

Ground motion predictive modelling based on genetic algorithms

S. Yilmaz

Department of Civil Engineering, Pamukkale University, Denizli, Turkey

Received: 25 March 2011 – Revised: 28 July 2011 – Accepted: 30 August 2011 – Published: 20 October 2011

Abstract. This study aims to utilise genetic algorithms for the estimation of peak ground accelerations (PGA). A case study is carried out for the earthquake data from south-west Turkey. The input parameters used for the development of attenuation relationship are magnitude, depth of earthquake, epicentral distance, average shear wave velocity and slope height of the site. Earthquake database compiled by the Earthquake Research Institute of Turkey was used for model development. An important contribution to this study is the slope/hill data included into the dataset. Developed empirical model has a good correlation ($R = 0.78$ and 0.75 for the training and overall datasets) between measured and estimated PGA values. The proposed model is also compared with local empirical predictive models and its results are found to be reasonable. The slope-hill effect found to be an important parameter for the estimation of PGA.

1 Introduction

Many seismic design codes utilise smoothed acceleration spectrums, which are scaled by spectrum coefficients for different seismic zones, for the determination of seismic loads on buildings. Spectrum coefficient is usually a measure of expected peak ground acceleration (PGA). On the other hand, seismic hazard estimation is also an important aspect of earthquake engineering. It is needed to determine seismic safety of existing building stocks for mitigation works. PGA values significantly affect the results of the hazard assessment studies.

Many ground motion predictive models for PGA are proposed in the existing literature. Various input parameters are used in different models. Magnitude of the event and distance between site and source are common parameters in

almost all models. However, attenuation behaviour is also affected by source type, properties of propagation media, local geology and topography (Kramer, 1996). On the other hand, many of the existing empirical predictive relations are obtained by regression analysis. Although, the attenuation behaviour is very complex due to uncertainties caused by the complexity of inputs, these empirical models have very simple formulations. The affect of the uncertainties may be reduced by the use of artificial intelligence (AI) techniques. In recent years, some new ground motion predictive models are proposed based on AI techniques (Güllü and Erçelebi, 2007; Cabalar and Cevik, 2009).

Genetic algorithm (GA) is a popular computing technique for the solution of optimization problems. It is a computer simulation of production of successive generations from a random initial population, which is a set of probable optimal solutions. As well as classical optimization problems, GA is utilised for curve fitting applications for the estimation of a depended variable (Ergun et al., 2008). In this approach, format of the aimed function is defined in the code and coefficients of the predefined parameters are determined by genetic operators. However, the success of regression for complex problems is limited to the imagination of the user, which defines the form of the curve. Instead of using predefined curve formats, GA can be programmed so that it selects the effective parameters and determines their coefficients. Such an application of GA is also available in the literature for the geotechnical engineering problems (Sen and Akyol, 2010).

Turkey is a seismically active region of the world and it has been affected by destructive earthquakes frequently. There is a need to develop local empirical predictive models for the region. There is a number of models developed for Turkey (Aydan, et al., 1996; Gülkan and Kalkan, 2002; Kalkan and Gülkan 2004; Ulusay et al., 2004). The recent models, which utilise dataset developed in previous studies, also employ AI approaches (Güllü and Erçelebi, 2007; Cabalar and Cevik, 2009) see Table 1.



Correspondence to: S. Yilmaz
(syilmaz@pau.edu.tr)

Table 1. Local empirical predictive relationships developed from Earthquakes in Turkey.

Author	Attenuation model
Aydan et al. (1996)	$PGA = 2.8(e^{0.9M_s} e^{-0.025R} - 1)$ M_s : surface wave magnitude, R : Epicentral distance
İnan et al. (1996)	$PGA = 10^{0.65M - 0.9 \log(R) - 0.44}$ M : magnitude, R : Epicentral distance
Gülkan and Kalkan (2002)	$\ln Y = b_1 + b_2(M - 6) + b_3(M - 6)^2 + b_5 \ln(r) + b_v \ln(V_S / V_A)$ Y : PGA or PSA; $b_1, b_2, b_3, b_4, b_5, b_v, V_A$: regression coefficients; r : Epicentral distance; V_S : Shear wave velocity
Kalkan and Gülkan (2004)	$\ln Y = b_1 + b_2(M - 6) + b_3(M - 6)^2 + b_5 \ln(\sqrt{r_{cl}^2 + h^2}) + b_v \ln(V_S / V_A)$ Y : PGA or PSA; $b_1, b_2, b_3, b_4, b_5, b_v, V_A$: regression coefficients; r : Epicentral distance; V_S : Shear wave velocity
Ulusay et al. (2004)	$PGA = 2.18e^{0.0218(33.3 M_w - R_e + 7.8427 S_A + 18.9282 S_B)}$ M_w : moment magnitude, R_e : Epicentral distance; S_A, S_B : Soil constants
Beyaz (2004)	$\log A = \beta_0 + (\beta_1 M_w^2) + (\beta_2 \log(R + 1))$ A : Peak acceleration at base-rock, $\beta_0, \beta_1, \beta_2$: regression coefficients; M_w : moment magnitude, R : Epicentral distance;
Cabalar and Cevik (2009)	$PGA = \left(\frac{5.7}{A}\right)^5 + \frac{B \cdot M_w \cdot \log(M_w)}{\sqrt[3]{R}}$ $A = \sqrt[4]{V_S} + \left(R - \sqrt[9]{V_S^4 + 238^5}\right)^2, \quad B = \sqrt[3]{4 \sqrt[4]{\frac{651}{V_S}} - \frac{\log R}{3}}$ M_w : moment magnitude, R : Epicentral distance; V_S : Shear wave velocity

In this study, south-west Turkey is selected as a case study for the development of GA based attenuation model. A dataset is prepared for the region. Recent earthquake data is also included in the dataset. It is well known that slope/hill affect is an important source of PGA variations. Due to the slope/hill affect, hazard level increased in some past earthquakes (Kaplan et al., 2008) in the region. Therefore, slope data of the recording stations is also used in the dataset.

2 Data used and method

2.1 Seismicity of the South-West Turkey

Turkey suffers significantly from earthquakes, being on Alpine orogenic system (Erdik et al., 1985). The African, Arabian and Eurasian plates interact with the Anatolian plate. Movement of African and Arabian plates to north-northeast forces the movement of the Anatolian plate to the Eurasian plate. Due to this movement, the west of Anatolian plate is squeezed and horst and graben systems are formed in south-west Anatolia (Fig. 1). These formations are the main source of the seismic activity in the region. For more information on the seismicity and tectonics of the region, Erdik et al. (1985) and Westaway (2003) are important references.

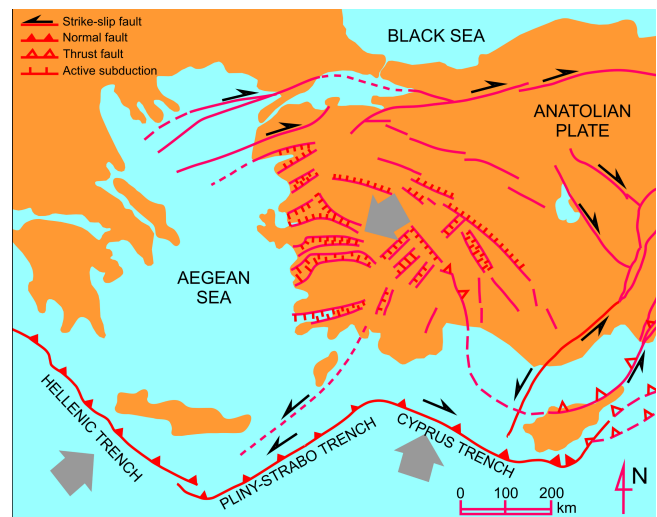


Fig. 1. Major tectonic structures in West Turkey (modified from Kocycigit, 1996).

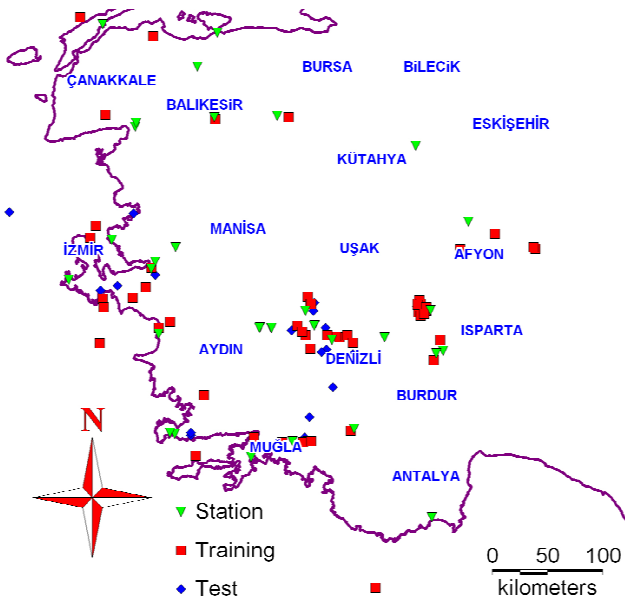


Fig. 2. Strong motion epicentres and Recording stations.

2.2 Empirical predictive models for Turkey

The first seismograph in Turkey was installed in 1973 in scope of a project carried out by Earthquake Research Institute of Turkey (today called Disaster and Emergency Management Presidency). The first record was taken during the 1976 Denizli Earthquake. Some of the records from the network were already used in some international datasets for empirical predictive modelling. Moreover, local predictive models based on earthquake data in Turkey have been developed in the last decade. Table 1 summarizes empirical predictive relations developed for Turkey. The relations developed before 2001 uses surface wave magnitude (M_s), which is developed for strong earthquakes. The most reliable magnitude scale is the moment magnitude (M_w). However, it has complex calculation procedures and calculated for a limited number of earthquakes. The recent models utilise moment magnitude. However, the dataset used in this study includes small to medium size earthquakes with records taken from a distance of up to 100 km to the source. As it has the simplest calculation procedure, duration magnitude is available for most of the earthquakes. Some empirical relations are also available in the literature for the conversion between different magnitude scales (Ulusay et al., 2004). However, such a conversion may introduce some extra errors due to development of a model by using the data obtained from another empirical model. Therefore, duration magnitude (M_d) is used for this study; instead of converting different magnitude scales to each other.

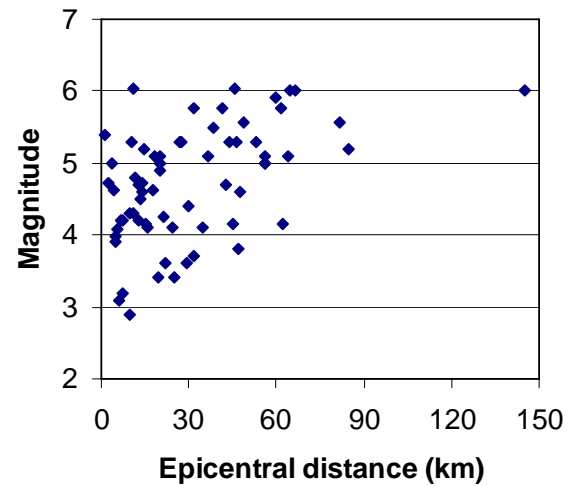


Fig. 3. Magnitude vs. epicentral distance for training set.

2.3 Ground motion dataset for south-west Turkey

Ground motion dataset used for the development of the empirical predictive model proposed in this study was divided into training and test datasets. Division is carried out by random selection. Researchers proposed to select 20 % to 30 % of the dataset for testing (Swingler, 1996; Nelson and Illingworth, 1990). In this study, 28 % of the overall dataset is selected as testing data. Tables 2 and 3 list the training and testing datasets, respectively. Station information can be accessed at Disaster and Emergency Management Presidency (DEMP). All of the destructive earthquakes ($M > 5.0$) were included in the database. However, to limit the affect of small PGA values, records with small PGA values were eliminated. The lower limit for the PGA is selected as 20 gals. Figure 2 shows geographic distribution of the epicentres of the earthquakes and recording stations used in this study.

Since the lower limit of 20 gals have been used for dataset creation, there exist many near source recordings for small magnitude earthquakes. Therefore, an epicentral distance limitation of 2 km is selected to filter cut-off near-source effects. When the magnitude increases, upper bound of the epicentral distance is also increases as illustrated in Fig. 3. Considering the earthquake with magnitude smaller than 4.0, there are only two records with epicentral distance greater than 30 kilometers (32 km and 47 km). On the other hand, for a greater magnitude number of records with greater epicentral distances than 65 kilometers, there is only three. Therefore, it seems to be rational that developed models can be valid for the first 30 and 65 km for magnitudes smaller and greater than 4.0, respectively.

2.4 Genetic Algorithms

Genetic Algorithm (GA) is an optimization and search method based on the principle of survival of the fittest.

Table 2. Training dataset.

Recording station	Date	PGA (gal)	Depth (km)	Epicentral distance (km)	M_d
Saraykoy	05.06.2006	22.84	11.1	16.07	4.10
Kuyucak	05.06.2006	66.78	11.1	24.23	4.10
Nazilli	05.06.2006	20.88	11.1	34.91	4.10
Finike	23.01.2005	25.00	12.1	85.03	5.20
Denizli	22.08.2005	33.36	11.7	7.38	3.20
Bornova	20.10.2005	31.92	15.4	59.87	5.90
Koycegiz	24.08.2004	27.98	10.5	9.85	2.90
Koycegiz	20.12.2004	27.40	12.5	36.61	5.10
Marmaris	20.12.2004	33.89	12.5	18.01	5.10
Saraykoy	23.07.2003	123.23	5.0	27.41	5.30
Kuyucak	23.07.2003	21.73	5.0	44.13	5.30
Nazilli	23.07.2003	25.94	5.0	53.26	5.30
Denizli	23.07.2003	45.84	5.0	46.11	5.30
Saraykoy	26.07.2003	47.54	5.0	20.34	5.00
Denizli	26.07.2003	25.79	4.3	38.58	5.50
Afyon	03.02.2002	113.50	5.0	66.38	6.00
Kutahya	03.02.2004	23.13	5.0	145.00	6.00
Afyon	03.08.2002	27.00	5.0	64.35	6.00
Afyon	03.02.2002	51.50	2.2	26.74	5.30
Burdur	03.04.2002	28.93	5.0	12.74	4.20
Burdur	02.02.2001	30.12	10.0	7.18	4.20
Bodrum	28.02.2000	42.94	15.4	47.01	3.80
Denizli	21.04.2000	27.56	14.7	14.76	5.20
Denizli	04.06.2000	40.22	5.9	19.48	3.40
Denizli	05.06.2000	27.16	5.0	25.02	3.40
Denizli	05.06.2000	36.87	1.2	4.85	3.90
Denizli	05.06.2000	21.58	5.0	29.19	3.60
Denizli	29.06.2000	22.25	5.0	6.26	3.10
Denizli	29.06.2000	28.38	5.0	21.69	3.60
Bodrum	08.12.1999	24.99	22.6	32.01	3.70
Edremit	05.03.1998	27.11	10.0	29.61	4.40
Cardak	04.04.1998	27.88	15.0	47.38	4.60
Dinar	04.04.1998	134.73	15.0	13.78	4.60
Bornova	09.07.1998	27.04	20.0	64.03	5.10
Kusadasi	29.05.1998	63.66	33.1	5.57	4.07
Gelibolu	25.10.1997	42.42	30.0	21.40	4.26
Kusadasi	20.02.1996	21.35	10.0	45.12	4.17
Kusadasi	02.04.1996	33.44	15.0	56.11	5.01
Dinar	29.06.1996	97.39	31.0	4.52	4.63
Dinar	26.09.1995	182.73	20.0	11.61	4.80
Dinar	27.09.1995	180.38	18.0	13.47	4.50
Dinar	28.09.1995	42.73	33.0	61.90	4.17
Cardak	01.10.1995	65.07	15.0	45.65	6.04
Dinar	01.10.1995	329.72	15.0	11.13	6.04
Cardak	01.10.1995	24.83	33.0	42.98	4.70
Dinar	01.10.1995	224.66	33.0	13.05	4.70
Dinar	03.10.1995	145.59	32.0	11.14	4.30
Dinar	04.10.1995	59.74	33.0	4.92	3.98
Dinar	06.10.1995	168.08	33.0	6.99	4.20
Izmir	16.12.1977	391.41	34.0	1.11	5.38
Dursunbey	18.07.1979	288.25	5.0	10.34	5.29
Edincik	05.07.1983	53.44	7.0	57.79	5.57
Edremit	05.07.1983	27.78	7.0	81.61	5.57
Gonen	05.07.1983	50.11	7.0	48.70	5.57
Balikesir	29.03.1984	223.89	12.0	2.39	4.73
Koycegiz	06.12.1985	114.46	8.0	14.23	4.73
Kusadasi	01.06.1986	94.43	21.0	15.33	4.17
Ilica	06.11.1992	37.81	17.0	61.76	5.76
Izmir	06.11.1992	38.34	17.0	31.93	5.76
Kusadasi	06.11.1992	83.49	17.0	41.45	5.76
Foca	24.05.1994	49.80	17.0	20.12	4.90
Foca	24.05.1994	57.65	16.0	19.91	5.10
Ilica	24.05.1994	26.88	16.0	56.10	5.10
Koycegiz	13.11.1994	96.51	15.0	17.79	4.63
Koycegiz	13.11.1994	57.16	10.0	9.85	4.30
Koycegiz	30.10.2007	21.12	28.9	56.01	5.01

Table 3. Test dataset.

Recording station	Date	PGA (gal)	Depth (km)	Epicentral distance (km)	M_d
Saraykoy	05.06.2006	22.32	10.4	21.85	3.90
Bornova	17.10.2007	22.52	18.6	57.26	5.80
Manisa	20.10.2005	22.00	15.4	83.22	5.90
Saraykoy	08.03.2004	35.69	5.0	13.39	3.60
Bodrum	04.08.2004	28.11	5.0	15.15	4.60
Bornova	10.04.2003	78.58	15.8	41.41	5.60
Saraykoy	26.07.2003	121.12	4.3	20.21	5.50
Kuyucak	26.07.2003	26.29	4.3	43.04	5.50
Nazilli	26.07.2003	27.16	4.3	53.13	5.50
Denizli	29.06.2000	20.39	6.1	10.35	3.40
Denizli	29.06.2000	21.36	1.0	43.42	3.30
Denizli	04.10.2000	66.38	8.4	12.72	4.70
Denizli	31.10.2000	44.13	14.2	23.52	3.40
Buldan	21.01.1997	38.51	18.0	11.09	4.91
Dinar	26.09.1995	81.02	24.0	13.56	4.30
Burdur	01.10.1995	43.92	15.0	48.26	6.04
Dinar	01.10.1995	171.79	33.0	6.53	4.35
Dinar	05.10.1995	128.84	33.0	5.71	4.40
Denizli	19.08.1976	348.53	3.0	14.75	5.10
Izmir	09.12.1977	272.97	19.0	6.57	4.91
Balikesir	05.07.1983	22.55	7.0	92.93	5.57
Foca	17.06.1984	24.17	15.0	98.03	5.10
Foca	04.08.1988	41.40	12.0	31.60	4.63
Koycegiz	13.11.1994	26.60	10.0	12.03	4.60
Koycegiz	13.11.1994	25.54	10.0	27.41	4.26
Çameli	29.10.2007	56.58	28.9	3.59	5.01

The solution process simulates natural selection mechanisms. The method has been effectively used in many engineering applications. GA was first introduced by Holland (1975) although it started to be widely used after Goldberg's book (1989).

A GA is initiated with a random population of many individuals. Each individual represents a probable solution to the problem of concern. Then, using these individuals, new offsprings are produced by crossovers and mutations in each generation. Among the population, specific selection rules are applied to the expanded population to reduce the size to that of the original population. These rules are based on the fitness of the individuals. The selection method is designed to increase the chance of selection of the fittest individuals. The fitness of individuals is a measure of optimality of the solution represented by that individual. The procedure of the GA generate better and better offsprings, in each generation to make the solutions close to the objective function (Tung et al., 2003). GA has been confirmed to offer many advantages with respect to the classical optimization methods.

Offspring chromosomes are generated by usually combining two parent chromosomes. This is called crossover operation. An offspring has features of both parents. Firstly, two individuals are selected for crossover and a random cut-off point is selected for a crossover. Then each chromosome is

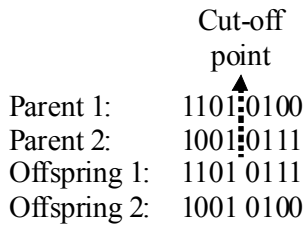


Fig. 4. Crossover Operation.

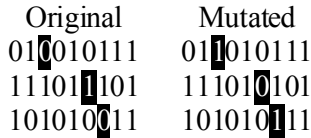


Fig. 5. Sample mutation operations.

cut at that point and right parts of the strings are swapped. This simplest crossover method is illustrated in Fig. 4. It is also possible to select many cut-off points for a crossover operation. The number of crossover operations in each generation is determined by crossover probability. Crossover probabilities up to 0.80 give satisfactory results in many applications.

Mutation is an essential operator of the GA. A mutation operation is simply carried out by changing 1 to 0 or vice versa at a randomly selected bit among all chromosomes. Premature loss of genetic information from the population is highly probable in especially small populations. This is prevented by mutations. Smaller mutation probabilities are ideal to satisfy stability of the population for most of the problems. Some sample mutation operations are shown in Fig. 5.

Increasing the population size is a problem for computer memory. Population size expanded by mutations and crossovers is reduced to original size by selection operators. Selection is based upon the fitness of individuals. Fittest ones have more chance to be selected to the next generation. Normally, less fit ones have less of a chance. Therefore, each generation has a greater average fitness value than the previous one. Fitness of a chromosome is calculated by fitness function. This function is a definition of the optimization problem. In case of curve fitting applications functions utilizing sum square error, root mean square error, etc., can give satisfactory results. These operations are carried out until the end criteria is satisfied (Fig. 6).

3 Empirical predictive models by GA

In most of the regression, a template formula is determined by the researchers. In this study, a standard formulation is not given to the code. An initial template is introduced to the code. The final form of the formula is determined after GA procedures are run. The initial template is given in Eq. (1).

$$PGA = X_0 + X_1 \cdot f_1(t_1^{x_2}) + X_3 \cdot f_2(t_2^{x_4}) + X_5 \cdot f_3(t_3^{x_6}) \quad (1)$$

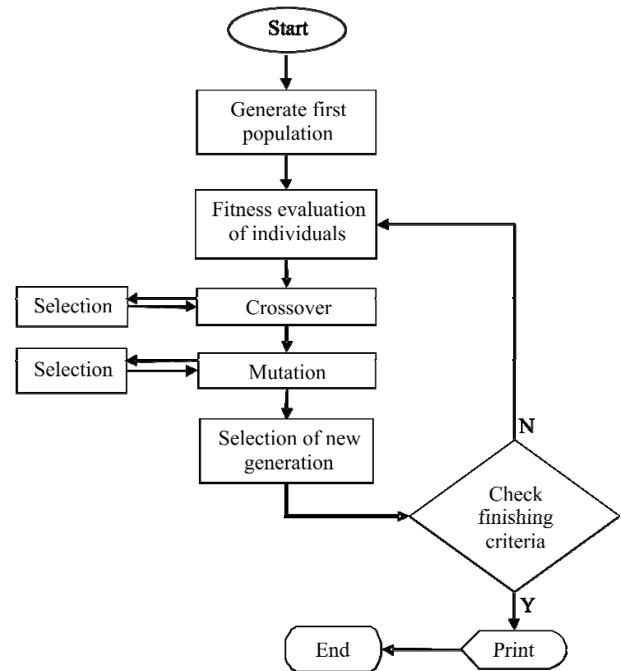


Fig. 6. Flowchart of the GA.

Table 4. Input parameters and function alternatives.

Function inputs	Solution domain
Coefficients, X_i	$-10.23 \leq X_i \leq 10.24$
Powers, X_i	$-5.11 \leq X_i \leq 5.12$
Functions, f_i	$\log(t), t, \sin(t), e^t$
Input parameters, t_i	$m, d, r, z, h, v_s, S1, S2, S3$

$$+ X_7 \cdot f_4(t_4^{x_8}) \cdot f_5(t_5^{x_9}) + X_{10} \cdot f_6(t_6^{x_{11}}) \cdot f_7(t_7^{x_{12}}) + X_{13} \cdot f_8(t_8^{x_{14}}) \cdot f_9(t_9^{x_{15}}) + X_{16} \cdot f_{10}(t_{10}^{x_{17}}) \cdot f_{11}(t_{11}^{x_{18}}) \cdot f_{12}(t_{12}^{x_{19}})$$

In this formulation, X_i denotes coefficients or powers of the terms. Function alternatives and input parameters to be selected by the GA are denoted as f_i and t_i , respectively. Upper and lower limits of the coefficients and powers and function alternatives and input parameters are summarized in Table 4. GA selects optimal coefficients, powers, input parameters and functions. For example, if the optimal solution for (X_1, f_1, t_1, X_2) is $(3, \log, r, 4)$, then the term $X_1 \cdot f_1(t_1^{x_2})$ becomes $3\log(r^4)$.

As the original inputs have very different intervals, changes in input parameters of mutation by a bit or a crossover may cause significant jumps at fitness value of the individual. For example, magnitude, M , changes between 2.9 and 6.0, whereas shear wave velocity, V_s , changes between 200 and 700 $m\ s^{-1}$ in the dataset. To prevent big jumps at fitness values due to change of parameters, all input

Table 5. Upper and lower limits of original data used for mapping

Original Parameter	Mapped parameter	P_{min}	P_{max}
Magnitude. M_d	m	2.5	6.5
Focal Depth. D	d	0	35
Epicentral distance. R	r	0	150
Focal distance. Z	z	0	150
Slope height. H	h	0	250
Shear wave velocity. V_S	v_s	200	700

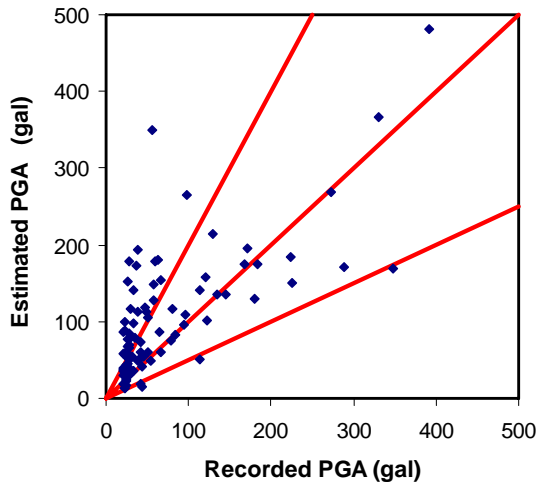


Fig. 7. Observed vs. Estimated PGA for the whole dataset.

parameters are mapped to similar intervals. Formulation of the mapping procedure for the inputs is given in Eq. (2). This formulation squeezes the variations in each parameter to interval [1, 5].

$$p = \frac{(P - P_{min})}{(P_{max} - P_{min})} \cdot 4 + 1 \tag{2}$$

P is the original data; P_{min} and P_{max} are upper and lower limits of the data corresponding to mapped limits of 1 and 5, respectively. p stands for the mapped data. Table 5 shows the upper and lower limits of the input parameters. For example, $M_d = 4$ is mapped to $m = 2.5$.

Fitness function used for the evaluation of the individuals is shown in Eq. (3), where PGA_i is the observed data, $PGA_{i,e}$ is the estimated value corresponding to i th record. Fitness evaluation takes care of absolute error and the ratio of greater estimated and real PGA values to those smaller. So it is penalized for small PGA values with a small absolute error, but estimations have 2 or 3 times the real values or vice versa. Lastly, as larger PGA values are more significant for engineering purposes, fitness term is multiplied by the real PGA value to increase the weight of larger PGA values in

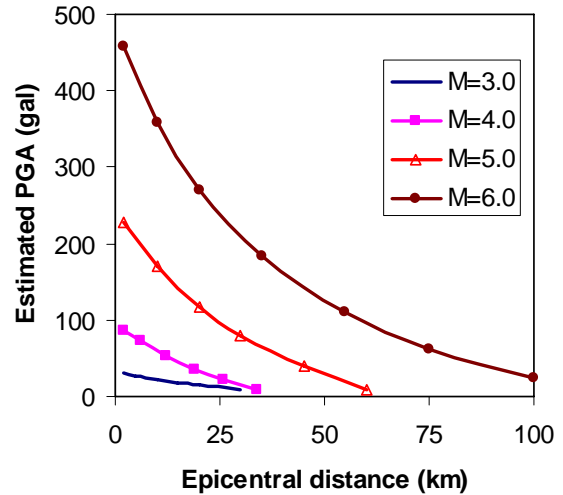


Fig. 8. Attenuation of the PGA with respect to epicentral distance.

the solution. Therefore, the domination of small PGA values is prevented.

$$\min f = \sum_{i=1}^n \sqrt{(PGA_i - PGA_{i,e})^2} \cdot \frac{\max(|PGA_i|, |PGA_{i,e}|)}{\min(|PGA_i|, |PGA_{i,e}|)} \cdot PGA_i \tag{3}$$

After running GA for the minimization of the objective function, an optimal curve is obtained as the solution of PGA estimation problem. The proposed empirical formula is shown by Eq. (4). The formula has a more complex structure than the existing empirical predictive models. However, an important advantage of the proposed model is that it counts for the slope affect. The formulation takes into account mapped earthquake parameters m, r, z, h and S_3 , which gets 1 for soft soils and 0 for others.

$$PGA = 4.47 - 8.96 \cdot \sin\left(e^{0.51 m^{0.97}}\right) + 10.2 \cdot S_3 - 8.8 \cdot e^{m^{0.63}} - 9.35 \log(m^{4.44}) \cdot \log(r^{4.72}) - 8 \sin(e^{0.3 m^{1.74}}) \cdot e^{z^{0.37}} + 7.6 \cdot m^{2.41} \cdot \log(e^{r^{-4.8}}) + 0.017 \cdot e^{2 \cdot h} \tag{4}$$

Estimated and recorded PGA values are plotted in Fig. 7. There exists a good correlation especially for high-PGA records, which are more important for the seismic design. The final model in Eq. (4) has PGA attenuation with the epicentral distance as shown in Fig. 8. It is quite reasonable that PGA reduces with the increase in epicentral distance.

Attenuation of the PGA were plotted for different slope heights for magnitudes of $M = 5$ and $M = 6$ in Fig. 9. According to developed model, slope heights up to 100 m yields a limited increase in PGA. After this limit, PGA increases rapidly. PGA increases about 12 gals between 100 and 150 m of slope height, which is also achieved between 150 and 170 m of slope height. Due to the limitations of the dataset the formulation may not yield reliable results for higher slopes.

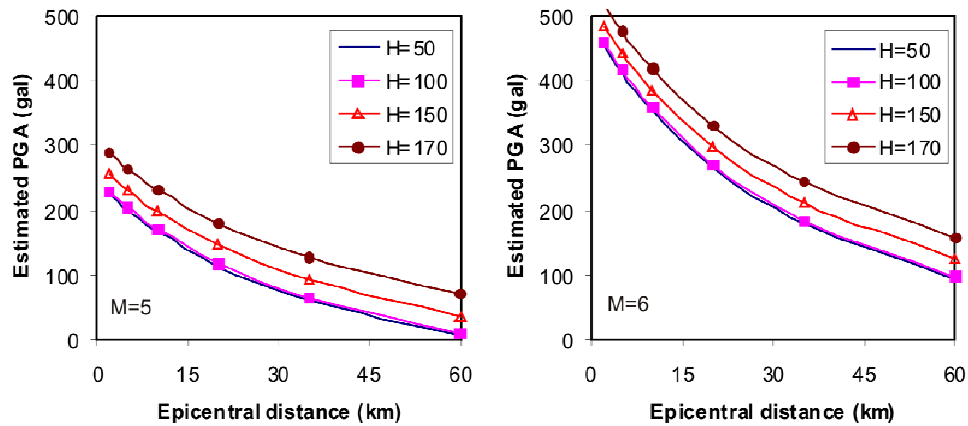


Fig. 9. Attenuation of the PGA with respect to slope height for $M = 5$ and $M = 6$

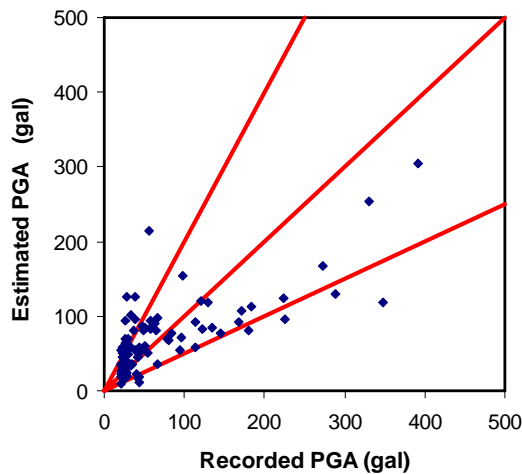


Fig. 10. Distribution diagrams for Kalkan and Gülkan (2004).

Table 6. Performance of empirical predictive models recently developed for Turkey in case of south-west Turkey earthquakes.

Dataset	Training		Test		Overall	
	R	RMSE	R	RMSE	R	RMSE
Model						
Gülkan and Kalkan (2002)	0.58	12.11	0.61	18.29	0.59	10.10
Kalkan and Gülkan (2004)	0.75	6.22	0.72	11.01	0.74	5.45
Ulusay et al. (2004)	0.67	7.22	0.63	12.40	0.66	6.26
Beyaz (2004)	0.73	7.97	0.46	14.24	0.44	29.47
Cabalar and Celik (2009)	0.23	19.28	0.27	25.63	0.24	15.57
This study	0.78	6.68	0.71	10.98	0.75	5.65

The only parameter related to soil behaviour is S_3 , which has a limited effect on the results. Mainly because of the small number of records available for medium and hard soils.

In addition to parametric studies, the developed model is also compared with existing models. The same dataset is

used in comparison with domestic relations. Results of those were compared with the proposed attenuation relationship in Table 6. According to the table, the attenuation model proposed in this study gives satisfactory results. Its performance is very similar to that of the model developed by Kalkan and Gülkan (2004). Both relations yield good correlation with low errors. Scatter diagram calculated for Kalkan and Gülkan (2004) relation is given in Fig. 10. When compared with Fig. 7, it is observed that both models overestimate low PGA values and underestimates PGA values higher than 100 gals. Deviation of the proposed formula is higher for low PGA values. On the other hand, deviation of estimations of proposed model is limited with respect to Kalkan and Gülkan (2004) formulation for higher PGA values. In Fig. 11, estimations of two models can be compared with the recorded data in descending order. Peak points for both models belong to the same recordings. Better performance of the proposed model for high-PGA values is also observed in this figure.

4 Conclusions

This study proposes an empirical model for the estimation of PGA based on genetic algorithms. The model is developed for the ground motion records obtained from south-west Turkey, with magnitudes ranging from 2.9 to 6.0. Besides, an evaluation of the existing attenuation models, developed with the records from Turkey, has been carried out. The dataset for the south-west Turkey was divided into training and test sets. The main results are discussed below.

Among the existing models, the empirical predictive relationship developed by Kalkan and Gülkan (2004) shows the best performance for the RMSE for the overall dataset, while the proposed model has a better correlation. However, the case is vice versa for the test dataset. In general, the proposed model in this study gives similar results with

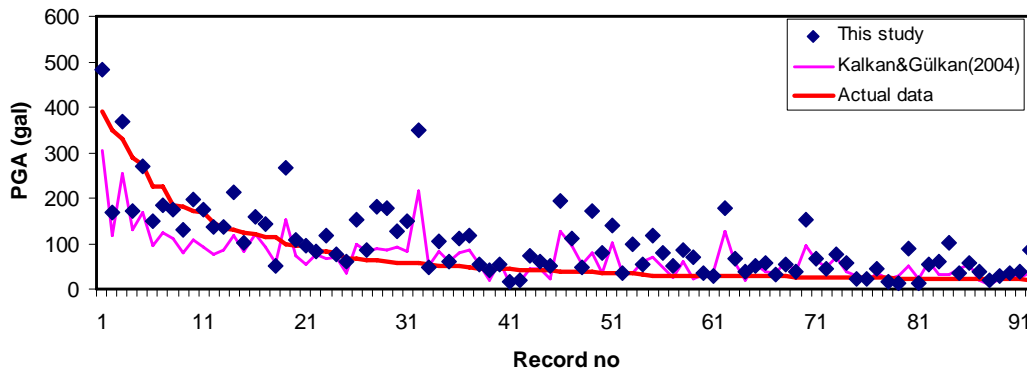


Fig. 11. Distribution diagrams for Kalkan and Gülkan (2004).

the Kalkan and Gülkan model (2004). Among other models proposed for Turkey, these two models generate more precise estimates of PGA. In the training dataset and overall datasets, the proposed model by this study has a better correlation than the Kalkan and Gülkan (2004) model. Contrary to this figure, RMSE of the proposed model is lower than the Kalkan and Gülkan model (2004) for the test dataset, but the training and overall datasets. Besides, the proposed model has better estimates for higher PGA values with respect to the Kalkan and Gülkan (2004) model. This is the effect of weighting PGA values in objective function. Due to such a weighting, errors for low PGA values becomes less important than those for high PGA values.

An important advantage of the proposed attenuation model is that it counts for the slope affects on PGA. It is a local attenuation model developed for the south-west Turkey, where many hilly settlement areas exist. The use of the model for these regions may produce more reliable results.

However, an important drawback of the models is that, the characterisation of the soil under the recording stations in Turkey is based on rough soil classifications. More precise soil parameters of each station are needed to improve the quality of the ground motion predictive models in terms of soil classification. Besides, very limited data is available for strong motions with magnitude greater than 6. Therefore, it should be noted that, results obtained from the proposed model may not be valid for higher magnitudes.

Acknowledgements. The author thanks his colleagues Gulmustafa Şen and Hayri Baytan Özmen for their valuable comments.

Edited by: M. E. Contadakis

Reviewed by: Y. Yardım and another anonymous referee

References

- Aydan, Ö., Sezaki, M., and Yazar, R.: The seismic characteristics of Turkish earthquakes, Eleventh World Conference on Earthquake Engineering, Acapulco, Mexico, 1996.
- Beyaz, T.: Development of a new attenuation relationship of seismic energy for Turkey using the strong motion records free of soil effects, PhD thesis, Ankara University, Ankara, 2004.
- Cabalar, A. F. and Cevik, A.: Genetic programming-based attenuation relationship: An application of recent earthquakes in Turkey, *Comp. & Geosciences*, 35, 1884–1896, 2009.
- DEMP: Disaster and Emergency Management Presidency of Turkey, <http://www.deprem.gov.tr/SarbisEng/Ist.Bilgi/ButunIst.aspx>, last access: 26 July 2011.
- Erdik, M., Doyuran, V., Gülkan, P., and Akkas, N.: A probabilistic assessment of the seismic hazard in Turkey, *Tectonophysics*, 117, 295–344, 1985.
- Ergun, E., Tasgetiren, S., and Topcu, M.: Determination of SIF for patched crack in aluminum plates by the combined finite element and genetic algorithm approach, *Fatigue Fract. Eng. M.*, 31, 929–936, 2008.
- Goldberg, D. E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison–Wesley Publishing, Reading MA, 1989.
- Gülkan P. and Kalkan, E.: Attenuation modelling of recent earthquakes in Turkey, *J. Seismol.*, 6, 397–409, 2002.
- Güllü H. and Erçelebi, E.: A neural network approach for attenuation relationships: an application using strong-ground-motion data from Turkey, *Eng. Geol.*, 93, 65–81, 2007
- Holland, J. H.: *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
- İnan, E., Çolakoğlu, Z., Koç, N., Bayülke, N., and Çoruh, E.: Earthquake Catalog for the records between 1976–1996, Disaster Affairs Directorate, Ankara, 1996 (in Turkish).
- Kalkan, E. and Gülkan, P.: Site-dependent spectra derived from ground-motion records in Turkey, *Earthq. Spectra*, 4, 1111–1138, 2004.
- Kaplan, H., Yilmaz, S., Akyol, E., Sen, G., Tama, Y. S., Cetinkaya, N., Nohutcu, H., Binici, H., Atımtay, E., and Sarisin, A.: 29 October 2007, Çameli earthquake and structural damages at unreinforced masonry buildings, *Nat. Hazards Earth Syst. Sci.*, 8, 919–926, doi:10.5194/nhess-8-919-2008, 2008.

- Kramer, S. L.: Geotechnical Earthquake Engineering, Prentice-Hall., New Jersey, 1996.
- Kocuyigit, A.: Superimposed basins and their relations to the recent strike-slip fault zone: a case study of the Refahiye superimposed basin adjacent to the north anatolian transform fault, northeastern Turkey, *Int. Geology Rev.*, 38, 701–713, 1996.
- Nelson, M. and Illingworth, W. T.: A Practical Guide to Neural Nets, Addison-Wesley, Reading MA, 1990.
- Sen, G. and Akyol, E.: A genetic-algorithm approach for assessing the liquefaction potential of sandy soils, *Nat. Hazards Earth Syst. Sci.*, 10, 685–698, doi:10.5194/nhess-10-685-2010, 2010.
- Swingler, K.: Applying Neural Networks: a Practical Guide, Academic Press, New York, 1996.
- Ulusay, R., Tuncay, E., Sonmez, H., and Gokceoglu, C.: An attenuation relationship based on Turkish strong motion data and iso-acceleration map of Turkey, *Eng. Geol.*, 74, 265–291, 2004.
- Tung, C. P., Tang, C. C., and Lin, Y. P.: Improving groundwater-flow modelling using optimal zoning methods, *Environ. Geol.*, 44, 627–638, 2003.
- Westaway, R.: Kinematics of the Middle East and eastern Mediterranean updated, *Turk. J. Earth Sci.*, 12, 5–46, 2003.