Article:

Gharehbaghi, S, Gandomi, AH, Achakpour, S et al. (2018) A hybrid computational approach for seismic energy demand prediction. Expert Systems with Applications, 110. C. pp. 335-351. ISSN: 0957-4174

https://doi.org/10.1016/j.eswa.2018.06.009

A HYBRID COMPUTATIONAL APPROACH FOR SEISMIC ENERGY DEMAND PREDICTION

S. Gharehbaghi[a], A.H. Gandomi[b,*], and S. Achakpour[c], M.N. Omidvar[d]

[a]*Department of Civil Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran*
[b]*School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA*
[c]*Department of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran*
[d]*Center of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, Birmingham B15 2TT, U.K.*

Abstract. In this paper, a hybrid genetic programming (GP) with multiple genes is implemented for developing prediction models of spectral energy demands. A multi-objective strategy is used for maximizing the accuracy and minimizing the complexity of the models. Both structural properties and earthquake characteristics are considered in prediction models of four demand parameters. Here, the earthquake records are classified based on soil type assuming that different soil classes have linear relationships in terms of GP genes. Therefore, linear regression analysis is used to connect genes for different soil types, which results in a total of sixteen prediction models. The accuracy and effectiveness of these models were assessed using different performance metrics and their performance was compared with several other models. The results indicate that not only the proposed models are simple, but also they outperform other spectral energy demand models proposed in the literature.

Keywords: Evolutionary computation; Genetic programming; Regression Analysis; Input energy; Hysteretic energy; Seismic energy spectra.

*Corresponding author.
Email:a.h.gandomi@stevens.edu
Tel.: +1 (201) 216-5029; Fax: 201-216-5385
URL: http://gandomi.beacon-center.org/

Table of Contents

1. Introduction

The approaches currently used for the seismic analysis and design of structures need to be improved through considering appropriate engineering demand parameters that would represent the characteristics of a structure and the design earthquake. In current approaches, either the structural members are designed based on satisfying the balance between the force demand and the corresponding strength supply while providing an adequate level of ductility (based on e.g. ASCE/SEI7 2010), or based on the concept whether they are force- or deformation-controlled (based on e.g. FEMA-356 2000). Such approaches disregard the frequency content and duration of earthquake ground motion, as well as the velocity response and hysteretic behavior (Gupta 1990). The importance of considering these factors lies in evidences suggesting that, for instance, dissipated hysteretic energy due to repeated inelastic excursions could result in a certain amount of seismic damage (Fajfar and Vidic 1994). In fact, in addition to the force and deformation, the energy demand is of great importance in capturing the mentioned seismic factors as the inelastic behavior is expected to occur due to the design and maximum earthquakes. Housner (1956) was first to introduce these factors through defining the energy concept. This concept requires that the energy dissipation capacitybe less than the input energy demand.

Both structural properties and earthquake characteristics affect the seismic energy demand. Determining the spectral values of energy demand is beneficial due to its connection tothe amount of structural damage (Fajfar and Vidic 1994, Gharehbaghi 2018). The design codes have not explicitly implemented the energy demand parameter in predicting seismic demands yet. Moreover, the priority of the energy-based design approach compared with the conventional strength-based design approach needs further studies. Although, previous studies (e.g. Housner 1956, Fajfar and Vidic 1994, Kalkan and Kunnath 2007, Akiyama 1985, Bertero and Uang 1988, Decanini and Mollaioli 2001, Manfredi 2001) have stipulated that the seismic energy parameters are of great importance in seismic design of structures. It was shown that the hystertic energy demand is directly connectedto the structural damage (Fajfar and Vidic 1994, Kalkan and Kunnath 2007, Akiyama 1985, Bertero and Uang 1988, Decanini and Mollaioli 2001, Manfredi 2001, Benavent-Climent et al. 2010, Gharehbaghi 2018). After more than two decades that the Housner’s proposal was almost neglected, it was received considerable attention among researchers (Akiyama 1985, Kuwamura and Galambos 1989), and became the key issue of a conference held in Bled city of Slovenia (Fajfar and Krawinkler 1992). It was recognized that input energy and hysteretic energy are the indicators of ground motion and have a correlation with the structural damage, and the quantity related to cumulative damage is the hysteretic energy (Fajfar and Vidic 1994, Bertero and Uang 1988, Decanini and Mollaioli 2001). Most recently, Deniz et al. (2017) also found that the most appropriate and reliable intensity measure for the seismic fragility analysis of buildings is the seismic energy demand.

Estimation of input and hysteretic energy demands using mathematical models could be considered as one of the important steps aligned with the extension of the energy-based seismic analysis and design. As previously mentioned, both structural and earthquake characteristics need to be accounted for the issue. Some earthquake characteristics such as soil type, earthquake magnitude, peak ground acceleration (PGA), peak ground velocity (PGV), cumulative energy index, fault type, distance from the hypocenter, were used by researchers in determining the energy spectra (e.g. Zahra and Hall 1984, Uang and Bertero 1988, Fajfar et al. 1989, Uang and Bertero 1990, Sucuoğlu and Nurtuğ 1995, Khashaee 2004). In addition to the earthquake characteristics, ductility ratio, damping ratio, and hysteretic behavior model (e.g. elastic-perfectly plastic, bilinear, pinching, Takeda, and Clough models) were the influential structural properties involved in the estimation of seismic energy demand spectra (e.g. Sucuoğlu and Nurtuğ 1995, Decanini and Mollaioli 2001, Benavent-Climent et al. 2011).Several works have been carried out on the estimation of the seismic energy demand parameters. Housner (1956) presented a model to determine input energy based on the spectral velocity of SDOF system. Kuwamura and Galambos (1989) presented energy demand spectra considering the soil type and dominant period of the earthquake. Chou and Uang (2000) estimated absorbed energy for an inelastic system by using an attenuation relation. They used nonlinear regression analysis considering both structural and earthquake variables. Manfredi (2001) proposed simple/efficient mathematical models to estimate input and hysteretic energy spectra. A dimensionless seismic index that is a function of PGA, PGV and cumulative energy was proposed to estimate the seismic energy spectra. Although the estimation models were simple and effective, the effect of soil behavior was not considered, and the number of earthquake ground motions was rather limited. Decanini and Mollaioli (2001) proposed the formulation of elastic seismic energy spectra. They also presented a comprehensive study to propose the design inelastic energy spectra by introducing the response modification factor for the input energy. Several structural variables (e.g. ductility ratio and hysteretic behavior) and earthquake characteristics such as soil type, source-to-site distance, and earthquake magnitude were considered in the proposed spectra. Arroyo and Ordaz (2007) estimated the hysteretic energy demand spectra from elastic response parameters in accordance with the earthquake events recorded in Mexico City. Their mathematical models were a function of pseudo-acceleration, velocity and displacement spectra. Elastic design input energy spectra based on Iranian earthquakes were also presented by Ghodrati Amiri et al. (2008). Recently, Dindar et al. (2015) proposed two regression-based simple mathematical models to estimate the input and hysteretic energy spectra. A database of earthquake ground motion records composed of near- and far-fault ones, PGA, soil types, earthquake magnitude, ductility ratio, and hysteretic behavior model was included in the proposed models. Using the regression analysis, Quinde et al. (2016) also proposed mathematical models to estimate the seismic energy spectra of inelastic systems located on the soft soil for Mexico City. They captured the effect of ductility ratio of inelastic systems and dominant period of the probable earthquakes on the presented models. More recently, Zhai et al. (2016) proposed an expression to account for the effect of after-shock on the input energy spectra using an equivalent velocity. Alıcı and Sucuoğlu (2016) carried out a regression analysis to estimate inelastic input energy spectrum. The prediction equations for the input energy spectra were expressed in terms of an equivalent velocity. Some crucial earthquake characteristics including soil type, epicentral distance, moment magnitude, and the fault type were considered in the proposed models. All the previously mentioned works use conventional regression methods to estimate their energy parameters of interest.

Based on the capability of soft computing approaches and their recent advances, it is worthwhile to use such efficient approaches for seismic demand prediction. Computational complexity of the conventional methods and their limitations has made soft computing techniques, such as evolutionary algorithms, artificial neural networks, support vector machines, and fuzzy logic, popular for solving complex engineering problems. A common application of these tools is in predictive analysis for modeling the nonlinear dependency of the input parameters to the output value(s) where the conventional approaches (e.g. regression analysis) fail or perform poorly (Khan et al. 2003, Gandomi and Roke 2015). Despite the success of artificial neural networks (ANNs) in prediction, they are inappropriate to develop practical intelligible equations. In addition to ANNs, support vector machines (SVMs) are another primary class of soft computing methods used to discover patterns and approximate relationships when large quantities of data is available. Although both ANNs and SVMs have received significant attention (e.g. Salajegheh and Heidari 2005, Gholizadeh and Salajegheh 2009, Papadopoulos et al. 2012, Gharehbaghi and Khatibinia 2015, Khatibinia et al. 2015, Yazdani et al. 2016), they require a pre-defined and initial structure for the equation and network architecture to be determined by the user. Genetic programming (GP), a learning algorithm originated from genetic algorithms, is another well-known and successful technique for developing nonlinear mathematical models for the complex problems. GP and its variants have been effectively used for solving various problems in civil engineering (e.g. Kayadelenet al. 2009, Alavi et al. 2011, Mirzahosseini et al. 2011, Gandomi et al. 2012, Vardhan et al. 2016; Lim et al. 2016). Several variants of GP have been proposed in the literature, such as gene expression programming (Ferreira 2006) and multi-stage genetic programming (Gandomi and Alavi 2011).One of the robust variants of GP is multi-gene genetic programming (MGGP) that adds the capability of conventional regression to the standard GP ability in parameter estimation. The effectiveness of MGGP has been proved in the works reported by Gandomi et al. (e.g. Gandomi and Alavi 2012a,b, Gandomi et al. 2013, Babanajad et al. 2013, Gandomi et al. 2016).

Structural and earthquake engineering has benefited from the soft computing techniques in different applications. For instance, ANNs and SVMs have been widely used for risk assessment, seismic response prediction, control and health monitoring (Tsompanakis and Topping 2011). In this paper, MOGP is used for predicting the seismic energy demand spectra considering both typical structural and earthquake characteristics. For this purpose, eighteen set of the single-degree-of-freedom (SDOF) systems with the structural properties of different hardening ratios of bilinear hysteretic behavior model, damping ratios, and ductility ratios are used to determine the energy demand spectra mentioned. Also, four different sets of earthquake ground motion records based on their soil types (soft, firm, stiff and rock) with the source-to-site distances of more than 17.5 km and the magnitudes of greater than 5.5 were used. It was assumed that the different soil classes have linear relationships in terms of GP genes which help to find one equation with different coefficients for different soil types. The records were scaled to two PGA levels 0.5g and 1.0g. Finally, four mathematical models corresponding to the four engineering demand parameters (EDPs) of spectral input and hysteretic energy, spectral hysteretic to input energy ratio, and spectral energy modification factor, are proposed using MOGP. Then, the effectiveness of the models is revealed using the performance metrics compared with those of available in the literature.

In this study, section 2 describes the seismic energy concept and its formulation. Also this section introduces the seismic energy based EDPs which can be useful in seismic design of inelastic structures. Section 3 express a hybrid computational approach based on genetic programming used as a predictive tool herein. Section 4 describes a framework for prediction of the EDPs. A set of mathematical models are proposed and their accuracy are examined using some performance metrics in section 5. Finally, the developed models are discussed and compared with some other methods proposed in the literature.

2. Seismic Energy Concept and Formulation

Housner (1956) first proposed the idea of the energy-based seismic design approach. When ground motion transmits energy into a structure, some of the energy is dissipated through the damping and inelastic behavior. The remained energy of the structure is stored in the form of kinetic energy and elastic strain energy. Housner stipulated that the energy supply should be more than the energy demand during an earthquake in the form of this principle that *energy supply < energy demand* for controlling and avoiding the structural collapse (Housner 1956).

When a structure is subjected to earthquake excitation, its governing equation for the dynamic behavior of an inelastic SDOF system could be written as (Chopra 2012):

$$m\ddot{u}(t) + c\dot{u}(t) + f_s(u(t), \dot{u}(t)) = -m\ddot{u}_g(t) \qquad (1)$$

where m, c and f_s represent the mass, damping coefficient and lateral resisting force of SDOF system, respectively; $\ddot{u}(t)$, $\dot{u}(t)$ and $u(t)$ are acceleration, velocity and displacement relative to the ground with t representing time in SDOF system, respectively; and $\ddot{u}_g(t)$ is the earthquake ground acceleration. Governing equation of energy equilibrium is obtained by the integration of Eq. (1) with respect to u (Uang and Bertero 1990, Chopra 2012):

$$\int m\ddot{u}(t)du + \int c\dot{u}(t)du + \int f_s(u(t),\dot{u}(t))du = -\int m\ddot{u}_g(t)du \tag{2}$$

In fact, Eq. (2) expresses the energy balance for a structural system during an earthquake while Eq. (1) explains the force balance. Substituting displacement unit (du) by velocity term and integrating it over the time of the earthquake ground motion t, Eq. (2) is expressed as:

$$\int m\ddot{u}(t)\dot{u}(t)dt + \int c\dot{u}(t)\dot{u}(t)dt + \int f_s(u(t),\dot{u}(t))\dot{u}(t)dt = -\int m\ddot{u}_g(t)\dot{u}(t)dt\ , \tag{3}$$

where t is the time of interest across the earthquake ground motion. Eq. (3) can also be written in a general form as follows:

$$E_K(t) + E_D(t) + E_A(t) = E_I(t)\ , \tag{4}$$

where $E_I(t)$, $E_K(t)$ and $E_D(t)$ are the input energy demand, kinetic energy, and damping energy, respectively; $E_A(t)$ is encompassed the recoverable elastic strain energy $E_S(t)$ and the irrecoverable plastic hysteretic energy $E_H(t)$. The amount of $E_H(t)$ is equal to zero in the elastic systems and is appeared in the inelastic systems. Therefore, Eq. (4) can be expressed as:

$$E_K(t) + E_D(t) + E_S(t) + E_H(t) = E_I(t)\ , \tag{5}$$

where $E_K(t)$ and $E_S(t)$ are cumulative during the earthquake ground motion and are vanished at the end of motion of the inelastic systems. In effect, these two terms are very small in comparison with $E_D(t)$ and $E_H(t)$. Since the most portion of energy demand is dissipated through the damping energy and hysteretic energy, the Eq. (5) can be approximately written as (Uang and Bertero 1990):

$$E_D(t) + E_H(t) \cong E_I(t)\ . \tag{6}$$

According to Eq. (6), the main portion of input energy demand is converted to the damping energy and hysteretic energy. Besides, if the structural system remains in the elastic range during an earthquake, $E_H(t)$ is trivial, and the energy-based analysis is not useful for seismic design (Uang and Bertero 1990).As mentioned before, for design basis earthquakes, it is expected that a structure will experience inelastic cyclic deformations resulting in hysteretic energy dissipation. As previously mentioned, the hysteretic energy dissipation (E_H) is directly attributed to the structural damage where the E_H/E_I ratio has been introduced as a good indicator of expected damage (Fajfar and Vidic 1994, Sucuoğlu and Nurtuğ 1995, Decanini and Mollaioli 2001, Manfredi 2001). For a given ductility ratio (μ), E_H/E_I ratio is defined as follow:

$$HI_{\mu} = \frac{E_{H\mu}}{E_{I\mu}}, \quad (7)$$

where the $E_{I\mu}$ and $E_{H\mu}$ are the input and hysteretic energy corresponding to the ductility ratio of μ. Another parameter that is of crucial importance for earthquake resistant design based on the energy concept could be the response modification factor of the input energy that can be expressed as follow:

$$RE_{I\mu} = \frac{E_I}{E_{I\mu}}. \quad (8)$$

The literature suggests that to have a practical energy based seismic design, the computation of the input and hysteretic energy ($E_{I\mu}$ and $E_{H\mu}$), hysteretic energy to input energy ratio (HI_{μ}), response modification factor of input energy ($RE_{I\mu}$), and energy dissipation capacity is useful. Since the design method requires inelastic dynamic analyses resulting in the expensive computational efforts, the use of soft computing techniques is of great importance in predicting the practical mathematical models (formulations). Except for the energy dissipation capacity that needs comprehensive experimental and theoretical studies, the spectral values of the mentioned energy-based EDPs were predicted using MOGP. The next section describes the MOGP.

3. Genetic Programming

There are two groups of models which can be used for modeling the complex nonlinear engineering systems: phenomenological and behavioral (Gandomi et al. 2016). Phenomenological models need a predefined structure obtained from the physical laws requiring a previous understanding about the system. Concerning the complex systems, sometimes it is hard to find such models. Unlike the phenomenological models, behavioral models can be simply generated by finding a reasonable approximate relation between input variables and the output value(s) for a collection of data (experimental or theoretical) irrespective of their governing physical principles. Although one of the main advantages of behavioral modeling techniques is their independence on prior knowledge about the governing physical relationships of input and outputs (Walter and Pronzato 1997, Metenidis 2004), most of these models need the user to pre-assign a formulation pattern requiring optimization of its unknown coefficients. Concerning the complex engineering systems, the use of conventional techniques such as regression analysis cannot be guaranteed to find an accurate and reliable behavioral model (Gandomi et al. 2016). It has been well recognized that most of the structural earthquake engineering problems such as determining earthquake response of inelastic structures could be considered as such complex problems.

Genetic programming (GP) (Koza 1990) is a novel behavioral modeling methodology with completely new characteristics. GP is an extension of the genetic algorithm capable of functionalizing data using tree structures. In fact, unlike classic regression models and ANNs, GP is capable of generating a prediction equation irrespective of a predefined structure. The successful application of GP and its variants have been reported in solving the various real-world problems (e.g. Gandomi and Alavi 2012a,b, Gandomi et al. 2013, Babanajad et al. 2013, Gandomi et al. 2016). One of the robust variants of GP is MGGP that adds the capability of conventional regression to the standard GP ability in parameter estimation, which has been proposed recently (Searson et al. 2007).The initial MGGP studies show that it can outperform other GP variants (Gandomi and Alavi 2012a,b). MGGP is described in the next subsection.

3.1. Multi-Gene Symbolic Regression

GP can generally be defined as a supervised machine learning technique that searches a program space instead of a parameter space. One of the useful variants of GP is Multi-gene genetic programming (MGGP) (Searson et al. 2007, Gandomi and Alavi2012a). MGGP is used to design mathematical model predictions which are inherently multi-gene, i.e., those models consisting of linear combinations of low order nonlinear transformations of the input variables. Unlike conventional GP that is based on an evaluation of single tree expression, MGGP uses a single GP particle swarm model selection program constructed from a number of genes where each gene has a tree expression (Searson et al. 2010). In effect, the model development procedure is decomposed by MGGP consisted of some relatively simple, fixed-depth sub-models.

To develop a population of trees, GP typically uses symbolic regression. Compared with the conventional GP, a weighted linear combination of outputs from a number of GP trees is considered as a symbolic model in which each of these trees could be considered as a "gene" (Searson et al. 2010). The number of genes and tree depth of any gene can be specified by the user which their maximum values depend on the complexity of the models developed by MGGP. The evolved models are linear combinations of low-order nonlinear transformations of the predictor variables (Searson et al. 2010). The ordinary least squares method is used to estimate the linear coefficients for each of the evolved genes of an individual. A more detailed explanation of MGGP is available in Refs. (Searson et al. 2007, Searson et al. 2010).

A fixed linear illustration associated with binary encoding of all parameters is used in GA, which results in a string of numbers as output. In GP however, the optimization problems are solved irrespective of a pre-defined solution structure. Depending on the problem domain, the first generation consisting of a population of possible solutions is randomly created by GP, and variation operators generate new candidate solutions then. During the evolutionary process, crossover selects a node from the parental individuals and exchanges the subtrees under the selected nodes randomly and creates a new individual. Mutation truncates and replaces one node of a tree with another randomly generated node from the same set, and then creates a new individual from an existing tree in the population. The individuals with higher fitness values have higher survival likelihood in the successive generation. To find the best fitting solution, the individual solution of a population is updated after executing a number of runs in each generation, and the parental individuals are selected from the population based on a good fitness function. To develop a population of genes, symbolic regression method can be implemented using standard GP in which a symbolic mathematical expression is directly encoded by each of the genes. Based on Figure 1 showing a typical multi-gene model, known as multi-gene symbolic regression, three input variables (x_1, x_2, and x_3) are used to predict the response. As shown in the figure, although the nonlinear terms such as "sin" and "log" are used, the overall model is a weighted linear combination of each gene utilizing the coefficients g_o, g_1, and g_2. Mathematically, the general formulation of the multi-gene symbolic regression model can be written as (Gandomi et al. 2016):

$$\hat{y}(\boldsymbol{x}, \boldsymbol{g}, \boldsymbol{\theta}) = g_o + \sum_{i=1}^{n} (g_i G_i(\boldsymbol{\theta}, \boldsymbol{x})) \tag{9}$$

where g_o is a bias term; g_i is the gene weight, and $G_i(\boldsymbol{\theta},\boldsymbol{x})$ is the outputs vector from the *i*th gene encompassing a multi-gene individual; $\boldsymbol{\theta}$ is the vector of the unknown parameters for each gene; and n is the number of genes. It should be noted that the algorithmic structure of MGGP, and standard GP is the same, except for crossover and mutation of multi-gene individuals. MGGP does not necessitate any simplifying assumption in the model development process, and it is more accurate and efficient than the standard GP for modeling complex nonlinear problems (Gandomi and Alavi 2012a,b). To construct an initial population in MGGP, random individuals are created by using different nonlinear functions, input variables, and a range of random constants and each individual includes 1 and G_{max} number of genes. The algorithm attempts to maximize diversity by ensuring that no individuals contain duplicate genes. The genes are randomly selected and the least squares normal equation is used to estimate the vector of unknown coefficients $\boldsymbol{g}$ as follows (Searson et al. 2007, Searson et al. 2010, Hii et al. 2011, Searson 2014):

$$\boldsymbol{g} = \left(\boldsymbol{G}^T \boldsymbol{G}\right)^{-1} \boldsymbol{G}^T \boldsymbol{y} \tag{10}$$

where $\boldsymbol{G} = [1\ G_1 \ldots G_n]$ is the gene response matrix. Since the columns of matrix $\boldsymbol{G}$ can be collinear, the Moore–Penrose pseudo-inverse $(\boldsymbol{G}^T\boldsymbol{G})^{\#}$ can be computed by means of the singular value decomposition instead of the standard matrix inverse $(\boldsymbol{G}^T\boldsymbol{G})^{-1}$.Genes can be employed or eliminated using a tree crossover operator (high-level crossover) during the MGGP run which is in addition to the standard GP sub-tree crossover (a low-level crossover). The low-level crossover chooses a gene randomly from each parent individual. Next, the standard sub-tree crossover is employed and the generated trees replace the parent trees in the otherwise unaltered individual in the next generation. The high-level crossover allows the exchange of one or more genes with another selected individual subject to the G_{max} constraint. The maximum number of genes of an individual is limited to G_{max}.If any individual contains more genes, the additional genes are randomly selected and deleted (Searson et al. 2007, Searson et al. 2010, Hii et al. 2011, Searson 2014).

3.2. Multi-Objective Genetic Programming (MOGP)

Generally, both tree-based GP and MGGP deal with a single objective optimization problem for each individual considering the defined fitness function in which, for symbolic regression, the goodness-of-fit to the training data is considered as the only objective to be maximized. Although MGGP yields more compacted models compared with standard GP (Searson 2014), ineffective genes may be acquired by multi-gene models and the single objective optimization problem results in the evolution of overly complex, impractical and non-robust models (Gandomi et al. 2016). The simplest solution to eliminate such shortcoming can be provided by limiting G in a model to G_{max} which is a hard-to-determine unique value for any given problem (Searson 2014). One good solution is the use of multi-objective concepts into symbolic regression which is commonly referred to as multi-objective genetic programming (MOGP). Using this methodology, both the goodness-of-fit and the complexity of the developed models can be optimized simultaneously by searching the so-called Pareto front (non-dominated solutions) set. Herein, the GPTIPS 2 toolbox (Searson 2014) associated with the related subroutines coded in MATLAB (2013) is used to solve the MOGP by using a non-dominated sorting technique (Deb et al. 2002). To sort the non-dominant solutions by their complexity and precision, the non-dominated sorting method is applied at the end of each generation of the MGGP algorithm. At first, the individuals are classified from both the new and old population based on their position on the Pareto front. Pareto front of each level encompasses a set of Pareto optimal solutions which other solutions do not dominate. In addition, the solutions that include Pareto front of each level are not dominated by any other solution, apart from those of in its previous Pareto front level. After that, a "crowding factor" (i.e., the average distance of a solution from the nearest solutions (either side) on the same Pareto front) is computed for separately all individual to increase population diversity, giving lower priority to the solutions that are crowded together during the ranking process. Finally, the position of solutions is used to rank them (those on level 1 are ranked above those on level 2, and so on),and a crowding factor of each solution is used to rank those within the same level. The top 50% of the population is remained to participate in the next generation, whilst the rest are eliminated (Searson 2014).

3.3. Accelerating GP Process

Typically, the data sets used in engineering studies are complex and do not include a very large number of records particularly for experimental studies (Gani et al. 2016). While the successful application of GP in modeling engineering systems has been reported in the literature (e.g., (Sajjadi et al. 2016)), it can be difficult to model the systems with big data using GP. The evolutionary approaches are often slower than statistical data mining. Since GP is usually used to find the structure of solution(s), it is one of the slowest evolutionary algorithms. In addition, the extra process of non-dominated sorting of a multi-objective GP magnifies the problem. To improve this weakness, in this paper, two strategies are used in the prediction process:

- 60% of data (2160 samples) were randomly selected and used for training process, and the rest (1440 samples) used as training set for each run;
- The final Pareto front was determined from merging the Pareto fronts for all runs.

In general, there are two classes of machine learning algorithms including trajectory based algorithms and population-based algorithms. ANNs and regression analysis are two well-known examples of trajectory algorithms. In contrast, GP is one of the mostly used population-based algorithms which it deals with a set of the solution in each generation. This feature makes it flexible to adopt with parallel processing, therefore, the paralleled computations can be used to accelerate MOGP procedure using a distributed computing machine in order to deal with the Big Data issue in GP. Although only twelve cores were used to evolve and evaluate new models herein, the number of cores can be increased up to the population size using this framework. The schematic of parallel processing in the GP process is shown in Figure 2.

4. Predicting Seismic Energy Demand Spectra Using MOGP

4.1. Preparing Exact Data

Since the inelastic responses of an SDOF system highly depend on both structural and earthquake ground motion variables, the most influential ones are contributed in predicting spectral seismic energy demand. The input variables are described in the next subsections in detail.

4.1.1. Inelastic SDOF Systems

The structural variables used in this study are the hardening ratio of bilinear hysteretic behavior model (η), damping ratio (ξ), and the displacement ductility ratio (μ) which are the structural properties used to determine the energy demand spectra mentioned. The values assumed for η were 0.0, corresponding to the elastic-perfectly plastic model, and 0.1 indicating bilinear model. Three values of 0.05, 0.10 and 0.15 were also used for ξ of the inelastic SDOF systems. In addition, three common values of 2, 4 and 6 were taken into account for μ. The periods range studied for the prediction of the energy spectra was between 0.01 to 5.0 second for every 0.05 second. These considered variables (η, ξ, and μ) resulted in 18 inelastic SDOF systems used for the prediction.

4.1.2. Earthquake Ground Motions

Three factors of the site class, source-to-site distance, and PGA are the three variables considered for the earthquake ground motion records used. Based on the shear wave velocities corresponding to the 30 m in depth ($V_{s,30}$) of more than 750, 360 to 750, 180 to 360 and less than 180 m/s, four soil types of S1, S2, S3, and S4 were assumed for the records used. Site types of S1, S2, S3, and S4 are respectively representing the soft, firm, stiff and rock soil types. To consider the source-to-site distance, the records having the Joyne-Boor distance (R_{JB}) in the range of more than 17.5km and less than 150km were used (known as far-fault records). In addition, two values of 0.5g and 1.0g were used for the PGA of the records. All records are non-pulse-like and have the magnitude (M) of greater than 5.5. All the records were downloaded from NGA-West-II project of PEER ground motion database (2017). The diversity of M, R_{JB} and $V_{s,30}$, and the number of records are shown in Figure 3. As shown in this figure, the records are selected in a way that they have a large variety of the properties mentioned. The individual pseudo-spectral acceleration of each record of each soil type and their mean spectra are also shown in Figure 4.

Four main seismic energy-based EDPs were chosen to be predicted: (i) EDP1: spectral $E_{I\mu}/m$; (ii) EDP2: spectral $E_{H\mu}/m$; (iii) EDP3: spectral HI_{μ}; and (iv) EDP4: spectral $RE_{I\mu}$. For this purpose, based on the values assumed for the structural variables, the 18 SDOF systems were modeled and subjected to the mentioned earthquake records. A large number of inelastic time history dynamic analyses (more than 1 million) were carried out, and the exact EDPs were determined. The entire process was simulated in MATLAB platform (2013).

For each EDP, individual spectral energy responses of the SDOF systems under each set of earthquake records were obtained. Then, based on the normal distribution, the mean plus one standard deviation (mean+σ) for each EDP under each set of the earthquake records were determined to be predicted.

4.2. Model Development Using MOGP

To develop powerful models, the suitable parameters ought to be utilized as a part of the MOGP predictive algorithm. To obtain the optimum MOGP models, basic arithmetic operators (+, -, ×, /) and mathematical functions (e.g. *tanh*, *ln*) were used. The models are formed by randomly combining the components from the functional set and the terminal set. The number of programs (solutions) in the population is determined by the population size, and the number of levels that the calculation would apply before the run ends are resolved as per the number of generations.The nature of thedata set, problem complexity, and the number of variables are the three determining factors for the population size and the number of generations. Note that, the upper bounds of an individual (G_{max}) and the maximum tree depth (D_{max}) need to be defined in order to restrict the complexity. To conduct a trade-off between the running time and the complexity of the evolved solutions, optimal values of 3 and 5 were respectively assumed for G_{max} and D_{max}. The parameter settings used for the MOGP implementation listed in Table 1 are based on the previously suggested values in the literature (Searson et al. 2007, Searson et al. 2010, Hii et al. 2011, Searson 2014) and employing a case-dependent trial-and-error process.

In order to generate new genes for individuals as well as to decrease the overall number of genes for one model and increase the total number of genes for the other, a "rate-based high-level crossover" through the use of a crossover rate parameter (CR) is employed herein. A uniform random number between 0 and 1 with a default value of 0.5 is generated separately for each gene in the parents. If *r* isless than *CR*, the corresponding gene is moved to the other individual. In case the exchanging process results in offspring that contain more genes than the G_{max}, the gene is randomly eliminated such that the constraint is no longer violated (Searson 2014). Two data sets are needed for the analyses. Therefore, data are randomly divided into two subsets for training and validation. The training dataset is used for learning and the validation set for determining the quality of the evolved programs on unseen data. Several combinations of training and validation sets were considered to determine a consistent data division. To evaluate the evolved expressions and finding the best-encoded one, the minimum of the root mean square error (*RMSE*) is used as the fitness function. *RMSE* can be expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - p_i)^2}\ , \tag{11}$$

where p_i and e_i are the predicted, and exact output values for the *i*th output, respectively; and *n* is the number of samples. Two objectives of maximizing the correlation coefficient (*R*) and minimizing the model complexity are used in MOGP approach in order to select the best final model.

5. Results and Discussion

Using MOGP, all EDPs (EDP1, EDP2, EDP3, and EDP4) were predicted, and their optimal mathematical models (formulations) were determined. Four cases (S1 to S4) based on different soil types of the earthquake records were considered in the prediction. Although it is quite possible that the obtained model formulations be different for different soil types, it is more practical to develop a unique mathematical model for an EDP with different coefficients. Therefore, in this paper, the complete database (which includes four soil types of S1 to S4) is employed to develop a unique prediction model for each EDP. The final mathematical model is selected based on a compromise between the prediction accuracy (as measured by the correlation coefficient *R*) and the model complexity (as measured by the number of input variables). After that, the complete database is divided into four groups based on the four soil types. Using each group of data, the predicted coefficients of the final model (g_i) are re-evaluated by conducting the regression analysis to reflect the influence of the soil type. Finally, four mathematical models including structural variables (and PGA of earthquake records) with four different groups of coefficients corresponding to the four cases mentioned above (S1 to S4) obtained are presented herein. The contribution of each input variable in the mathematical prediction models and the Pareto front obtained by using a nondominated sorting method at the end of an MOGP run are presented.

The results of all EDP models developed by MOGP for EDP1 to EDP4 have been shown in Figures 5(a)-(d). The Pareto front sets are shown in green circles and the rest of the models are shown in solid blue circles. As mentioned earlier, the Pareto front set is obtained by using a non-dominated sorting of populations at the end of all MOGP runs. This process simultaneously optimizes the accuracy and the complexity of all developed models. The final model in each Pareto front set is selected and highlighted in a red circle.

In order to benchmark the MOGP models, they were compared with gene expression programming (GEP) models as a well-known and widely used GP algorithm. The GEP model also uses multiple gene structure, which makes it similar to the MGGP in this respect. GEP requires 10 times more generations to converge. It is because the MOGP algorithm converges quickly since it uses regression analysis beside the evolutionary process. GEP's parameter setting is similar to that of MOGP (shown in Table 1). The final results of GEP are presented in Figure 5. The results show that none of the models found by GEP are among the Pareto front sets for any of the EDP1 to EDP4 problems.

The contribution of each input variable in the mathematical prediction models can be investigated through their frequencies where a frequency value of 1.0 for a variable indicates that it has the maximum contribution within the best-generated models (Gandomi et al. 2010). It was assumed that the models with $R^2 > 0.8$ are the best-generated models. The frequency histograms of the input variables for all predicted EDPs are shown in Fig. 6. As shown, for the selected database of EDP1, *PGA* and η respectively show the most and the least statistically significant contributions in the best-generated MOGP models. For the collected database of EDP2, although all the input variables almost have significant contributions in the best generated MOGP models, T and ξ are the most and the least influential variables on the EDP2 prediction model. According to the literature (e.g. Kuwamura and Galambos 1989, Manfredi 2001, Dindar et al. 2015), there is no report about the role of *PGA* on EDP3 (hysteretic to input energy ratio). Moreover, based on the physics of the problem, when the damping ratio is increased, the portion of hysteretic energy dissipation of the imparted seismic input energy is decreased. These issues have been confirmed by the frequencies of ξ and *PGA* for EDP3 where they have the largest and smallest statistically significant contributions, respectively. Moreover, as can be seen in the figure, for the collected database relevant to EDP4, *T* and *PGA* have the largest and smallest statistically significant contributions in the best-generated MOGP models, respectively.

5.1. Mathematical Model for EDP1

The mathematical model obtained for EDP1 is expressed as follows:

$$\frac{E_{I\mu}}{m} = \left(a_0 + a_1 X_1 + a_2 X_2\right) S_V \,, \tag{12-a}$$

$$X_1 = PGA\,.\tanh(\tanh(T))\sqrt{\frac{\xi}{T}} \,, \tag{12-b}$$

$$X_2 = \frac{\xi\,(T - \mu^{1.5})PGA}{\mu \mathrm{e}^{\eta}} \,, \tag{12-c}$$

where a_0, is the bias term, a_1 and a_2 are the gene weight for the EDP1 prediction model. These coefficients are listed in Table 2. S_V is the spectral velocity of the elastic SDOF system.

5.2. Mathematical model for EDP2

The mathematical model derived for EDP2 is expressed as follows:

$$\frac{E_{H\mu}}{m} = \left(b_0 + b_1 Y_1 + b_2 Y_2\right) S_D \,, \tag{13-a}$$

$$Y_1 = PGA \,, \tag{13-b}$$

$$Y_2 = -e^{-2T}\,(PGA + 2\eta)(\xi - \mu) \,, \tag{13-c}$$

where b_0, is the bias term, b_1 and b_2 are the gene weight for EDP2. These coefficients are listed in Table 3. S_D is the spectral displacement of the elastic SDOF system.

5.3. Mathematical model for EDP3

The mathematical model derived for EDP3 is expressed as follows:

$$HI_{\mu} = \frac{E_{H\mu}}{E_{I\mu}} = c_0 + c_1 Z_1 + c_2 Z_2 , \tag{14-a}$$

$$Z_1 = e^{-2\eta} e^{-0.151T-\xi-\mu} , \tag{14-b}$$

$$Z_2 = \log\left(\xi(T + 7.75)\tanh(\frac{T}{\xi}) \right), \tag{14-c}$$

476 where c_0 is the bias term, c_1 and c_2 are the gene weight for EDP3. These coefficients are listed in Table
477 4.

478 ***5.4. Mathematical model for EDP4***

479 The mathematical model obtained for EDP4 is expressed as follows:

$$RE_{I\mu} = \frac{E_I}{E_{I\mu}} = d_0 + d_1 U_1 + d_2 U_2 , \tag{15-a}$$

$$U_1 = (\eta + \mu + e^{-T} + \mu\xi - 0.422)(T\,\xi - 1.4074)\tanh(T) , \tag{15-b}$$

$$U_2 = \frac{1}{T}\ln(\eta + T\mu^2 + \sqrt{\mu}) , \tag{15-c}$$

480 where d_0 is the bias term, d_1 and d_2 are the gene weight for EDP4. These coefficients are listed in Table
481 5.

482 It should be noted that EDP2 can also be determined by the following relationship:

$$E_{H\mu} = \frac{E_{I\mu}}{m} H_{I\mu} \tag{16}$$

483 In fact, there are two ways to compute EDP2: one is by using Eq. (13) directly (called as EDP2-1), and
484 another is by using Eq. (16) (called EDP2-2).

485 ***5.5. Accuracy of the Models***

486 It should be noted that all the models are only valid in the range of actual database (discussed in section
487 4.1). In order to investigate the effectiveness and accuracy of the EDP models in the range of our
488 database, the common performance metrics, including the mean absolute percentage error (*MAPE*), the
489 relative root mean square error (*RRMSE*), the linear correlation coefficient (*R*), the performance index
490 (*PI*), coefficient of determination (r^2), coefficient of efficiency (*E*), and the index of agreement (*d*)

corresponding to the predicted formulation of each EDP are obtained. *MAPE, RRMSE, R*, *PI*, r^2, *E*, and *d* are expressed as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i - p_i}{e_i}\right|, \tag{16}$$

$$RRMSE = \frac{1}{|\bar{e}|}\sqrt{\frac{\sum_{i=1}^{n}(e_i - p_i)^2}{n}}, \tag{17}$$

$$R = \frac{\sum_{i=1}^{n}(e_i - \bar{e}_i)(p_i - \bar{p}_i)}{\sqrt{\sum_{i=1}^{n}(e_i - \bar{e}_i)^2 \sum_{i=1}^{n}(p_i - \bar{p}_i)^2}}, \tag{18}$$

$$PI = \frac{RRMSE}{R+1}, \tag{19}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(e_i - p_i)^2}{\sum_{i=1}^{n}(p_i)^2}, \tag{20}$$

$$E = 1 - \frac{\sum_{i=1}^{n}(e_i - p_i)^2}{\sum_{i=1}^{n}(e_i - \bar{e}_i)^2}, \tag{21}$$

$$d = 1 - \frac{\sum_{i=1}^{n}(e_i - p_i)^2}{\sum_{i=1}^{n}(|e_i - \bar{e}_i| + |p_i - \bar{e}_i|)^2}, \tag{22}$$

where $\bar{e}_i$ and $\bar{p}_i$ are the average values of the exact and predicted outputs, respectively; and *n*, e_i, and p_i were defined before. Lower *MAPE* and *RRMSE*, higher *R* (or R^2), r^2, *E* and *d* indicate the accuracy and effectiveness of the prediction model used. Based on Eq. (19), higher *R* values and lower *RRMSE* values result in lower *PI* and, subsequently, indicate a more precise model. It should be noted that *PI* varies from 0 to ∞ and its values close to 0 indicate the model fits very well to the exact (actual) values. It is worth noting that two sources of complexity affect the accuracy of the models. First, the structures used are inelastic which leads to high nonlinearity, and second, the earthquake ground motion records have some influential characteristics, such as frequency content, which make a structure to experience different cyclic excursions associated with complex behavior. These problems are accentuated when PGA is increased.

The abovementioned performance metrics of the predicted EDP1, EDP2 (including EDP2-1 and EDP2-2), EDP3 and EDP4 models using MOGP are listed in Table 6. As can be seen in this table, EDP1 and EDP2-2 have *MAPE* values respectively less than almost 16% and 17%, and it is less than 33.1% for EDP2-1 all of which are in an acceptable/reasonable range of the performance metric for inelastic and complex systems. The *MAPE* values are very low for both EDP3 and EDP4. Lower *RRMSE* values (near-zero usually less than 50%) also confirm the accuracy of the mathematical models. According to Table 6, except for EDP2-1 with soil type of S4 which has an approximate *RRMSE* of 54%, EDP1, EDP2-1 and EDP2-2 respectively have *RRMSE* values less than 36.3%, 44%, and 42.1%, and the remaining EDPs have *RRMSE* values less than 6.4%, indicating the accuracy of the predicted models. The higher accuracy of predicted EDP3 and EDP4 is evident. The R^2, E, d and r^2 values near 1.0 (e.g. more than 0.8) indicate a good correlation, efficiency and agreement of the predicted values with the exact ones. Based on Table 6, the R^2, E, d, and r^2 values are more than 0.8 for EDP1, EDP2-2, EDP3, and EDP4. The minimum values of R^2, E, d, and r^2 of EDP2-1are almost equal to 0.68, 0.55, 0.83 and 0.65 which belong to the soil type S4 whilst for other soil types almost all of them are more than 0.8. Regarding the *PI*, that is a combination of R and *RRMSE*, the values less than 0.3 and 0.2 indicate a good and an excellent prediction, respectively. The *PI* values are less than 0.2 for EDP1, less than 0.3 for EDP2-1, and less than 0.22 for EDP2-2 which indicate a good prediction capability of their corresponding proposed MOGP models. This index is less than 0.04 for both EDP3 and EDP4, confirming the excellent prediction capability of the proposed MOGP models.

To make a more informative and general evaluation of the proposed MOGP models, the average values of the performance metrics of all soil types, for each of EDPs, are presented in Table 6. The average results of EDP2 demonstrate that EDP2-2 has a better performance than EDP2-1. In fact, *MAPE*, *RRMSE,* and *PI* of EDP2-2 model have about 39%, 34.5%, and 37% smaller values as well as R^2, E, d and r^2 have almost 17%, 21%, 5.6%, and 13% larger values as compared to those of EDP2-1. Finally, considering the stochastic nature of earthquake engineering problems and the nonlinear relations governing the inelastic SDOFs behavior, the results indicate the good capability of MOGP prediction models for EDP1, EDP2-1 and EDP2-2, and excellent capability of MOGP prediction models for EDP3 and EDP4.

5.6. Comparative study

5.6.1. Assumptions

In the previous section, it was shown that all the mathematical models precisely predict each EDP of interest. In this section, the accuracy of the proposed mathematical models of all EDPs (EDP1 to EDP4) is compared with those of some available models from the relevant literature. The works presented by Housner (1956), Kuwamura and Galambos (1989), Fajfar and Vidic (1994), Manfredi (2000), Bakhshi and Tavallali (2006), and Dindar et al. (2015) are selected to make the comparison. All the works selected herein have dealt with the inelastic SDOF systems having elastoplastic behavior model ($\eta = 0$) and damping ratio (ξ) of 0.05. The PGAs of 0.5 and 1.0g and displacement ductility ratios (μ) of 2, 4 and 6, are considered for the comparison. The well-known performance metrics including *MAPE*, *RRMSE*, R^2, *PI*, *E*, *d*, and r^2 are used for comparison. To make an informative comparison, all the performance metrics are listed in separated Tables for EDP1, EDP2, (including EDP2-1 and EDP2-2), EDP3 and EDP4 (Tables 7-10 respectively).

In order to show the varying trend of the predicted models using MOGP and the best model from literature compared with the exact results, a graphical comparison is also made. As mentioned in the Introduction section and as can be concluded in the above comparison, Manfredi (2000) model is one of the best models presented in the literature. Therefore, this is the only model selected to make a more clear comparison for EDP1, EDP2 (including EDP2-1 and EDP2-2) and EDP3. Moreover, Bakhshi and Tavallali (2006) model is also used for making a comparison for EDP4. In addition, the inelastic SDOF systems having $\eta = 0$, $\mu = 6$ and $\xi = 0.05$ are used and are subjected to earthquake records corresponding to all soil types S1 to S4 with PGAs of 0.5 and 1.0g. The exact and predicted energy-based spectrum using MOGP and another selected model from the literature are shown in Figures 7-10 respectively for EDP1 to EDP4.

5.6.2. Comparison

The mathematical model of Eq. (12), proposed for prediction of EDP1 using MOGP, is compared with the seismic input energy equation presented by Housner (1956), Kuwamura and Galambos (1989), Manfredi (2000), and Dindar et al. (2015). As shown in Table 7, Eq. (12) and Kuwamura and Galambos (1989) models almost have similar performance for the soil type of S1 which are better than the other models. Although Eq. (12) works very well for the soil type of S2, performance metric values of Manfredi (2000) model, excepting *MAPE*, show that it works better than Eq. (12) and the other models. Concerning the soil types of S3 and S4, Eq. (12) has an outperformance compared with the other models in total. Based on the average values of performance metrics listed in Table 7, Eq. (12) has lower *MAPE*, *RRMSE* and *PI*, higher R^2 and E compared with the other models (and slightly lower d and r^2 compared with Manfredi (2000) model), indicating that the predicted model using MOGP totally outperforms the other models available in the literature. The varying trend of the Eq. (12) and Manfredi (2000) model compared with the exact results is also shown in Figure 7. As shown, the varying trend of Eq. (12) is almost conform to the exact results trend excepting a difference for soil type of S1 at medium-periods (see Fig. 7(a)) which is not considerable. This is also true for Manfredi (2000) model except for the long-periods of soil types of S2 and S4 shown in Fig. 7(d).

The mathematical models of Eq. (13) and Eq. (16), proposed for prediction of EDP2 using MOGP, are compared with the presented models by Kuwamura and Galambos (1989), Manfredi (2000), and Dindar et al. (2015). The results of the comparison are listed in Table 8. As shown in this table, MOGP-2 model (Eq. (16)) significantly outperforms MOGP-1 model (Eq. (13)). Eq. (16) is resulted to lower *MAPE*, *RRMSE* and *PI*, and higher R^2, E, d and r^2 compared with those of for other models with the exception of *RRMSE*, *PI*, R^2, E and r^2 of soil type of S2 and r^2 of soil type of S3 for Manfredi (2000) model and soil type of S1 for Kuwamura and Galambos (1989) which are slightly better than those of for Eq. (16). To make an overall comparison, the average values of the performance metrics are listed in Table 8. These results confirm the superiority of the EDP2 prediction model using MOGP-2 (Eq. (16)) as compared to those from the literature. The varying trend of the Eq. (13), Eq. (16) and Manfredi (2000) models compared with the exact results is also shown in Figure 8. As shown, all models relatively have a consistent varying trend to the exact results except for Eq. (13) at medium- and long-periods and for Manfredi model at long-periods of soil types of S1 and S4. Moreover, as shown in Figs. 7 and 8, the effect of PGA on both EDP1 and EDP2 is obvious as it is confirmed by their frequencies in MOGP models depicted in Fig. 6(a) and Fig.6 (b).

A similar comparison was also carried out for EDP3 prediction model of Eq. (14) using MOGP. The predicting equations presented in the works by Kuwamura and Galambos (1989), Fajfar and Vidic (1994), Manfredi (2000), and Dindar et al. (2015) were selected for comparison. The results of this comparison are shown in Table 9. According to the table, although some of the mentioned works show relatively appropriate results, the performance metrics of the prediction model based on Eq. (14) show lower *MAPE*, *RRMSE* and *PI*, and higher R^2, E, d and r^2 as compared to the other presented models, except for R^2 for soil type S1 of the model used in Manfredi (2000). This indicates the novel capability of MOGP for accurate prediction of the EDP3 as compared with those presented in the literature. To make a general comparison on the performance of the MOGP prediction model of EDP3, the average values of the performance metrics are listed in Table 9. The results show that Eq. (14) highly outperforms the other presented models in the literature. Figure 9 also shows the varying trend of Eq. (14) and Manfredi (2000) model as compared to exact values. As depicted in the figure, the varying trend of Eq. (14) is in good agreement with exact values except for soil type S1 which has inconsiderable differences. Despite the exact values and Eq. (14), the Manfredi (2000) model has a constant trend which has significant differences in the majority of periods. It should be noted that, as depicted in Fig. 9, EDP3 is not affected by PGA as evidenced by its very low frequency in MOGP prediction model for EDP3 shown in Fig. 6(c).

EDP4 prediction model of Eq. (15) using MOGP was also compared with the Bakhshi and Tavallali (2006) model. The comparative results are shown in Table 10. According to the table, for all soil types, lower *MAPE*, *RRMSE* and *PI*, and higher R^2, E, d and r^2 is obtained for Eq. (15) with respect to the Bakhshi and Tavallali (2006) model. As shown in the table, average performance metrics are computed for making a general comparison, demonstrating the superiority of Eq. (15) using MOGP in order to predict EDP4 rather than the Bakhshi and Tavallali (2006) model. Figure 10 shows the varying trend of Eq. (15) and Bakhshi and Tavallali (2006) model compared with exact values. The shown graphs confirm the considerable differences between the Bahshi and Tavalali (2006) model and the exact values. In contrast, the MOGP model of Eq. (15) has the closest fit to the exact results. In addition, as shown in this figure, EDP4 is not influenced by PGA as evidenced by its very low frequency in MOGP prediction model for EDP4 shown in Fig. 6(d).

It should be noted that most models in the literature are developed for a system with a limited number of SDOF and a low number of earthquake ground motion records. However, the proposed MOGP-based models can deal with SDOF systems with a wide range of structural features subjected to moderate-to-severe earthquake ground motions..

6. Summary and Conclusion

Formulation of the seismic energy demand of inelastic SDOF systems is one of the main steps of extending the energy-based seismic analysis and design approach. A comprehensive study was carried out to propose accurate and simple mathematical models for predicting seismic energy demand spectra. Multi-objective genetic programming (MOGP) was employed to formulate some main energy-based EDPs, i.e., spectral input and hysteretic energy, spectral hysteretic to input energy ratio, and spectral energy modification factor. Maximizing the accuracy and minimizing the complexity of the predictive models were considered as two objectives of the multi-objective optimization procedure. Both structural and earthquake characteristics were included in the proposed mathematical models. Regarding each EDP, one equation with different coefficients was proposed for various soil types assuming that different soil types have linear relationships. The frequency of each input variable in the best-generated models was also presented to measure the importance of the variable.

Finally, the capability of the proposed models was examined using several common performance metrics. The results indicate the accuracy and effectiveness of the proposed mathematical models in predicting the seismic energy demand spectra compared with some of the models presented in the literature.

References

Akiyama, H., 1985. *Earthquake-resistant limit-state design for buildings*, The University of Tokyo Press, Tokyo, Japan.

Alavi, A.H., Aminian, P., Gandomi, A.H., and Esmaeili, M.A., 2011. Genetic-based modeling of uplift capacity of suction caissons, *Expert Systems with Applications* **10**, 12608–12618.

Alıcı, F.S., and Sucuoğlu, H., 2016. Prediction of input energy spectrum: attenuation models and velocity spectrum scaling, *Earthquake Engineering and Structural Dynamics* **45**, 2137–2161.

Amiri, G.G., Darzi, G.A., and Amiri J.V., 2008. Design elastic input energy spectra based on Iranian earthquakes, *Canadian Journal of Civil Engineering* **35**, 635–646.

ANSI/AISC, 2010. *Specification for Structural Steel Buildings*, American Institute of Steel Construction, Chicago, IL.

Arroyo, D., and Ordaz, M., 2007. On the estimation of hysteretic energy demands for SDOF systems, *Earthquake Engineering and Structural Dynamics* **36**, 2365–82.

Babanajad, S.K., and Gandomi, A.H., Mohammadzadeh, D., Alavi, A.H., 2013. Numerical modeling of concrete strength under multiaxial confinement pressures using linear genetic programming, *Automation in Construction* **36**, 136–144.

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

Bertero, V.V., and Uang, C.M., 1988. Implications of recorded earthquake ground motions on seismic design of building structures, *Research Report*, *UCB/EERC-88/13*, University of California at Berkeley, Los Angeles, CA.

Cabalar, A. F., and Cevik, A., 2011. Triaxial behavior of sand–mica mixtures using genetic programming. *Expert Systems with Applications* **38**, 10358–10367.

Chopra, A.K., 2012. *Dynamics of structures: theory and applications to earthquake engineering*, Fourth Edition, Prentice Hall, CA.

Chou, C.C., and Uang, C.M., 2000 Establishing absorbed energy spectra – an attenuation approach,*Earthquake Engineering and Structural Dynamics* **29**, 1441–1455.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., 2002. A fast and elitist multi objective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation***6**, 182–197.

Decanini, L.D., and Mollaioli, F., 2001. An energy-based methodology for the assessment of seismic demand, *Soil Dynamic and Earthquake Engineering* **21**, 113–137.

Deniz, D., Song, J., and Hajjar, J.F., 2017. Energy-based seismic collapse criterion for ductile planar structural frames, *Engineering Structures* **141**, 1–13.

Dindar, A.A., Yalçın, C., Yüksel, E., Özkaynak, H., and Büyüköztürke, O., 2015. Development of earthquake energy demand spectra, *Earthquake Spectra* **31**, 1667–1689.

Fajfar, P., and Vidic, T., 1994. Consistent inelastic design spectra: hysteretic and input energy, *Earthquake Engineering and Structural Dynamics* **23**, 523–537.

Fajfar, P., Vidic, T., and Fischinger, M., 1989. Seismic demand in medium- and long-period structures, *Earthquake Engineering and Structural Dynamics* **18**, 1133–1144.

Fajfar. P., and Krawinkler. H., 1992. *Nonlinear seismic analysis and design of reinforced concrete buildings*, Elsevier Applied Science, New York.

Federal Emergency Manage Agency (FEMA), 2000. *Prestandard and commentary for the seismic rehabilitation of buildings*. *FEMA-356*, prepared by the American Society of Civil Engineers, Washington, D.C.

Ferreira, C., 2006. *Gene expression programming: mathematical modeling by an artificial intelligence* (Vol. 21). Springer.

Gandomi, A. H., and Alavi, A. H., 2011. Multi-stage genetic programming: a new strategy to nonlinear system modeling. *Information Sciences* **181**, 5227-5239.

Gandomi, A. H., Babanajad, S. K., Alavi, A. H., and Farnam, Y. (2012). Novel approach to strength modeling of concrete under triaxial compression. *Journal of Materials in Civil Engineering* **24**, 1132–1143.

Gandomi, A.H., Alavi, A.H., Mirzahosseini, M.R., and Nejad, F.M., 2010. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures, *Journal of Materials in Civil Engineering* **23**, DOI: 10.1061/(ASCE)MT.1943-5533.0000154.

Gandomi, A.H., and Alavi, A.H., 2012a. A new multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems, *Neural Computing and Applications* **21**, 171–187.

Gandomi, A.H., and Alavi, A.H., 2012b. A new multi-gene genetic programming approach to nonlinear system modeling. Part II: geotechnical and earthquake engineering problems, *Neural Computing and Applications* **21**, 189–201.

Gandomi, A.H., and Roke, D.A., 2015. Assessment of artificial neural network and genetic programming as predictive tools, *Advances in Engineering Software* **88**, 63–72.

Gandomi, A.H., Roke, D.A., and Sett, K., 2013. Genetic programming for moment capacity modeling of ferrocement members, *Engineering Structures* **57**, 169–176.

Gandomi, A.H., Sajedi, S., Kiani, B., and Huang, Q., 2016. Genetic programming for experimental big data mining: A case study on concrete creep formulation, *Automation in Construction* **70**, 89–97.

Gani, A., Siddiqa, A., Shamshirband, S.,and Hanum, F., 2016. A survey on indexing techniques for big data: taxonomy and performance evaluation, *Knowledge and Information Systems* **46**, 241–284.

Gharehbaghi, S., 2018. Damage controlled optimum seismic design of reinforced concrete framed structures, *Structural Engineering and Mechanics* **65**, 53–68.

Gharehbaghi, S., and Khatibinia, M., 2015. Optimal seismic design of reinforced concrete structures subjected to time–history earthquake loads using an intelligent hybrid algorithm, *Earthquake Engineering and Engineering Vibration* **14**, 97–109.

Gholizadeh, S., and Salajegheh, E., 2009. Optimal design of structures for time history loading by swarm intelligence and an advanced metamodel, *Computer Methods in Applied Mechanics and Engineering* **198**, 2936–49.

Gupta, A.K., 1990. *Response spectrum method in seismic analysis and design of structures*. CRC Press, Boca Raton, FL.

Hii, C., Searson, D.P., and Willis, M., 2011. Evolving toxicity models using multigene symbolic regression and multiple objectives, *International Journal of Machine Learning and Computing* **1**, 30–35.

Housner, G.W., 1956. Limit design of structures to resist earthquakes, in *Proceedings*, *of the First World Conference on Earthquake Engineering*, **5**, 1–13.

Kalkan, E., and Kunnath, S.K., 2007. Effective cyclic energy as a measure of seismic demand effective cyclic energy as a measure of seismic demand, *Journal of Earthquake Engineering* **11**,725–751.

Kayadelen, C., Günaydın, O., Fener, M., Demir, A., &Özvan, A. (2009). Modeling of the angle of shearing resistance of soils using soft computing systems. *Expert Systems with Applications* **36**, 11814-11826.

Khan, S.A., Shahani, D.T., and Agarwala, A.K., 2003. Sensor calibration and compensation using artificial neural network, *ISA Trans* **42**, 337–352.

Khashaee, P., 2004. Energy–based seismic design and damage assessment for structures, *Ph.D. Dissertation*, Department of Civil Engineering, Southern Methodist University, USA.

Khatibinia, M., Gharehbaghi, S., and Moustafa, A., 2015. Seismic reliability-based design optimization of reinforced concrete structures including soil-structure interaction effects, *Earthquake Engineering- From*

Engineering Seismology to Optimal Seismic Design of Engineering Structures, Chapter 11, Moustafa A (ed.), InTech, 267–304.

Koza, J. R., 1990. *Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems* (Vol. 34). Stanford, CA: Stanford University, Department of Computer Science.

Kuwamura, H., and Galambos, T.V., 1989. Earthquake load for structural reliability, *Journal of Structural Engineering* **115**, 1446-1462.

Manfredi, G., 2001. Evaluation of Seismic Energy Demand, *Earthquake Engineering and Structural Dynamics* **30**, 485–499.

MATLAB, 2010. The language of technical computing, Math Works Inc.

Metenidis, M.F., Witczak, M., and Korbicz, J., 2004. A novel genetic programming approach to nonlinear system modelling: application to the DAMADICS benchmark problem, *Engineering Applications of Artificial Intelligence***17**, 363–370.

Mirzahosseini, M.R., Aghaeifar, A., Alavi, A.H., Gandomi, A.H., and Seyednour, R., 2011. Permanent deformation analysis of asphalt mixtures using soft computing techniques, *Expert Systems with Applications* **38**, 6081–6100.

Papadopoulos, V., and Giovanis, D.G., Lagaros, N.D., Papadrakakis, M., 2012. Accelerated subset simulation with neural networks for reliability analysis, *Computer Methods in Applied Mechanics and Engineering* **223–224**, 70–80.

PEER Strong Motion Database. 2017; *http://ngawest2.berkeley.edu/.*

Quinde, P., and Reinoso, E., Terán-Gilmore, A., 2016. Inelastic seismic energy spectra for soft soils: Application to Mexico City, *Soil Dynamics and Earthquake Engineering* **89**,198–207.

Sajjadi, S., Shamshirband, S., Alizamir, M., Yee, L., Mansor, Z., Manaf, A.A. et al. 2016, Extreme learning machine for prediction of heat load in district heating systems, *Energy and Buildings* **122**, 222–227.

Salajegheh, E., and Heidari, A., 2005. Optimum design of structures against earthquake by wavelet neural network and filter banks, *Earthquake Engineering and Structural Dynamic* **34**, 67–82.

Searson, D., Willis, M., and Montague, G., 2007. Co-evolution of non-linear PLS model components, *Journal of Chemometrics* **21**, 592–603.

Searson, D.P., 2014, *GPTIPS 2: an open-source software platform for symbolic data mining*, (arXiv preprint arXiv:14124690).

Searson, D.P., Leahy, D.E., and Willis, M.J., 2010. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression, in *Proceedings of the International Multiconference of Engineers and Computer Scientists: Citeseer*, 77–80.

Sucuoğlu, H., and Nurtuğ, A., 1995. Earthquake ground motion characteristics and seismic energy dissipation, *Earthquake Engineering and Structural Dynamics* **24**, 1195–1213.

Tsompanakis, Y., and Topping, B.H.V., 2011. *Soft computing methods for civil and structural engineering*, Saxe-Coburg Publications, Stirlingshire, UK.

Uang, C.M., and Bertero, V.V., 1988. Implications of recorded earthquake ground motions on seismic design of building structures, *Earthquake Engineering Research Center*, *Report No. UCB/EERC-88/13*, University of California, Berkeley.

Uang, C.M., and Bertero, V.V., 1990. Evaluation of seismic energy in structures, *Earthquake Engineering and Structural Dynamics* **19**, 77–90.

Vardhan, H., Garg, A., Li, J., and Garg, A., 2016. Measurement of stress dependent permeability of unsaturated clay, *Measurement* **91**, 371–376.

Walter, E., and Pronzato, L., 1997. *Identification of parametric models: from experimental data*, Communications and Control Engineering, Springer, 413 pp.

Yazdani, H., Khatibinia, M., Gharehbaghi, S., and Hatami, K., 2016. Probabilistic optimum seismic design of reinforced concrete structures considering soil-structure interaction effects, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **3**, G4016004–1–12; DOI: 10.1061/AJRUA6.0000880.

Zahra, T.F., and Hall W.J., 1984. Earthquake energy absorption in SDOF structures, *Journal of Structural Engineering* **110**, 1757–1772.

Zhai, C., Ji, D., Wen, W., Lei, W., Xie, L., and Gong, M., 2016. The inelastic input energy spectra for main shock–aftershock sequences *Earthquake Spectra* **32**, 2149–2166.