

A Comparative Study of the Least Squares Method and the Genetic Algorithm in Deducing Peak Ground Acceleration Attenuation Relationships

Ching-Yun Kao¹, Jen-Kuang Chung², and Yeong-Tein Yeh^{1,*}

¹*Institute of Industrial Safety and Disaster Prevention, Chia Nan University of Pharmacy and Science, Tainan, Taiwan, ROC*

²*Institute of Geoinformatics and Disaster Reduction Technology, Ching Yun University, Zhongli, Taiwan, ROC*

Received 14 September 2009, accepted 10 February 2010

ABSTRACT

In engineering applications, the development of attenuation relationships in a seismic hazard analysis is a useful way to plan for earthquake hazard mitigation. However, finding an optimal solution is difficult using traditional mathematical methods because of the nonlinearity of many relationships. Furthermore, using unweighted regression analysis in which each recording carries an equal weight is often problematic because of the non-uniform distribution of the data with respect to distance. In this study, the least squares method (LSM) and a genetic algorithm (GA) were employed as optimization methods for an attenuation model to compare the robustness and prediction accuracy of the two methods. Different (equal and unequal) weights of each recording were used to compare the adaptability of the weighting for practical application. The unequal weights of each recording were defined as functions of the hypocentral distance or the shortest distance from a station to the fault on the earth's surface. Finally, regression analysis of horizontal peak ground acceleration (PGA) attenuation model in southwest Taiwan was shown.

Key words: Attenuation law, Regression analysis, Peak ground acceleration, Genetic algorithm

Citation: Kao, C. Y., J. K. Chung, and Y. T. Yeh, 2010: A comparative study of the least squares method and the genetic algorithm in deducing peak ground acceleration attenuation relationships. *Terr. Atmos. Ocean. Sci.*, 21, 869-878, doi: 10.3319/TAO.2010.02.10.01(T)

1. INTRODUCTION

The development of an attenuation relationship is very important in seismic hazard analysis for earthquake hazard mitigation. There are several well-defined empirical attenuation models that relate given ground-motion parameters (e.g., peak ground acceleration, PGA) to several seismological parameters of an earthquake, such as earthquake magnitude, source-to-site distance, style of faulting, and local site conditions (Campbell 1981; Campbell 1991; Abrahamson and Silva 1997; Sadigh et al. 1997).

Empirical equations for predicting strong ground motion are typically fit into a strong-motion data set by least square methods (LSM)(Campbell 1981; Campbell 1991). Campbell (1981) used weighted least squares in an attempt to compensate for the non-uniform distribution of data with respect to distance. Campbell (1991) also analyzed peak horizontal accelerations recorded during the 18 October 1989 Loma Prieta, California earthquake to determine their

dependence on distance, azimuth, and generalized site conditions. The coefficients in peak ground motion attenuation law were estimated from a nonlinear least squares regression algorithm. The LSM is fast in execution if the initial guess of the solution is adequate. However, a LSM has some drawbacks, including that the initial guess of the solution is difficult to determine; the solution may be an optimal local solution and a gradient of objective function should be exist (Draper and Smith 1966).

A GA (Genetic Algorithm) is a robust search technique based on the principles of evolution (Goldberg 1989). Extensive research has been performed exploiting the properties of genetic algorithms and demonstrating their capabilities across a broad range of problems. These evolutionary methods have gained recognition as general problem solving techniques in many applications, including image processing, neural networks, etc. Unlike traditional regression analysis methods, a GA searches for a global optimal solution and does not need to calculate the gradient of the objective function and thereby makes GA a highly promising tool

* Corresponding author
E-mail: yteychna@gmail.com

for regression analysis. Although a GA has many positive features, it is slow in execution and is best applied to difficult problems.

The feasibility of applying a GA to estimate coefficients in peak ground motion attenuation law also has received attention. Tavakoli and Pezeshk (2005) predicted the ground-motion relationship for eastern North America from the ground-motion relationship for western North America based on a hybrid-empirical model. In their study, a composite functional attenuation form was defined, and, in turn, an optimization was performed by using a genetic algorithm for a wide range of magnitudes and distances. Scherbaum et al. (2006) developed a systematic way of estimating minimum-misfit stochastic models from empirical ground-motion prediction equations. In that study, a genetic algorithm search was used to determine the model parameters. Although both LSM and GA have been applied to estimate coefficients in peak ground motion attenuation law, there are no studies comparing the two methods in this field.

Most studies (e.g., Jean et al. 2006) that developed peak ground motion attenuation law in Taiwan used the ground motion data for the entire area of Taiwan. Peak ground motion attenuation is very complicated because it is affected by many factors, such as the characteristics of seismic source, attenuation of seismic waves from the seismic source to a recording site, and soil characteristics at the recording site. Therefore, we believe that a feasible way to enhance the accuracy of predicted peak ground motion attenuation in Taiwan is to separate the Taiwan into several smaller areas, and then develop peak ground motion attenuation law for each area.

In this study, the robustness and predicted accuracy of the LSM and the GA were compared. In addition, different (equal and unequal) weights of each recording were used to compare the adaptability of the weights for practical application. The regression methods used in this study include the least squares method (LSM) and a genetic algorithm (GA).

2. DATA AND OPTIMIZATION OF THE ATTENUATION MODEL

2.1 Data

The Taiwan strong-motion seismic network consists of about 700 free-field strong-motion stations. Each station includes triaxial accelerometers, a digital recorder, a power supply, and a GPS timing system. Most of the digital accelerometers used in these stations are ± 2 g full scale, 200 or higher samples per second, and 16-bit or better resolution with up to 20-second pre-event recording (Liu et al. 1999). The strong-motion stations are spaced approximately 5 km apart in nine metropolitan regions. The Taiwan strong-motion seismic network has collected a large amount of high quality strong-motion data and provided much useful infor-

mation for seismology and earthquake engineering. These ground motion data offer a good opportunity to study attenuation models. Earthquake data collected from Taiwan strong-motion seismic network were used in this study to develop the empirical horizontal PGA attenuation model. This study focused on analyzing strong-motion recordings collected at sites with soil conditions of classes C and D in the southwest region (Lee et al. 2001). Because of the near-source data recorded within a distance of 10 km can be crucial in regression of PGA attenuation, some extra stations in central and eastern Taiwan (shown as green triangles in Fig. 1) were used to provide additional observations. However, the stations underlain by unconsolidated soft soil (class E) were excluded to avoid unpredictable changes in amplitude due to significant site effects.

The database consisted of 5323 recordings collected from 208 earthquake events of local magnitude $4.5 \leq M_L \leq 7.3$ and hypocentral distance or shortest distance from a station to the fault on earth surface $R_h \leq 300$ km from 1993 to 2006. Figure 1 shows the free-field strong-motion stations and the epicenter of earthquakes used in this study. Figure 2 shows the distribution of recordings with respect to magnitude and distance.

2.2 Optimization of the Attenuation Model

2.2.1 Attenuation Model

Since Campbell's attenuation form (Campbell 1981) can reasonably predict the characteristic of ground motion attenuation from the Taiwan strong seismic network (Jean et al. 2006), the same approach was applied in this study. Campbell's form is expressed as follows

$$Y(g) = b_1 e^{b_2 M_L} (R_h + b_4 e^{b_5 M_L})^{-b_3} \quad (1)$$

The parameter Y is the geometric average of two horizontal PGA, which is abbreviated as horizontal PGA below, M_L is local magnitude (Richter's magnitude). Since both the PGA and M_L are controlled by short period seismic waves, M_L is better than M_W (Moment magnitude), which is controlled by the amount of fault displacement, for estimating attenuation of PGA. R_h is defined by the distance from the energy source to the recording site. In this study, hypocentral distance was used as R_h for earthquakes were source dimension, relative to source-site distance, was small and the shortest distance from a station to the long fault on earth surface was adopted to approximate R_h (for the Chi-Chi earthquake)(Jean et al. 2006).

2.2.2 LSM

The general form of nonlinear regression models can be written as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{e} \tag{2}$$

where the vector of dependent variables $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$; the vector of independent variables $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$; the vector of parameters to be estimated $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_p]^T$; the vector of errors $\mathbf{e} = [e_1 \ e_2 \ \dots \ e_n]^T$; $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_n]^T$ where f_i ($i = 1, 2, \dots, n$) is a nonlinear function. In a least squares

method, by minimizing S in Eq. (2) the parameters of a non-linear regression model can be estimated:

$$S = \frac{1}{2}[\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})]^T [\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})] \tag{3}$$

In contrast with linear models, analytical solution methods are not sufficient in solving for parameters in nonlinear

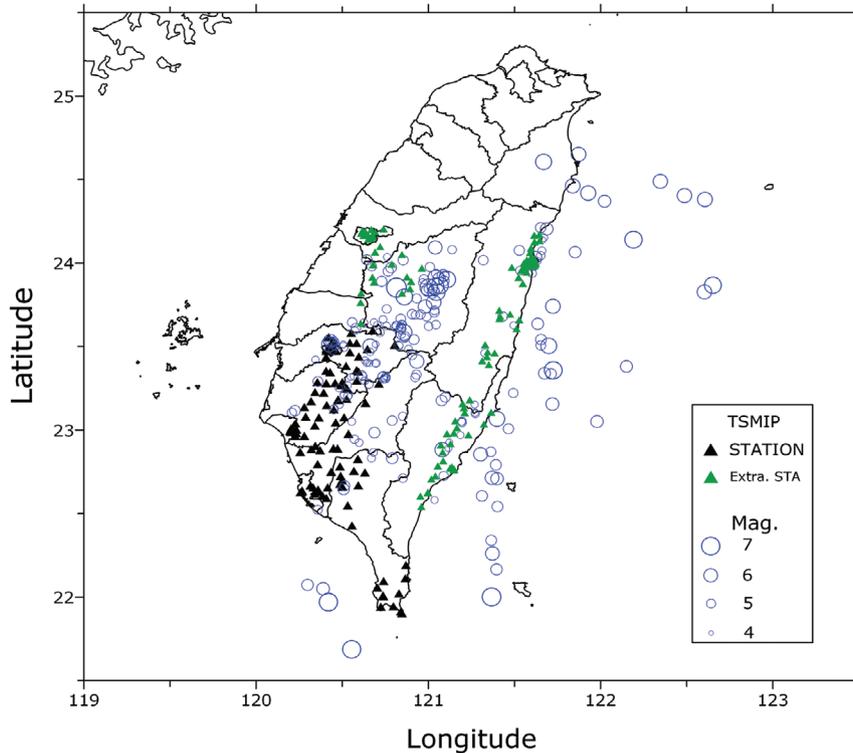


Fig. 1. The free-field strong-motion stations and the epicenter of earthquakes used in this study.

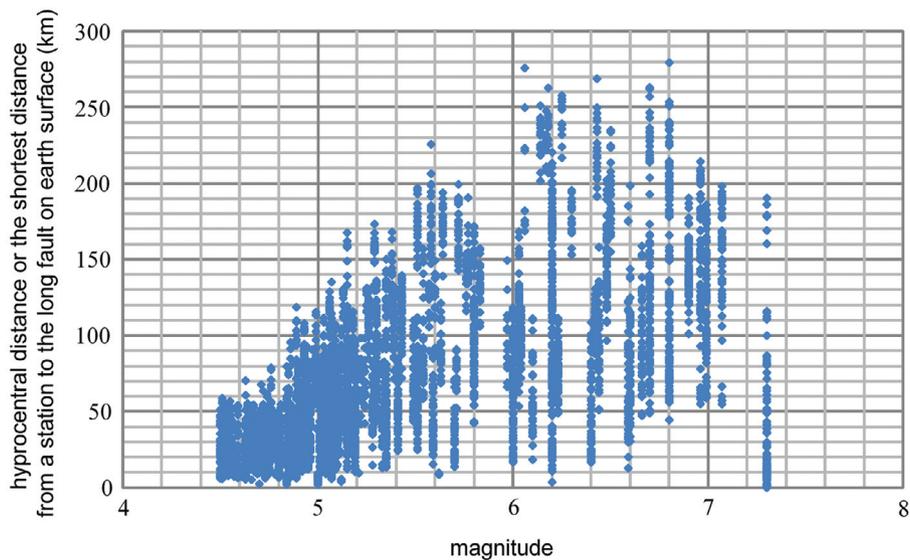


Fig. 2. The distribution of strong-motion recordings with respect to magnitude and distance.

models, and therefore, iterative numerical search methods, such as the linearization method, steepest descent method, and the Levenberg-Marguardt method, are often used to solve this problem by taking derivative of S in Eq. (3) with respect to θ_i (Draper and Smith 1966).

2.2.3 GA

Recently, computational intelligence methods have been applied to a broad range of problems. Computational intelligence methods, such as neural networks and GA, are highly adaptive methods originating from the laws of nature and biology. Unlike mathematical methods, one of the important characteristics of computational intelligence methods is their effectiveness and robustness in coping with uncertainty, insufficient information, and noise.

The method of string representation in the GA used this study, is shown in Fig. 3. In this method of string representation, the value of each parameter is represented by a sub-string of k -bit binary integers. In a simple GA, a string is composed of sequentially connecting all the sub-strings. Figure 4 shows the typical string representations in a simple GA using binary bits (each parameter is encoded in an 8-bit binary string in Fig. 4). The binary bits for the parameters are sequentially concatenated in a simple GA.

To generate the fitter string, a GA reproduces the population according to their relative fitness; the strings with higher fitness have a better chance of passing their genes to the next generation. The proportional method is used to select the members of the next generation. A pair of parents is selected by the roulette wheel method. The slots on the perimeter of the roulette wheel are assigned to the individuals in proportion to their relative fitness functions. After reproduction a one-point crossover, with the probability of p_c , is performed to evolve new offspring. In addition, to inhibit premature convergence during reproduction and crossover, mutation, with the probability of p_m , is implemented to maintain the genetic variability of the string.

Figure 5 shows the essential elements of a simple GA, which starts with a randomly generated population of in-

dividual possible solutions scattered over a pre-determined search space (the region in which the best answer is thought to lie). The relative fitness of these individuals is determined and a stochastic selection process biased towards the fitter individuals is used to select parents for mating. In mating, attributes of the parents are mixed to form offspring which may or may not be fitter than one or both of the parents. In forming offspring, occasional random mutations can occur and also have the possibility of leading to a fitter individual. The process of selection, mating and mutation is repeated over a number of generations to allow the solution to evolve towards an optimum.

3. RESULTS AND DISCUSSION

Six objective functions were used in this paper and were defined as follows:

$$\sum_{p=1}^P (Y_{m,p} - Y_{e,p})^2 / P \tag{4}$$

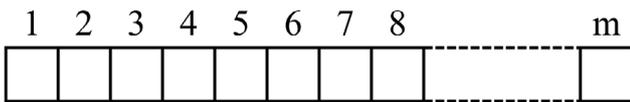


Fig. 3. String representation in simple GA.

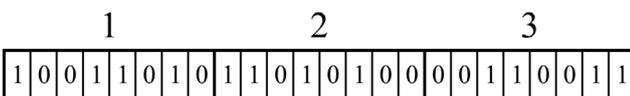


Fig. 4. Example of string representation scheme in simple GA.

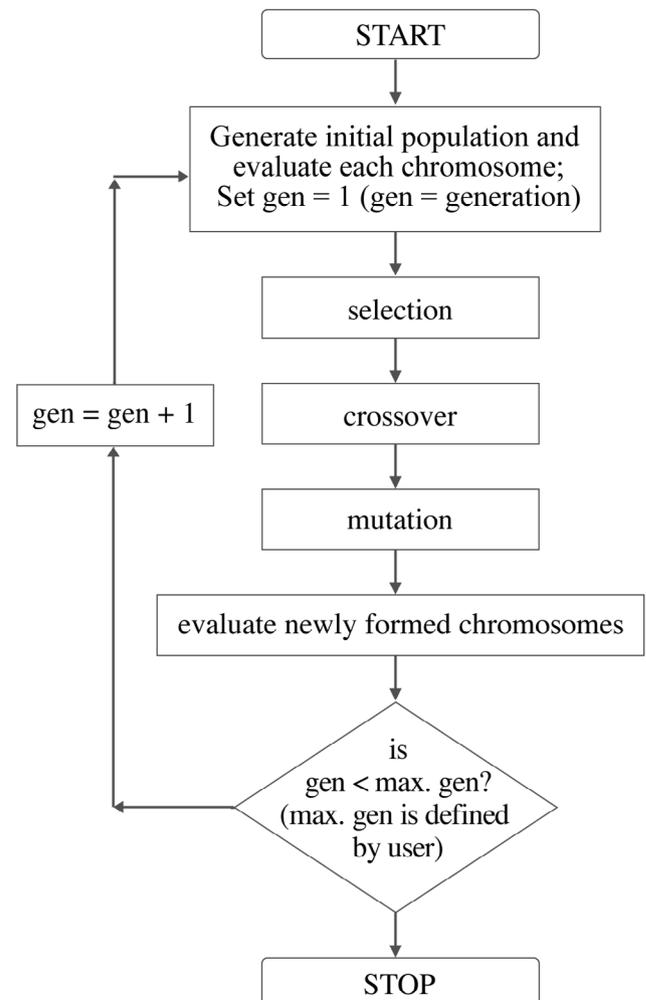


Fig. 5. Elements of simple GA.

$$\sum_{p=1}^P \frac{1}{\sqrt{R_{h,p}}} (Y_{m,p} - Y_{e,p})^2 / P \tag{5}$$

$$\sum_{p=1}^P \frac{1}{R_{h,p}} (Y_{m,p} - Y_{e,p})^2 / P \tag{6}$$

$$\sum_{p=1}^P (\ln Y_{m,p} - \ln Y_{e,p})^2 / P \tag{7}$$

$$\sum_{p=1}^P \frac{1}{\sqrt{R_{h,p}}} (\ln Y_{m,p} - \ln Y_{e,p})^2 / P \tag{8}$$

$$\sum_{p=1}^P \frac{1}{R_{h,p}} (\ln Y_{m,p} - \ln Y_{e,p})^2 / P \tag{9}$$

where $Y_{m,p}$ and $Y_{e,p}$ are the measured and estimated PGA of the p th instance, respectively. P is the total number of occurrences, and $R_{h,p}$ is the hypocentral distance of the p th instance. The first three objective functions [as shown in Eqs. (4) to (6)] are defined in terms of the output error, which is a function of the difference between the measured and estimated PGA. The last three objective functions [as shown in Eqs. (7) to (9)] are defined in terms of the output error, which is a function of the difference between the measured and estimated natural logarithm of PGA. The weight of each recording in Eqs. (4) and (7) is 1. It is obvious that the strong motion data at short distances are more important than those collected at longer distances in earthquake hazard assessment. Moreover, the farther seismic waves travels, greater wave scattering would be expected. The PGA measurements of long distance would be more dispersed

than those of short distance. Hence, the unequal weights of each recording can be defined as functions of hypocentral distance or the shortest distance from a station to the fault on the earth's surface. Thus, the weight of each recording in Eqs. (5) and (8) is defined as $1/\sqrt{R_{h,p}}$ and Eqs. (6) and (9) is $1/R_{h,p}$. Therefore, the regression method with functions defined in Eqs. (5), (6), (8), and (9) seem to be more suitable for practical application.

3.1 Comparison of the Robustness and Accuracy of the LSM and the GA

First, to compare the robustness and accuracy of the LSM and the GA, the LSM and the GA with the first three objective functions were used for optimization. The LSM with an objective function defined in Eqs. (4), (5), and (6) were denoted as LSM1, LSM2, and LSM3, respectively. The methods of the GA with objective function defined as Eqs. (4), (5), and (6) were denoted as GA1, GA2, and GA3, respectively. The optimization of the GA maximizes the fitness function. However, the optimization of GA1, GA2, and GA3 minimize the objective functions defined as Eqs. (4), (5), and (6), respectively. Therefore, the fitness functions of GA1, GA2, and GA3 can be defined as a constant (25 was used in this study) minus objective functions defined as Eqs. (4), (5), and (6), respectively. The number of iteration was set to be 5000 and the crossover probability (p_c) and mutation probability (p_m) were set to be 0.8 and 0.05, respectively for all of GA1, GA2, and GA3.

The predicted coefficients in Campbell's form are listed in Table 1. It is worthwhile to note that these coefficients predicted by the LSM are largely influenced by the initial guess. The coefficients predicted by LSM1 were

Table 1. The predicted coefficients, the standard deviation of the natural logarithm of the horizontal PGA, $\sigma_{\ln(err)}$, and the value of objective function for Campbell's form.

Method	Regression coefficients					$\sigma_{\ln(err)}$	Value of objective function
	b_1	b_2	b_3	b_4	b_5		
LSM1	0.0397	3.3146	4.0560	0.7320	0.7867	0.8285	3.2932
LSM2	/	/	/	/	/	/	/
LSM3	/	/	/	/	/	/	/
GA1	0.1000	3.4770	0.9871	0.7577	4.3725	0.8370	3.2954
GA2	0.0929	2.4235	1.0945	0.6698	3.2544	0.7509	1.0657
GA3	0.0687	1.0359	1.0958	0.4513	1.6485	0.7076	0.6149
GA4	0.0095	1.2143	1.5632	0.6575	0.5282	0.7459	/
GA5	0.0127	1.1678	1.4948	0.7705	0.4697	0.6370	/
GA6	0.0188	1.2019	1.6211	0.9482	0.4745	0.6441	/

Note: "/": no data available.

easily found after a few trials for initial guesses. However, these coefficients were not found by LSM2 and LSM3 after 245 trials of initial guesses, which explains the difficulty of initial guess determination of the LSM. Among the 245 initial guesses for LSM2, one is the initial guess in LSM1 optimization, another is one of the initial guesses in GA2 optimization, and the rest includes b_1 is increased from 0.09 to 0.1 every 0.005, b_2 from 2.4 to 2.5 every 0.05, b_3 from 1.0 to 1.1 every 0.05, b_4 from 0.6 to 0.7 every 0.05, and b_5 from 3.2 to 3.3 every 0.05. Among the 245 initial guesses for LSM3, one is the initial guess in LSM1 optimization, another is one of the initial guesses in GA3 optimization, and the rest includes b_1 is increased from 0.06 to 0.07 every 0.005, b_2 from 1.0 to 1.1 every 0.05, b_3 from 1.0 to 1.1 every 0.05, b_4 from 0.4 to 0.5 every 0.05, and b_5 from 1.6 to 1.7 every 0.05. Compared with the LSM, these coefficients were much more easily found by the GA. These coefficients were found by GA1, GA2, and GA3 using any initial guess which reveals that the GA was a much more robust optimization method compared with the LSM.

Table 1 lists the standard deviation of the natural logarithm of the peak ground acceleration, representing the dispersion about their respective median value, and objective function values of these methods. The standard deviation of the natural logarithm of the peak ground acceleration, $\sigma_{\ln(err)}$, is defined as follows

$$\sigma_{\ln(err)} = \sqrt{\sum_{p=1}^P (\ln Y_{m,p} - \ln Y_{e,p})^2 / (P - n_c)} \quad (10)$$

where n_c is the number of coefficients to be estimated in attenuation model, that is $n_c = 5$. Our results show that the standard deviation of the natural logarithm of PGA and the objective function value of LSM1 and GA1 are about the same. A smaller value of objective function implies a more accurate prediction. Although predicted accuracies of LSM1 were better than those of GA1, the predicted accuracy of a GA method may increase with the increasing number of iterations. The standard deviations of the natural logarithm of PGA, from the smallest value to the largest value, are those of GA3, GA2, LSM1, and GA1, respectively. This shows that the GA has a better chance to arrive at a more accurate prediction compared with the LSM. Furthermore, weighting the objective function with the factors defined as the reciprocals of square root of source-site distance ($1/\sqrt{R_{h,p}}$) or source-site distance ($1/R_{h,p}$) makes the GA regression procedure easier to reach an optimum solution.

3.2 Estimation of the Optimum Solution Using the GA

Second, to find the optimum solution that minimizes the standard deviation of the natural logarithm of PGA, three different objective functions, defined in Eqs. (7), (8),

and (9), were used for optimization. Remarkably, the objective function of a LSM can not be defined using Eqs. (7) to (9). Methods of the GA with objective functions defined in Eqs. (7), (8), and (9) were denoted as GA4, GA5, and GA6, respectively. The predicted coefficients using Campbell's form and the standard deviation of the natural logarithm of PGA are listed in Table 1. The standard deviation of the natural logarithm of the PGA of GA4 is smaller than that of GA1, GA5 is smaller than GA2, and GA6 is smaller than GA3 as well. Our results show that the GA with objective functions defined in Eqs. (7) to (9) are much more adequate than those with objective functions defined in Eqs. (4) to (6) for optimization that minimizes standard deviation of the natural logarithm of PGA. The standard deviation of the natural logarithm of PGA of GA5 was about the same as that of GA6 but smaller than that of GA4. Therefore, GA5 and GA6 would be better than GA4 to estimate the optimum solution of the coefficients in Campbell's form for current earthquake data (nonuniform distribution of the data with respect to distance). Generally speaking, near-source ($R_h \leq 50$ km) damage is more serious than far-source ($R_h > 50$ km) damage during large earthquakes; therefore, the optimum solution that minimizes the standard deviation of near-source data should be ideal for practical application. Table 2 lists the standard deviation of the natural logarithm of near-source and far-source PGA data of GA4, GA5, and GA6. The results show that the standard deviation of the natural logarithm of near-source PGA data of GA5 and GA6 were almost the same and both of them are smaller than that of GA4.

The results of far-source data present the same situation. However, the results (Tables 1 and 2) indicate that GA5 is better than GA6. That is, GA5 is more suitable for practical application than GA4 and GA6. The weighting factor as the reciprocal of square root of source-to-site distance allows more freedom for long distance data fitting. Figure 6 shows the comparison of the earthquake data with the attenuation model predicted by GA4, GA5, and GA6.

To apply the constrained PGA attenuation model by Campbell (1981), two constraints were made. The first constraint was that the far-source attenuation rate b_3 of Eq. (1)

Table 2. Near-source and far-source standard deviation of the natural logarithm of the horizontal PGA, $\sigma_{\ln(err)}$, of GA4, GA5, and GA6 for Campbell's form.

Method	$\sigma_{\ln(err)}$	
	$R_h \leq 50$ km	$R_h > 50$ km
GA4	0.7717	0.7272
GA5	0.7140	0.5752
GA6	0.7141	0.5886

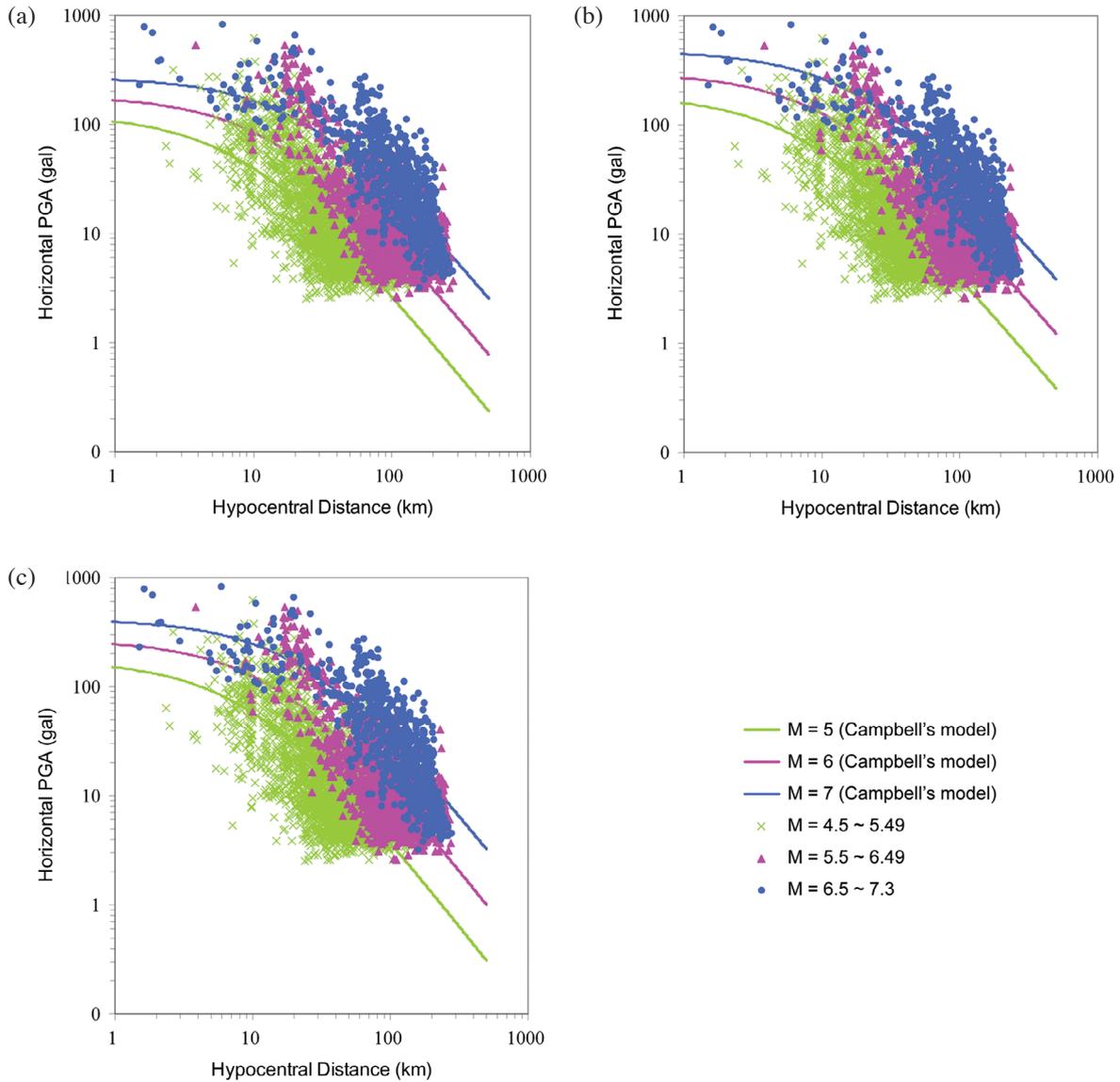


Fig. 6. Comparison of the horizontal PGA with the attenuation form predicted by (a) GA4, (b) GA5, and (c) GA6.

was constrained to a value of 1.75 in order for the model results to be consistent with far-source data and give a more realistic value at larger distances. The second constraint was a near-source constraint involving distances nearer than 3 to 5 km from fault, where strong-motion data are extremely limited. Many seismologists and geophysicists currently believe that at or very near the rupture surface peak acceleration become essentially independent of earthquake magnitude (Brune 1970; Trifunac 1973; Ambraseys 1978; Hanks 1979; Hadley and Helmberger 1980; McGarr et al. 1981). Based on this argument, the coefficient b_5 of Eq. (1) was constrained to be $b_2/1.75$. Campbell's constrained form can be expressed as follows:

$$Y(g) = b_1 e^{b_2 M_L} (R_h + b_4 e^{b_2/1.75 M_L})^{-1.75} \quad (11)$$

The predicted coefficients and the standard deviation of the natural logarithm of PGA of GA4, GA5, and GA6 using Campbell's constrained form are listed in Table 3. The results show that the three standard deviations of the natural logarithm are about equal. Compared with the results in Table 1, the standard deviation of the natural logarithm of the PGA of GA4 for the constrained form was smaller than that for an unconstrained form. The standard deviations of the natural logarithm of the PGA of GA5 and GA6 for the constrained form were slightly larger than those for unconstrained model. Table 4 lists near-source and far-source standard deviation of the natural logarithm of the horizontal PGA of GA4, GA5, and GA6 for the constrained form. These results also show that the standard deviations of the natural logarithm of near-source PGA data of GA5 and GA6 were almost the same and both were smaller than those of GA4

Table 3. The predicted coefficients and the standard deviation of the natural logarithm of the horizontal PGA, $\sigma_{\ln(err)}$, for Campbell's constrained form.

Method	Regression coefficients					$\sigma_{\ln(err)}$
	b_1	b_2	b_3	b_4	b_5	
GA4	0.0126	1.3996	1.75	0.1963	0.7998	0.6525
GA5	0.0048	1.5492	1.75	0.0916	0.8853	0.6481
GA6	0.0031	1.6120	1.75	0.0602	0.9211	0.6565

using the constrained form. Compared to the results listed in Table 2 (Campbell's unconstrained form) reveal that all near-source standard deviations of the natural logarithm of the PGA of GA4, GA5, and GA6 for the constrained form were smaller than those for unconstrained form.

In fact, the differences in near-source standard deviations of the natural logarithm of the PGA of GA5 and GA6 for unconstrained and constrained forms are small. The far-source standard deviations of the natural logarithm of the PGA of GA5 and GA6 using the unconstrained form were smaller than those of constrained form. Figure 7 shows the comparison of the unconstrained and constrained forms predicted by GA4, GA5, and GA6 for earthquakes with magnitudes of 5, 6, and 7. At near distances, the peak acceleration is independent of earthquake magnitude for constrained form compared with the unconstrained form. Interestingly, Fig. 7 shows that the predicted near-source PGA values, by the constrained GA4 and GA5, both converged toward the values predicted for magnitude 6 while those predicted by constrained GA6 converged toward higher values for magnitude 7. It should be acknowledged that hazard estimations, at least in probabilistic analysis, are influenced not only by the level of ground motion predicted but also by the variability of the predicted ground motion expressed by the standard deviation of the attenuation relation. Comparing the observed data (Fig. 6) with the constrained prediction models, the constrained GA6 may be more adequate for predicting the PGA of large earthquake than the constrained GA4 and GA5.

In practice, for seismic hazard assessments in terms of PGA, the results are controlled by the occurrence rate of earthquakes with various magnitudes and within a short distance (less than 50 km). In general, the earthquake occurrence rate increases with decreasing magnitude. Hence, for a certain region, the hazard estimated using constrained GA6 should be higher than that estimated by using unconstrained GA6 because constrained GA6 predicts higher PGAs for small magnitude earthquakes. However, the hypothesis that PGAs at near-source distances become essentially independent of earthquake magnitude is not confirmed here due to lack of near-source observations in southwestern Taiwan. Based on the results listed in Tables 1 and 2, the PGA at-

Table 4. Near-source and far-source standard deviation of the natural logarithm of the horizontal PGA, $\sigma_{\ln(err)}$, of GA4, GA5, and GA6 for Campbell's constrained form.

Method	$\sigma_{\ln(err)}$	
	$R_h \leq 50$ km	$R_h > 50$ km
GA4	0.7280	0.5923
GA5	0.6952	0.6122
GA6	0.6867	0.6342

tenuation relation of GA5 is suggested as the most practical for application to southwestern Taiwan.

4. CONCLUSIONS

This study compared the robustness and prediction accuracy of the LSM and the GA. A GA is a much more robust optimization method than a LSM because a LSM can't find the optimal solution of coefficients if the initial guess is not sufficiently near the ideal solution. Our results show that a GA has a better chance to obtain a more accurate model compared with using a LSM.

Different objective functions were used in this study for optimization. We found that the objective function defined as a function of the difference between the observed and estimated PGA of the natural logarithm was better than that defined as a function of the difference between the observed and estimated PGA for optimization that minimizes the standard deviation of the natural logarithm of PGA. Notably, the fitness function of a GA method is easy to define as a function of the difference between the observed and estimated PGA of the natural logarithm. However, the objective function of a LSM can only be defined as a function of the difference between the observed and estimated PGA.

Furthermore, different (equal and unequal) weights for each earthquake record were used to compare the adaptability of the weights for practical application. Our results indicate that the optimization procedure reaches the optimum solution more easily when the objective function is weight-

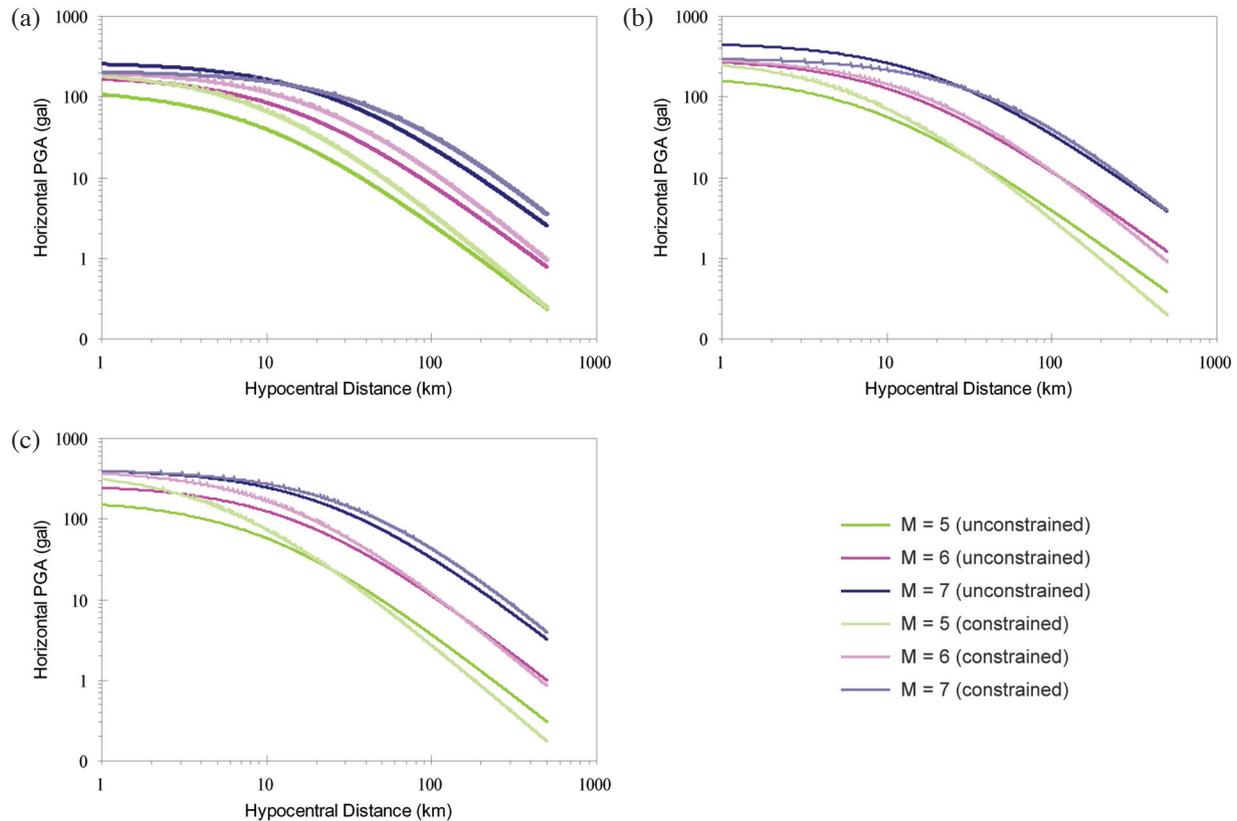


Fig. 7. Comparison of Campbell's unconstrained and constrained forms for magnitudes of 5, 6, and 7 of (a) GA4, (b) GA5, and (c) GA6.

ed with the factors defined as the reciprocals of square root of source-site distance ($1/\sqrt{R_{h,p}}$) or source-site distance ($1/R_{h,p}$).

We used the unconstrained and constrained attenuation models of Campbell (1981) to analyze the PGA attenuation relationships for southwestern Taiwan. Overall our results show that a useful horizontal peak ground acceleration (PGA) attenuation model in southwest Taiwan deduced by the GA is

$$Y(g) = 0.0127 e^{1.1678 M_L} (R_h + 0.7705 e^{0.4697 M_L})^{-1.4948} \quad (12)$$

Acknowledgements This research was supported by the Taiwan Earthquake Research Center (TEC) funded through National Science Council (NSC) with grant number NSC 96-2119-M-041-001 and NSC 97-2116-M-041-003. Appreciation is also extended to the Central Weather Bureau in Taiwan for providing earthquake data. The TEC contribution number for this article is 00068.

REFERENCES

- Abrahamson, N. A. and W. J. Silva, 1997: Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismol. Res. Lett.*, **68**, 94-127.
- Ambraseys, N. N., 1978: Preliminary analysis of European strong-motion data 1965-1978. *Bull. Europ. Assoc. Earthq. Eng.*, **4**, 17-37.
- Brune, J. N., 1970: Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res.*, **75**, 4997-5009, doi: 10.1029/JB075i026p04997. [Link]
- Campbell, K. W., 1981: Near-source attenuation of peak horizontal acceleration. *Bull. Seismol. Soc. Am.*, **71**, 2039-2070.
- Campbell, K. W., 1991: An empirical analysis of peak horizontal acceleration for the Loma Prieta, California, earthquake of 18 October 1989. *Bull. Seismol. Soc. Am.*, **81**, 1838-1858.
- Draper, N. R. and H. Smith, 1966: Applied Regression Analysis, John Wiley and Sons, 505-566.
- Goldberg, D. E., 1989: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 432 pp.
- Hadley, D. M. and D. V. Helmberger, 1980: Simulation of strong ground motions. *Bull. Seismol. Soc. Am.*, **70**, 617-630.
- Hanks, T. C., 1979: b values and $\omega^{-\gamma}$ seismic source models: Implications for tectonic stress variations along active crustal fault zones and the estimation of high-frequency strong ground motion. *J. Geophys. Res.*, **84**, 2235-2242, doi: 10.1029/JB084iB05p02235. [Link]

- Jean, W. Y., Y. W. Chang, K. L. Wen, and C. H. Loh, 2006: Early estimation of seismic hazard for strong earthquakes in Taiwan. *Nat. Hazards*, **37**, 39-53, doi: 10.1007/s11069-005-4655-y. [[Link](#)]
- Lee, C. T., C. T. Cheng, C. W. Liao, and Y. B. Tsai, 2001: Site classification of Taiwan free-field strong-motion stations. *Bull. Seismol. Soc. Am.*, **91**, 1283-1297, doi: 10.1785/0120000736. [[Link](#)]
- Liu, K. S., T. C. Shin, and Y. B. Tsai, 1999: A free-field strong motion network in Taiwan: TSMIP. *Terr. Atmos. Ocean. Sci.*, **10**, 377-396.
- McGarr, A., R. W. E. Green, and S. M. Spottiswoode, 1981: Strong ground motion of mine tremors: Some implications for near-source ground motion parameters. *Bull. Seismol. Soc. Am.*, **71**, 295-319.
- Sadigh, K., C. Y. Chang, J. A. Egan, F. Makdisi, and R. R. Youngs, 1997: Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismol. Res. Lett.*, **68**, 180-189.
- Scherbaum, F., F. Cotton, and H. Staedtke, 2006: The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations. *Bull. Seismol. Soc. Am.*, **96**, 427-445, doi: 10.1785/0120050015. [[Link](#)]
- Tavakoli, B. and S. Pezeshk, 2005: Empirical-stochastic ground-motion prediction for eastern north America. *Bull. Seismol. Soc. Am.*, **95**, 2283-2296, doi: 10.1785/0120050030. [[Link](#)]
- Trifunac, M. D., 1973: Analysis of strong earthquake ground motion. *Int. J. Earthq. Eng. Struct. Dyn.*, **2**, 59-69.