



UNIVERSITY OF LEEDS

This is a repository copy of *A hybrid computational approach for seismic energy demand prediction*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/156298/>

Version: Accepted Version

Article:

Gharehbaghi, S, Gandomi, AH, Achakpour, S et al. (1 more author) (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>

© 2018, Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A HYBRID COMPUTATIONAL APPROACH FOR SEISMIC ENERGY DEMAND PREDICTION

S. Gharehbaghi^a, A.H. Gandomi^{b,*}, and S. Achakpour^c, M.N. Omidvar^d

^aDepartment of Civil Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran

^bSchool of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA

^cDepartment of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran

^dCenter 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/>

25	
26	
27	1. Introduction
28	2. Seismic Energy Concept and Formulation
29	3. Genetic Programming
30	3.1. Multi-Gene Symbolic Regression
31	3.2. Multi-Objective Genetic Programming (MOGP)
32	3.3. Accelerating GP Process
33	4. Predicting Seismic Energy Demand Spectra Using MOGP
34	4.1. Preparing Exact Data
35	4.1.1. Inelastic SDOF Systems
36	4.1.2. Earthquake Ground Motions
37	4.2. Model Development Using MOGP
38	5. Results and Discussion
39	5.1. Mathematical Model for EDP1
40	5.2. Mathematical Model for EDP2
41	5.3. Mathematical Model for EDP3
42	5.4. Mathematical Model for EDP4
43	5.5. Accuracy of the Models
44	5.6. Comparative Study
45	5.6.1. Assumptions
46	5.6.2. Comparison
47	6. Summary and Conclusion
48	

49 **1. Introduction**

50 The approaches currently used for the seismic analysis and design of structures need to be improved
51 through considering appropriate engineering demand parameters that would represent the
52 characteristics of a structure and the design earthquake. In current approaches, either the structural
53 members are designed based on satisfying the balance between the force demand and the corresponding
54 strength supply while providing an adequate level of ductility (based on e.g. ASCE/SEI7 2010), or
55 based on the concept whether they are force- or deformation-controlled (based on e.g. FEMA-356
56 2000). Such approaches disregard the frequency content and duration of earthquake ground motion, as
57 well as the velocity response and hysteretic behavior (Gupta 1990). The importance of considering
58 these factors lies in evidences suggesting that, for instance, dissipated hysteretic energy due to repeated
59 inelastic excursions could result in a certain amount of seismic damage (Fajfar and Vidic 1994). In fact,
60 in addition to the force and deformation, the energy demand is of great importance in capturing the
61 mentioned seismic factors as the inelastic behavior is expected to occur due to the design and maximum
62 earthquakes. Housner (1956) was first to introduce these factors through defining the energy concept.
63 This concept requires that the energy dissipation capacity be less than the input energy demand.

64 Both structural properties and earthquake characteristics affect the seismic energy demand.
65 Determining the spectral values of energy demand is beneficial due to its connection to the amount of
66 structural damage (Fajfar and Vidic 1994, Gharehbaghi 2018). The design codes have not explicitly
67 implemented the energy demand parameter in predicting seismic demands yet. Moreover, the priority
68 of the energy-based design approach compared with the conventional strength-based design approach
69 needs further studies. Although, previous studies (e.g. Housner 1956, Fajfar and Vidic 1994, Kalkan
70 and Kunnath 2007, Akiyama 1985, Bertero and Uang 1988, Decanini and Mollaioli 2001, Manfredi
71 2001) have stipulated that the seismic energy parameters are of great importance in seismic design of
72 structures. It was shown that the hysteretic energy demand is directly connected to the structural damage
73 (Fajfar and Vidic 1994, Kalkan and Kunnath 2007, Akiyama 1985, Bertero and Uang 1988, Decanini
74 and Mollaioli 2001, Manfredi 2001, Benavent-Climent et al. 2010, Gharehbaghi 2018). After more than
75 two decades that the Housner's proposal was almost neglected, it was received considerable attention

76 among researchers (Akiyama 1985, Kuwamura and Galambos 1989), and became the key issue of a
77 conference held in Bled city of Slovenia (Fajfar and Krawinkler 1992). It was recognized that input
78 energy and hysteretic energy are the indicators of ground motion and have a correlation with the
79 structural damage, and the quantity related to cumulative damage is the hysteretic energy (Fajfar and
80 Vidic 1994, Bertero and Uang 1988, Decanini and Mollaioli 2001). Most recently, Deniz et al. (2017)
81 also found that the most appropriate and reliable intensity measure for the seismic fragility analysis of
82 buildings is the seismic energy demand.

83 Estimation of input and hysteretic energy demands using mathematical models could be considered
84 as one of the important steps aligned with the extension of the energy-based seismic analysis and design.
85 As previously mentioned, both structural and earthquake characteristics need to be accounted for the
86 issue. Some earthquake characteristics such as soil type, earthquake magnitude, peak ground
87 acceleration (PGA), peak ground velocity (PGV), cumulative energy index, fault type, distance from
88 the hypocenter, were used by researchers in determining the energy spectra (e.g. Zahra and Hall 1984,
89 Uang and Bertero 1988, Fajfar et al. 1989, Uang and Bertero 1990, Sucuoğlu and Nurtuğ 1995,
90 Khashae 2004). In addition to the earthquake characteristics, ductility ratio, damping ratio, and
91 hysteretic behavior model (e.g. elastic-perfectly plastic, bilinear, pinching, Takeda, and Clough models)
92 were the influential structural properties involved in the estimation of seismic energy demand spectra
93 (e.g. Sucuoğlu and Nurtuğ 1995, Decanini and Mollaioli 2001, Benavent-Climent et al. 2011). Several
94 works have been carried out on the estimation of the seismic energy demand parameters. Housner
95 (1956) presented a model to determine input energy based on the spectral velocity of SDOF system.
96 Kuwamura and Galambos (1989) presented energy demand spectra considering the soil type and
97 dominant period of the earthquake. Chou and Uang (2000) estimated absorbed energy for an inelastic
98 system by using an attenuation relation. They used nonlinear regression analysis considering both
99 structural and earthquake variables. Manfredi (2001) proposed simple/efficient mathematical models
100 to estimate input and hysteretic energy spectra. A dimensionless seismic index that is a function of
101 PGA, PGV and cumulative energy was proposed to estimate the seismic energy spectra. Although the
102 estimation models were simple and effective, the effect of soil behavior was not considered, and the

103 number of earthquake ground motions was rather limited. Decanini and Mollaioli (2001) proposed the
104 formulation of elastic seismic energy spectra. They also presented a comprehensive study to propose
105 the design inelastic energy spectra by introducing the response modification factor for the input energy.
106 Several structural variables (e.g. ductility ratio and hysteretic behavior) and earthquake characteristics
107 such as soil type, source-to-site distance, and earthquake magnitude were considered in the proposed
108 spectra. Arroyo and Ordaz (2007) estimated the hysteretic energy demand spectra from elastic response
109 parameters in accordance with the earthquake events recorded in Mexico City. Their mathematical
110 models were a function of pseudo-acceleration, velocity and displacement spectra. Elastic design input
111 energy spectra based on Iranian earthquakes were also presented by Ghodrati Amiri et al. (2008).
112 Recently, Dindar et al. (2015) proposed two regression-based simple mathematical models to estimate
113 the input and hysteretic energy spectra. A database of earthquake ground motion records composed of
114 near- and far-fault ones, PGA, soil types, earthquake magnitude, ductility ratio, and hysteretic behavior
115 model was included in the proposed models. Using the regression analysis, Quinde et al. (2016) also
116 proposed mathematical models to estimate the seismic energy spectra of inelastic systems located on
117 the soft soil for Mexico City. They captured the effect of ductility ratio of inelastic systems and
118 dominant period of the probable earthquakes on the presented models. More recently, Zhai et al. (2016)
119 proposed an expression to account for the effect of after-shock on the input energy spectra using an
120 equivalent velocity. Alici and Sucuoğlu (2016) carried out a regression analysis to estimate inelastic
121 input energy spectrum. The prediction equations for the input energy spectra were expressed in terms
122 of an equivalent velocity. Some crucial earthquake characteristics including soil type, epicentral
123 distance, moment magnitude, and the fault type were considered in the proposed models. All the
124 previously mentioned works use conventional regression methods to estimate their energy parameters
125 of interest.

126 Based on the capability of soft computing approaches and their recent advances, it is worthwhile to
127 use such efficient approaches for seismic demand prediction. Computational complexity of the
128 conventional methods and their limitations has made soft computing techniques, such as evolutionary
129 algorithms, artificial neural networks, support vector machines, and fuzzy logic, popular for solving

130 complex engineering problems. A common application of these tools is in predictive analysis for
131 modeling the nonlinear dependency of the input parameters to the output value(s) where the
132 conventional approaches (e.g. regression analysis) fail or perform poorly (Khan et al. 2003, Gandomi
133 and Roke 2015). Despite the success of artificial neural networks (ANNs) in prediction, they are
134 inappropriate to develop practical intelligible equations. In addition to ANNs, support vector machines
135 (SVMs) are another primary class of soft computing methods used to discover patterns and approximate
136 relationships when large quantities of data is available. Although both ANNs and SVMs have received
137 significant attention (e.g. Salajegheh and Heidari 2005, Gholizadeh and Salajegheh 2009, Papadopoulos
138 et al. 2012, Gharehbaghi and Khatibinia 2015, Khatibinia et al. 2015, Yazdani et al. 2016), they require
139 a pre-defined and initial structure for the equation and network architecture to be determined by the
140 user. Genetic programming (GP), a learning algorithm originated from genetic algorithms, is another
141 well-known and successful technique for developing nonlinear mathematical models for the complex
142 problems. GP and its variants have been effectively used for solving various problems in civil
143 engineering (e.g. Kayadelen et al. 2009, Alavi et al. 2011, Mirzahosseini et al. 2011, Gandomi et al.
144 2012, Vardhan et al. 2016; Lim et al. 2016). Several variants of GP have been proposed in the literature,
145 such as gene expression programming (Ferreira 2006) and multi-stage genetic programming (Gandomi
146 and Alavi 2011). One of the robust variants of GP is multi-gene genetic programming (MGGP) that
147 adds the capability of conventional regression to the standard GP ability in parameter estimation. The
148 effectiveness of MGGP has been proved in the works reported by Gandomi et al. (e.g. Gandomi and
149 Alavi 2012a,b, Gandomi et al. 2013, Babanajad et al. 2013, Gandomi et al. 2016).

150 Structural and earthquake engineering has benefited from the soft computing techniques in different
151 applications. For instance, ANNs and SVMs have been widely used for risk assessment, seismic
152 response prediction, control and health monitoring (Tsompanakis and Topping 2011). In this paper,
153 MOGP is used for predicting the seismic energy demand spectra considering both typical structural and
154 earthquake characteristics. For this purpose, eighteen set of the single-degree-of-freedom (SDOF)
155 systems with the structural properties of different hardening ratios of bilinear hysteretic behavior model,
156 damping ratios, and ductility ratios are used to determine the energy demand spectra mentioned. Also,

157 four different sets of earthquake ground motion records based on their soil types (soft, firm, stiff and
158 rock) with the source-to-site distances of more than 17.5 km and the magnitudes of greater than 5.5
159 were used. It was assumed that the different soil classes have linear relationships in terms of GP genes
160 which help to find one equation with different coefficients for different soil types. The records were
161 scaled to two PGA levels 0.5g and 1.0g. Finally, four mathematical models corresponding to the four
162 engineering demand parameters (EDPs) of spectral input and hysteretic energy, spectral hysteretic to
163 input energy ratio, and spectral energy modification factor, are proposed using MOGP. Then, the
164 effectiveness of the models is revealed using the performance metrics compared with those of available
165 in the literature.

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

172

173 **2. Seismic Energy Concept and Formulation**

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

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

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

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

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

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

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

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

$$E_K(t) + E_D(t) + E_A(t) = E_I(t), \quad (4)$$

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

$$E_K(t) + E_D(t) + E_S(t) + E_H(t) = E_I(t), \quad (5)$$

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

$$E_D(t) + E_H(t) \cong E_I(t). \quad (6)$$

200 According to Eq. (6), the main portion of input energy demand is converted to the damping energy
 201 and hysteretic energy. Besides, if the structural system remains in the elastic range during an

202 earthquake, $E_H(t)$ is trivial, and the energy-based analysis is not useful for seismic design (Uang and
203 Bertero 1990). As mentioned before, for design basis earthquakes, it is expected that a structure will
204 experience inelastic cyclic deformations resulting in hysteretic energy dissipation. As previously
205 mentioned, the hysteretic energy dissipation (E_H) is directly attributed to the structural damage where
206 the E_H/E_I ratio has been introduced as a good indicator of expected damage (Fajfar and Vidic 1994,
207 Sucuoğlu and Nurtuğ 1995, Decanini and Mollaioli 2001, Manfredi 2001). For a given ductility ratio
208 (μ), E_H/E_I ratio is defined as follow:

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

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

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

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

220 3. Genetic Programming

221 There are two groups of models which can be used for modeling the complex nonlinear engineering
222 systems: phenomenological and behavioral (Gandomi et al. 2016). Phenomenological models need a
223 predefined structure obtained from the physical laws requiring a previous understanding about the
224 system. Concerning the complex systems, sometimes it is hard to find such models. Unlike the

225 phenomenological models, behavioral models can be simply generated by finding a reasonable
226 approximate relation between input variables and the output value(s) for a collection of data
227 (experimental or theoretical) irrespective of their governing physical principles. Although one of the
228 main advantages of behavioral modeling techniques is their independence on prior knowledge about
229 the governing physical relationships of input and outputs (Walter and Pronzato 1997, Metenidis 2004),
230 most of these models need the user to pre-assign a formulation pattern requiring optimization of its
231 unknown coefficients. Concerning the complex engineering systems, the use of conventional
232 techniques such as regression analysis cannot be guaranteed to find an accurate and reliable behavioral
233 model (Gandomi et al. 2016). It has been well recognized that most of the structural earthquake
234 engineering problems such as determining earthquake response of inelastic structures could be
235 considered as such complex problems.

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

246 **3.1. Multi-Gene Symbolic Regression**

247 GP can generally be defined as a supervised machine learning technique that searches a program space
248 instead of a parameter space. One of the useful variants of GP is Multi-gene genetic programming
249 (MGGP) (Searson et al. 2007, Gandomi and Alavi 2012a). MGGP is used to design mathematical model
250 predictions which are inherently multi-gene, i.e., those models consisting of linear combinations of low
251 order nonlinear transformations of the input variables. Unlike conventional GP that is based on an

252 evaluation of single tree expression, MGGP uses a single GP particle swarm model selection program
253 constructed from a number of genes where each gene has a tree expression (Searson et al. 2010). In
254 effect, the model development procedure is decomposed by MGGP consisted of some relatively simple,
255 fixed-depth sub-models.

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

265 A fixed linear illustration associated with binary encoding of all parameters is used in GA, which
266 results in a string of numbers as output. In GP however, the optimization problems are solved
267 irrespective of a pre-defined solution structure. Depending on the problem domain, the first generation
268 consisting of a population of possible solutions is randomly created by GP, and variation operators
269 generate new candidate solutions then. During the evolutionary process, crossover selects a node from
270 the parental individuals and exchanges the subtrees under the selected nodes randomly and creates a
271 new individual. Mutation truncates and replaces one node of a tree with another randomly generated
272 node from the same set, and then creates a new individual from an existing tree in the population. The
273 individuals with higher fitness values have higher survival likelihood in the successive generation. To
274 find the best fitting solution, the individual solution of a population is updated after executing a number
275 of runs in each generation, and the parental individuals are selected from the population based on a
276 good fitness function. To develop a population of genes, symbolic regression method can be
277 implemented using standard GP in which a symbolic mathematical expression is directly encoded by
278 each of the genes. Based on Figure 1 showing a typical multi-gene model, known as multi-gene

279 symbolic regression, three input variables (x_1 , x_2 , and x_3) are used to predict the response. As shown in
 280 the figure, although the nonlinear terms such as “sin” and “log” are used, the overall model is a weighted
 281 linear combination of each gene utilizing the coefficients g_0 , g_1 , and g_2 . Mathematically, the general
 282 formulation of the multi-gene symbolic regression model can be written as (Gandomi et al. 2016):

$$\hat{y}(\mathbf{x}, \mathbf{g}, \boldsymbol{\theta}) = g_0 + \sum_{i=1}^n (g_i G_i(\boldsymbol{\theta}, \mathbf{x})) \quad (9)$$

283 where g_0 is a bias term; g_i is the gene weight, and $G_i(\boldsymbol{\theta}, \mathbf{x})$ is the outputs vector from the i th gene
 284 encompassing a multi-gene individual; $\boldsymbol{\theta}$ is the vector of the unknown parameters for each gene; and n
 285 is the number of genes. It should be noted that the algorithmic structure of MGGP, and standard GP is
 286 the same, except for crossover and mutation of multi-gene individuals. MGGP does not necessitate any
 287 simplifying assumption in the model development process, and it is more accurate and efficient than
 288 the standard GP for modeling complex nonlinear problems (Gandomi and Alavi 2012a,b). To construct
 289 an initial population in MGGP, random individuals are created by using different nonlinear functions,
 290 input variables, and a range of random constants and each individual includes 1 and G_{\max} number of
 291 genes. The algorithm attempts to maximize diversity by ensuring that no individuals contain duplicate
 292 genes. The genes are randomly selected and the least squares normal equation is used to estimate the
 293 vector of unknown coefficients \mathbf{g} as follows (Searson et al. 2007, Searson et al. 2010, Hii et al. 2011,
 294 Searson 2014):

$$\mathbf{g} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y} \quad (10)$$

295 where $\mathbf{G} = [1 \ G_1 \dots G_n]$ is the gene response matrix. Since the columns of matrix \mathbf{G} can be collinear, the
 296 Moore–Penrose pseudo-inverse $(\mathbf{G}^T \mathbf{G})^\#$ can be computed by means of the singular value decomposition
 297 instead of the standard matrix inverse $(\mathbf{G}^T \mathbf{G})^{-1}$. Genes can be employed or eliminated using a tree
 298 crossover operator (high-level crossover) during the MGGP run which is in addition to the standard GP
 299 sub-tree crossover (a low-level crossover). The low-level crossover chooses a gene randomly from each
 300 parent individual. Next, the standard sub-tree crossover is employed and the generated trees replace the
 301 parent trees in the otherwise unaltered individual in the next generation. The high-level crossover allows

302 the exchange of one or more genes with another selected individual subject to the G_{\max} constraint. The
303 maximum number of genes of an individual is limited to G_{\max} . If any individual contains more genes,
304 the additional genes are randomly selected and deleted (Searson et al. 2007, Searson et al. 2010, Hii et
305 al. 2011, Searson 2014).

306 **3.2. Multi-Objective Genetic Programming (MOGP)**

307 Generally, both tree-based GP and MGGP deal with a single objective optimization problem for each
308 individual considering the defined fitness function in which, for symbolic regression, the goodness-of-
309 fit to the training data is considered as the only objective to be maximized. Although MGGP yields
310 more compacted models compared with standard GP (Searson 2014), ineffective genes may be acquired
311 by multi-gene models and the single objective optimization problem results in the evolution of overly
312 complex, impractical and non-robust models (Gandomi et al. 2016). The simplest solution to eliminate
313 such shortcoming can be provided by limiting G in a model to G_{\max} which is a hard-to-determine unique
314 value for any given problem (Searson 2014). One good solution is the use of multi-objective concepts
315 into symbolic regression which is commonly referred to as multi-objective genetic programming
316 (MOGP). Using this methodology, both the goodness-of-fit and the complexity of the developed models
317 can be optimized simultaneously by searching the so-called Pareto front (non-dominated solutions) set.
318 Herein, the GPTIPS 2 toolbox (Searson 2014) associated with the related subroutines coded in
319 MATLAB (2013) is used to solve the MOGP by using a non-dominated sorting technique (Deb et al.
320 2002). To sort the non-dominant solutions by their complexity and precision, the non-dominated sorting
321 method is applied at the end of each generation of the MGGP algorithm. At first, the individuals are
322 classified from both the new and old population based on their position on the Pareto front. Pareto front
323 of each level encompasses a set of Pareto optimal solutions which other solutions do not dominate. In
324 addition, the solutions that include Pareto front of each level are not dominated by any other solution,
325 apart from those of in its previous Pareto front level. After that, a “crowding factor” (i.e., the average
326 distance of a solution from the nearest solutions (either side) on the same Pareto front) is computed for
327 separately all individual to increase population diversity, giving lower priority to the solutions that are
328 crowded together during the ranking process. Finally, the position of solutions is used to rank them

329 (those on level 1 are ranked above those on level 2, and so on),and a crowding factor of each solution
330 is used to rank those within the same level. The top 50% of the population is remained to participate in
331 the next generation, whilst the rest are eliminated (Searson 2014).

332 **3.3. Accelerating GP Process**

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

- 341 • 60% of data (2160 samples) were randomly selected and used for training process, and the rest
342 (1440 samples) used as training set for each run;
- 343 • The final Pareto front was determined from merging the Pareto fronts for all runs.

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

353 **4. Predicting Seismic Energy Demand Spectra Using MOGP**

354 **4.1. Preparing Exact Data**

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

358 **4.1.1. Inelastic SDOF Systems**

359 The structural variables used in this study are the hardening ratio of bilinear hysteretic behavior model
360 (η), damping ratio (ξ), and the displacement ductility ratio (μ) which are the structural properties used
361 to determine the energy demand spectra mentioned. The values assumed for η were 0.0, corresponding
362 to the elastic-perfectly plastic model, and 0.1 indicating bilinear model. Three values of 0.05, 0.10 and
363 0.15 were also used for ξ of the inelastic SDOF systems. In addition, three common values of 2, 4 and
364 6 were taken into account for μ . The periods range studied for the prediction of the energy spectra was
365 between 0.01 to 5.0 second for every 0.05 second. These considered variables (η , ξ , and μ) resulted in
366 18 inelastic SDOF systems used for the prediction.

367 **4.1.2. Earthquake Ground Motions**

368 Three factors of the site class, source-to-site distance, and PGA are the three variables considered for
369 the earthquake ground motion records used. Based on the shear wave velocities corresponding to the
370 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
371 S1, S2, S3, and S4 were assumed for the records used. Site types of S1, S2, S3, and S4 are respectively
372 representing the soft, firm, stiff and rock soil types. To consider the source-to-site distance, the records
373 having the Joyne-Boor distance (R_{JB}) in the range of more than 17.5km and less than 150km were used
374 (known as far-fault records). In addition, two values of 0.5g and 1.0g were used for the PGA of the
375 records. All records are non-pulse-like and have the magnitude (M) of greater than 5.5. All the records
376 were downloaded from NGA-West-II project of PEER ground motion database (2017). The diversity
377 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
378 are selected in a way that they have a large variety of the properties mentioned. The individual pseudo-
379 spectral acceleration of each record of each soil type and their mean spectra are also shown in Figure
380 4.

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

387 For each EDP, individual spectral energy responses of the SDOF systems under each set of
388 earthquake records were obtained. Then, based on the normal distribution, the mean plus one standard
389 deviation ($\text{mean} + \sigma$) for each EDP under each set of the earthquake records were determined to be
390 predicted.

391 **4.2. Model Development Using MOGP**

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

406 In order to generate new genes for individuals as well as to decrease the overall number of genes for
407 one model and increase the total number of genes for the other, a “rate-based high-level crossover”

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

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

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

421 **5. Results and Discussion**

422 Using MOGP, all EDPs (EDP1, EDP2, EDP3, and EDP4) were predicted, and their optimal
423 mathematical models (formulations) were determined. Four cases (S1 to S4) based on different soil
424 types of the earthquake records were considered in the prediction. Although it is quite possible that the
425 obtained model formulations be different for different soil types, it is more practical to develop a unique
426 mathematical model for an EDP with different coefficients. Therefore, in this paper, the complete
427 database (which includes four soil types of S1 to S4) is employed to develop a unique prediction model
428 for each EDP. The final mathematical model is selected based on a compromise between the prediction
429 accuracy (as measured by the correlation coefficient R) and the model complexity (as measured by the
430 number of input variables). After that, the complete database is divided into four groups based on the
431 four soil types. Using each group of data, the predicted coefficients of the final model (g_i) are re-
432 evaluated by conducting the regression analysis to reflect the influence of the soil type. Finally, four

433 mathematical models including structural variables (and PGA of earthquake records) with four different
434 groups of coefficients corresponding to the four cases mentioned above (S1 to S4) obtained are
435 presented herein. The contribution of each input variable in the mathematical prediction models and the
436 Pareto front obtained by using a nondominated sorting method at the end of an MOGP run are presented.

437 The results of all EDP models developed by MOGP for EDP1 to EDP4 have been shown in Figures
438 5(a)-(d). The Pareto front sets are shown in green circles and the rest of the models are shown in solid
439 blue circles. As mentioned earlier, the Pareto front set is obtained by using a non-dominated sorting of
440 populations at the end of all MOGP runs. This process simultaneously optimizes the accuracy and the
441 complexity of all developed models. The final model in each Pareto front set is selected and highlighted
442 in a red circle.

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

450 The contribution of each input variable in the mathematical prediction models can be investigated
451 through their frequencies where a frequency value of 1.0 for a variable indicates that it has the maximum
452 contribution within the best-generated models (Gandomi et al. 2010). It was assumed that the models
453 with $R^2 > 0.8$ are the best-generated models. The frequency histograms of the input variables for all
454 predicted EDPs are shown in Fig. 6. As shown, for the selected database of EDP1, PGA and
455 η respectively show the most and the least statistically significant contributions in the best-generated
456 MOGP models. For the collected database of EDP2, although all the input variables almost have
457 significant contributions in the best generated MOGP models, T and ξ are the most and the least
458 influential variables on the EDP2 prediction model. According to the literature (e.g. Kuwamura and
459 Galambos 1989, Manfredi 2001, Dindar et al. 2015), there is no report about the role of PGA on EDP3

460 (hysteretic to input energy ratio). Moreover, based on the physics of the problem, when the damping
 461 ratio is increased, the portion of hysteretic energy dissipation of the imparted seismic input energy is
 462 decreased. These issues have been confirmed by the frequencies of ξ and PGA for EDP3 where they
 463 have the largest and smallest statistically significant contributions, respectively. Moreover, as can be
 464 seen in the figure, for the collected database relevant to EDP4, T and PGA have the largest and smallest
 465 statistically significant contributions in the best-generated MOGP models, respectively.

466 **5.1. Mathematical Model for EDP1**

467 The mathematical model obtained for EDP1 is expressed as follows:

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

$$X_1 = \text{PGA} \cdot \tanh(\tanh(T)) \sqrt{\frac{\xi}{T}}, \quad (12-b)$$

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

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

470 **5.2. Mathematical model for EDP2**

471 The mathematical model derived for EDP2 is expressed as follows:

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

$$Y_1 = \text{PGA}, \quad (13-b)$$

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

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

474 **5.3. Mathematical model for EDP3**

475 The mathematical model derived for EDP3 is expressed as follows:

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

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

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

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:

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

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

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

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:

$$E_{H\mu} = \frac{E_{I\mu}}{m} H_{I\mu} \quad (16)$$

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)

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

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

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

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

$$\text{PI} = \frac{\text{RRMSE}}{R+1}, \quad (19)$$

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

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

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

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

503 The abovementioned performance metrics of the predicted EDP1, EDP2 (including EDP2-1 and
504 EDP2-2), EDP3 and EDP4 models using MOGP are listed in Table 6. As can be seen in this table,
505 EDP1 and EDP2-2 have MAPE values respectively less than almost 16% and 17%, and it is less than
506 33.1% for EDP2-1 all of which are in an acceptable/reasonable range of the performance metric for
507 inelastic and complex systems. The MAPE values are very low for both EDP3 and EDP4. Lower
508 RRMSE values (near-zero usually less than 50%) also confirm the accuracy of the mathematical models.
509 According to Table 6, except for EDP2-1 with soil type of S4 which has an approximate RRMSE of
510 54%, EDP1, EDP2-1 and EDP2-2 respectively have RRMSE values less than 36.3%, 44%, and 42.1%,
511 and the remaining EDPs have RRMSE values less than 6.4%, indicating the accuracy of the predicted
512 models. The higher accuracy of predicted EDP3 and EDP4 is evident. The R^2 , E , d and r^2 values near
513 1.0 (e.g. more than 0.8) indicate a good correlation, efficiency and agreement of the predicted values
514 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-
515 2, EDP3, and EDP4. The minimum values of R^2 , E , d , and r^2 of EDP2-1 are almost equal to 0.68, 0.55,
516 0.83 and 0.65 which belong to the soil type S4 whilst for other soil types almost all of them are more
517 than 0.8. Regarding the PI, that is a combination of R and RRMSE, the values less than 0.3 and 0.2
518 indicate a good and an excellent prediction, respectively. The PI values are less than 0.2 for EDP1, less
519 than 0.3 for EDP2-1, and less than 0.22 for EDP2-2 which indicate a good prediction capability of their
520 corresponding proposed MOGP models. This index is less than 0.04 for both EDP3 and EDP4,
521 confirming the excellent prediction capability of the proposed MOGP models.

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

529 prediction models for EDP1, EDP2-1 and EDP2-2, and excellent capability of MOGP prediction
530 models for EDP3 and EDP4.

531 **5.6. Comparative study**

532 5.6.1. Assumptions

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

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

554 5.6.2. Comparison

555 The mathematical model of Eq. (12), proposed for prediction of EDP1 using MOGP, is compared with
556 the seismic input energy equation presented by Housner (1956), Kuwamura and Galambos (1989),
557 Manfredi (2000), and Dindar et al. (2015). As shown in Table 7, Eq. (12) and Kuwamura and Galambos
558 (1989) models almost have similar performance for the soil type of S1 which are better than the other
559 models. Although Eq. (12) works very well for the soil type of S2, performance metric values of
560 Manfredi (2000) model, excepting MAPE, show that it works better than Eq. (12) and the other models.
561 Concerning the soil types of S3 and S4, Eq. (12) has an outperformance compared with the other models
562 in total. Based on the average values of performance metrics listed in Table 7, Eq. (12) has lower MAPE,
563 RRMSE and PI, higher R^2 and E compared with the other models (and slightly lower d and r^2 compared
564 with Manfredi (2000) model), indicating that the predicted model using MOGP totally outperforms the
565 other models available in the literature. The varying trend of the Eq. (12) and Manfredi (2000) model
566 compared with the exact results is also shown in Figure 7. As shown, the varying trend of Eq. (12) is
567 almost conform to the exact results trend excepting a difference for soil type of S1 at medium-periods
568 (see Fig. 7(a)) which is not considerable. This is also true for Manfredi (2000) model except for the
569 long-periods of soil types of S2 and S4 shown in Fig. 7(d).

570 The mathematical models of Eq. (13) and Eq. (16), proposed for prediction of EDP2 using MOGP,
571 are compared with the presented models by Kuwamura and Galambos (1989), Manfredi (2000), and
572 Dindar et al. (2015). The results of the comparison are listed in Table 8. As shown in this table, MOGP-
573 2 model (Eq. (16)) significantly outperforms MOGP-1 model (Eq. (13)). Eq. (16) is resulted to lower
574 MAPE, RRMSE and PI, and higher R^2 , E, d and r^2 compared with those of for other models with the
575 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)
576 model and soil type of S1 for Kuwamura and Galambos (1989) which are slightly better than those of
577 for Eq. (16). To make an overall comparison, the average values of the performance metrics are listed
578 in Table 8. These results confirm the superiority of the EDP2 prediction model using MOGP-2 (Eq.
579 (16)) as compared to those from the literature. The varying trend of the Eq. (13), Eq. (16) and Manfredi
580 (2000) models compared with the exact results is also shown in Figure 8. As shown, all models
581 relatively have a consistent varying trend to the exact results except for Eq. (13) at medium- and long-

582 periods and for Manfredi model at long-periods of soil types of S1 and S4. Moreover, as shown in Figs.
583 7 and 8, the effect of PGA on both EDP1 and EDP2 is obvious as it is confirmed by their frequencies
584 in MOGP models depicted in Fig. 6(a) and Fig.6 (b).

585 A similar comparison was also carried out for EDP3 prediction model of Eq. (14) using MOGP.
586 The predicting equations presented in the works by Kuwamura and Galambos (1989), Fajfar and Vidic
587 (1994), Manfredi (2000), and Dindar et al. (2015) were selected for comparison. The results of this
588 comparison are shown in Table 9. According to the table, although some of the mentioned works show
589 relatively appropriate results, the performance metrics of the prediction model based on Eq. (14) show
590 lower MAPE, RRMSE and PI, and higher R^2 , E, d and r^2 as compared to the other presented models,
591 except for R^2 for soil type S1 of the model used in Manfredi (2000). This indicates the novel capability
592 of MOGP for accurate prediction of the EDP3 as compared with those presented in the literature. To
593 make a general comparison on the performance of the MOGP prediction model of EDP3, the average
594 values of the performance metrics are listed in Table 9. The results show that Eq. (14) highly
595 outperforms the other presented models in the literature. Figure 9 also shows the varying trend of Eq.
596 (14) and Manfredi (2000) model as compared to exact values. As depicted in the figure, the varying
597 trend of Eq. (14) is in good agreement with exact values except for soil type S1 which has inconsiderable
598 differences. Despite the exact values and Eq. (14), the Manfredi (2000) model has a constant trend
599 which has significant differences in the majority of periods. It should be noted that, as depicted in Fig.
600 9, EDP3 is not affected by PGA as evidenced by its very low frequency in MOGP prediction model for
601 EDP3 shown in Fig. 6(c).

602 EDP4 prediction model of Eq. (15) using MOGP was also compared with the Bakhshi and Tavallali
603 (2006) model. The comparative results are shown in Table 10. According to the table, for all soil types,
604 lower MAPE, RRMSE and PI, and higher R^2 , E, d and r^2 is obtained for Eq. (15) with respect to the
605 Bakhshi and Tavallali (2006) model. As shown in the table, average performance metrics are computed
606 for making a general comparison, demonstrating the superiority of Eq. (15) using MOGP in order to
607 predict EDP4 rather than the Bakhshi and Tavallali (2006) model. Figure 10 shows the varying trend
608 of Eq. (15) and Bakhshi and Tavallali (2006) model compared with exact values. The shown graphs

609 confirm the considerable differences between the Bahshi and Tavalali (2006) model and the exact
610 values. In contrast, the MOGP model of Eq. (15) has the closest fit to the exact results. In addition, as
611 shown in this figure, EDP4 is not influenced by PGA as evidenced by its very low frequency in MOGP
612 prediction model for EDP4 shown in Fig. 6(d).

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

617 **6. Summary and Conclusion**

618 Formulation of the seismic energy demand of inelastic SDOF systems is one of the main steps of
619 extending the energy-based seismic analysis and design approach. A comprehensive study was carried
620 out to propose accurate and simple mathematical models for predicting seismic energy demand spectra.

621 Multi-objective genetic programming (MOGP) was employed to formulate some main energy-based
622 EDPs, i.e., spectral input and hysteretic energy, spectral hysteretic to input energy ratio, and spectral
623 energy modification factor. Maximizing the accuracy and minimizing the complexity of the predictive
624 models were considered as two objectives of the multi-objective optimization procedure. Both
625 structural and earthquake characteristics were included in the proposed mathematical models.
626 Regarding each EDP, one equation with different coefficients was proposed for various soil types
627 assuming that different soil types have linear relationships. The frequency of each input variable in the
628 best-generated models was also presented to measure the importance of the variable.

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

633

634 **References**

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

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

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

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

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

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

647 Babanajad, S.K., and Gandomi, A.H., Mohammadzadeh, D., Alavi, A.H., 2013. Numerical modeling of concrete
648 strength under multiaxial confinement pressures using linear genetic programming, *Automation in Construction*
649 **36**, 136–144.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

683 Gandomi, A.H., Alavi, A.H., Mirzahosseini, M.R., and Nejad, F.M., 2010. Nonlinear genetic-based models for
684 prediction of flow number of asphalt mixtures, *Journal of Materials in Civil Engineering* **23**, DOI:
685 10.1061/(ASCE)MT.1943-5533.0000154.

686 Gandomi, A.H., and Alavi, A.H., 2012a. A new multi-gene genetic programming approach to nonlinear system
687 modeling. Part I: materials and structural engineering problems, *Neural Computing and Applications* **21**, 171–
688 187.

689 Gandomi, A.H., and Alavi, A.H., 2012b. A new multi-gene genetic programming approach to nonlinear system
690 modeling. Part II: geotechnical and earthquake engineering problems, *Neural Computing and Applications* **21**,
691 189–201.

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

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

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

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

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

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

705 Gholizadeh, S., and Salajegheh, E., 2009. Optimal design of structures for time history loading by swarm
706 intelligence and an advanced metamodel, *Computer Methods in Applied Mechanics and Engineering* **198**, 2936–
707 49.

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

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

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

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

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

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

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

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

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

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

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

730 Manfredi, G., 2001. Evaluation of Seismic Energy Demand, Earthquake Engineering and Structural Dynamics
731 **30**, 485–499.

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

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

736 Mirzahosseini, M.R., Aghaeifar, A., Alavi, A.H., Gandomi, A.H., and Seyednour, R., 2011. Permanent
737 deformation analysis of asphalt mixtures using soft computing techniques, Expert Systems with Applications **38**,
738 6081–6100.

739 Papadopoulos, V., and Giovanis, D.G., Lagaros, N.D., Papadrakakis, M., 2012. Accelerated subset simulation
740 with neural networks for reliability analysis, Computer Methods in Applied Mechanics and Engineering **223–224**,
741 70–80.

742 PEER Strong Motion Database. 2017; <http://ngawest2.berkeley.edu/>.

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

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

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

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

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

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

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

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

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

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

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

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

769 Yazdani, H., Khatibinia, M., Gharehbaghi, S., and Hatami, K., 2016. Probabilistic optimum seismic design of
770 reinforced concrete structures considering soil-structure interaction effects, ASCE-ASME Journal of Risk and
771 Uncertainty in Engineering Systems, Part A: Civil Engineering **3**, G4016004–1–12; DOI:
772 10.1061/AJRUA6.0000880.

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

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