Intelligent Modeling and Prediction of Elastic Modulus of Concrete Strength via Gene Expression Programming

Amir Hossein Gandomi[1], Amir Hossein Alavi[2], T.O. Ting[3,*], and Xin-She Yang[4]

[1] Department of Civil Engineering, The University of Akron, Akron, OH 44325-3905, USA
[2] Department of Civil and Environmental Engineering, Engineering Building, Michigan State University, East Lansing, MI, 48824, USA
[3] Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, P.R. China
[4] School of Science and Technology, Middlesex University Hendon Campus, London, UK
ag72@uakron.edu, alavi@msu.edu, toting@xjtlu.edu.cn, x.yang@mdx.ac.uk

Abstract. The accurate prediction of the elastic modulus of concrete can be very important in civil engineering applications. We use gene expression programming (GEP) to model and predict the elastic modulus of normal-strength concrete (NSC) and high-strength concrete (HSC). The proposed models can relate the modulus of elasticity of NSC and HSC to their compressive strength, based on reliable experimental databases obtained from the published literature. Our results show that GEP can be an effective method for deriving simplified and precise formulations for the elastic modulus of NSC and HSC. Furthermore, the comparison study in the present work indicates that the GEP predictions are more accurate than other methods.

Keywords: Tangent elastic modulus, Normal and High strength concrete, Gene expression programming, Compressive strength, Formulation.

1 Introduction

In many civil engineering applications, to estimate the material properties such as elastic modulus is very important to meet design requirements. For example, the elastic modulus of normal and high strength concrete is a key parameter in structural engineering, and this parameter helps to determine the static and time-dependent deformation and system behaviour. It is also related to the assessment of other key processes such as creep, shrinkage, crack propagation and control in both reinforced concrete and prestressed concrete [1,2]. From the slope of a stress-strain curve of a given concrete material, we can estimate the elastic modulus of the sample.

Despite its importance, the elastic modulus is not usually measured in situ as it is time-consuming and expensive. The common practice is to estimate it using empirical relationships, based on various codes of practice. Such models often link the elastic

---

* Corresponding author.

modulus with compressive strength, which essentially eliminate the need for going through laborious and time-consuming direct measurements from load-deformation curve [2, 3].

In recent years, techniques such as pattern recognition systems have received much attention in civil engineering applications. These systems are trained based on empirical data and thus can extract various discriminators. Loosely speaking, in the context of engineering applications, Artificial Neural Networks (ANNs), Fuzzy Logic (FL), Adaptive Neuro Fuzzy Inference System (ANFIS), and Support Vector Machine (SVM) can all be referred to as pattern recognition methods. Not surprisingly, these techniques have been used in predicting the elastic modulus of normal and high strength concrete (NSC and HSC) [4-6]. Although ANNs, FL, ANFIS, and SVM are successful in prediction, they cannot produce explicit equations for predictions, and thus limiting their usage.

In this paper, we present an alternative approach to produce explicit equations for elastic modulus of concrete materials by using genetic programming (GP), and this partly overcomes the limitations of ANNs, FL, ANFIS, and SVM for this type of applications. To achieve this goal, we investigate a relatively new variant of GP, namely gene expression programming (GEP) [7] that have been used to solve civil engineering applications such as concrete modeling [2,8,9]. In our predictions and model formulation, we have used reliable databases of previously published test results. A comparative study is carried out between the results obtained by GEP and those obtained from the buildings codes [10-13], compatibility aided [14, 15], FL [4], and ANN [5] models. The rest of the paper is organized as follows: Section 2 provides a brief description of the gene expression programming. In Section 3, a detailed study of model prediction of concrete strength and parameters using GEP is presented. Further, Section 4 provides the performance comparison and analysis and finally we draw brief conclusions in Section 5.

2   Gene Expression Programming

Genetic programming is a branch of artificial intelligence techniques that creates computer programs to solve a problem by mimicking the evolution of living or biological organisms [16]. In essence, the main aim of this method is to use inputs and their corresponding output data samples so as to create a computer program that connects them with the minimum fitting or prediction errors. The major difference between GP and genetic algorithms (GA) is the way of representing the solutions. In GA, a solution is represented by a string of numbers, either binary or real, while in the classical GP, solutions are represented as computer programs in terms of tree structures and are the expressed in a functional programming language (such as LISP) [2, 8]. In GP, a random set or population of individuals (computer programs) are created and evolved in an iterative manner to achieve sufficient diversity. A comprehensive description of GP can be found in Koza (1992) [16]. GEP is a new variant of GP first proposed by Ferreira [17]. GEP has five main components: function set, terminal set, fitness function, control parameters, and termination condition. GEP uses a fixed length of character strings to represent solutions in a

domain of interest, which are then expressed as parse trees with different sizes and shapes. These trees are called GEP expression trees (ET). A main advantage of the GEP technique is that its creation of genetic diversity in solution is simplified and carried out using genetic operators that work at the chromosome level. In GEP, individuals are selected and copied into the next generation according to their fitness by the so-called roulette wheel sampling technique, together with elitism. This essentially guarantees the survival and cloning of the best individual to the next generation, which may speed up the overall convergence rate. Variations in the population are introduced by applying single or several genetic operators on selected chromosomes, and these genetic operators include crossover, mutation and rotation [7, 18]. The GEP algorithm has four main steps until it reaches one of the stop criteria [9, 17]:

I.   Randomly generating the fixed-length chromosomes as initial population.
II.   Expressing chromosomes as expression trees and evaluating fitnesses.
III.   Selecting the best individuals according to their fitnesss to reproduce with modification.
IV.   Repeating the steps II an III until a termination condition is reached.

3   GEP-Based Modelling of Elastic Modulus of NSC and HSC

The main goal of this study is to obtain the prediction equations for elastic modulus (E_c) of NSC and HSC in terms of compressive strength (f_c) in the following generic form:

$$E_c = f(f_c)   \tag{1}$$

Hence, there is only one parameter that has been used for the GEP models as the input variable. Using reliable databases for the NSC and HSC, two different GEP-based formulas for the elastic modulus of NSC and HSC can be obtained. In this study, basic arithmetic operators and mathematical functions are utilized to obtain the optimum GEP models. The actual number of generation depends on the number of possible solutions and complexity of the problem. However, it must be set properly before the runs. A large number of generations has to be tested so as to find the models with minimum errors. The program is run iteratively until there is no longer significant improvement in the performance of the models, or a specified number of iterations is reached. The values of the other parameters are selected, based on some previously suggested values [7, 18] or determined by a trial and error approach. For the GEP-based analysis, we adopted the computer software known as GeneXproTools [19]. The best GEP model is chosen on the basis of a multi-objective strategy as below:

i.   The simplicity of the model, although this is not a predominant factor.
i.   The goodness of the best fit on the training set of data.
iii.   The best fitness value on a test set of unseen data.

The first objective can be controlled by the user through the parameter settings (e.g., head size or number of genes), while for the other two objectives, the following

objective function (Obj) is constructed as a measure of how well the model fits the experimental data [2]. The selection criterion of the best GEP model is based on the minimization of the following function:

$$Obj = \left(\frac{N_{Train} - N_{Test}}{N_{All}} \right) \frac{MAE_{Train}}{R_{Train}^2} + \frac{2N_{Test}}{N_{All}} \frac{MAE_{Test}}{R_{Test}^2} \tag{2}$$

where N_{Train}, N_{Test} and N_{All} are the numbers of training, testing and whole, respectively, of data. R and MAE are the correlation coefficient and mean absolute error, respectively. The above objective function has taken into account the changes of R and MAE together. Higher R values and lower MAE values result in lower Obj and, consequently, corresponds to a more precise model. In addition, the above function has also taken into account the effects of different data divisions between the training and testing data.

3.1   Experimental Database

The experimental database of previously published test results consist of 89 and 70 test results for the elastic modulus of HSC and NSC, respectively [7]. Descriptive statistics of the variables used in the model development are given in Fig. 1.
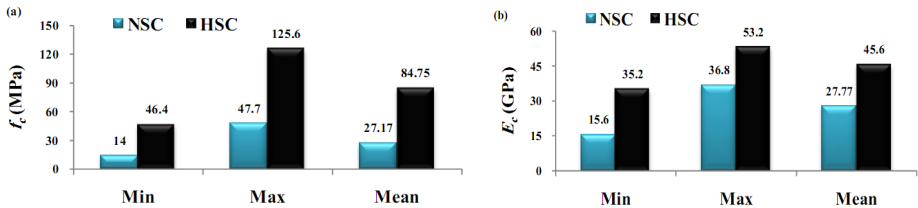

Fig. 1. Descriptive statistics of the variables

For the analysis, the data sets have been divided into training and testing subsets. The training data are applied in the learning process via genetic evolution whereas the validation data were used to measure the prediction capability of the obtained models on data that played no role in building the models. Out of 89 data sets for HSC, 69 values were taken for training of the GEP algorithm and the remaining 20 values are used for the testing and prediction. For NSC, 57 values are taken for the training process and the remaining 13 values are used for testing of the models. Out of a the total 159 data sets for HSC and NSC, 126 values were used for the training, 33 values were used for the testing of the generic model for both HSC and NSC. From these simulation, training and multiple runs, the main results can be summarized in the following sections.

3.2   Explicit Formula for Elastic Modulus of HSC and NSC

The GEP-based formulation of the *Ec* of HSC in terms of f_c is as given below:

$$E_{c,GEP} = 7\left(\sqrt[3]{7 + f_c} + 2\right) \tag{3}$$

This proposed model for the *Ec* of HSC gives a value of 5.462 (*Obj*=5.462). The expression tree of the above formulation is given in Fig. 2. The comparisons of the GEP predicted values against experimental elastic modulus of HSC are shown in Fig. 3.
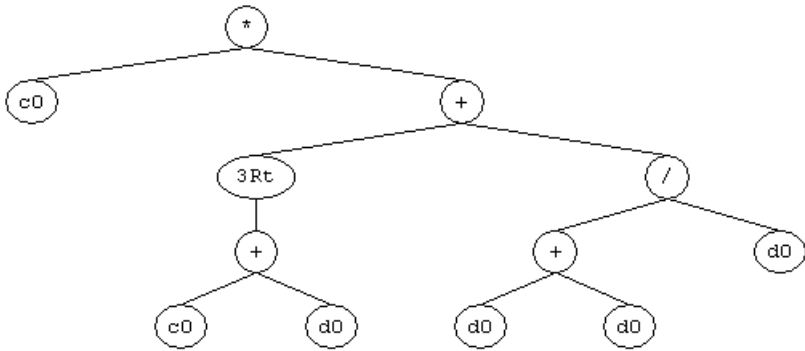

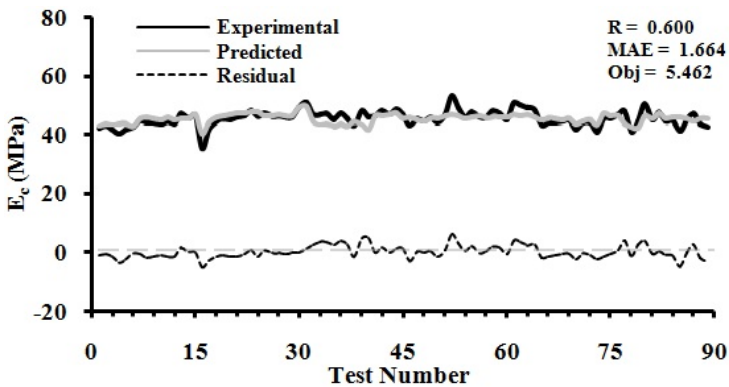**Fig. 2.** Expression tree for Ec of HSC (d0 = f_c)

Fig. 3. Predicted versus experimental E_c of HSC using the GEP model

The GEP-based formulation of the E_c of NSC in terms of f_c can be written as

$$E_{c,GEP} = \sqrt[3]{875 f_c - 2100}, \tag{4}$$

which yields an Obj value of 7.841. The expression tree of the above formulation is given in Fig. 4. Comparisons of the GEP predicted values against experimental elastic modulus of NSC are shown in Fig. 5.
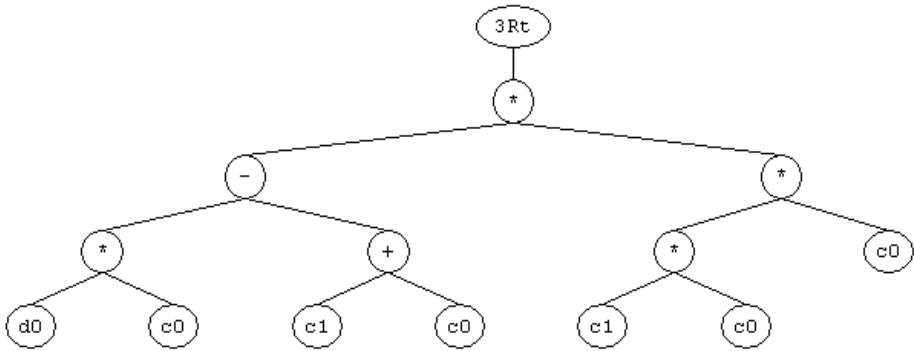

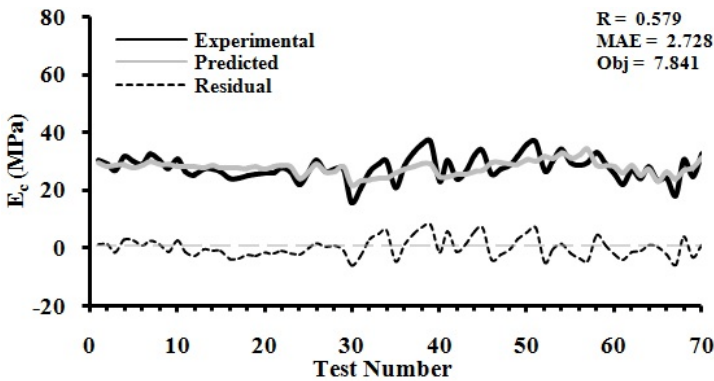**Fig. 4.** Expression tree for E_c of HSC and NSC (d0 = f_c)

Fig. 5. Predicted versus experimental E_c of NSC using the GEP model

4   Performance Analysis

Table 1 shows the prediction performance of the GEP models, Iranian (NBS) [10], American (ACI 318-95) [11], Norwegian (NS 3473) [12], and Turkish (TS 500) [13] codes, two compatibility aided model [14, 15], FL [6], and ANN [7] models for the E_c of NSC and HSC, respectively. It can be clearly seen from this table that the proposed GEP models provide more accurate predictions than the available codes and models for the elastic modulus of HSC and NSC. However, the exception is the FL and ANN models for HSC provide better results than the GEP models.

Table 1. Comparisons between the GEP models and other models in the literature

HSC			NSC		
Model	MAE (%)	R	Model	MAE (%)	R
FL [6]	0.0368	0.6130	FL [6]	0.1031	0.5536
ANN [7]	0.0365	0.6354	ANN [7]	0.1032	0.5151
ACI [11]	0.1808	0.6024	ACI [11]	0.1327	0.5784
NS [12]	0.2124	0.5916	NBS [10]	0.1057	0.5719
[14]	0.0412	0.5577	TS [13]	0.1411	0.5693
[15]	0.1354	0.6002	[15]	0.1028	0.5839
GEP	0.0374	0.6005	GEP	0.0982	0.5795

Numerically, although the ANN and FL models have a good performance, they do not give any explicit function or formula. ANN has only final synaptic weights to obtain the outcome in a parallel manner. The determination of the fuzzy rules in FL is also a non-trivial task [8]. In addition, the ANN and FL approaches are appropriate to be used as a part of a computer program and may not be suitable for practical calculations such as *in situ* applications.

5   Conclusion

We have adopted a relatively new technique, GEP, to obtain best-fit equations for predicting the elastic modulus of HSC and NSC. Two design formulas for the elastic modulus have been obtained via GEP using a reliable database of previously published elastic modulus test results. The database is used for the training and testing of the prediction models. The GEP models can indeed give reliable estimations of the elastic modulus of HSC and NSC. The obtained formulas and proposed approach can outperform the other existing models in nearly all cases. In addition to the advantages of the acceptable accuracy, the GEP-based prediction equations are really simple to use, and can thus be used reliably for practical pre-planning and pre-design purposes by simple calculations. Such simple models for estimating elastic moduli are advantageous due to the demand in carrying out destructive, sophisticated and time-consuming laboratory tests. Further studies can focus on the extension of the proposed approach to model prediction equations for other time-consuming tasks and key parameters in engineering applications with reliable databases.

References

1. Mesbah, H.A., Lacherni, M., Aitcin, P.C.: Determination of elastic properties of high performance concrete at early age. ACI Material Journal 99(1), 37–41 (2002)
2. Gandomi, A.H., Alavi, A.H., Sahab, M.G., Arjmandi, P.: Formulation of Elastic Modulus of Concrete Using Linear Genetic Programming. Journal of Mechanical Science and Technology 24(6), 1011–1017 (2010)
3. ASTM C 469. Standard test method for static modulus of elasticity and poisson's ratio of concrete in compression. Annual Book of ASTM standards (1994)

4. Demir, F.: A new way of prediction elastic modulus of normal and high strength concrete–fuzzy logic. Cement and Concrete Research 35, 1531–1538 (2005)
5. Demir, F.: Prediction of elastic modulus of normal and high strength concrete by artificial neural networks. Construction and Building Materials 22, 1428–1435 (2008)
6. Yan, K., Shi, C.: Prediction of elastic modulus of normal and high strength concrete by support vector machine. Construction and Building Materials 24(8), 1479–1485 (2010)
7. Alavi, A.H., Gandomi, A.H.: A robust data mining approach for formulation of geotechnical engineering systems. Engineering Computations 28(3), 242–274 (2011)
8. Gandomi, A.H., Babanajad, S.K., Alavi, A.H., Farnam, Y.: A Novel Approach to Strength Modeling of Concrete under Triaxial Compression. Journal of Materials in Civil Engineering-ASCE 24(9), 1132–1143 (2012)
9. Gandomi, A.H., Alavi, A.H.: Expression Programming Techniques for Formulation of Structural Engineering Systems. In: Gandomi, A.H., et al. (eds.) Metaheuristic Applications in Structures and Infrastructures, ch. 18. Elsevier, Waltham (2013)
10. NBS. Analysis and Design of Reinforced Concrete Buildings, National Building Standard, Part 9, Iran (2006)
11. ACI 318-95, Building code requirements for structural concrete. ACI Manual of Concrete Practice Part 3: Use of concrete in Buildings –Design, Specifications, and Related Topics. Detroit, Michigan (1996)
12. NS 3473. Norwegian Council for Building Standardization. Concrete Structures Design Rules. Stockholm (1992)
13. TS 500. Requirements for design and construction of reinforced concrete structures. Ankara: Turkish Standardization Institute (2000)
14. Wee, T.H., Chin, M.S., Mansur, M.A.: Stress–strain relationship of high-strength concrete in compression. Journal of Materials in Civil Engineering-ASCE 8(2), 70–76 (1994)
15. Mostofinejad, D., Nozhati, M.: Prediction of the modulus of elasticity of high strength concrete. Iranian Journal of Science & Technology, Transaction B: Engineering 29(B3), 85–99 (2005)
16. Koza, J.: Genetic programming, on the programming of computers by means of natural selection. MIT Press, Cambridge (1992)
17. Ferreira, C.: Gene expression programming: a new adaptive algorithm for solving problems. Complex Syst. 2001 13(2), 87–129 (1988)
18. Gandomi, A.H., Alavi, A.H., Mirzahosseini, M.R., Moqhadas Nejad, F.: Nonlinear Genetic-Based Models for Prediction of Flow Number of Asphalt Mixtures. Journal of Materials in Civil Engineering-ASCE 23(3), 248–263 (2011)
19. GEPSOFT. GeneXproTools. Version 4.0 (2006), http://www.gepsoft.com