Selection and Scaling of Ground Motion Time Histories for Structural Design Using Genetic Algorithms

Farzad Naeim,[a] M.EERI, **Arzhang Alimoradi,**[b] M.EERI, **and Shahram Pezeshk,**[b] M.EERI

This paper presents a new approach to selection of a set of recorded earthquake ground motions that in combination match a given site-specific design spectrum with minimum alteration. The scaling factors applied to selected ground motions are scalar values within the range specified by the user. As a result, the phase and shape of the response spectra of earthquake ground motions are not tampered with. Contrary to the prevailing scaling methods where a preset number of earthquake records (usually between a single component to seven pairs) are selected first and scaled to match the design spectrum next, the proposed method is capable of searching a set consisting of thousands of earthquake records and recommending a desired subset of records that match the target design spectrum. This task is achieved by using a genetic algorithm (GA), which treats the union of 7 records and corresponding scaling factors as a single "individual." The first generation of individuals may include a population of, for example, 200 records. Then, through processes that mimic mating, natural selection, and mutation, new generations of individuals are produced and the process continues until an optimum individual (seven pairs and scaling factors) is obtained. The procedure is fast and reliable and results in records that match the target spectrum with minimal tampering and the least mean square of deviation from the target spectrum. [DOI: 10.1193/1.1719028]

INTRODUCTION

Nonlinear time-history analysis is becoming more common in seismic analysis and design of structures. Code provisions governing design of seismic isolated structures, for example, have included nonlinear time-history analysis provisions for over a decade (see Naeim and Kelly, 1999). Modern seismic evaluation guidelines such as *FEMA-356* (ASCE 2000) contain detailed and elaborate provisions for performing nonlinear analysis for all kinds of building structures.

Since traditionally the seismic hazard at a site for design purposes has been represented by design spectra, virtually all seismic design codes and guidelines require scaling of selected ground motion time histories so that they match or exceed the controlling design spectrum within a period range of interest (e.g., ICC 2000, ASCE 2000).

[a] John A. Martin & Associates, Inc., 1212 S. Flower Street, Los Angeles, CA 90015
[b] The University of Memphis, Civil Engineering Department, Memphis, TN 38152

Earthquake Spectra, Volume 20, No. 2, pages 413–426, May 2004;

A typical code or guideline provision would require scaling of the two horizontal components of each ground motion (called a data set) such that the average square root of the sum of the squares (SRSS) of the 5% damped response spectra of the data set used does not fall below α times the 5% damped design spectrum for periods between T_0 and T_n. Typical value of α is either 1.3 or 1.4. For conventional buildings, T_0 and T_n are usually assigned values such as $0.2T$ and $1.5T$ where T is the fundamental period of the structure. For seismic isolated structures, some codes provide for a narrower range of matching around the fundamental period, T. Generally, for nuclear power plants and other critical facilities a broader range of matching is used.

Several methods of scaling time histories have been proposed. These include frequency-domain methods where the frequency content of the recorded ground motions are manipulated in order to obtain a match (Gasparini and Vanmarcke 1976, Silva and Lee 1987, Bolt and Gregor 1993, Department of the Army 2000, Carballo and Cornell 2000) and time-domain methods which limit themselves to manipulating only the amplitude of the recorded ground motions (Kircher 1993, Naeim and Bhatia 2000).

Regardless of the method (frequency-domain or time-domain), in virtually all of the existing approaches, the processes of selecting earthquake ground motions and their scaling to match the design spectrum are separate and distinct. In other words, first one or more time histories are selected, and then appropriate scaling mechanisms for spectrum matching are applied. This is not the case for the method proposed in this paper where, as will be illustrated later, the search for appropriate time histories and corresponding scaling factors are completely intertwined and parts and parcels of a single process.

Genetic algorithms (GA) have proven themselves as reliable computational search and optimization procedures for complex objectives involving large number of variables. In structural and earthquake engineering, during the past decade, genetic algorithms have been used in design optimization of nonlinear structures (Pezeshk et al. 1999, 2000) active structural control (Alimoradi 2001), and performance-based design (Foley and Schinler 2001, Foley et al. 2003). Therefore, using genetic algorithms to scale earthquake ground motions for design is but a natural continuation of such applications and parallels the recent use of neural networks to achieve the same objective (Ghaboussi and Lin 1998, Kim and Ghaboussi 1999).

This paper presents a scaling procedure based on genetic algorithms for the purpose of closely approximating a given target spectrum over a range of periods and tolerances specified by the user.

GENETIC ALGORITHM BASICS

A genetic algorithm is a computer simulation of the natural evolutionary processes in order to solve search and optimization problems. The early thoughts of simulating adaptable systems on machines go back to the premature stages of computer software and hardware development (see Levy, 1992). It has taken a long time, however, for this subject to become mature enough to be used as a practical tool. The pioneering work by Goldberg (1989) and others and the availability of high-speed computers have paved the way for application of genetic algorithms in engineering.

The power of the genetic algorithm is inherent in its capability to adapt. In natural systems, species adapt to the environment through successive interactions and generations subject to the environment. After several consecutive generations, only those species that can adapt well to the environment survive and the rest disappear. In mathematical terms, individuals are analogous to problem variables and environment is the stated problem. The final generation of the variable strings that can adapt to the problem is the solution. Genetic algorithms provide the necessary tools to mimic this natural process.

The basic elements of a genetic algorithm as applied to this problem are as follows:

1. *Population*: This is a set of assumed solution variables. In most applications there are tens to thousands of "individuals" in the population. These individuals are binary strings that are evaluated after decoding to real or integer numbers that represent the problem variables for "natural parent selection." The initial population is usually produced randomly. The offspring generations are reproduced by applying the genetic algorithm operators (crossover and mutation) to the population of parents.
2. *Fitness function*: This is a mathematical expression to evaluate the fitness of individuals in a generation. The basic rule in defining a fitness function is that it should yield higher values for individuals that are closer to the optima. As a result, those individuals that are fitter would receive a higher chance of being picked as a parent for the next generation.
3. *Crossover*: This is the procedure by which two individuals mate to reproduce the offspring individuals. This is done by switching and sharing segments of the parent characteristics. Several patterns of crossover have been introduced such as single point, multiple point, and uniform crossover (see Camp et al., 1998; Pezeshk et al., 2000). In this study a single-point crossover pattern is used.
4. *Mutation*: This is the necessary mechanism to ensure diversity in the population. When an individual is selected randomly to undergo mutation (by enabling the mutation probability), the algorithm flips a randomly selected bit along the length of a sub-entity from zero to one or vice versa to prevent a fixed pattern of solutions from being propagated through all forthcoming generations. This is essential in providing a broader search within the whole search space. Very high mutation probabilities, however, can elongate the process of adaptation and convergence to the optima.
5. *Natural parent selection*: This is a probabilistic method of selection based on the fitness of the individuals. To ensure the survival of the fittest, the individuals that have higher fitness function values receive a higher chance of being selected as the parents of the offspring generation.

SCALING EARTHQUAKE GROUND MOTIONS USING A GENETIC ALGORITHM

The objective of this research is twofold. The first is to use a genetic algorithm to find the best combination of strong ground motion records and the corresponding time-domain scaling factors from a large database of earthquake records to minimize the difference between a given design spectrum (target) and the average of scaled ground mo-

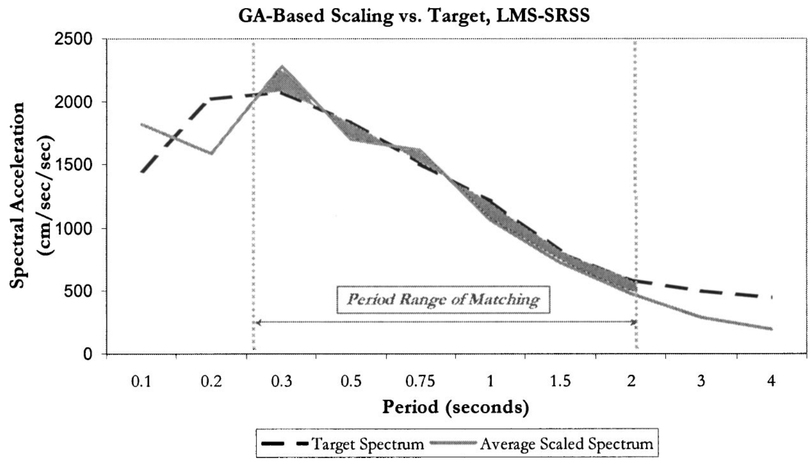

Figure 1. Graphical representation of the first optimization problem (minimization of the hatched areas).

tions. The deviation from the target is measured by the mean square of error between the square root of the sum of the squares (LMS-SRSS) of the average scaled spectrum and the target (see Figure 1). The second is to modify this approach in order to select earthquake records and scaling factors that result in the scaled average spectrum that is above the target in the period range of T_0 to T_n. The search process is to obtain the best seven pairs of ground motion and corresponding scaling factors. There is, however, no built-in limitation on the number of earthquake record and scaling factor pairs that the algorithm may select.

The first problem is formulated as the minimization of the error function, Z, between the averaged scaled spectra and the target spectrum in a range of T_0 to T_n. The error function is defined as

$$Z = \min \left\{ \sum_{T=T_o}^{T_n} \left(\sqrt{\frac{\sum_{i=1}^{7} [S_i \cdot SA_i(T)]^2}{\sum_{i=1}^{7} S_i^2}} - F_T(T) \right)^2 \right\} \tag{1a}$$

in which

T = the fundamental vibration period of the structure

S_i = the scaling factor corresponding to record number i

$SA_i(T)$ = value of the spectral acceleration of record number i at period T

$F_T(T)$ = value of the target design spectrum at period T

T_o = initial period to consider (i.e., $0.2T$)

T_n=final period to consider (i.e., $1.5T$)

The optimization procedure is subject to

$$S_{\min} \leqslant S_i \leqslant S_{\max}$$

and

$$S_{\min}, S_{\max} > 0 \tag{1b}$$

where

$S_{\min}$=the lower bound of the acceptable scaling factors, and

$S_{\max}$=the upper bound of the acceptable scaling factors.

As mentioned earlier, this formulation does not guarantee that the final solution would not fall below the target in the period range under consideration; instead it would merely attempt to minimize the deviation of the solution from the target. The second formulation achieves this objective by adding another penalty function or constraint to the optimization problem:

$$\sqrt{\frac{\sum_{i=1}^{7} [S_i \cdot SA_i(T)]^2}{\sum_{i=1}^{7} S_i^2}} - F_T(T) \geqslant 0 \quad \text{for all periods} \quad T_o \leqslant T \leqslant T_n \tag{2}$$

A search space of earthquake records is needed for the genetic algorithm to select from. For this paper a set of 1,496 horizontal strong ground motion components were selected from the database compiled by Naeim and Anderson (1996). Obviously, any appropriate set of records may be used for the same purpose. It is worth mentioning that for a database of this size, the search space is huge. Setting aside the scaling factors, 1,496 records may be combined in groups of 7 records in more than 3×10^{18} different ways. Clearly, the use of conventional optimization techniques such as nonlinear programming would take an enormous number of computations and would not be feasible. Conversely, a genetic algorithm as demonstrated here can converge with a reasonable computing effort and a rather short computing time.

The operators of genetic algorithm are selected as follows:

- *Solution Variables/Population of Individuals*: Any arbitrary combination of 7 records and 7 scaling factors is defined as a single "individual" or chromosome (see Figure 2). The objective is to create the best individual using the pool of earthquake records in the database and scaling factors within the acceptable range specified by the user. Therefore, each individual has fourteen subdivisions to represent each variable (seven for identification of records in the database and seven for identification of the corresponding scale factors). We assigned a length of 10 binary digits to each subdivision making the total length of each individual equal to 140 binary digits. This, of course, can be changed and longer binary strings may be used to accommodate larger earthquake record databases. The

The binary strings: variables 1 through 14, Individual No. 100

Generation 1:

```
1001101000|0111100110|1110011100|0100110101|0111101011|0000010101|0000000101|
0001010111|1010110011|1011010111|1001101001|1000001101|1000000111|0101011000|
```

Generation 150:

```
0010101011|1111001010|1001001000|1100100101|0000111110|1101110000|1111111000|
0101010100|1011001001|0010111111|0101011111|1100000111|1100011011|1111111000|
```

Generation 300:

```
1000011100|0110110101|0010100010|1011010011|0001011010|0000111101|1001001011|
0010001100|1001111110|1010100111|1100001100|0011111110|1001001000|1011111101|
```

Set of decoded variables or solution vector: variables 1 through 14, Individual No. 100

Generation 1:

```
1,090 |  959  | 1,400 |  782  |  964  |  494  |  478  |
0.59  | 1.18  |  1.21 | 1.10  | 1.01  | 1.01  | 0.84  |
```

Generation 150:

```
644   | 1,440 | 1,060 | 1,280 |  535  | 1,350 | 1,490 |
0.83  |  1.20 |  0.69 |  0.84 | 1.26  | 1.28  | 1.49  |
```

Generation 300:

```
1,010 |  910  |  635  | 1,200 |  563  |  534  | 1,060 |
0.64  | 1.12  | 1.16  |  1.26 | 0.75  | 1.07  |  1.25 |
```

Figure 2. Binary strings (chromosomes) and decoded representation of an "individual."

first seven binary substrings provide the positions of the seven records in the database. The remaining seven substrings represent the corresponding scaling factors (Figure 2). Since the record numbers are integers and the scaling factors are real, the optimization method required a mixed integer-real process.

- *Fitness Function*: This function is defined as the reciprocal of the objective function (Equations 1a, 1b, and 2). Therefore, the lesser the error function for a given combination of selected records and scaling factors, the higher the fitness of the individual. The individuals may be penalized if their average scaled spectrum falls below the target. For these cases, a penalty function is defined to lower the fitness of the individual. The penalty function is proportional to the area under the target for the specific individual.

$$Fitness\_Function(j) = \frac{C_1}{\left\{ \sum_{T=T_o}^{T_n} \left(\sqrt{\frac{\sum_{i=1}^{7} [S_i \cdot SA_i(T)]^2}{\sum_{i=1}^{7} S_i^2}} - F_T(T) \right)^2 \right\} + [C_2 \cdot (A^-_{T_o \to T_n})]^2} \tag{3}$$

where

j = the individual number

$A^-_{T_o \to T_n}$ = the area of the scaled spectrum under the target

C_1=fitness scaling

C_2=0 if simple LMS or 1 if LMS and constraint on negative values is used.

We adapted and modified the backbone genetic algorithm routines from the LGADOS code placed in public domain by Coley (2001, 2002). We programmed the graphical user interface in Visual Basic and the interface between the user interface and the genetic algorithm in Fortran. An overall flowchart of the program operation is presented in Figure 3.

The input data consist of the ordinates of the target acceleration design spectrum, the period range for the matching, lower- and upper-bound acceptable values for scaling factors, and a set of GA parameters. The GA parameters consist of a population size, number of generations, crossover ratio, and mutation ratio. We have successfully used the default values, although other values may also prove promising.

- Acceptable scale factor range: 0.5 to 1.5
- Population of individuals: 200
- Number of generations: 300
- Crossover ratio: 0.65
- Mutation ratio: 0.025

The program is very fast and it takes only a few seconds for it to converge to an optimum solution on a typical personal computer. A typical screen showing the selected input and obtained results and the match between the target and the selected individual is presented in Figure 4.

EXAMPLES

Two examples are presented. The first example illustrates the application of the proposed method to select 7 records and corresponding scale factors to match a given target spectrum in the database of 1,496 records (Naeim and Anderson 1996). This example also illustrates the stability of the genetic algorithm in adapting itself to unusual target spectrum shapes. The second example illustrates the application of the proposed method to the limited task of only selecting the appropriate scale factors for a preselected set of seven pairs of earthquake records. This permits a comparison of the accuracy of the proposed method with other traditional time-domain approaches for scaling earthquake records.

EXAMPLE 1

The target spectrum for this example is shown in Figure 5. A building period of 1.26 seconds was assumed with a period range of 0.25 to 1.89 seconds for matching the target. A genetic search of a 200-individual population over 300 generations with a crossover ratio of 65.0% and a mutation probability of 2.5% was utilized. Acceptable range of scale factors was from to 0.5 to 1.5. The genetic algorithm selected 7 records and the corresponding scale factors shown in Table 1 as representing the best match.

The mean square of error between the average spectra of the scaled records and the target in the range of 0.25 to 1.89 seconds is 3.12%. This represents an excellent match

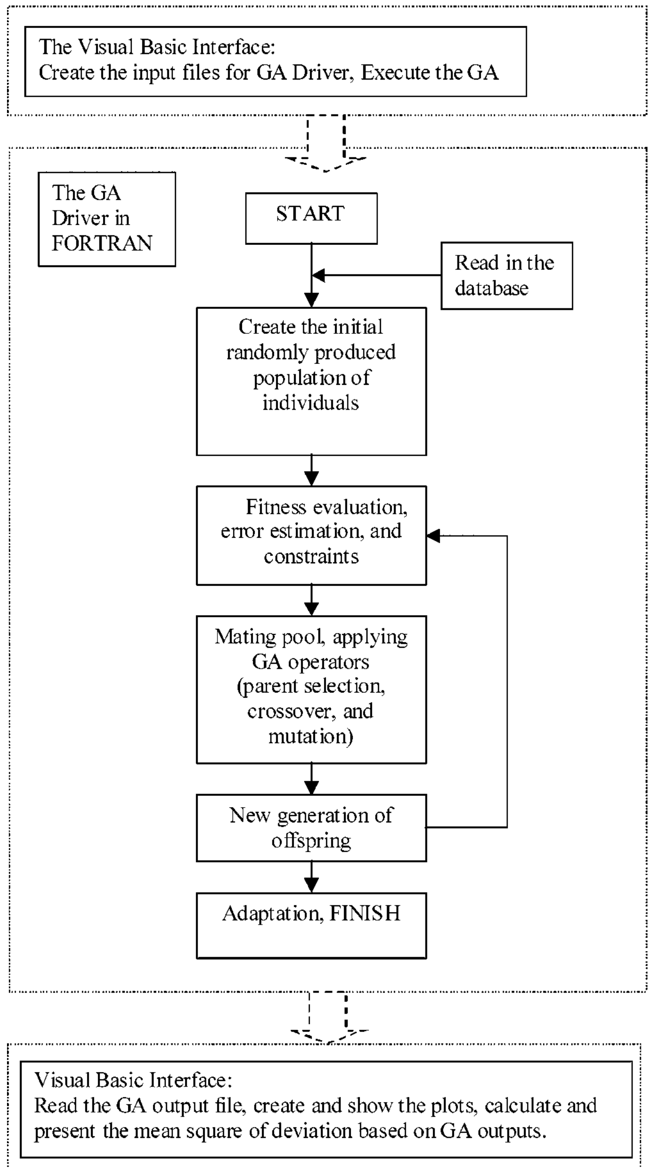

Figure 3. The program flowchart.

as can be observed in Figure 6 where the spectrum of individual scaled records is shown with narrow lines, the average of scaled records is indicated by a solid thick line and the target is represented by a hatched thick line. Notice also that only objective function of Equation 1 was utilized here and therefore the average of scaled records falls below the target at certain locations. The fitness transition curve as a function of successive generations is shown in Figure 7.

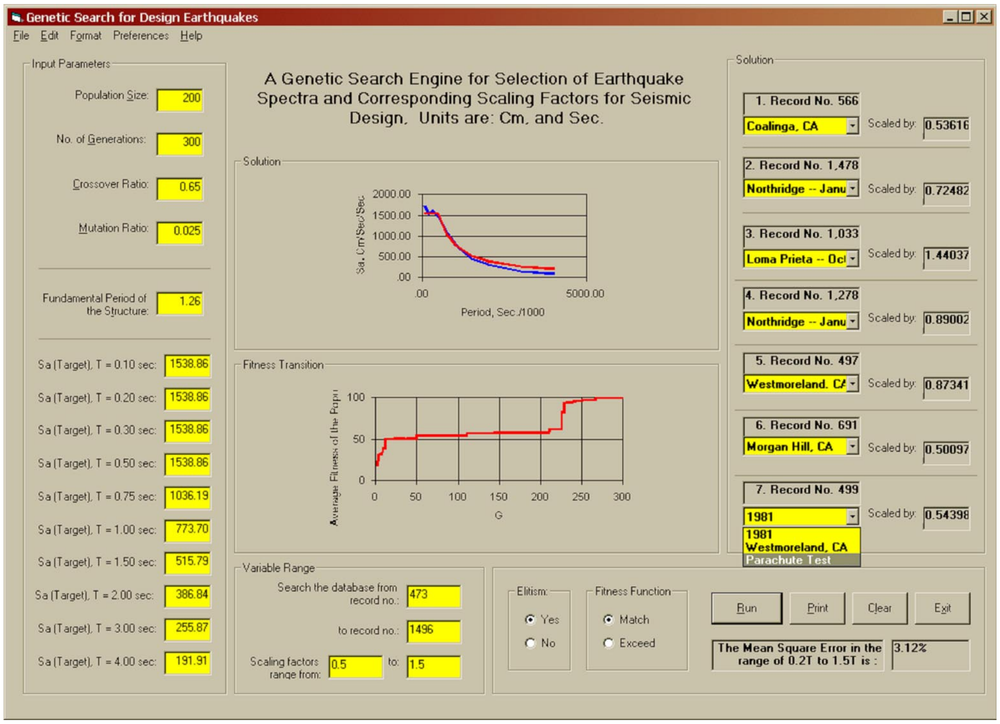

Figure 4. A computer display summarizing the input parameters and output results.

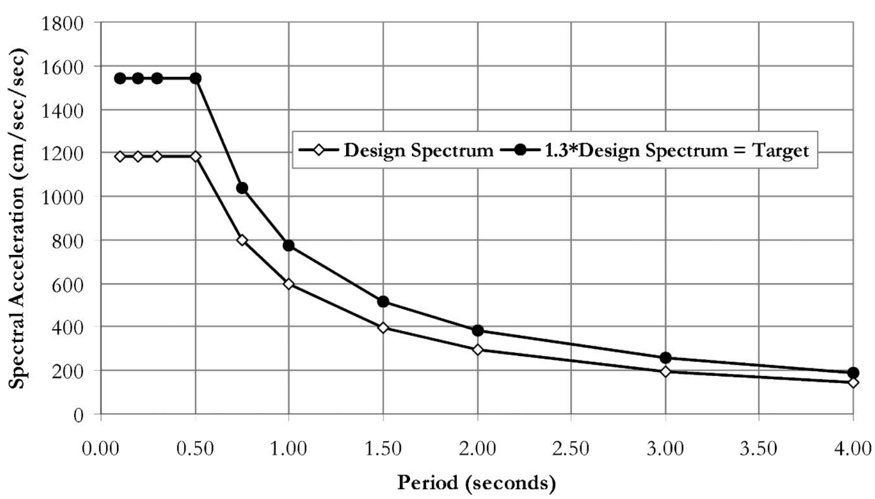

Figure 5. Target design spectrum for Example 1.

Table 1. The records and scaling factors selected by the genetic algorithm for Example 1

No.	Year	Earthquake Name	Station and Component	Scale Factor
1	1983	Coalinga, CA	Parkfield, Stone Corral 4 E, 0°	0.54
2	1994	Northridge, CA	LA, Wadsworth V.A. Hospital, 235°	0.72
3	1989	Loma Prieta, CA	Hollister—South & Pine, 0°	1.44
4	1994	Northridge, CA	Tarzana—Cedar Hill Nursery, 90°	0.89
5	1981	Westmoreland, CA	Niland, 0°	0.87
6	1984	Morgan Hill, CA	Coyote Lake Dam, 285°	0.50
7	1981	Westmoreland, CA	Parachute Test Facility, 0°	0.54

EXAMPLE 2

For this example the seven pairs of ground motions are already selected by the user and are not subject to change. The genetic algorithm is executed to select appropriate scaling factors for the given set of seven ground motion pairs. A more relaxed range for acceptable scaling factors is used (0.20 to 2.50). The results are shown in Figure 8 and Table 2 where the genetic algorithm results are compared to a manual solution obtained by the first author. Notice that the genetic algorithm was not constrained to produce solutions that stay above the target over the entire range of periods. The fit of the genetic algorithm solution with the target is excellent.

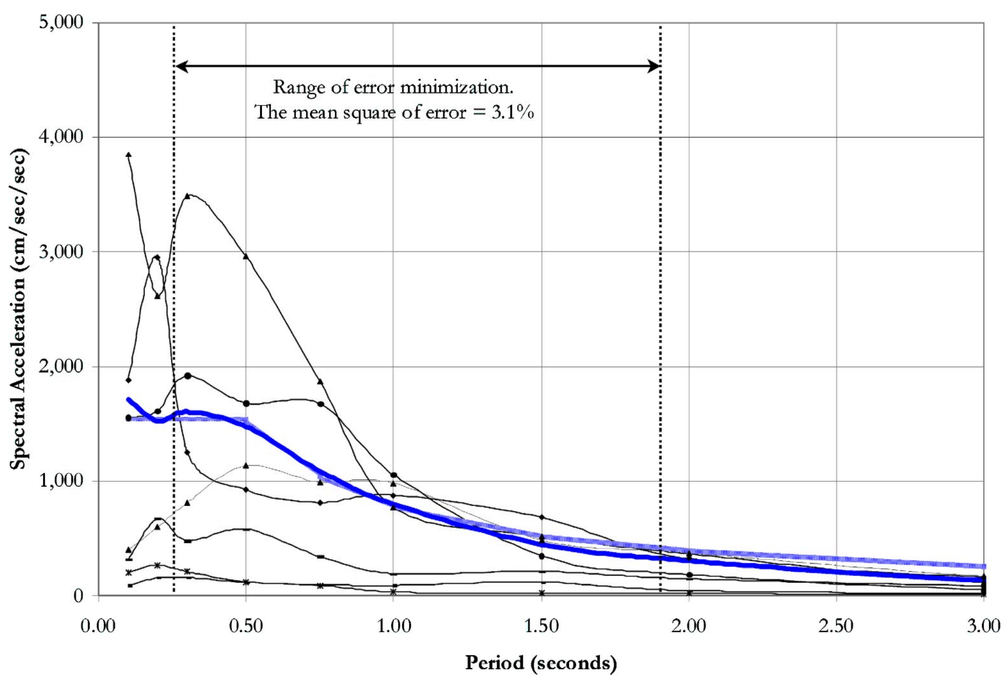

Figure 6. Spectrum matching of target in Example 1 using only Equation 1 as objective.

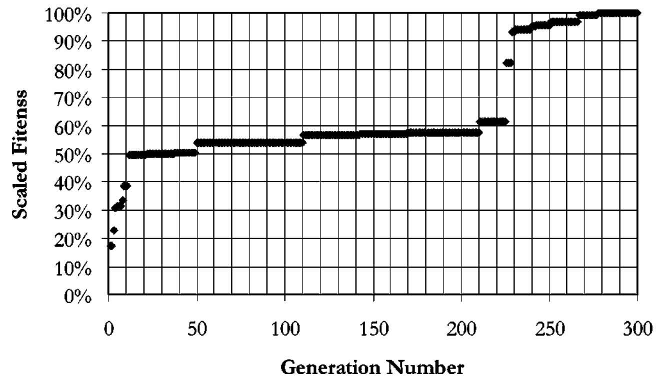

Figure 7. Fitness transition of solutions as a function of generations for Example 1.

As mentioned before, the minimization of the error between the scaled spectrum and the target does not guarantee that the scaled spectrum does not fall below the target at some point in the period range of interest. Enabling the penalty constraint of Equation 2 achieves this objective as seen in Figure 9. Notice that activating the penalty function has resulted in a new set of scale factors (see Table 3).

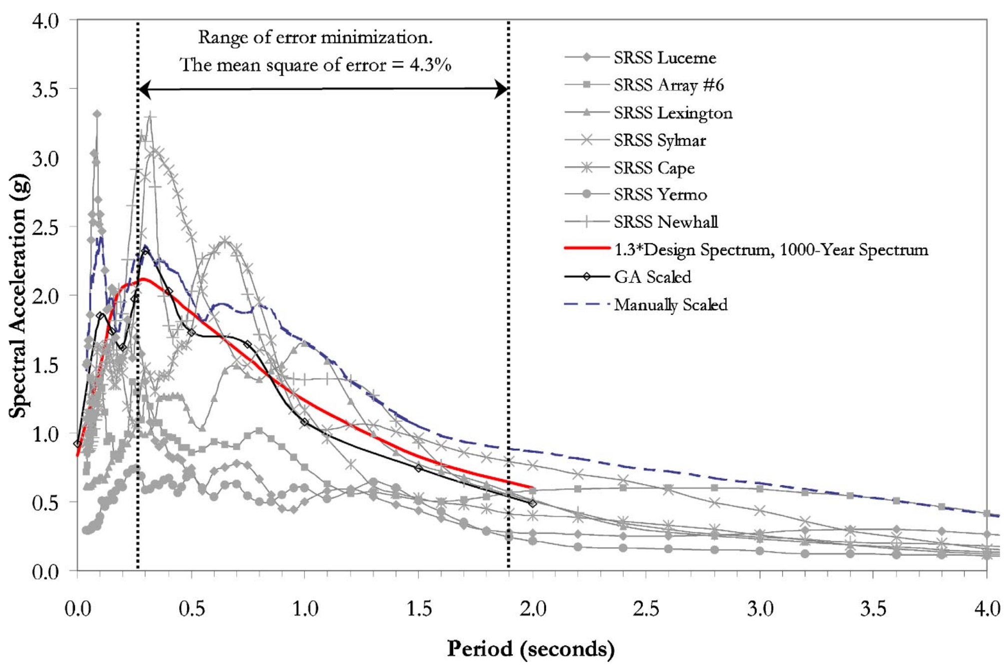

Figure 8. Result of selection of scale factors for a preselected seven pairs of time histories by the genetic algorithm.

Table 2. The records and scaling factors selected by the genetic algorithm for Example 2

Earthquake Name	Scaling factors applied to SRSS of two horizontal components of each record
1992 Landers at Lucerne Valley	2.48
1979 Imperial Valley at Array #6	0.20
1989 Loma Prieta at Lexington Dam	0.63
1994 Northridge at Sylmar County Hospital Parking Lot	1.77
1992 Cape Mendocino at Petrolia	1.80
1992 Landers at Yermo Fire Station	0.20
1994 Northridge at Newhall Fire Station	2.45

CONCLUSION

A new method for selection of earthquake ground motions that in combination match a given site-specific design spectrum was presented. This method uses a genetic algorithm that treats any random union of 7 records and corresponding scale factors as a single "individual" with 14 variables (7 record identifiers and 7 scale factors). The first generation of individuals is modified through the processes that mimic mating, natural selection, and mutation. The process continues until an optimum individual (seven pairs, and scaling factors) is obtained. The procedure is fast and reliable and results in records that match the target spectrum with minimal tampering and the least mean square of deviation from the target.

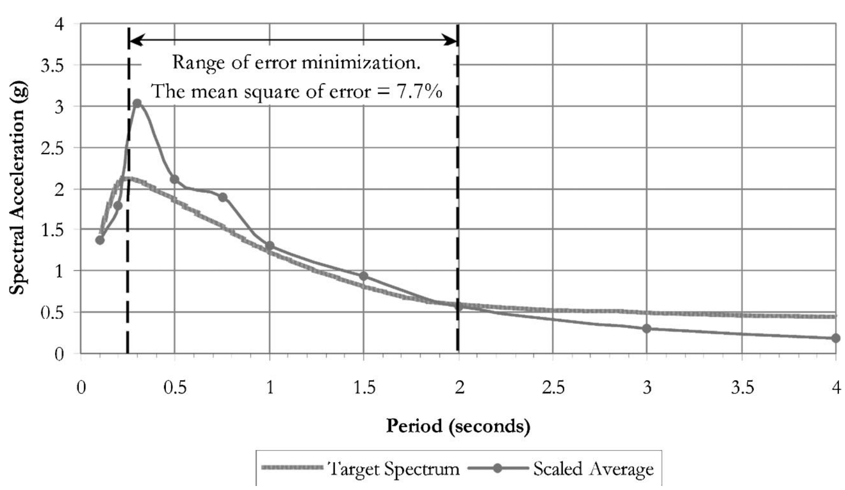

Figure 9. Result of selection of scale factors for a preselected seven pairs of time histories by the genetic algorithm using the penalty function.

Table 3. The records and scaling factors selected by the genetic algorithm for Example 2 after application of the penalty function

Earthquake Name	Scaling factors applied to SRSS of two components of each record
1992 Landers at Lucerne Valley	0.20
1979 Imperial Valley at Array #6	0.20
1989 Loma Prieta at Lexington Dam	0.20
1994 Northridge at Sylmar County Hospital Parking Lot	1.35
1992 Cape Mendocino at Petrolia	0.20
1992 Landers at Yermo Fire Station	0.20
1994 Northridge at Newhall Fire Station	2.50

The proposed procedure was applied using a large database of earthquake records to illustrate its efficiency. In practice, it may be prudent for the user to restrict selection of ground motions to some magnitude-distance bin appropriate for the site as defined from the deaggregated seismic hazard study, site soil conditions, and other relevant parameters.

ACKNOWLEDGMENTS

The Mid-America Earthquake Center Professional Student Interaction program provided a travel grant to the second author to travel to Los Angeles and stay with John A. Martin & Associates for completion of this research project.

REFERENCES

Alimoradi, A., 2001. Performance study of a GA-based active/hybrid control system under near source strong ground motion, *2001: A Structural Engineering Odyssey, Proceedings of the 2001 Structures Congress and Exposition, Washington, D.C.*, ASCE.

American Society of Civil Engineers (ASCE), 2000. *Prestandard and Commentary for the Seismic Rehabilitation of Buildings*, prepared for the SAC Joint Venture, published by the Federal Emergency Management Agency, *FEMA-356*, Washington, D.C.

Bolt, B. A., and Gregor, N. J. 1993. Synthesized strong ground motions for the seismic condition assessment of the eastern portion of the San Francisco Bay Bridge, *Report UCB/EERC-93/12*, University of California, Earthquake Engineering Research Center, Berkeley, CA.

Camp, C. V., Pezeshk, S., and Cao, G., 1998. Optimized design of two-dimensional structures using a genetic algorithm, *J. Struct. Eng.* **124** (5), 551–559.

Carballo, J. E., and Cornell, C. A., 2000. *Probabilistic Seismic Demand Analysis: Spectrum Matching and Design*, Department of Civil and Environmental Engineering, Stanford University, *Report No. RMS-41*.

Coley, D. A., 2001. *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific.

Coley, D. A., 2002. David Coley's GA: ⟨http://www.ex.ac.uk/cee/ga/software.htm⟩.

Department of the Army, 2000. *Dynamic Analysis of Concrete Hydraulic Structures*, Circular No. EC 1110-2-6051, Washington, D.C.

Foley, C. M., and Schinler D., 2001. Optimized design of partially and fully-restrained steel frames using distributed plasticity, *2001: A Structural Engineering Odyssey: Proceedings of the 2001 Structures Congress and Exposition, Washington, D.C.*, ASCE.

Foley, C. M., Pezeshk, S., and Alimoradi, A., 2003. State of the art in performance-based design optimization, *2003 Proceedings of the ASCE Structures Congress, Seattle, Washington*.

Gasparini, D., and Vanmarcke, E. H. 1976. *SIMQKE: A Program for Artificial Motion Generation*, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Ghaboussi, J., and Lin, C. J., 1998. New method of generating spectrum compatible accelerograms using neural networks, *Earthquake Eng. Struct. Dyn.* **27** (4), 377–396.

Goldberg, D. E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.

International Code Council (ICC), 2000. *International Building Code*, Falls Church, VA.

Kim, Y., and Ghaboussi, J., 1999. A new method of reduced-order feedback control using genetic algorithms, *Earthquake Eng. Struct. Dyn.* **28** (3), 235–254.

Kircher, C., 1993. Personal communication with Farzad Naeim and Marshall Lew.

Levy, S., 1992. *Artificial Life*, Vintage Books, New York.

Naeim, F., and Anderson, J. C., 1996. *Design Classification of Horizontal and Vertical Earthquake Ground Motion (1933–1994)*, John A. Martin & Associates, Inc. Report 7738.68-96.

Naeim, F., and Bhatia, H., 2000. Personal communication.

Naeim, F., and Kelly, J. M., 1999. *Design of Seismic Isolated Structures—From Theory to Practice*, John Wiley and Sons, New York.

Pezeshk, S., Camp, C. V., and Chen, D., 1999. Genetic algorithm for design of nonlinear framed structures, *Proceedings of the 1999 Structures Congress, New Orleans, Louisiana*, ASCE.

Pezeshk, S., Camp, C. V., and Chen, D., 2000. Design of framed structures by genetic optimization, *J. Struct. Eng.* **126** (3), 382–388.

Silva, W., and Lee, K. 1987. State-of-the-art for assessing earthquake hazards in the United States, *Report 24, WES RASCAL Code for Synthesizing Earthquake Ground Motions, Miscellaneous Paper S-73-1*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

(Received 8 January 2003; accepted 3 November 2003)