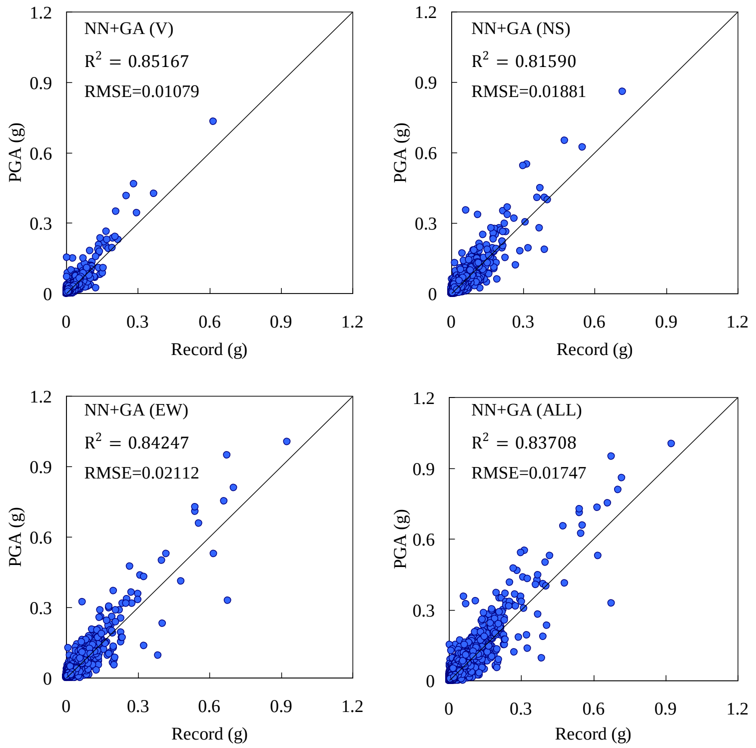

Figure 2. Relationship of PGA estimations with seismic records for all checking stations.

For the unmeasured site in each of the subdivision zones, it requires a model to calculate PGA at this site. An easy way to perform the task is by distributing the estimated PGA results from nearby checking stations, and summing up with a weighting factor to each checking station. Alternatively, a better way to estimate the PGA at an unchecked site as used in this study, is to take a new set of seismic data (same Mg and De, but new Di for each of seismic records) and a new set of geological conditions (weight-based soil test results of SPT-N and Vs) from known checking stations nearby. Then, insert the data set in a NN+GA model developed for each known checking station. By summing the results with weighting factors, the final estimation is obtained for the unmeasured site.

The ability of using NN+GA model to predict PGA is proved for all checking stations, and for the case of PGA prediction at an unmeasured site, the result from the weight-based model is comparing with available microtremor measurement at a specified location (subdivision B6) as shown in Figure 3. The bar chart reveals that the present NN+GA model do perform better than that of using NN model in previous studies (Kerh and Chu, 2002; Kerh and Ting, 2005), as the present prediction results exhibit to closer to PGA transformed from microtremor measurement. This comparison result may provide the reliability and confidence of using NN+GA model for predicting PGA at other unmeasured sites.

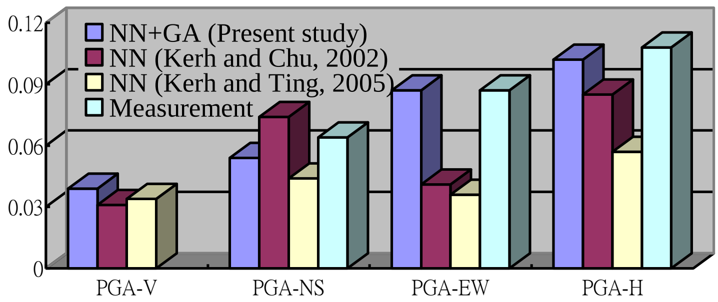

Figure 3. Comparison of PGA predictions with microtremor measurement (unit: g).

To check the potential hazardous locations in the research area from the present weight-based NN+GA model, the PGA prediction results for the 24 subdivision zones is shown in Figure 4. For the seismic zone A, it can be found that there are 6 subdivision zones (A4, A5, A7, A8, A9, and A10) exhibit to have a higher horizontal PGA than that of the design value (0.33g). However, this directly predicted result seems too conservative, a modified result by using square root of the sum of the square (Hong, 2004), for checking stations within each of the subdivision zones, are calculated and included in the plot. It can be seen that there are 3 subdivision zones (A7, A8, and A9) exhibit a higher horizontal PGA than the design value, and the tendency is also similar to previous research (Kerh et al., 2013) Therefore, this modified result is believed to have a more reliability for the case studied herein, and the identified potential hazardous subdivision zones should pay more precautious in relevant engineering applications. For the seismic zone B, as the PGAs obtained from both NN+GA and NN models are lower than that of the design value (0.23g), so all of prediction results comply with design standard in building code.

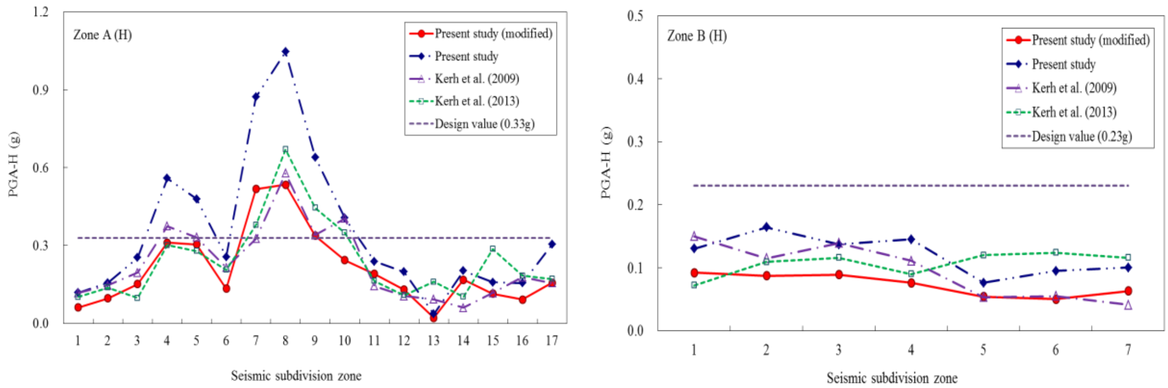

Figure 4. Comparison of PGA prediction result with design standard at 24 subdivision zones.

5. CONCLUSION

The recently developed neural network model can be applied to analyse seismic data sets for predicting an important seismic parameter PGA. Whereas, the drawback of this approach may lead converge to a local minimum. Therefore, this study attempts to incorporate genetic algorithm, which has a global search capability into neural network calculation process, for developing a more reliable prediction model. Except of using 3 major seismic parameters (Mg, Di, De) obtained from 86 checking stations in Taiwan region, this study also includes 2 seismic related geological conditions (SPT-N, V_s) in the training process to represent site response more adequately.

The comparison results based on the evaluation indices of correlation coefficient and root mean square error show that the NN+GA model has a better performance than that of the NN model solely. For the unmeasured sites at 24 seismic subdivision zones in the research area, the PGA predictions by a weight-based model exhibit that there exist 3 potential hazardous zones, as the horizontal PGA exceeds the design value as required in the building code.

Note that the geological conditions for each of the unmeasured sites are obtained from the nearby soil boring test result with a weighting factor, and that may have an influence on the accuracy of prediction model. However, this study combining genetic algorithm with neural network, and by inputting both seismic parameters and soil test data to develop a prediction model, may provide a new approach to solve this type of earthquake related nonlinear problem, and may be applied to other areas of interest around the world.

ACKNOWLEDGEMENTS - The financial support from Ministry of Science and Technology (#103-2918-I-020-001) is greatly appreciated. The authors also gratefully acknowledge the Central Weather Bureau Seismological Center and National Center for Research on Earthquake Engineering of Taiwan for providing historical seismic records and geological surveys, respectively.

REFERENCES

Bailey IW, Becker TW and en-Zion Y (2009) Patterns of co-seismic strain computed from southern California focal mechanisms, *Geophysical Journal International*, 177(3): 1015-1036

Bartels SA and VanRooyen MJ (2012) Medical complications associated with earthquakes, *Lancet*, 379:748-757

Central Geological Survey (2013) Taiwan faults, Ministry of Economic Affairs, http://www.moeacgs.gov.tw/ app/index.jsp

Chang FJ and Chang LC (2010) Introduction to artificial neural networks: principle and application, Tsanghai Publishing, Taichung, Taiwan

Chen TH and Chang WY (2012) Analysis of earthquake activities in Taiwan area (iv), Central Weather Bureau, MOTC-CWB-101-E-12, 62: 325-346

Construction and Planning Agency (2006) Building code, Ministry of the Interior, Taiwan, http://www.cpami.gov.tw

García SR, Romo MP and Mayoral JM (2007) Estimation of peak ground accelerations for Mexican subduction zone earthquakes using neural networks, *Geofísica Internacional*, 46(1): 51-63

Hong TG (2004) Dynamics of Structures, Tienyu Publishing Incorporation, Taipei, Taiwan, 8-12

Kerh T and Chu D (2002) Neural network approach and microtremor measurement in estimating peak ground acceleration due to strong motion, *Advance in Engineering Software*, 31: 367-375

Kerh T, Gunaratnam D and Chan Y (2010) Neural computing with genetic algorithm in evaluating potentially hazardous metropolitan area result from earthquake, *Neural Computing and Applications*, 19(4): 521-529

Kerh T, Lai JS, Gunaratnam D and Saunders R (2008) Evaluation of seismic design values in the Taiwan building code by using artificial neural network, *Computer Modeling in Engineering and Sciences*, 26(1): 1-12

Kerh T, Lin Y and Saunders R (2013) Seismic design value evaluations based on checking records and site geological condition using artificial neural networks, *Abstract and Applied Analysis*, http://dx.doi.org/10.1155/2013/242941, 1-12

Kerh T and Ting SB (2005) Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system, *Engineering Applications of Artificial Intelligence*, 18: 857-866

Mandic D and Chambers J (2001) recurrent neural networks for prediction: architectures, learning algorithms and stability, Wiley, New York, USA

Rhoades DA and Evison FF (2004) Long-range earthquake forecasting with every earthquake a precursor on its own scale, *Pure and Applied Geophysics*, 161: 47-72

Shanga Z (2005) Application of artificial intelligence CFD based on neural network in vapor-water two-phase flow, *Engineering Applications of Artificial Intelligence*, 18: 663-671

User’s Guide (2004) GeneHunter targets optimal solutions with genetic, Ward Systems Group Incorporation, Maryland, USA

Vonk E, Jain LC and Johnson RP (1997) Automatic generation of neural network architecture using evolutionary computation, World Scientific, Singapore

Wu S, Chou L, Lee S and Chang BZ (2003) MATLAB neural networks simulation and application, Science Publishing Company, China

Wikipedia (2014) Genetic algorithm, Wikipedia Foundation Incorporation, http:// en.wikipedia.org/wiki/ Genetic_algorithm

Wikipedia (2014) Standard penetration test, Wikimedia Foundation Incorporation, http://en.wikipedia.org/wiki/ Standard_penetration_test

Wikipedia (2014) S-wave, Wikimedia Foundation Incorporation, http://en.wikipedia. org/wiki/S-wave

Wun GL, Chien WY and Chang YW (2004) Site effect due to strong ground motion in Taiwan area, Central Weather Bureau, MOTC-CWB-93-E-09, 39: 127-176

Yeh YC (2004) Applied artificial neural networks, Scholar Books Company Limited, Taipei, Taiwan, 3-7

Yuen KV and Mu HQ (2011) Peak ground acceleration estimation by linear and nonlinear models with reduced order Monte Carlo simulation, *Computer-Aided Civil and Infrastructure Engineering*, 26(1): 30-47

Zobin VM, Bretón M and Navarro C (2014) Similarity in seismic source scaling relations for tectonic and volcanic processes, *Physics of the Earth and Planetary Interiors*, 231: 65-73