4th International Conference on
Earthquake Geotechnical Engineering
June 25-28, 2007
Paper No. 1313

GENETIC ALGORITHM (GA): A NEW APPROACH IN ESTIMATING STRONG GROUND MOTION ATTENUATION RELATIONS

Gholamreza SOBHANINEJAD [1], Assadollah NOORZAD[2], Anooshirvan ANSARI[3]

ABSTRACT

In establishing different attenuation relationships, the conventional procedure is to use prevalent optimization methods mainly based on calculation of gradient vector. The main disadvantage of such methods is that all of them result in local minimums of error function. However, the main objective is to find the global minimum of error function. The objective of this paper is to compensate the main disadvantage of conventional optimization methods and to introduce a procedure for finding the global minimum of error function. Approaching this goal, the Genetic Algorithm (GA) is used to overcome this problem. In this paper, the coefficients of attenuation relationship are determined using the principles of Genetic Algorithm for horizontal and vertical peak ground acceleration based on 589 earthquake records in Europe and Middle East. In addition, the relationships obtained through the implementation of GA and those that are results of the conventional methods are compared.

Keywords: Earthquake, Attenuation Relation, Nonlinear Regression, Genetic Algorithm

INTRODUCTION

Empirical equations for predicting strong ground motion are typically fitted to the strong-motion data by means of the statistical regression methods which mainly are based on the optimizing a linear or a non-linear error function. The most prevalent methods proposed and used in (Joyner and Boore,1993a), (Joyner and Boore,1993b), (Boore et al.,1993) and (Boore et al.,1997) could be summarized as follows:

- one or two stage regression
- one or two stage Maximum Likelihood
- one or two stage weighted regression
- one or two stage weighted Maximum Likelihood

In aforementioned methods, the first step is defining an error measure for regression. Usually Mean Square Error is used as an error measure. In this step, an optimization problem must be solved respect to unknown coefficients of attenuation relation. There are two approaches in solving an optimization problem: local and global optimization (Nelles, 2001). In the previous works, for estimating ground motion attenuation relation local optimization techniques are used which mainly are based on the gradient. Two predominant disadvantages of local optimization techniques are first, these techniques are highly dependent on initial value of unknown parameters, and secondly some of those methods need the linearization of the error function that usually is done by means of Taylor expansion around

[1] M.Sc. Student Civil Engineering Department, University of Tehran, Tehran, Iran, Email: grsobhani@gmail.com
[2] Assistant Professor, Earthquake Research Center, University of Tehran, Tehran, Iran, Email: Noorzad@ut.ac.ir
[3] PhD Candidate, Earthquake Research Center, University of Tehran, Tehran, Iran, Email: hansari@ut.ac.ir

an unknown point. Regarding the above disadvantages, the method may result in finding a local minimum of the error function.
To overcome the above disadvantages a global optimization technique (Genetic Algorithm (GA)) is utilized here which is independent of initial values and linearization of error function. A detailed explanation of the GA could be found in (Nelles,2001). In order to prove the potential of proposed method the data and attenuation relation forms used by Ambrasyes et al. in (Ambraseys and Douglas,2005a) and (Ambraseys and Douglas,2005b) are employed.

AN OVERVIEW ON GENETIC ALGORITHM

In the 1950s and 1960s, several computer scientists independently studied evolutionary systems with the idea that evolution could be used as an optimization tool for engineering problems. In Goldberg's short history of evolutionary computation ((Goldberg,1989), Chapter 4), the names of Box (Box,1957), (Fraser,1957a), (Fraser,1957b), Friedman (Friedman,1959), Bledsoe (Bledsoe,1961), and Bremermann (Bremermann,1962) are associated with a variety of work in the late 1950s and early 1960s, some of which presages the later development of GAs. These early systems contained the rudiments of evolution in various forms all had some kind of selection of the "t-test". Some had population-based schemes for selection and variation, and some, like many GAs, had binary strings as abstractions of biological chromosomes.

In the later 1960s, Rechenberg introduced "evolution strategies", a method first designed to optimize real-valued parameters (Rechenberg,1973). This idea was further developed by Schwefel (Schwefel,1975). (Schwefel,1977), and the field of evolution strategies has remained an active area of research, developing in parallel to GA research, until recently when the two communities have begun to interact. For a review of evolution strategies, see (Back et al.,1991). Also in the 1960s Fogel, Owens, and Walsh developed "evolutionary programming" (Back et al.,1991). Candidate solutions to given tasks are represented as finite-state machines, and the evolutionary operators are selection and mutation. Evolutionary programming also remains an area of active research. For a recent description of the work of Fogel et al., see (Fogel,1992).

GAs as they are known today were first described by John Holland in the 1960s and further developed by Holland and his students and colleagues at the University of Michigan in the 1960s and 1970s. Holland's 1975 book Adaptation in Natural and Artificial Systems (Holland,1977) presents the GA as an abstraction of biological evolution and gives a theoretical framework for adaptation under the GA. Holland's GA is a method for moving from one population of "chromosomes" (e.g. bit strings representing organisms or candidate solutions to a problem) to a new population, using selection together with the genetic operators of crossover, mutation, and inversion. Each chromosome consists of "genes" (e.g., bits), with each gene being an instance of a particular "allele" (e.g., 0 or 1 or real value). Selection chooses those chromosomes in the population that will be allowed to reproduce, and decides how many offspring each is likely to have, with the fitter chromosomes producing on average more offspring than less fit ones. Crossover exchanges subparts of two chromosomes (roughly mimicking sexual recombination between two single-chromosome organisms); mutation randomly changes the values of some locations in the chromosome; and inversion reverses the order of a contiguous section of the chromosome, thus rearranging the order in which genes are arrayed in the chromosome. Inversion is rarely used in today's GAs, at least partially because of the implementation expense for most representations. A simple form of the GA (without inversion) works as follows:
1. Start with a randomly generated population of chromosomes (e.g., candidate solutions to a problem).
2. Calculate the fitness of each chromosome in the population.
3. Apply selection and genetic operators (crossover and mutation) to the population to create a new population.
4. Go to step two.

This process is iterated over many time steps, each of which is called a "generation". After several generations, the result is often one or more highly fit chromosomes in the population. It should be noted that the above description leaves out many important details. For example, selection can be implemented in different ways it can arbitrarily eliminate the least 50% of the population and replicate every other individual once, it can replicate individuals in direct proportion to their fitness (fitness-proportionate selection), or it can scale the fitness and replicate individuals in direct proportion to their scaled fitness. For implementation details such as these, see (Goldberg,1989) and (Nelles,2001). In continuation, the used data and the proposed attenuation model in estimating attenuation relation are presented.

PROPOSED MODEL AND USED DATA

EARTHQUAKE DATA

The choice of which records to include and which to exclude from the analysis is one of the most substantial decisions in deriving ground motion's attenuation relations. Since the main goal of this paper is proposing a method for estimating ground motion's attenuation relation, the database that Ambraseys et al. used in (Ambraseys and Douglas, 2005a) is applied for this study. Since the reliable data source was unavailable for six records, only 589 records from 595 records are used here which Ambraseys used in (Ambraseys and Douglas, 2005a) and (Ambraseys and Douglas, 2005b). Six records that omitted from database are tabulated in Table 1; five of them are for Syria (SY) and one for Greek (GR). Complete database is published in (Ambraseys and Douglas, 2005a) .

Table 1: Omitted data from database

Date	Time (UTC)	Co.	h (Km)	Mw	Mechanism	Station	Co.	Site	d (Km)
15/06/1995	0:15:51	GR	10	6.5	N	Patra-San Dimitrios Church	GR	A	35
24/12/1996	22:16:26	SY	29	5.5	S	Souana	SY	S	55
22/01/1997	17:57:20	TU	15	5.7	O	Batrach	SY	S	52
22/01/1997	17:57:20	TU	15	5.7	O	Karkour	SY	S	66
22/01/1997	18:24:51	TU	10	5.1	O	Batrach	SY	S	36
22/01/1997	18:24:51	TU	10	5.1	O	Karkour	SY	S	49

In following section a brief description about database used in this article presented and complementary information could be found in (Ambraseys and Douglas, 2005a).

Magnitude

The moment magnitude (Mw) based on Kanamori relation (Kanamori, 1977) is used as magnitude scale in this study. The moment magnitude is computed from seismic moment (M_0) based on Equation (1).

$$M_W = 2/3\log(M_O) - 10.7 \tag{1}$$

Only earthquakes with published seismic moment are used in this study. Because of inherent uncertainties in other equation that convert other magnitude scales (M_s, M_l, etc.) to moment magnitude, they are neglected. Since the moment magnitude usually is used for measuring moderate to large earthquakes, the magnitude of earthquakes which is considered here are those larger than 5 ($M_W \geq 5$). The other reason for selecting this lower limit is that earthquake with less magnitude than 5 are unimportant in engineering viewpoint.

Distance

The distance to the surface projection of the fault (Joyner and Boore, 1981), r_{jb} (also known as fault distance or Joyner–Boore distance) is used as the distance metric. For earthquakes where the location of the causative fault has not been reported, mainly earthquakes with $M_w \leq 6$, epicentral distance, r_{seis} is used instead. For small earthquakes, r_{seis} and $_{rjb}$ are similar because of the small rupture planes of such earthquakes. The definition of r_{jb} and r_{seis} is illustrated in Figure 1. The lower limit of distance is 1 km and upper bound is 100 km.

Faulting mechanism and Local site condition

For classifying faulting mechanism the definition presented by Frohlich et al.(1992) at (Frohlich, 1992) is used, and the fault mechanism is categorized in four categories: Normal (N), Strike Slip(S), Thrust (T) and the unknown mechanism named Odd (Odd). Also for site classification the method based on mean shear velocity at 30 m depth ($V_{s,30}$) is used which is proposed by (Boore et al., 1993). In table 2 the ranges of $V_{s,30}$ is given for soil categories.

Table 2: Local site condition

Site Class	Rock(R)	Stiff Soil(A)	Soft Soil(S)	Very Soft Soil(L)
$V_{s,30}$ range	$V_{s,30} > 750\ m/s$	$360 < V_{s,30} \leq 750\ m/s$	$180 < V_{s,30} \leq 360\ m/s$	$V_{s,30} \leq 180\ m/s$

More information about site classification and faulting mechanism could be found in (Ambraseys and Douglas, 2005a). In addition, those earthquakes with unknown site class or fault mechanism are not considered in the database.

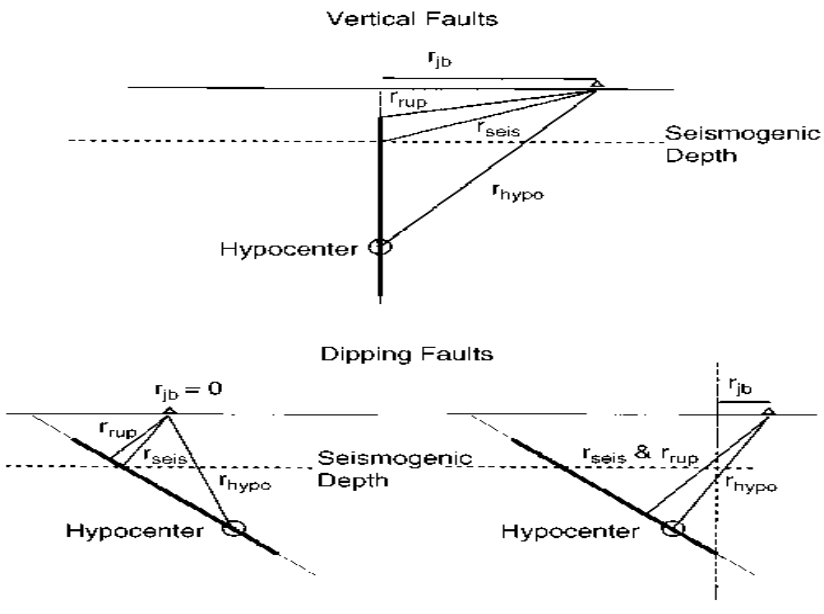

Figure 1: Definition of r_{jb} and r_{seis}

Summary of used data

In this section, a summary of selected record presented. Adding up, 589 records are used for estimating horizontal and vertical peak ground acceleration. In Table 3 a list of countries and number of records, participating in attenuation relation is presented. The distribution of records with respect to magnitude (M_w) and distance (r_{jb}) are illustrated in Figure 2 and 3.

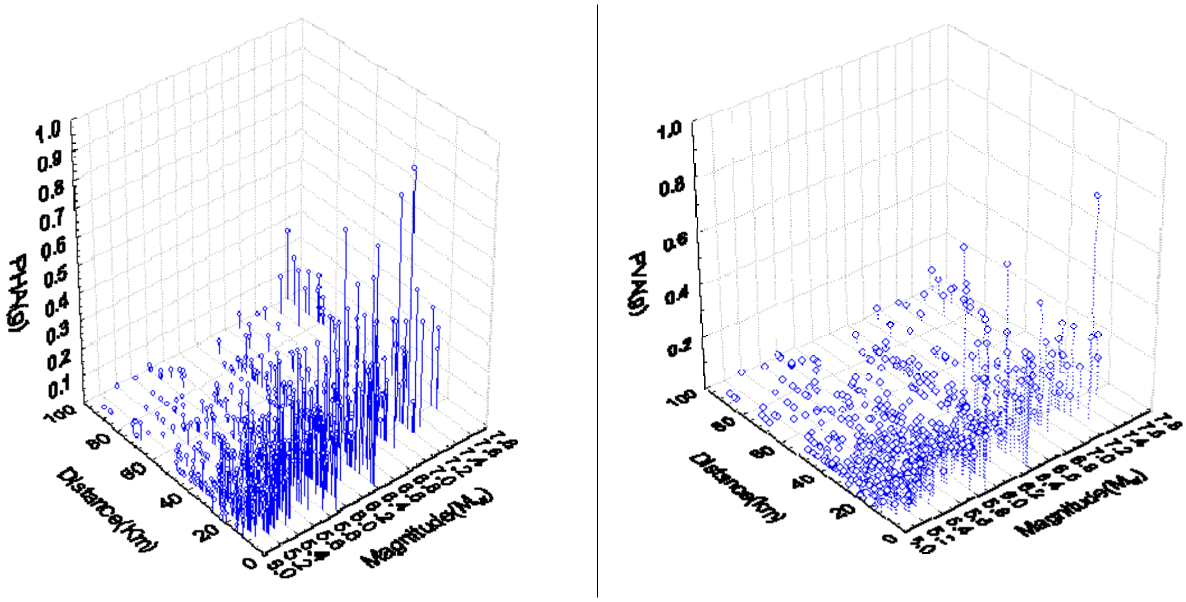

Figure 2: Distribution of PHA and PVA

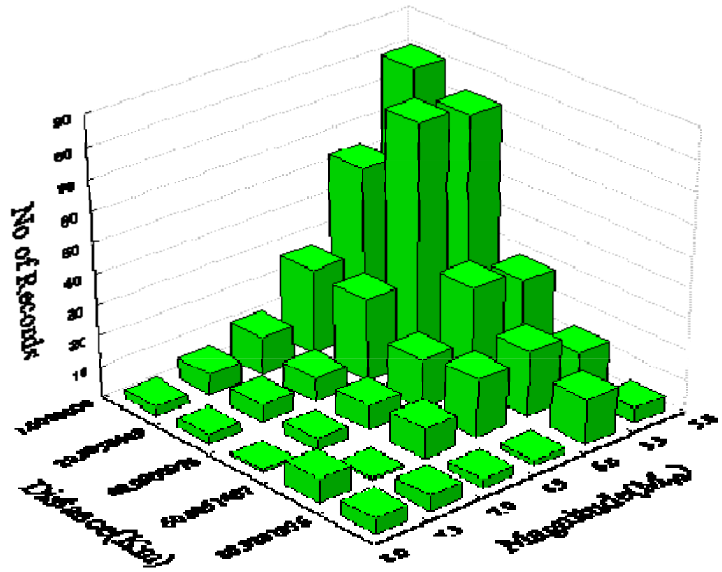

Figure 3: Distribution of Records

FUNCTIONAL FORM AND FITNESS FUNCTION

As mentioned before the main purpose of this study is to present a method for estimation coefficients of strong ground motion attenuation relation; therefore we used the proposed functional form, which is suggested by Ambrasyes et al. in (Ambraseys and Douglas, 2005a) and (Ambraseys and Douglas,2005b). The Equation (2) illustrates the attenuation functional form used in this study and (Ambraseys and Douglas, 2005a), (Ambraseys and Douglas,2005b).

$$\log y = a_1 + a_2 M_w + (a_3 + a_4 M_w)\log\sqrt{{r_{jb}}^2 + a_5^2} + a_6 S_S + a_7 S_A + a_8 F_N + a_9 F_T + a_{10} F_O \quad (2)$$

where S_S =1 for soft soil sites and 0 otherwise, S_A =1 for stiff soil sites and 0 otherwise, F_N =1 for normal faulting earthquakes and 0 otherwise, F_T =1 for thrust faulting earthquakes and 0 otherwise and F_O =1 for odd faulting earthquakes and 0 otherwise. The r_{jb}, defined previously, is Joyner–Boore distance measure and M_w is moment magnitude. Y is ground motion parameter that considered in this study as Peak Horizontal Acceleration (PHA) or Peak Vertical Acceleration (PVA).

For measuring the precision of the regression, the error function (cost function or fitness function) is defined as Unbiased Mean Square Error (UMSE) (Equation (3)).

$$E = \frac{1}{P-v}\sum_{p=1}^{P}(Y_p - \overline{Y}_p)^2 \qquad (3)$$

Where P is the number of data (here is 589), v is the regression model's degree of freedom (regard to the assumed functional form v is 10), Y_P is the real value of the ground motion parameter (PVA or PHA) and $\tilde{Y}_P$ is the estimated value of ground motion parameter by Equation (2). In next section, the error function (Equation (3)) is minimized and the a_i coefficients in Equation (2) are estimated by means of GA.

RESULTS AND COMPARISON

The GA program prepared by Sobhaninejad (Sobhaninejad, 2006) is used here to compute coefficients of Equation (2). Total number of generation is selected such that error measure stands constant for half of the generations. Hence, the 5000 generations are selected as a maximum number of generations for PHA and PVA. Twenty chromosomes are chosen as an initial population of the genetic algorithm and the best-fitness selection method is applied for offspring generation. For mutation of chromosomes, the uniform random selection is used. The Heuristic crossover algorithm is employed here for generation new offspring. The coefficients of Equation (2), estimated by proposed method and Ambrasyes (Ambraseys and Douglas, 2005a and Ambraseys and Douglas, 2005b), are illustrated in Table 4.

Table 4: Coefficients of Equation (2)

Coefficients of Equation (2)		a_1	a_2	a_3	a_4	a_5
This Study	PHA	-0.703	0.392	-0.598	-0.100	-7.063
Ambrasyes (Ambraseys and Douglas, 2005a)	PHA	2.522	-0.144	-3.18	0.314	7.6
This Study	PVA	0.495	0.027	-2.83	0.235	7.181
Ambrasyes (Ambraseys and Douglas, 2005b)	PVA	0.835	0.083	-2.49	0.206	5.6
Coefficients of Equation (2)		a_6	a_7	a_8	a_9	a_{10}
This Study	PHA	0.186	0.125	0.082	0.012	-0.038
Ambrasyes (Ambraseys and Douglas, 2005a)	PHA	0.137	0.05	-0.084	0.062	-0.044
This Study	PVA	1.150	1.103	-0.074	0.065	-0.170
Ambrasyes (Ambraseys and Douglas, 2005b)	PVA	0.078	0.046	-0.126	0.005	-0.082

The Unbiased Mean Square Error (UMSE) of the proposed method is compared in Table 5 with UMSE of the method used by Ambraseys in (Ambraseys and Douglas, 2005a). The proposed method has less UMSE than proposed equations in (Ambraseys and Douglas,2005a) and (Ambraseys and Douglas,2005b). Regarding to Table 5, the UMSE of PHA component 26 percent and PVA component 16.66 percent decreased. In addition, the mean of the error measure for each component (PHA/PVA) for both equations is estimated in table 5. As it is clear, the falling rate of this measure is significant. From Table 5 could be concluded that the proposed method(based on using GA as optimizer) is more efficient than usual methods used in estimating coefficients of ground motion attenuation (Gradient Based Methods).

Table 5: Comparison of error measures in propose method and Ambrasyes (Ambraseys and Douglas,2005a) , (Ambraseys and Douglas,2005b)

		UMSE	Average OF ERROR
This Study	PHA	0.937	0.023
Ambrasyes (Ambraseys and Douglas,2005a)		1.267	0.0328
Error Decreased		%26	%30
This Study	PVA	0.577	0.002
Ambrasyes (Ambraseys and Douglas,2005b)		0.692	0.174
Error Decreased		%16.66	%99

The attenuation relation for rock site and strike slip faulting mechanism is illustrated in Figure (4) in comparison with the equations proposed by Ambraseys (Ambraseys and Douglas, 2005a, Ambraseys and Douglas, 2005b). As mentioned before, the suggested method is able to find the global minimum of the error measure function. In addition, from the instances in Table 4 and 5 could be concluded the method used in (Ambraseys and Douglas, 2005a) and (Ambraseys and Douglas, 2005b) for determination of coefficients in Equation (2) may be entrapped into local minimum of the error function. The effect of proposed method on the attenuation form is shown in Figure (4).

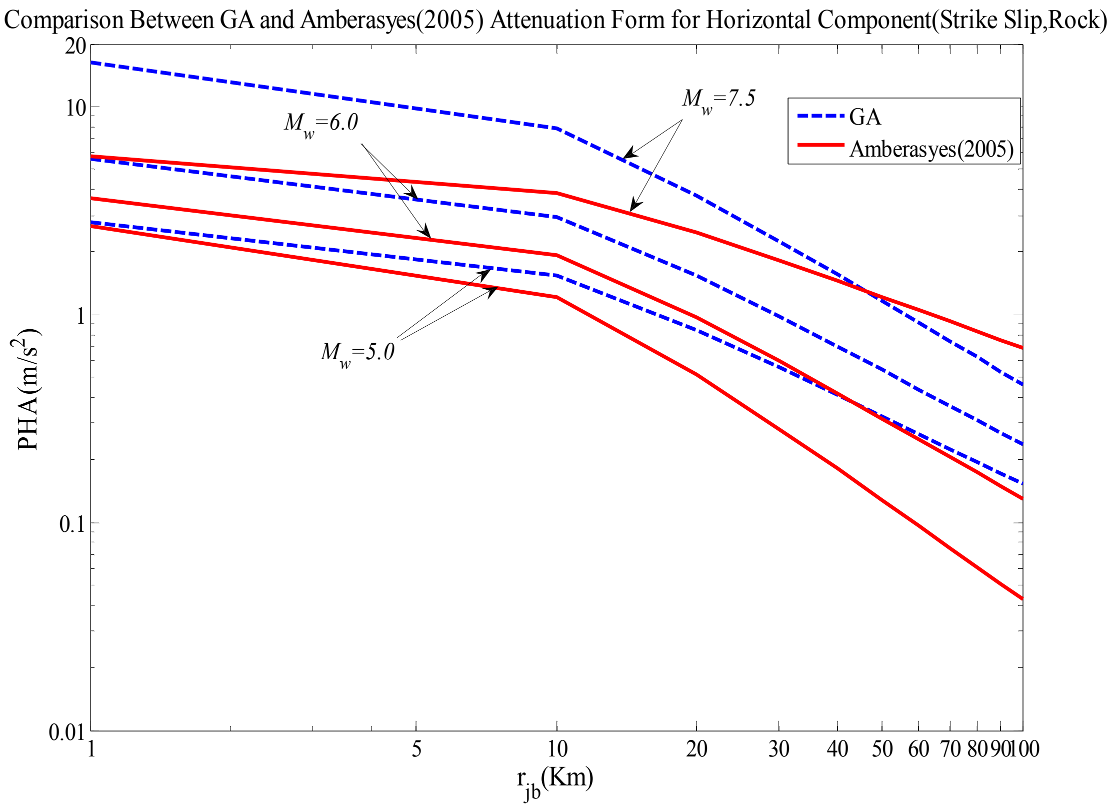

Figure 4: PHA attenuation for Rock and Strike Slip Fault

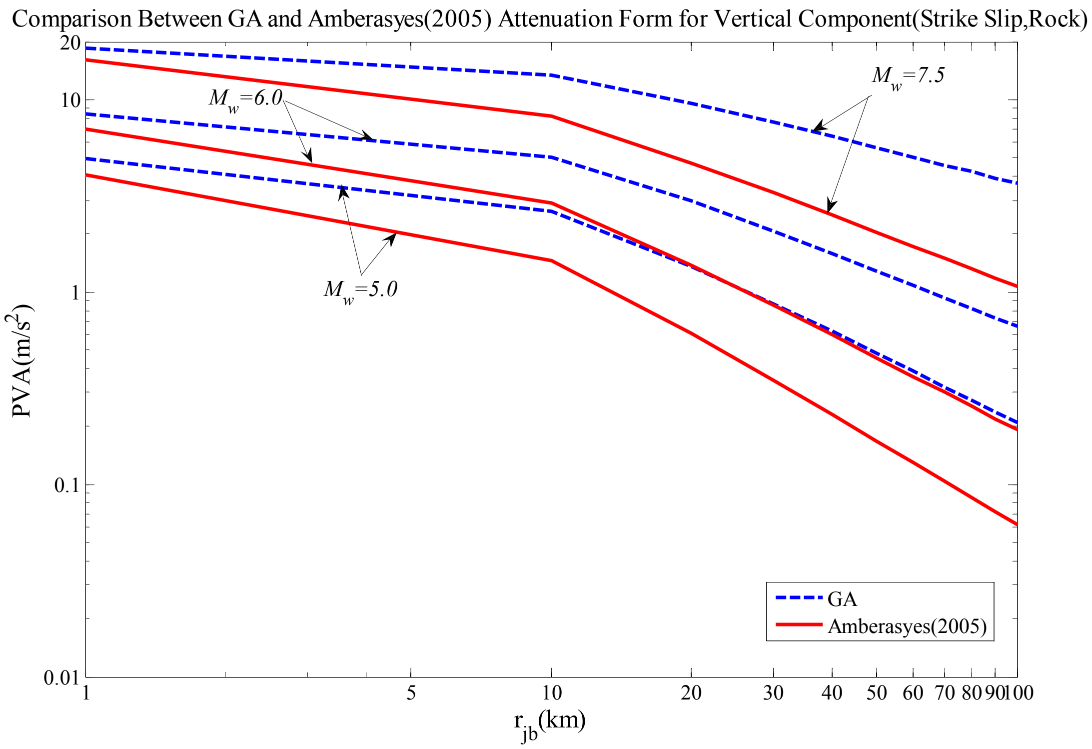

Figure 5: PVA attenuation for Rock and Strike Slip Fault

CONCLUSION

In this study, a method is proposed for estimating coefficients of ground motion's attenuation relation based on global minimization concept (Genetic Algorithm). For comparison the proposed method with prevalent methods, a total number of 589 records from Europe and Middle East caused by shallow crustal earthquakes with magnitude Mw≥5 and distance to the surface projection of the fault less than 100 km is selected which was used by Ambraseys in (Ambraseys and Douglas, 2005a) and (Ambraseys and Douglas, 2005b). The GA is applied on 589 data sets for horizontal and vertical components to estimate the coefficients of attenuation relation. The results show 26 percent decrease in UMSE error for horizontal component and 16.66 percent decrease in UMSE error for vertical component in comparison with those equations proposed by Ambraseys et al. in (Ambraseys and Douglas, 2005a) and (Ambraseys and Douglas, 2005b). In addition to decreasing the error measure, the coefficients of the relation significantly altered (see Table 4). The coefficients estimated by means of local optimization techniques overestimate the global ones somewhere in magnitudes and distances, and underestimate elsewhere (Figure 4).

REFERENCES

Ambraseys, N.N. and Douglas, J,"The Equations for the Estimation of Strong Ground Motions from Shallow Crustal Earthquakes Using Data from Europe and the Middle East: Vertical Peak Ground Acceleration and Spectral Acceleration." Bulletin of Earthquake Engineering 3:55–73, 2005.

Ambraseys, N.N. and Douglas, J., "The Equations for the Estimation of Strong Ground Motions from Shallow Crustal Earthquakes Using Data from Europe and the Middle East: Vertical Peak Ground Acceleration and Spectral Acceleration", Bulletin of Earthquake Engineering 3:55–73, 2005.

A.S. Fraser., "Simulation of genetic systems by automatic digital computers: IEE soft linkage on rates of advance under selection", Australian Journal of Biological Science, 10:492{499}, 1957.

A.S. Fraser.,”Simulation of genetic systems by automatic digital computers: I: introduction”, Australian Journal of Biological Science, 10:484{491}, 1957.

Boore, D. M., Joyner, W. B., & Fumal, T. E. , “Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report.”, U.S. Geological Survey Open-File Report 93-509, 70 pages, 1993.

Boore, D. M., Joyner, W. B., & Fumal, T. E., “Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work.” , Seismological Research Letters, 68(1), 128–153. 1997.

D. B. Fogel, “Evolving Artificial Intelligence” PhD thesis, University of California, San Diego, CA, 1992.

D. E. Goldberg. “Genetic Algorithms in Search, Optimization, and Machine Learning”, Addison-Wesley, Reading, MA, 1989.

Frohlich, C. and Apperson, K.D., “Earthquake focal mechanisms, moment tensors, and the consistency of seismic activity near plate boundaries”, Tectonics 11(2), 279–296, 1992.

G. E. P. Box. “Evolutionary operation: A method for increasing industrial productivity”, Journal of the Royal Statistical Society C, 6(2):81{101, 1957.

G. J. Friedman, “Digital simulation of an evolutionary process”, General Systems Year-book, 4:171{184, 1959.

Gh. Sobhaninejad, “Using Soft Computation Techniques(ANFIS) and Biological Optimization(GA) for Estimating Strong Ground Motion Attenuation Relation”, M.Sc. Theses, Civil Engineering Department, University of Tehran, Tehran, Iran, 2006

H. J. Bremermann, “Optimization through evolution and recombination”, In M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, editors, Self-organizing systems, pages 93{106}, Washington, D.C., 1962. Spartan Books.

H.-P. Schwefel, “Evolutionsstrategie und numerische Optimierung”, PhD thesis, Technische University at Berlin, Berlin, 1975.

H.-P. Schwefel, “Numerische Optimierung von Computer-Modellen mittels der Evolu-tionsstrategie”, volume 26 of Interdisciplinary Systems Research. Birkhauser, Basel, 1977.

H. Kanamori, “The energy release in great earthquakes”. Journal of Geophysical Research 82(20), 2981–2987, 1977.

I. Rechenberg, “Evolutionstrategie: optimierung technischer systeme nach prinzipien der biologischen evolution”, Frommann-Holzboog, Stuttgart, 1973.

J. H. Holland, “Adaptation in Natural and Artificial Systems”, MIT Press, Cambridge,MA, 1992. Second edition (First edition, 1975).

Joyner, W.B. and Boore, D.M., “Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake”, Bulletin of the Seismological Society of America 71(6), 2011–2038, 1981.

Joyner, W. B., & Boore, D. M. ,“Methods for regression analysis of strong-motion data.” , Bulletin of the Seismological Society of America, 83(2), 469–487, 1993.

Joyner, W. B., & Boore, D. M. , “ Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake.”, Bulletin of the Seismological Society of America, 71(6), 2011–2038, 1993.

O. Nelles, “Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models”, Springer-Verlag Berlin Heidelrberg New York Publication, Germany, 2001

T. Back, F. Homeister, and H.-P. Schwefel, “A survey of evolution strategies”, In R. K.Belew and L. B. Booker, editors, Proceedings of the Fourth International Conference on Genetic Algorithms, pages 2{9, San Mateo, CA, 1991. Morgan Kaufmann.

W. W. Bledsoe, “The use of biological concepts in the analytical study of systems”, November 1961. Paper presented at the ORSA-TIMS National Meeting, San Francisco, CA.