Engineering Applications of Artificial Intelligence 24 (2011) 717–732

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A hybrid computational approach to derive new ground-motion prediction equations

Amir Hossein Gandomi [a], Amir Hossein Alavi [b,*], Mehdi Mousavi [c], Seyed Morteza Tabatabaei [d]

[a] College of Civil Engineering, Tafresh University, Tafresh, Iran
[b] School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran
[c] Department of Civil Engineering, Faculty of Engineering, Arak University, Arak, Iran
[d] Department of Civil Engineering, The Institute of Higher Education of Eqbal Lahoori, Mashhad, Iran

ARTICLE INFO

Article history:
Received 13 May 2010
Received in revised form
9 December 2010
Accepted 10 January 2011

Keywords:
Time-domain ground-motion parameters
Prediction equations
Genetic programming
Orthogonal least squares
Nonlinear modeling

ABSTRACT

A novel hybrid method coupling genetic programming and orthogonal least squares, called GP/OLS, was employed to derive new ground-motion prediction equations (GMPEs). The principal ground-motion parameters formulated were peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD). The proposed GMPEs relate PGA, PGV and PGD to different seismic parameters including earthquake magnitude, earthquake source to site distance, average shear-wave velocity, and faulting mechanisms. The equations were established based on an extensive database of strong ground-motion recordings released by Pacific Earthquake Engineering Research Center (PEER). For more validity verification, the developed equations were employed to predict the ground-motion parameters of the Iranian plateau earthquakes. A sensitivity analysis was carried out to determine the contributions of the parameters affecting PGA, PGV and PGD. The sensitivity of the models to the variations of the influencing parameters was further evaluated through a parametric analysis. The obtained GMPEs are effectively capable of estimating the site ground-motion parameters. The equations provide a prediction performance better than or comparable with the attenuation relationships found in the literature. The derived GMPEs are remarkably simple and straightforward and can reliably be used for the pre-design purposes.

1. Introduction and background

Seismic hazard analysis is one of the fundamental steps in engineering phase. The seismological characteristics of earthquakes commonly include magnitude, distance, faulting type, and soil effects. The engineering parameters of an earthquake can be divided into two main classes: (1) time-domain parameters and (2) response-domain parameters. Peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD) are the major time-domain class parameters. In the response domain, pseudo spectral acceleration (PSA) is the well-known parameter. Both classes of the time- and response-domain parameters can be applied to the risk assessment of structures. It has been demonstrated that the spectral parameters are more efficient than the time-domain parameters (Luco and Cornell, 2007). On the other hand, application of the time-domain parameters is more convenient due to their independency from the considered structures. Thus, PGA, PGV and PGD are commonly used in the seismic hazard studies. Various methods may be employed to estimate these elements such as on-site investigation and physical modeling. Implementing these methods is usually extensive, cumbersome and costly (Güllü and Erçelebi, 2007). Much effort is often made to describe limited observations through the physical modeling of the earthquake process. In this approach, the observations are used for the calibration of the physical model. Such models are usually developed in the context of stochastic modeling approach and random vibration theory (see e.g., Papageorgiou and Aki, 1983). More advanced physical models try to model the realistic process of faulting through the numerical modeling of crack and wave propagation (e.g., Krishnan et al., 2006).

Ground-motion prediction equations (GMPEs) play a critical role in the seismic hazard analysis. The GMPEs relate the ground-motion parameters to various independent variables such as earthquake magnitude, distance from source to site, local site conditions, earthquake source characteristics, and wave propagation (Kramer, 1996; Douglas, 2003; Güllü and Erçelebi, 2007). Other physical parameters such as stress drop, rupture propagation, directivity, basin effects, and nonlinear soil behavior are not generally used in the predictive models (Somerville and Graves, 2003). Correlating

* Corresponding author.
E-mail addresses: a.h.gandomi@gmail.com (A.H. Gandomi), ah_alavi@hotmail.com (A.H. Alavi), m.mousavi@ut.ac.ir (M. Mousavi).

0952-1976/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2011.01.005

PGA, PGV and PGD with the predictor parameters in a mathematical form is not an easy task due to high nonlinearity in the relationships. A conventional way to construct the GMPEs from the recorded strong-motion data is to use the regression analysis (e.g., Ambraseys et al., 1996; Boore et al., 1997; Boore and Atkinson, 2007; Campbell and Bozorgnia, 2007). In each form, a simple statistical model may be developed to describe the tendency of the ground-motion parameters with other parameters at the station. In addition to the physical aspects (Douglas, 2003, 2004; Güllü and Erçelebi, 2007), the significant limitations of the statistical techniques strongly affect the capabilities of the regression-based GMPEs. Most commonly used regression analyses can have large uncertainties. It has major drawbacks for idealization of complex processes, approximation, and averaging widely varying prototype conditions. Another important issue is due to the limitation of this method. The regression analysis tries to model the nature of the corresponding problem by a pre-defined linear or nonlinear equation (Gandomi et al., 2010a). Another major constraint in application of the regression analysis is the assumption of normality of residuals. Thus, the developed attenuation models are often limited in their ability to reliably simulate the complex behavior of the ground-motion parameters (Güllü and Erçelebi, 2007). The issues raised above suggest the necessity of employing more comprehensive methods to decrease errors for the ground-motion parameters estimates.

Artificial neural networks (ANNs) are the well-known pattern recognition methods. Recently, considerable researches have been carried out to estimate the strong-motion characteristics via ANNs. Kerh and Chu (2002) employed ANNs to estimate PGA at two main line sections of Kaohsiung Mass Rapid Transit in Taiwan. They also compared the ANN-based results with those of available empirical formulas in the literature. Chu et al. (2003) developed a neural network model by employing historical seismic records to analyze the strong-motion characteristics around the Kaohsiung area of Taiwan. Kerh and Ting (2005) used back-propagation neural networks to predict PGA along the high-speed rail system in Taiwan. Güllü and Erçelebi (2007) and Gunaydın and Gunaydın (2008) developed PGA prediction models using ANNs based on the strong-motion data from Turkey. However, ANNs are not usually able to produce practical prediction equations. Moreover, they require the structure of the network to be identified a priori. The ANN method is mostly appropriate to be used as a part of a computer program.

Genetic programming (GP) (Koza, 1992) is a new approach with completely new characteristics and traits. GP is an extension of genetic algorithms. It may generally be defined as a supervised machine learning technique that searches a program space instead of a data space. The programs generated by GP are represented as tree structures and expressed in the functional programming language (Koza, 1992). The main advantage of the GP-based approaches over the regression and ANN techniques is their ability to generate prediction equations without assuming prior form of the relationship. The developed equations can be easily manipulated in practical circumstances (Gandomi et al., 2010a, 2011). Many researchers have employed GP and its variants to find out any complex relationships among the experimental data (e.g., Johari et al., 2006; Gandomi et al., 2010a,b, 2011; Alavi et al., 2010; Alavi and Gandomi, 2010). Some of the limited scientific efforts directed at applying GP to the analysis of the ground-motion parameters include prediction of PGA using the strong ground-motion data from Turkey (Cabalar and Cevik, 2009), and developing new predictive equations for the ratio of PGV to PGA (Kermani et al., 2009).

Orthogonal least squares (OLS) algorithm (Billings et al., 1988) is an effective algorithm to determine which terms are significant in a linear-in-parameters model. The OLS algorithm introduces the error reduction ratio, which is a measure of the decrease in the variance of output by a given term. Madár et al. (2005) combined GP and OLS to make a hybrid algorithm with better efficiency. It was shown that introducing this strategy into the GP process results in more robust and interpretable models. Some of the limited researches with the specific objective of applying the GP/OLS method to civil engineering problems have recently conducted by Gandomi and Alavi (in press) and Gandomi et al. (2010a).

In this paper, the hybrid GP/OLS technique is utilized to derive new linear-in-parameters GMPEs. GP/OLS is useful in deriving prediction equations for PGA, PGV and PGD by directly extracting the knowledge contained in the experimental data. The predictor variables included in the analysis were earthquake magnitude, earthquake source to site distance, average shear-wave velocity, and types of faulting mechanisms (strike-slip, normal, and reverse). The performance of the developed GMPEs were compared with that of the previously published empirical models (e.g., Sadigh and Egan, 1998; Nowroozi, 2005; Campbell and Bozorgnia, 2007; Rajabi et al., 2010). The GMPEs were developed based on a comprehensive database of the strong ground-motions assembled by PEER (Power et al., 2006).

2. Genetic programming

GP is a modern regression technique with a great ability to automatically evolve computer programs. The evolutionary process followed by the GP algorithm is inspired from the principle of Darwinian natural selection. GP was introduced by Koza (1992) after experiments on symbolic regression. This classical GP technique is also called tree-based GP (Koza, 1992; Alavi et al., 2011). The main difference between the GA and GP approaches is that in GP the evolving programs (individuals) are parse trees rather than fixed-length binary strings. The traditional optimization techniques, like GA, are generally used in parameter optimization to evolve the best values for a given set of model parameters. GP, on the other hand, gives the basic structure of the approximation model together with the values of its parameters. GP optimizes a population of computer programs according to a fitness landscape determined by a program ability to perform a given computational task. GP is relatively a new field of pattern recognition methods in contrast with GA. A survey of the literature reveals the growing interest of the research community in GP (Alavi et al., 2011).

In GP, a random population of individuals (trees) is created to achieve high diversity. The symbolic optimization algorithms present the potential solutions by structural ordering of several symbols (Gandomi et al., 2010a). A population member in GP is a hierarchically structured tree comprising functions and terminals. The functions and terminals are selected from a set of functions and a set of terminals. For example, the function set F can contain the basic arithmetic operations ($+$, $-$, $\times$, /, etc.), Boolean logic functions (AND, OR, NOT, etc.), or any other mathematical functions. The terminal set T contains the arguments for the functions and can consist of numerical constants, logical constants, variables, etc. The functions and terminals are chosen at random and constructed together to form a computer model in a tree-like structure with a root point with branches extending from each function and ending in a terminal. An example of a simple tree representation of a GP model is illustrated in Fig. 1 (Gandomi et al., 2010a).

The creation of the initial population is a blind random search for solutions in the large space of possible solutions. Once a population of models has been created at random, the GP algorithm evaluates the individuals, selects individuals for reproduction, generates new individuals by mutation, crossover, and direct reproduction, and finally creates the new generation in all

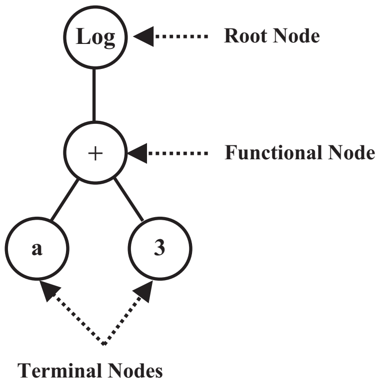

Fig. 1. The tree representation of a GP model (Log(a+3)).

iterations (Koza, 1992; Gandomi et al., 2010a). During the crossover procedure, a point on a branch of each solution (program) is selected at random and the set of terminals and/or functions from each program are then swapped to create two new programs as can be seen in Fig. 2 (Gandomi et al., 2010a). The evolutionary process continues by evaluating the fitness of the new population and starting a new round of reproduction and crossover. During this process, the GP algorithm occasionally selects a function or terminal from a model at random and mutates it (see Fig. 3). The best program that appeared in any generation, the best-so-far solution, defines the output of the GP algorithm (Koza, 1992). In the following subsections, the coupled algorithm of GP and OLS, GP/OLS, is briefly described.

2.1. Genetic programming for linear-in-parameters models

In general, GP creates not only nonlinear models but also linear-in-parameters models. In order to avoid parameter models, the parameters must be removed from the set of terminals. That is, it contains only variables: $T=\{x_0(k), \ldots, x_i(k)\}$, where $x_i(k)$ denotes the ith repressor variable. Hence, a population member represents only F_i nonlinear functions (Pearson, 2003). The parameters are assigned to the model after "extracting" the F_i function terms from the tree, and determined using a least squares (LS) algorithm (Reeves, 1997). A simple technique for the decomposition of the tree into function terms can be used. The sub-trees, representing the F_i function terms, are determined by decomposing the tree starting from the root as far as reaching nonlinear nodes (nodes which are not "+" or "−"). As can be seen in Fig. 4 (Gandomi et al., 2010a), the root node is a "+" operator; therefore, it is possible to decompose the tree into two sub-trees of "A" and "B". The root node of the "A" tree is again a linear operator; therefore, it can be decomposed into "C" and "D" trees. As the root node of the "B" tree is a nonlinear node (/), it cannot be decomposed. The root nodes of "C" and "D" trees are also nonlinear. Consequently, the final decomposition procedure results in three sub-trees: "B", "C", and "D". According to the results of the decomposition, it is possible to assign parameters to the functional terms represented by the obtained sub-trees. The resulted linear-in-parameters model for this example is y: $p_0+p_1(x_2+x_1)/x_0+p_2x_0+p_3x_1$ (Gandomi et al., 2010a).

GP can be used for selecting from special model classes, such as a polynomial model. To achieve it, the set of operators must be restricted and some simple syntactic rules must be introduced. For instance, if the set of operators is defined as $F=\{\times, +\}$ and there is a syntactic rule that exchanges the internal nodes that are below a "$\times$"-type internal nodes to "$\times$"-type nodes, GP will generate only polynomial models (Madár et al., 2004, 2005; Gandomi et al., 2010a).

2.2. Orthogonal least squares algorithm

The great advantage of using linear-in-parameter models is that the LS method can be used for identifying the model parameters. This is much less computationally demanding than other nonlinear optimization algorithms, because the optimal $p=[p_1, \ldots, p_m]^{\mathrm{T}}$ parameter vector can analytically be calculated:

$$p=(U^{-1}U)^{\mathrm{T}}U_y \tag{1}$$

in which $y=[y(1), \ldots, y(N)]^{\mathrm{T}}$ is the measured output vector and the U regression matrix is

$$U=\begin{pmatrix} U_1(x(1)) & \cdots & U_M(x(1)) \\ \vdots & \ddots & \vdots \\ U_1(x(N)) & \cdots & U_M(x(N)) \end{pmatrix} \tag{2}$$

The OLS algorithm (Billings et al., 1988) is an effective algorithm for determining which terms are significant in a linear-in-parameters model. The OLS technique introduces the error reduction ratio (err), which is a measure of the decrease in the variance of output by a given term. The matrix form corresponding to the linear-in-parameters model is

$$y=U_p+e \tag{3}$$

where the U is the regression matrix, p is the parameter vector, and e is the error vector. The OLS method transforms the columns of the U matrix into a set of orthogonal basis vectors to inspect the individual contributions of each term (Cao et al., 1999; Gandomi et al., 2010a). It is assumed in the OLS algorithm that the regression matrix U can be orthogonally decomposed as $U=WA$, where A is an $M\times M$ upper triangular matrix (i.e., $A_{ij}=0$ if $i>j$). W is an $N\times M$ matrix with orthogonal columns in the sense that $WTW=D$ is a diagonal matrix (N is the length of the y vector and M is the number of repressors). After this decomposition, the OLS auxiliary parameter vector g can be calculate as

$$g=D^{-1}W^{\mathrm{T}}y \tag{4}$$

where g_i represents the corresponding element of the OLS solution vector. The output variance $(y^{\mathrm{T}}y)/N$ can be described as

$$y^{\mathrm{T}}y=\sum_{i=1}^{M} g_i^2 w_i^{\mathrm{T}} w_i+e^{\mathrm{T}}e \tag{5}$$

Therefore, the error reduction ratio $[\mathrm{err}]_i$ of the U_i term can be expressed as

$$[\mathrm{err}]_i=\frac{g_i^2 w_i^{\mathrm{T}} w}{y^T y} \tag{6}$$

This ratio offers a simple mean for order and selects the model terms of a linear-in-parameters model on the basis of their contribution to the performance of the model.

2.3. Hybrid genetic programming-orthogonal least squares algorithm (GP/OLS)

The application of OLS to the GP algorithm leads to significant improvements in the performance of GP. The main feature of this hybrid approach is to transform the trees to simpler trees which are more transparent, but their accuracies are close to the original trees (Gandomi et al., 2010a). In this coupled algorithm, GP generates a lot of potential solutions in the form of a tree structure during the GP operation. These trees may have better and worse terms (sub-trees) that contribute more or less to the accuracy of the model represented by the tree. OLS is used to estimate the contribution of the branches of the tree to the accuracy of the model, whereas, using the OLS, one can select the less significant terms in a linear regression problem.

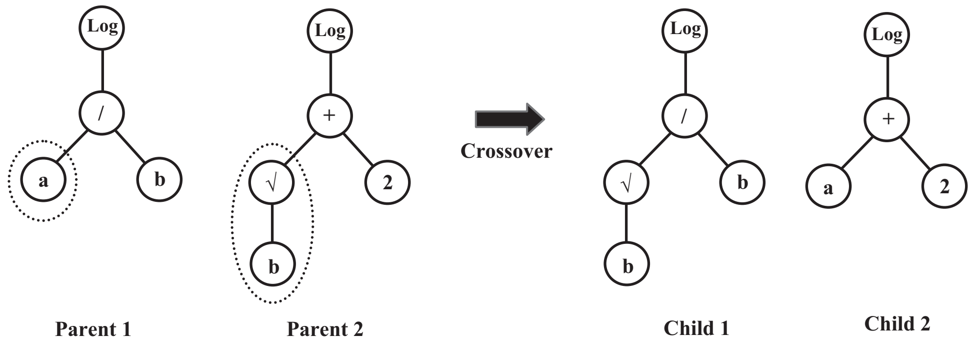

Fig. 2. Typical crossover operation in GP.

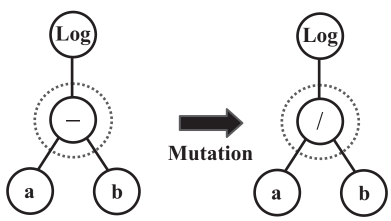

Fig. 3. Typical mutation operation in GP.

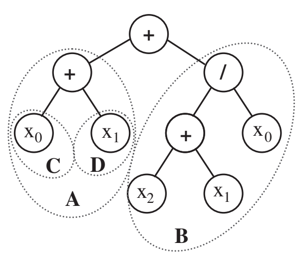

Fig. 4. Decomposition of a tree to function terms (Madár et al., 2004).

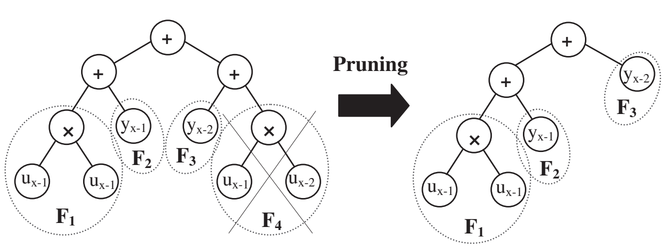

Fig. 5. Pruning of a tree with OLS.

According to this strategy, terms (sub-trees) having the smallest error reduction ratio are eliminated from the tree (Pearson, 2003). This "tree pruning" approach is realized in every fitness evaluation before the calculation of the fitness values of the trees. Since GP works with the tree structure, the further goal is to preserve the original structure of the trees as far as it possible. The GP/OLS method always guarantees that the elimination of one or more function terms of the model can be done by pruning the corresponding sub-trees, so there is no need for structural rearrangement of the tree after this operation. The way the GP/OLS method works on its basis is simply demonstrated in Fig. 5 (Gandomi et al., 2010a). Assume that the function which must be identified is $y(x)=0.8(u_{x-1})^2+1.2y_{x-1}-0.9y_{x-2}-0.2$. As can be seen in Fig. 5, the GP algorithm finds a solution with four terms: $(u_{x-1})^2$, y_{x-1}, y_{x-2}, $u_{x-1}\times u_{x-2}$. Based on the OLS algorithm, the sub-tree with the least error reduction ratio ($F_4=u_{x-1}\times u_{x-2}$) is eliminated from the tree. Subsequently, the error reduction ratios and mean square error values (and model parameters) are calculated again. The new model (after pruning) may have a higher mean square error but it obviously has a more adequate structure (Gandomi et al., 2010a).

3. Modeling of time-domain strong ground-motion characteristics

The damage potential of earthquakes is dependent on the ground-motion characteristics and local site condition. Amplitude, frequency content and duration of motion are the significant characteristics of earthquake motion (Kramer, 1996). The ground-motion parameters have essential roles in explaining the characteristics of strong ground-motions. PGA, PGV and PGD are the commonly used time-domain parameters of ground-motions. These parameters are frequently presented in quantitative forms as functions of different seismic variables. The most significant seismological aspects that influence the ground-motion parameters are the source effect, path effect and site effect. The source effect is related to numerous parameters such as the level of stress drop in the earthquake event, mechanism of faulting, and direction of faulting. The influence of these parameters is not similar and strongly depends on the distance of the desired site from the fault. Especially, for relatively distant sites, it is not needed to consider the complex form of the faulting procedure (e.g., the hanging-wall effect). The path effect is related to the distance of the site from the fault. Different definitions are proposed for the distance in the literature (e.g., closest distance, Joyner–Boore distance, etc.) in different GMPEs (Boore and Atkinson, 2007). The site effect is also reflected in the attenuation relations as a significant element. Some of the models consider the site effect in a generic way (i.e., soil or rock) and the others uses the soil shear velocity as an indicator of the site effects. However, the modern GMPEs are mainly in terms of the earthquake magnitude, source to site distance, geotechnical condition of site, and faulting mechanism (e.g., Boore et al., 1997; Boore and Atkinson, 2007; Campbell and Bozorgnia, 2007). Other affecting physical parameters such as stress drop, rupture propagation, directivity, and nonlinear soil behavior significantly reflect the uncertainties and therefore, they are not generally considered in the development of the predictive equations (Somerville and Graves, 2003; Cabalar and Cevik, 2009).

To enhance the precision of the conventional regression-based analyses, several novel methods have recently been employed. The main purpose of this study is to derive alternative prediction equations for PGA, PGV and PGD using the GP/OLS approach. The most important factors representing the ground-motion parameters behavior were selected based on the literature review (e.g., Douglas, 2003; Ambraseys et al., 1996; Boore et al., 1997; Boore and Atkinson, 2007; Campbell and Bozorgnia, 2007; Güllü and Erçelebi, 2007; Boore and Atkinson, 2008; Cabalar and Cevik, 2009; Kermani et al., 2009; Rajabi et al., 2010). Consequently, the formulations of PGA (cm/s^2), PGV (cm/s) and PGD (cm) were considered to be as follows:

$$\ln(\mathrm{PGA}),\ \ln(\mathrm{PGV}),\ \ln(\mathrm{PGD}) = f(M_w,\ \ln(R_{jb}),\ \ln(V_{s30}), F) \quad (7)$$

where M_w is the earthquake magnitude (moment magnitude); R_{jb} (km) is the closest distance to the surface projection of the fault plane (Joyner–Boore distance). R_{jb} is approximately equal to the epicentral distance for events of $M_w < 6$ (Boore and Atkinson, 2007); V_{s30} (m/s) is the average shear-wave velocity over the top 30 m of site

F is the coefficient representing different fault types:

- Reverse (dip slip with hanging-wall side up).
- Normal (dip slip with hanging-wall side down).
- Strike-slip (horizontal slip).

The significant influence of the above predictor variables in determining PGA, PGV and PGD is well understood. Indicator variables representing the style of faulting are defined in terms of rake angle. The rake angle is described as the average angle of slip in degrees measured in the plane of rupture between the strike direction and the slip vector (Campbell and Bozorgnia, 2007). Majority of the existing GMPEs introduce empirically derived coefficients for different fault types rather than directly including the faulting mechanism categories in the model development (see Douglas, 2004; Campbell and Bozorgnia, 2007). A similar procedure is followed in this study. PGA, PGV and PGD were first formulated in terms of M_w, R_{jb}, and V_{s30} using GP/OLS. Thereafter, empirical coefficients for different faulting mechanisms were obtained using GAs.

3.1. Strong-motion database and data preprocessing

Source of the strong ground-motion data employed for the developing the GMPEs was the database compiled in the PEER-NGA project by Power et al. (2006) (NGA Flatfile V 7.3). The database is comprised of shallow crustal earthquakes recorded data at active tectonic regions of the world. The database covers a broad range of magnitude and distance. It is optional for researchers to use the entire database or to limit their analyses to selected subsets. In this study, a part of the database was excluded from the analysis considering some of the filtering strategies presented by Boore and Atkinson (2007). Data sets missing the required information and also the duplicate records were excluded from the analysis. Finally, out of the total of 3551 records, 2777 records for reverse (dip slip with hanging-wall side up), normal (dip slip with hanging-wall side down), and strike-slip (horizontal slip) fault types were employed for the model development. The predictor variables included in the analysis were M_w, R_{jb} (km), V_{s30} (m/s), and *F*. PGA, PGV and PGD were the ground-motion parameters to be formulated. Fig. 6 shows the distribution of the data used to develop the predictive equations. For more visualization, the data are presented by frequency histograms (Fig. 7). The ranges and statistics of different parameters involved in the modeling process are given in Table 1. As can be seen in this table, the ranges of the predictor variables is

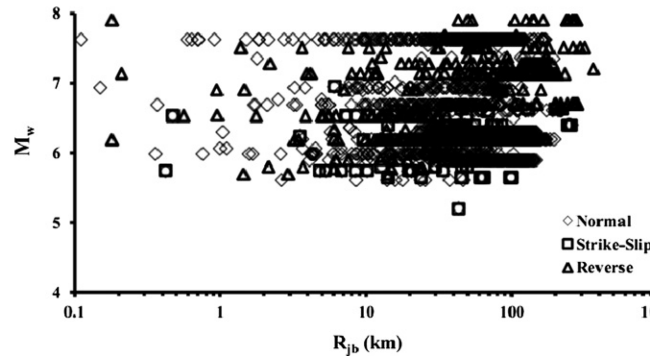

Fig. 6. Distribution of the data used for the model development, differentiated by fault type.

relatively wide, particularly for magnitude and distance. The wideness of the ranges has an important role in the GP-based seismic hazard analysis.

For the GP/OLS analyses, the data sets were randomly divided into training (learning and validation) and testing subsets. The learning data were taken for the training of the algorithm (genetic evolution). The validation data were used to specify the generalization capability of the models on data they did not train on (model selection). Thus, both of the learning and validation data were involved in the modeling process and were categorized into one group referred to as "training data" (Alavi et al., 2011). The testing data were finally employed to measure the performance of the models obtained by GP/OLS on the data that played no role in building the models. In order to obtain a consistent data division, several combinations of the training and testing sets were considered. The selection was such that the maximum, minimum, mean and standard deviation of parameters were consistent in the training and testing data sets. Of the 2777 data, 2361 data were used for the training process (1945 for learning data and 416 for validation) and 416 data sets were taken for the testing of the generalization capability of the models.

3.2. Performance measures

The best GMPEs were chosen on the basis of a multi-objective strategy as below

(i) The simplicity of the model, although this was not a predominant factor.
(ii) Providing the best fitness value on a learning set of data.
(iii) Providing the best fitness value on a validation set of data.

The first objective can be controlled by the user through the parameter settings (e.g., maximum program depth). For the other objectives, the following objective function (OBJ) was constructed as a measure of how well the model predicted output agrees with the measured output. The selections of the best models were deduced by the minimization of the following function:

$$\mathrm{OBJ} = \left(\frac{\mathrm{No}_{\mathrm{Learning}} - \mathrm{No}_{\mathrm{Validation}}}{\mathrm{No}_{\mathrm{Training}}}\right)\frac{\mathrm{RMSE}_{\mathrm{Learning}} + \mathrm{MAE}_{\mathrm{Learning}}}{R^2_{\mathrm{Learning}}} + \frac{2\mathrm{No}_{\mathrm{Validation}}}{\mathrm{No}_{\mathrm{Training}}}\frac{\mathrm{RMSE}_{\mathrm{Validation}} + \mathrm{MAE}_{\mathrm{Validation}}}{R^2_{\mathrm{Validation}}} \quad (8)$$

where $\mathrm{No}_{\mathrm{Learning}}$, $\mathrm{No}_{\mathrm{Validation}}$ and $\mathrm{No}_{\mathrm{Training}}$ are, respectively, the number of learning, validation and training data; *R*, RMSE and MAE are, respectively, correlation coefficient, root mean squared error and mean absolute error given in the form of formulas as

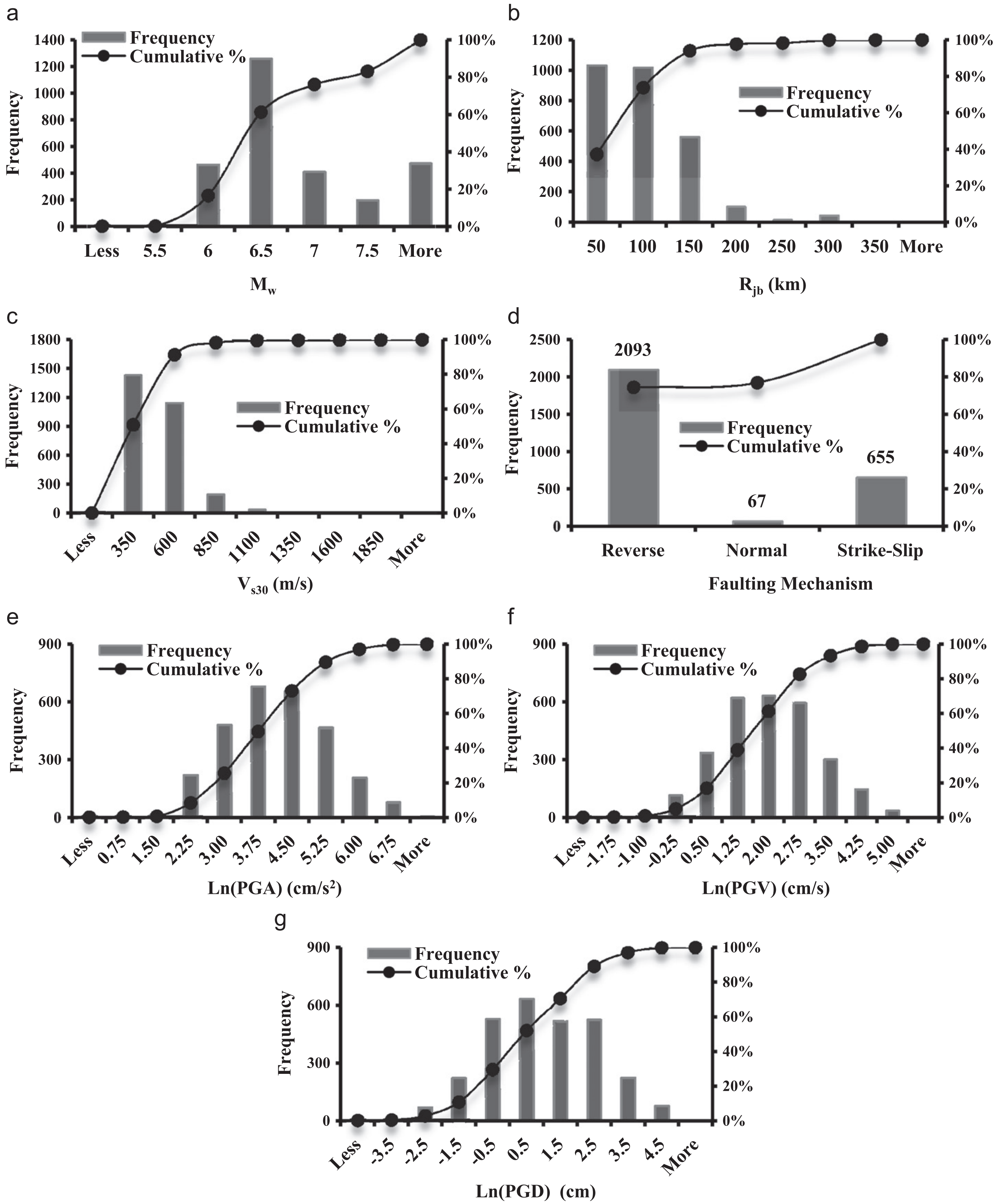

Fig. 7. Histograms of the input and output variables.

Table 1
Descriptive statistics of the variables used for the model development.

Parameter	M_w	R_{jb} (km)	V_{s30} (m/s)	F	PGA (cm/s^2)	PGV (cm/s)	PGD (cm)
Mean	6.55	73.34	386.07	–	78.84	9.38	4.93
Standard error	0.01	0.96	3.28	–	2.08	0.24	0.18
Median	6.30	63.49	345.42	1	42.19	4.89	1.45
Standard deviation	0.59	50.55	172.73	–	109.42	12.65	9.57
Sample variance	0.35	2555.43	29837.37	–	11972.19	160.09	91.66
Kurtosis	−0.70	3.18	10.23	–	33.88	14.47	62.32
Skewness	0.82	1.38	2.12	–	4.43	3.28	5.65
Range	2.70	365.11	1899.78	2	1628.83	117.04	188.30
Minimum	5.20	0.03	116.35	1	1.14	0.10	0.01
Maximum	7.90	365.14	2016.13	3	1629.97	117.14	188.32

follows:

$$R=\frac{\sum_{i=1}^{n}(h_i-\overline{h}_i)(t_i-\overline{t}_i)}{\sqrt{\sum_{i=1}^{n}(h_i-\overline{h}_i)^2\sum_{i=1}^{n}(t_i-\overline{t}_i)^2}} \quad (9)$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(h_i-t_i)^2}{n}} \quad (10)$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|h_i-t_i| \quad (11)$$

in which h_i and t_i are, respectively, the actual and predicted output values for the ith output, $\overline{h}_i$ is the average of the actual outputs, and n is the number of sample. It is well-known that the R value alone is not a good indicator of prediction accuracy of a model. This is because that by shifting the output values of a model equally, the R value will not change. The constructed objective function takes into account the changes of R, RMSE and MAE together. Higher R values and lower RMSE and MAE values result in lower OBJ and, consequently, indicate a more precise model. In addition, the above function considers the effects of different data divisions for the learning and validation data.

3.3. Model development using GP/OLS

The available database was used for generating the PGA, PGV and PGD prediction equations. Various parameters are involved in the GP/OLS predictive algorithm. The parameter selection significantly affects the generalization capability of the derived GP/OLS-based GMPEs. The parameter settings for the GP/OLS algorithm are shown in Table 2. In order to obtain simple and straightforward formulas, four basic arithmetic operators (+, −, ×, /) were utilized in the analysis. The number of programs in the population that GP/OLS will evolve is set by the population size. A run will take longer with a larger population size. The number of generation sets the number of levels the algorithm will use before the run terminates. The proper numbers of population and generation depend on the complexity of the problem. In this study, a fairly large number of initial population and generations were tested to find models with minimum error. The program was run until the runs terminated automatically. Mutation and crossover rate are the probabilities that an offspring will be subject to the mutation and crossover operations. The values of both of these parameters for the optimal models were 50%. The maximum tree depth directly influences the size of the search space and the number of solutions explored within the search space. The success of the GP/OLS algorithm usually increases with increasing these parameters. In this case, the complexity of the evolved function

Table 2
Parameter settings for the GP/OLS algorithm.

Parameter	Settings
Function set	+, −, ×, /
Population size	500–1000
Maximum tree depth	12
Maximum number of evaluated individuals	250
Maximum number of evaluated individuals	250
Generation	100
Type of selection	Roulette-wheel
Type of mutation	Point-mutation
Type of crossover	One-point (2 parents)
Type of replacement	Elitist
Probability of crossover	0.5
Probability of mutation	0.5
Probability of changing terminal–non-terminal nodes (vice versa) during mutation	0.25

increases and the speed of the algorithm decreases. The maximum tree depth was set to an optimal value of 12 as a tradeoff between the running time and the complexity of the evolved solutions. The other involved parameters values were selected based on some previously suggested values (Gandomi et al., 2010a; Gandomi and Alavi, 2010; Madár et al., 2004) and also after a trial and error approach.

3.3.1. GP/OLS-based ground-motion prediction equations

The peak ground acceleration (PGA), velocity (PGV) and displacement (PGD) prediction equations, for the best results by the GP/OLS algorithm, are as given below

$$\ln(\mathrm{PGA})(\mathrm{cm}/s^2)=7.673-1.28(\ln(V_{s30})^{(\ln(R_{jb}))/M_w})+F_{\mathrm{PGA}} \quad (12)$$

$$\ln(\mathrm{PGV})(\mathrm{cm}/s)=-3.37+1.059M_w-0.02(\ln(R_{jb}) +\ln(R_{jb})^2\ln(V_{s30}))+F_{\mathrm{PGV}} \quad (13)$$

$$\ln(\mathrm{PGD})(\mathrm{cm})=-1.973+1.917M_w+7.82\frac{\ln(R_{jb})}{\ln(V_{s30})}+F_{\mathrm{PGD}} \quad (14)$$

where M_w, R_{jb}, and V_{s30}, respectively, denote the earthquake magnitude, earthquake source to site distance, and average shear-wave velocity. F_{PGA}, F_{PGV}, and F_{PGD} are the empirical coefficients derived for different fault types. These coefficients are presented in Table 3. Comparisons of the measured and predicted PGA, PGV and PGD values are, respectively, shown in Figs. 8–10.

4. Performance analysis and model validity

Based on a logical hypothesis (Smith, 1986), if a model gives $R>0.8$, and the error values (e.g., RMSE and MAE) are at the minimum, there is a strong correlation between the predicted and measured values. The model can therefore be judged as very good. It can be observed from Figs. 8–10 that the GP/OLS models with high R and low RMSE and MAE values are able to predict the target values to an acceptable degree of accuracy. Meanwhile, it is noteworthy that the RMSE and MAE values are not only low but also as similar as possible for the training and testing sets. This suggests that the proposed models have both predictive ability (low values) and generalization performance (similar values) (Pan et al., 2009).

It is known that the models derived using the ANNs, GP, or other soft computing tools, in most cases, have a predictive capability within the data range used for their development. This is because of the nature of these techniques which distinguishes them from the other conventional techniques. Thus, the amount of data used for the modeling process is an important issue, as it bears heavily on the reliability of the final models. To cope with this limitation, Frank and Todeschini (1994) argue that the minimum ratio of the number of objects over the number of selected variables for model acceptability is 3. They also suggest that considering a higher ratio equal to 5 would be safer. In the present study, this ratio is much higher and is equal to 2777/3 = 925.7. The PGD prediction results are slightly better than those of PGA and PGV. Furthermore, new criteria recommended by Golbraikh and Tropsha (2002) were checked for the external validation of the models on the test data sets. It is suggested that at least one slope of regression lines (k or k') through the origin should be close to 1. Also, the performance indexes of m and n should be lower than 0.1. Either the squared correlation coefficient (through the origin) between predicted and experimental values (R_o^2), or the coefficient between experimental and predicted values (R'^2_o) should be close to 1 (Alavi et al., 2011). The considered validation criteria and the relevant results obtained by the models are presented in Table 4. As it is seen, the proposed models satisfy the required conditions. This confirms that the derived GMPEs are strongly valid, have the prediction power and are not chance correlations.

Table 3
The coefficients derived for different fault types.

Fault type	F		
	PGA	PGV	PGD
Reverse	0.046	0.001	−0.037
Normal	−0.059	−0.371	−0.372
Strike-Slip	−0.101	0.080	0.236

5. Sensitivity and parametric analyses

Sensitivity analysis is of utmost concern for selecting the important input variables. The contribution of each predictor variable in the GP/OLS-based GMPEs was evaluated through a sensitivity analysis. For this aim, frequency values of the input variables were obtained. A frequency value equal to 1.00 for an input indicates that this variable has been appeared in 100% of the best thirty programs evolved by GP/OLS. This methodology is a common approach in the GP-based analyses (Gandomi et al., 2010a, 2011). The frequency values of the predictor variables are presented in Fig. 11. According to these results, it can be found that the ground-motion parameters are more sensitive to M_w followed by R_{jb} and V_{s30}. These are expected cases and go arm in arm with the ground-motions behavior.

A parametric analysis was also performed in this study to verify the robustness of the GP/OLS prediction equations. The parametric analysis investigates the response of the predicted ground-motion parameters from the GMPEs to a set of hypothetical input data. The methodology is based on the change of only one predictor variable at a time while the other seismic variables are kept constant at the average values of their entire data sets. A set of synthetic data for the single varied parameter is generated by increasing the value of this in increments. These variables are presented to the prediction equations and PGA, PGV and PGD are calculated. This procedure is repeated using another variable until the model response is tested for all input variables. The robustness of the design equations is determined by examining how well the predicted PGA, PGV and PGD values agree with the underlying physical behavior of the investigated system (Kuo et al., 2009). Figs. 12–14, respectively, present the tendency of the PGA, PGV and PGD predictions to the variations of the seismic parameters, M_w, R_{jb}, and V_{s30}. The results of the parametric analysis indicate that PGA, PGV and PGD continuously increase due to increasing M_w and decrease with increasing R_{jb} and V_{s30}. The parametric analysis results are soundly expected cases from a seismological viewpoint (Ambraseys et al., 1996;

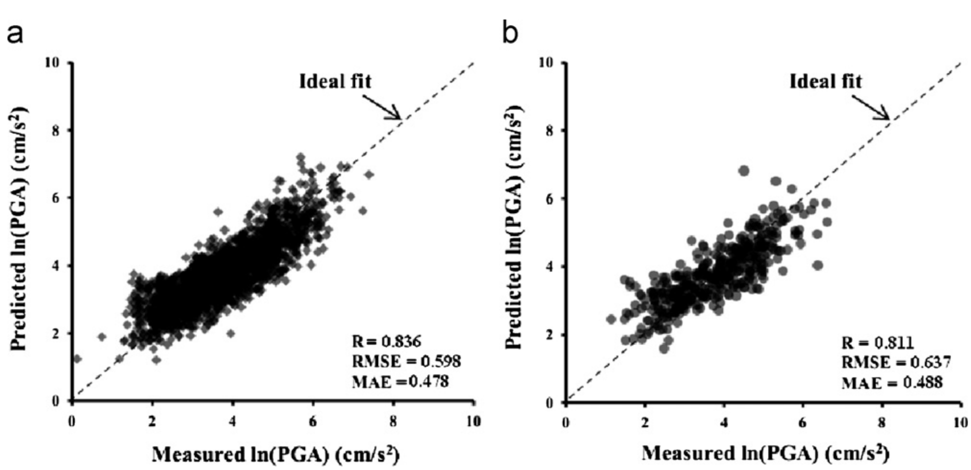

Fig. 8. Measured versus predicted PGA values using the GP/OLS model: (a) training data and (b) testing data.

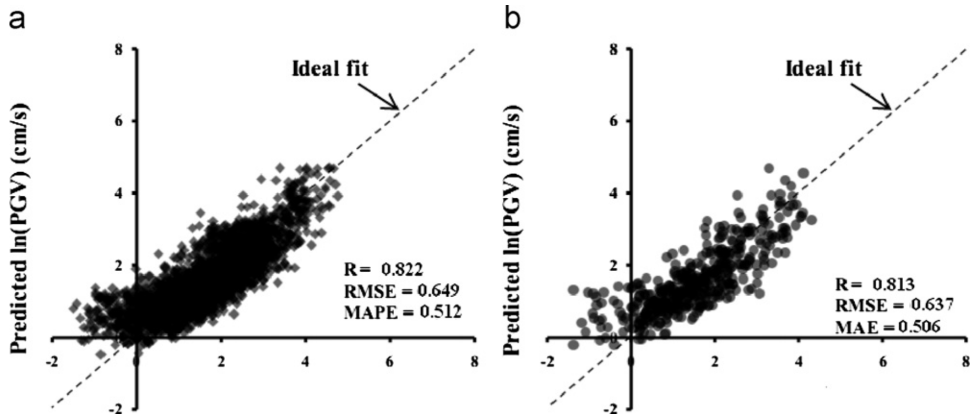

Fig. 9. Measured versus predicted PGV values using the GP/OLS model: (a) training data and (b) testing data.

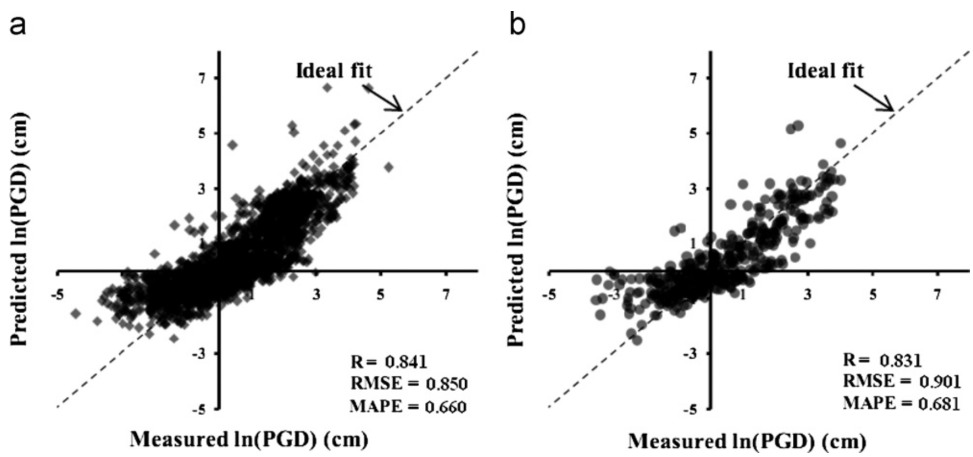

Fig. 10. Measured versus predicted PGD values using the GP/OLS model: (a) training data and (b) testing data.

Table 4
Statistical parameters of the GP/OLS models.

Item	Formula	Condition	PGA	PGV	PGD
1	R	$0.8 < R$	0.811	0.813	0.831
2	$k = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{h_i^2}$	$0.85 < k < 1.15$	0.999	0.901	1.002
3	$k' = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{t_i^2}$	$0.85 < k' < 1.15$	0.975	0.978	0.702
4	$m = \frac{R^2 - R_o{}^2}{R^2}$	$m < 0.1$	-0.519	-0.453	-0.448
5	$n = \frac{R^2 - R_o'^2}{R^2}$	$n < 0.1$	-0.500	-0.511	-0.307
where	$R_o^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - h_i^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t}_i)^2}$ and $h_i^o = k \times t_i$		1.000	0.960	1.000
	$R_o'^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - t_i^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h}_i)^2}$ and $t_i^o = k' \times h_i$		0.987	0.999	0.903

Boore et al., 1997; Boore and Atkinson, 2007; Campbell and Bozorgnia, 2007). The above results confirm that the proposed design equations are robust and can confidently be used for predictive purposes in seismic hazard studies.

6. Comparison of the ground-motion prediction equations

A comparison of the GP/OLS-based GMPEs developed in this study with all of the existing ground-motion models is beyond the scope of this study. The results obtained in this research are compared with those provided by the well-known models of Sadigh and Egan (1998)

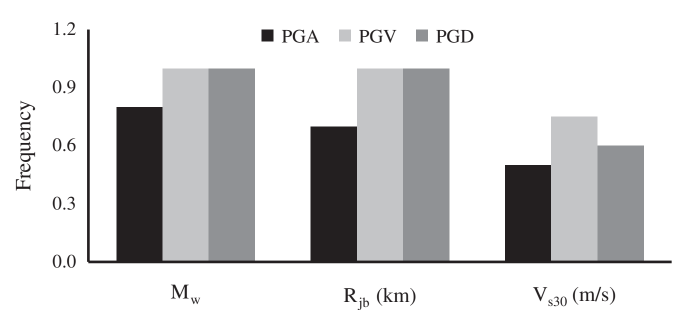

Fig. 11. Contributions of the predictor variables in the GP/OLS-based GMPEs.

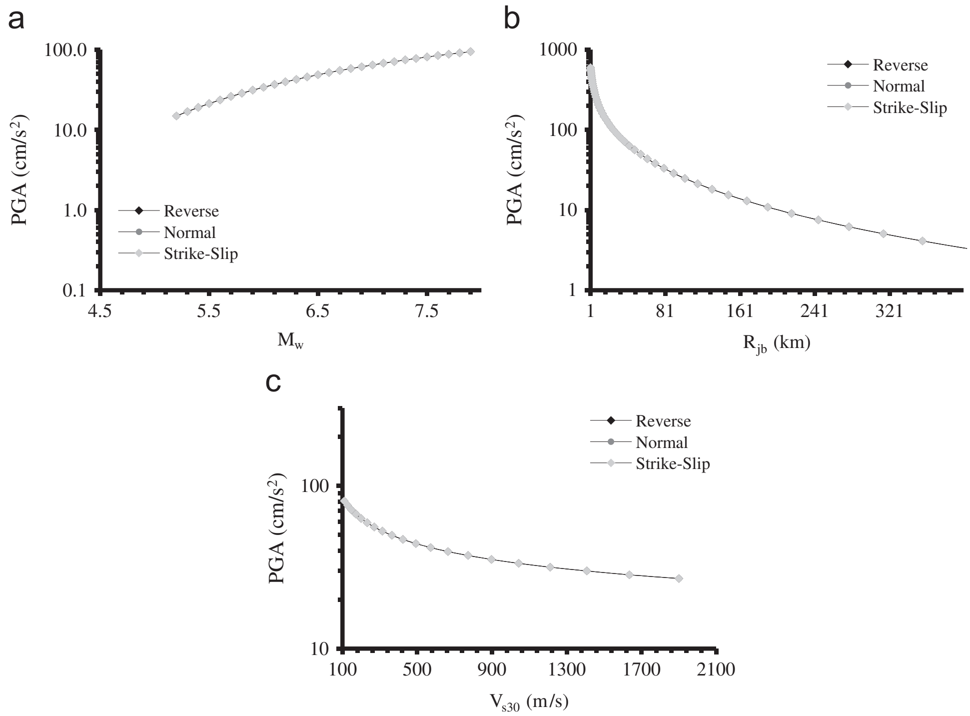

Fig. 12. Parametric analysis of PGA in the GP/OLS-based GMPE.

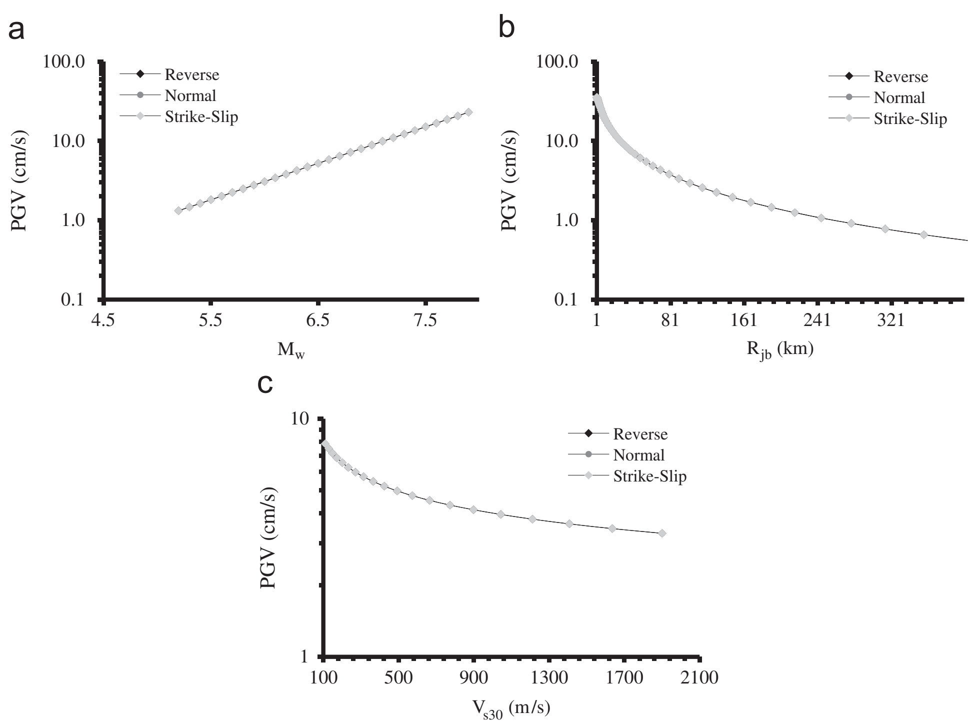

Fig. 13. Parametric analysis of PGV in the GP/OLS-based GMPE.

for the prediction of PGA, PGV and PGD. As a part of the PEER Next Generation Attenuation of Ground Motion (NGA) Project, Campbell and Bozorgnia (2007) and Boore and Atkinson (2008) developed updated prediction equations for PGA, PGV and PGD. The results of these recent researches have also been included in the comparative study. It is notable that the Boore and Atkinson's models were established for the PGA and PGV prediction. The performance of the models was evaluated on a well-known database provided by Haselton and Deierlein (2007). The database contains 78 records for different earthquakes. A reduced subset of these records has been used in the Applied Technology Council Project 63 (ATC, 2008), which is focused on developing a procedure to validate seismic provisions for structural design. The performance statistics of the PGA, PGV and PGD prediction models are visualized in Figs. 15–17. As can be observed from Fig. 15, the proposed prediction equation for PGA is able to reach a prediction performance comparable with the

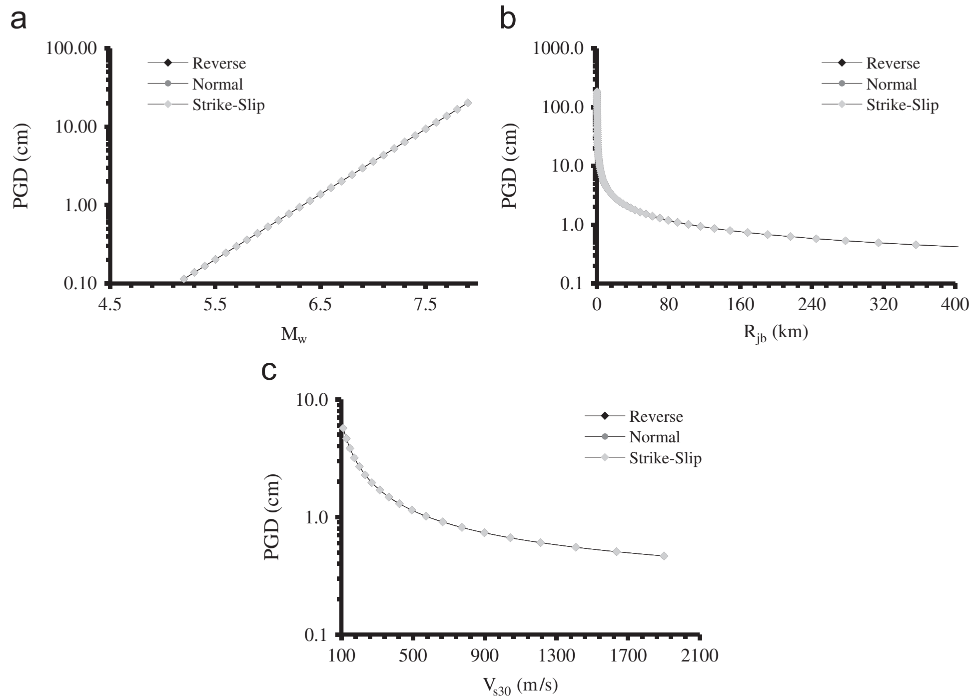

Fig. 14. Parametric analysis of PGD in the GP/OLS-based GMPE.

NGA's GMPEs and the model proposed by Sadigh and Egan (1998). Considering the results for the PGV prediction, it can be seen in Fig. 16 that the GP/OLS-based GMPE with RMSE and MAE values equal to 0.565 and 0.455 performs superior than the other models. As shown in Fig. 17, the performance of the proposed PGD prediction equation (RMSE=0.696, MAE=0.570) is considerably better than the NGA's GMPEs and Sadigh and Egan's model.

Despite the good performance the NGA models, they are complicated equations with long linear or nonlinear terms. These models include several predictor (independent) variables. The involved variables are moment magnitude, one or more of the fault distance measures, indicator variables for style of faulting, hanging-wall parameters, shear-wave velocity, one or more of the sediment depth parameters, and the depth to the top of coseismic rupture. On the other hand, the high-precision GMPEs obtained by means of GP/OLS are significantly short and straightforward. This is mainly because of the important role of the tree pruning process in the GP/OLS algorithm. Note that the GP/OLS equations are developed using only four predictor variables (M_w, R_{jb}, V_{s30} and F) and therefore they can easily be used via hand calculations. As more data become available, including those for other main-shocks or up-to-date NGA strong-motion data, these GMPEs can be improved to make more accurate predictions for a wider range. Also, the user physical insight and the shape of the classical GMPEs can be regarded in making propositions on the elements and structure of the evolved functions.

Most of the existing GMPEs are derived by performing the multiple regression analysis (e.g., the NGA models). The significant limitations of the conventional regression analysis have previously described. Such models often assume linear, or in some cases nonlinear, relationships between the ground-motion parameters and the predictor variables, which is not always true. In most cases, the best models developed using the commonly used regression approach are obtained after controlling just some equations established in advance. Thus, they cannot efficiently consider the interactions between the dependent and independent variables. One of the major advantages of the GP/OLS approach over the traditional regression analysis is its ability to derive explicit relationships for PGA, PGV and PGD without assuming prior forms of the existing relationships. The best solutions (equations) evolved by this technique are determined after controlling numerous preliminary models, even millions of linear and nonlinear models (Alavi et al., 2011).

However, it is notable that the GP-based methods are extremely parameter sensitive, especially when difficult experimental training data sets like the one used in this paper are employed. Using any form of optimally controlling the parameters of the run (e.g., GAs), can significantly improve the performance of their algorithms. Also, one of the main goals of introducing expert systems, such as the GP-based approaches, into the design processes is better handling of the information in the pre-design phase. The initial steps of design are based on imprecise and incomplete information about the features and properties of targeted output or process (Kraslawski et al., 1999; Alavi et al., 2011). Nevertheless, it is idealistic to have some initial estimates of the outcome before performing any extensive laboratory or field work. The GP/OLS approach employed in this research is based on the data alone to determine the structure and parameters of the model. Thus, the derived GMPEs are considered to be mostly valid for use in preliminary design stages and should cautiously be used for final decision-making. For more reliability, the results of the GP/OLS-based analyses are suggested to be compared with those obtained using deterministic methods.

7. Validity verification

To further verify the validity of the GP/OLS-based GMPEs, they were employed to predict the ground-motion parameters for additional records from the Iranian plateau earthquakes. The verification data covers a fairly wide range of magnitudes and distances for rock and soil sites in Iran. Comparisons of the PGA, PGV and PGD predictions made by the GP/OLS models and attenuation models developed by Zare et al. (1999), Khademi (2002), Nowroozi (2005), and Rajabi et al. (2010) for Iran are presented in Fig. 18. As can be seen in these figures, the proposed GMPEs give a sound basis for simulation of the ground-motions parameters for earthquakes of

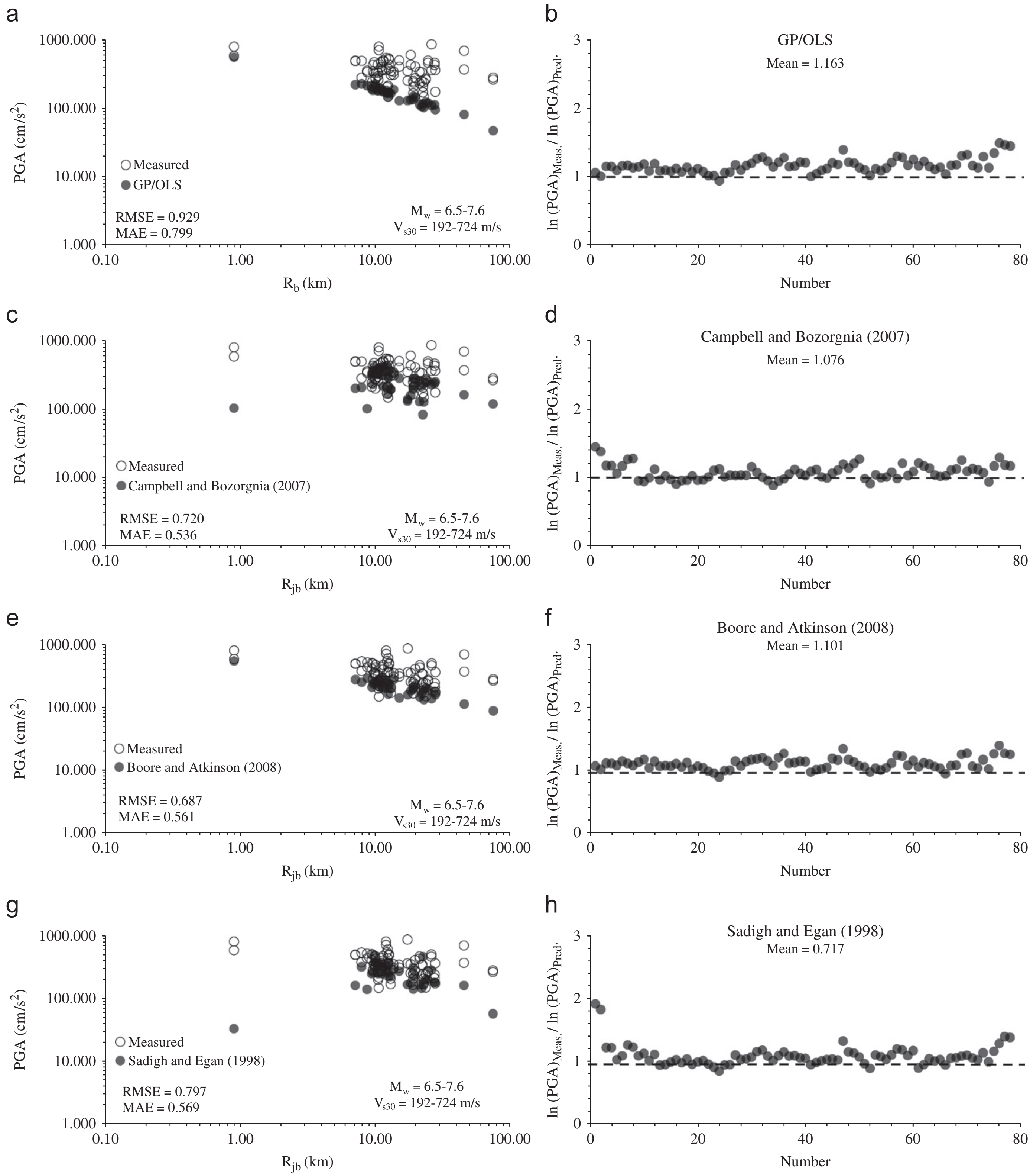

Fig. 15. Comparison of the PGA predictions made by different models.

magnitude 5.0–7.7, and distances ranging from 0–123 km. The PGD prediction equation provides the best performance followed by the PGV and PGA equations. The GP/OLS model for PGA prediction remarkably outperforms the previously published empirical models which are specifically developed for the Iranian earthquakes.

8. Summary and conclusion

In this research, high-precision GMPEs were derived using the hybrid GP/OLS method. The proposed GMPEs were developed based on an extensive database of thousands of records compiled

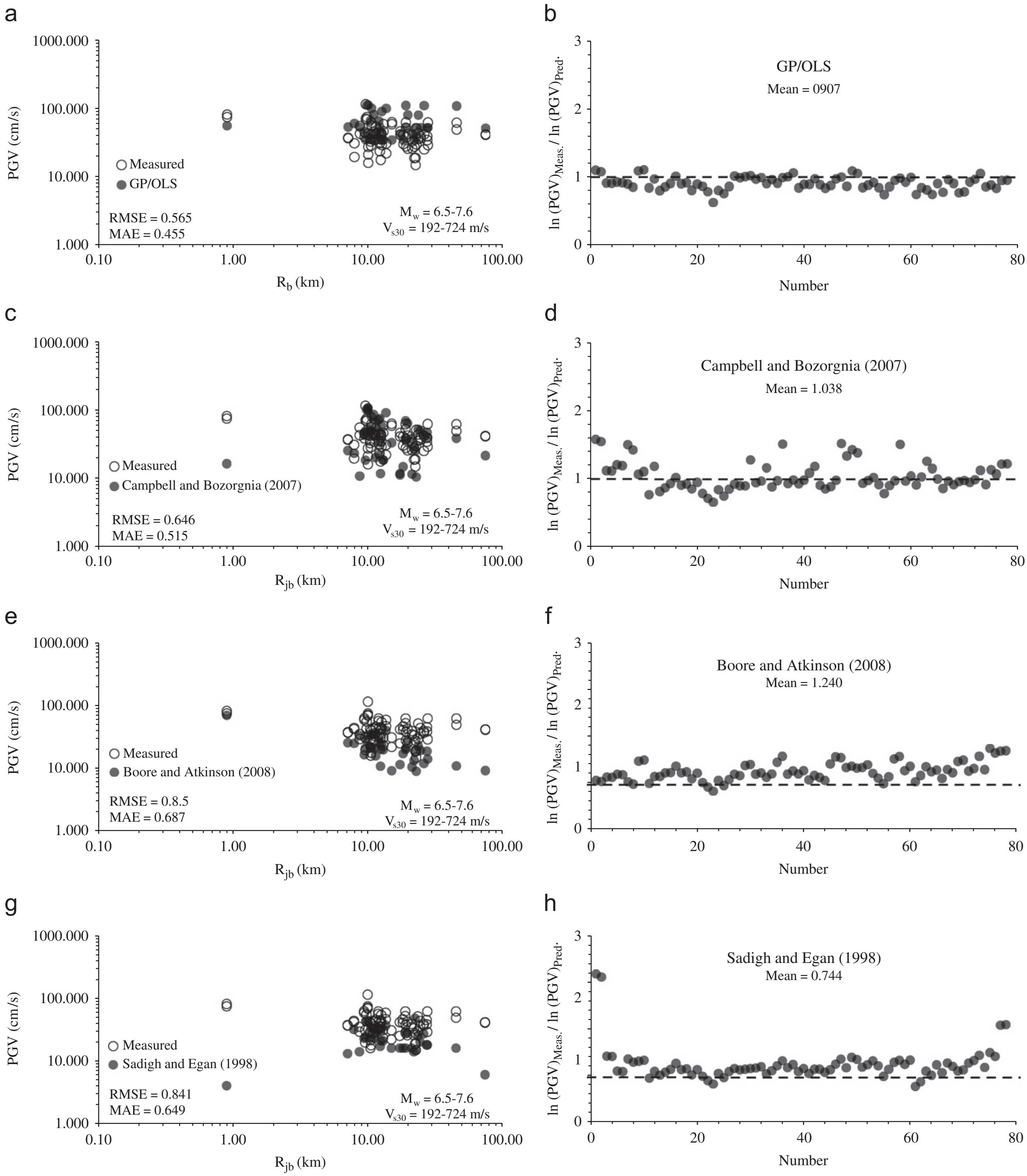

Fig. 16. Comparison of the PGV predictions made by different models.

in the PEER-NGA project. The following conclusions may be drawn based on the results presented:

(i) The developed GMPEs give reliable estimates of the PGA, PGV and PGD values. The PGD prediction equation provides slightly better results compared with the PGA and PGV models. The models efficiently satisfy the conditions of different criteria considered for their external validation.
(ii) M_w, R_{jb}, V_{s30}, and F were found to be efficient representative of behavior of the strong ground-motion parameters.

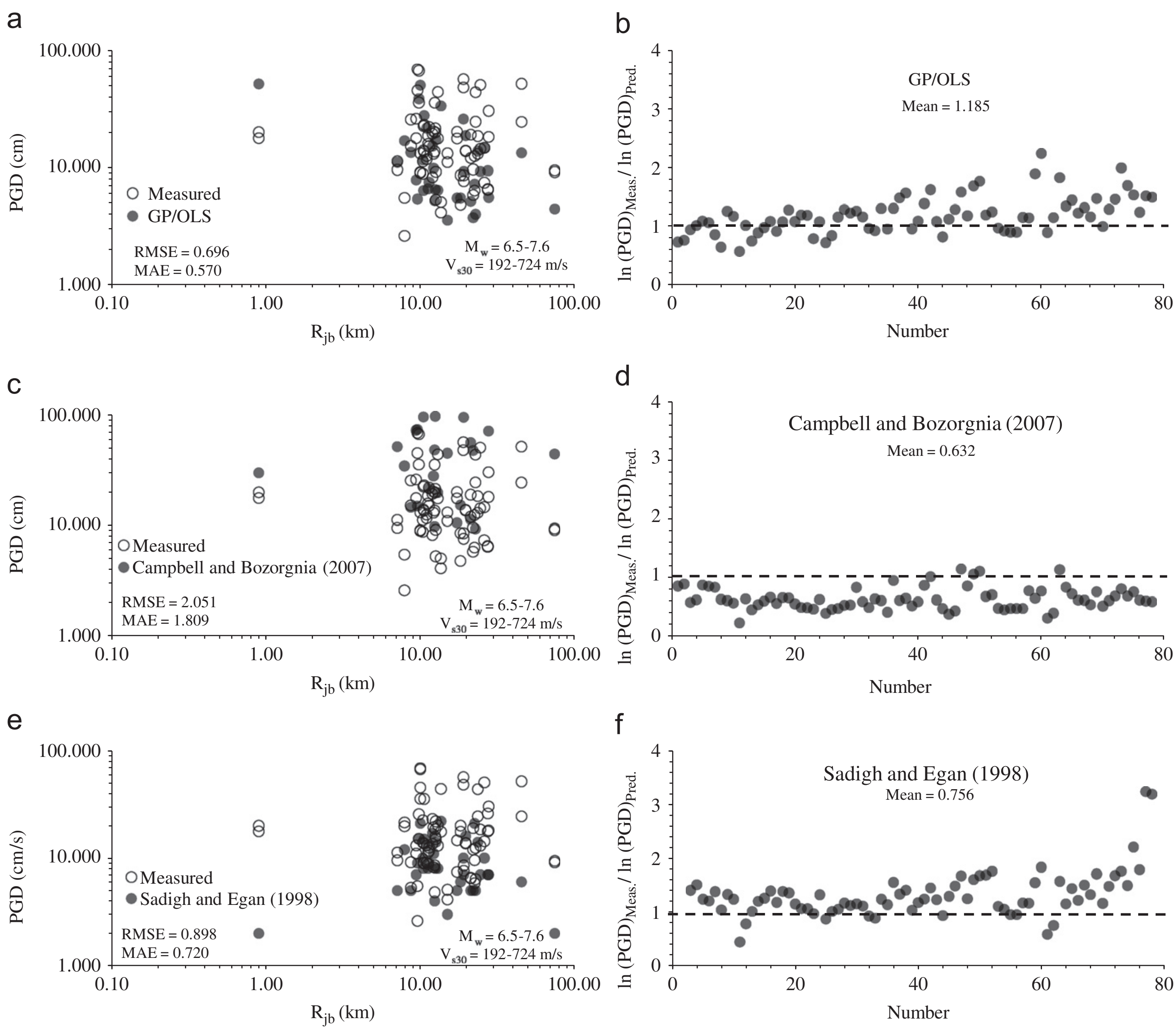

Fig. 17. Comparison of the PGD predictions made by different models.

(iii) The sensitivity and parametric analyses guarantee that the proposed GMPEs efficaciously take into account the underlying physical relations governing the system. The sensitivity and parametric results clearly indicate that the derived GMPEs are not mere combinations of the predictor variables which best fit the experimental results.

(iv) Further verification was done by benchmarking the proposed equations against the GMPEs proposed by Sadigh and Egan (1998), Campbell and Bozorgnia (2007) and Boore and Atkinson (2008). The nonlinear GP/OLS equation for the PGA prediction provides a prediction performance comparable with these regression-based models. The proposed PGV and PGD prediction equations produce notably better outcomes over the existing GMPEs.

(v) The GP/OLS-based GMPEs can reliably be used for practical pre-planning and design purposes since they were developed upon on a comprehensive database with wide range properties. The proposed equations are remarkably simple and provide useful alternatives to the more complicated equations provided by NGA. The GMPEs are considered to be mostly valid for use in the western United States and in other similar tectonically active regions of shallow crusting faulting worldwide.

(vi) The data recorded during some real earthquakes at different Iranian sites were used to validate the derived GMPEs. The verification phase confirms the efficiency of the GMPEs for their general application to the simulation of the ground-motions.

Further research can be focused on identifying other predictor variables and incorporating them into the model development. For instance, the rake angle or dummy variables representing styles of faulting may directly be included into the analysis. The energy input is another important response parameter closely related to the damage potential of the earthquake (Zahrah and Hall, 1984; Benavent-Climent et al., 2010). Therefore, this parameter can be studied as an effective intensity measure in probabilistic seismic hazard analysis procedure. Also, the earthquake source effect and the geometrical spreading are two critical issues that can modify the current empirical approach.

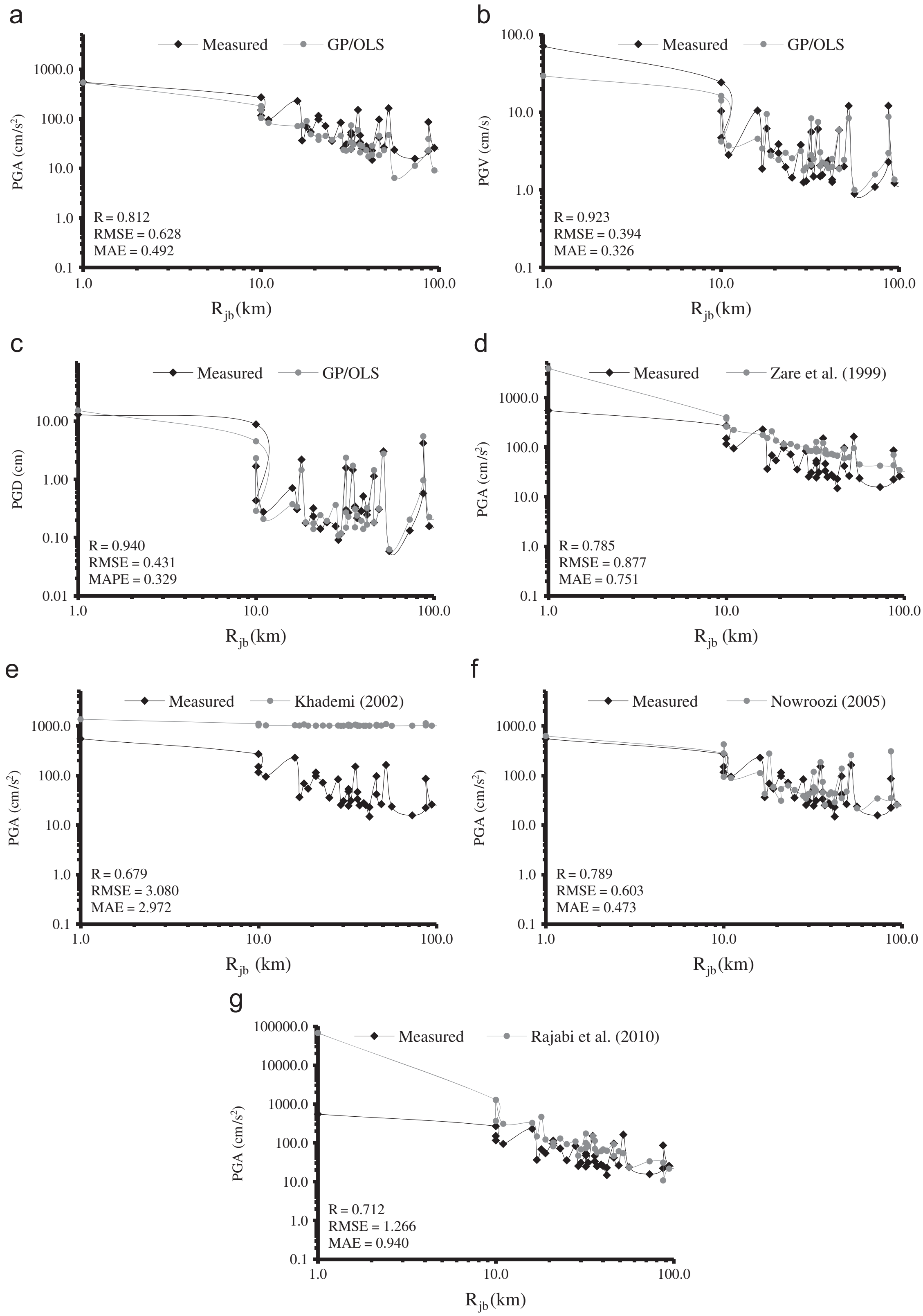

Fig. 18. Comparison of the PGA, PGV and PGD predictions made by different models for the verification data.

References

Alavi, A.H., Gandomi, A.H. A robust data mining approach for formulation of geotechnical engineering systems. International Journal of Computer Aided Methods in Engineering—Engineering Computations, in press.

Alavi, A.H., Gandomi, A.H., Sahab, M.G., Gandomi, M., 2010. Multi expression programming: a new approach to formulation of soil classification. Engineering with Computers 26 (2), 111–118.

Alavi, A.H., Ameri, M., Gandomi, A.H., Mirzahosseini, M.R., 2011. Formulation of flow number of asphalt mixes using a hybrid computational method. Construction and Building Materials 25 (3), 1338–1355.

Ambraseys, N.N., Simpson, K.A., Bommer, J.J., 1996. Prediction of horizontal response spectra in Europe. Earthquake Engineering and Structural Dynamics 25 (4), 371–400.

ATC-63, 2008. Recommended methodology for quantification of building system performance and response parameters. 90% Draft Report, Applied Technology Council, Redwood City, CA.

Benavent-Climent, A., Lopez-Almansa, F.D., Bravo-Gonzalez, A., 2010. Design energy input spectra for moderate-to-high seismicity regions based on Colombian earthquakes. Soil Dynamics and Earthquake Engineering 30, 1129–1148.

Billings, S., Korenberg, M., Chen, S., 1988. Identification of nonlinear outputaffine systems using an orthogonal least-squares algorithm. International Journal of Systems Science 19 (8), 1559–1568.

Boore, D.M., Atkinson, G.M., 2007. Boore–Atkinson NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. PEER Report 2007/01, Pacific Engineering Research Center, University of California, Berkeley, California.

Boore, D.M., Joyner, W.B., Fumal, T.E., 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work. Seismological Research Letters KNZ, 128–153.

Boore, M.D., Atkinson, M.G., 2008. Equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. Earthquake Spectra 24 (1), 99–138.

Cabalar, A.F., Cevik, A., 2009. Genetic programming-based attenuation relationship: an application of recent earthquakes in turkey. Computers & Geosciences 35 (9), 1884–1896.

Campbell, K.W., Bozorgnia, Y., 2007. Campbell–Bozorgnia NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. PEER Report 2007/02, Pacific Engineering Research Center, University of California, Berkeley, California.

Cao, H., Yu, J., Kang, L., Chen, Y., 1999. The kinetic evolutionary modelling of complex systems of chemical reactions. Computers & Chemical Engineering 23 (1), 143–151.

Chu, D., Kerh, T., Wu, C.H., 2003. Analysis of strong motion information at checking stations of Kaohsiung area by backpropagation neural networks. Journal of Civil Engineering Technology 7 (2), 45–62.

Douglas, J., 2003. Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. Earth-Science Reviews 61 (1–2), 43–104.

Douglas, J., 2004. Ground motion estimation equations 1964–2003. Research Report Number: 04-001-SM, Imperial College.

Frank, I.E., Todeschini, R., 1994. The Data Analysis Handbook. Elsevier, Amsterdam.

Gandomi, A.H., Alavi, A.H. Hybridizing genetic programming with orthogonal least squares for modeling of soil liquefaction. Computational Collective Intelligence and Hybrid Systems Concepts and Applications. IGI Global Publishing, in press.

Gandomi, A.H., Alavi, A.H., Arjmandi, P., Aghaeifar, A., Seyednour, R., 2010a. Modeling of compressive strength of hpc mixes using a combined algorithm of genetic programming and orthogonal least squares. Journal of Mechanics of Materials and Structures 5 (5), 735–753.

Gandomi, A.H., Alavi, A.H., Mirzahosseini, R., Moghdas Nejad, F., 2011. Nonlinear genetic–based models for prediction of flow number of asphalt mixtures. Journal of Materials in Civil Engineering ASCE 23 (3), 1–18.

Gandomi, A.H., Alavi, A.H., Sahab, M.G., 2010b. New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming. Materials and Structures 43 (7), 963–983.

Golbraikh, A., Tropsha, A., 2002. Beware of q^2!. Journal of Molecular Graphics and Modelling 20 (4), 269–276.

Güllü, H., Erçelebi, E., 2007. A neural network approach for attenuation relationships: an application using strong-ground-motion data from Turkey. Engineering Geology 93 (3-4), 65–81.

Gunaydın, K., Gunaydın, A., 2008. Peak ground acceleration prediction by artificial neural networks for Northwestern Turkey. Mathematical Problems in Engineering ID 919420. 10.1155/2008/919420.

Haselton, C.B., Deierlein, G.G., 2007. Assessing seismic collapse safety of modern reinforced concrete frame. PEER Report 2007/08, Pacific Engineering Research Center, University of California, Berkeley, California.

Johari, A., Habibagahi, G., Ghahramani, A., 2006. A Prediction of Soil–Water Characteristic Curve Using Genetic Programming. Journal of Geotechnical and Geoenvironmental Engineering ASCE 132, 661–665.

Kerh, T., Chu, D., 2002. Neural networks approach and microtremor measurements in estimating peak ground acceleration due to strong motion. Advances in Engineering Software 33 (11–12), 733–742.

Kerh, T., Ting, S.B., 2005. Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system. Engineering Applications of Artificial Intelligence 18 (7), 857–866.

Kermani, E., Jafarian, Y., Baziar, M.H., 2009. New Predictive Models for the vmax/amax Ratio of Strong Ground-motions using Genetic Programming. International Journal of Civil Engineering 7 (4), 236–249.

Khademi, M.H., 2002. Attenuation of peak and spectral accelerations in the Persian plateau. In: Proceedings of the 12th European Conference on Earthquake Engineering, Paper Reference 330.

Koza, J., 1992. Genetic programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA.

Kramer, S.L., 1996. Geotechnical Earthquake Engineering. Prentice-Hall, New Jersey.

Kraslawski, A., Pedrycz, W., Nyström, L., 1999. Fuzzy neural network as instance generator for case-based reasoning system: an example of selection of heat exchange equipment in mixing. Neural Computing & Applications 8 (2), 106–113.

Krishnan, S., Chen, J., Komatitsch, D., Tromp, J., 2006. Case studies of damage to tall steel moment-frame buildings in Southern California during large San Andreas earthquakes. Bulletin of the Seismological Society of America 96, 1523–1537.

Kuo, Y.L., Jaksa, M.B., Lyamin, A.V., Kaggwa, W.S., 2009. ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. Computers and Geotechnics 36 (3), 503–516.

Luco, N., Cornell, C.A., 2007. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. Earthquake Spectra 23 (2), 357–392.

Madár, J., Abonyi, J., Szeifert, F., 2004. Genetic programming for system identification. In: Proceedings of the Intelligent Systems Design and Applications (ISDA 2004) Conference, Budapest, Hungary.

Madár, J., Abonyi, J., Szeifert, F., 2005. Genetic Programming for the Identification of Nonlinear Input-Output Models. Industrial & Engineering Chemistry Research 44 (9), 3178–3186.

Nowroozi, A.A., 2005. Attenuation relations for peak horizontal and vertical accelerations of earthquake ground motion in Iran: a preliminary analysis. Journal of Sustainable Energy & Environment 7 (2), 109–128.

Pan, Y., Jiang, J., Wang, R., Cao, H., Cui, Y., 2009. A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. Journal of Hazardous Materials 168 (2-3), 962–969.

Papageorgiou, A.S., Aki, K., 1983. A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion, Part I: description of the model. Bulletin of the Seismological Society of America 73, 693–722.

Pearson, R., 2003. Selecting nonlinear model structures for computer control. Journal of Process Control 13 (1), 1–26.

Power, M.B., Chiou, N.A., Abrahamson, N., Roblee, C., 2006. The next generation of ground motion attenuation models, NGA project: an overview. In: Proceedings of the Eighth National Conference on Earthquake Engineering, San Francisco, Paper No. 2022. 〈http://peer.berkeley.edu/nga/flatfile.html〉.

Rajabi, A.M., Khamehchiyan, M., Mahdavifar, M.R., Del Gaudio, V., 2010. Attenuation relation of Arias intensity for Zagros Mountains region (Iran). Soil Dynamics and Earthquake Engineering 30, 110–118.

Reeves, C.R., 1997. Genetic algorithm for the operations research. INFORMS Journal on Computing 9, 231–250.

Sadigh, R.K., Egan, J.A., 1998. Updated relationships for horizontal peak ground velocity and peak ground displacements for shallow crustal earthquakes. In: Proceedings of the Sixth US National Conference on Earthquake Engineering, Seattle, Paper No. 317.

Smith, G.N., 1986. Probability and Statistics in Civil Engineering. Collins, London.

Somerville, P.G., Graves, R.W., 2003. Characterization of earthquake strong ground motion. Pure and Applied Geophysics 160, 1811–1828.

Zahrah, T.F., Hall, W.J., 1984. Earthquake energy absorption in SDOF systems. Journal of Structural Engineering 110, 1757–1772.

Zare, M., Ghafory-Ashtiany, M., Bard, P.Y., 1999. Attenuation law for the strong-motions in Iran. In: Proceedings of the Third International Conference on Seismology and Earthquake Engineering, Tehran, pp. 345–354.