International Journal of Civil Engineering

Geotechnical Engineering

Evolutionary-based approaches for settlement prediction of shallow foundations on cohesionless soils

H. Shahnazari[1,]*, M.A. Shahin[2], M.A. Tutunchian[1]

Received: February 2013, Revised: May 2013, Accepted: June 2013

Abstract

Due to the heterogeneous nature of granular soils and the involvement of many effective parameters in the geotechnical behavior of soil-foundation systems, the accurate prediction of shallow foundation settlements on cohesionless soils is a complex engineering problem. In this study, three new evolutionary-based techniques, including evolutionary polynomial regression (EPR), classical genetic programming (GP), and gene expression programming (GEP), are utilized to obtain more accurate predictive settlement models. The models are developed using a large databank of standard penetration test (SPT)-based case histories. The values obtained from the new models are compared with those of the most precise models that have been previously proposed by researchers. The results show that the new EPR and GP-based models are able to predict the foundation settlement on cohesionless soils under the described conditions with R^2 values higher than 87%. The artificial neural networks (ANNs) and genetic programming (GP)-based models obtained from the literature, have R^2 values of about 85% and 83%, respectively which are higher than 80% for the GEP-based model. A subsequent comprehensive parametric study is further carried out to evaluate the sensitivity of the foundation settlement to the effective input parameters. The comparison results prove that the new EPR and GP-based models are the most accurate models. In this study, the feasibility of the EPR, GP and GEP approaches in finding solutions for highly nonlinear problems such as settlement of shallow foundations on granular soils is also clearly illustrated. The developed models are quite simple and straightforward and can be used reliably for routine design practice.

Keywords: *Shallow foundations, Settlement prediction, Evolutionary polynomial regression, Genetic programming, Gene expression programming, Cohesionless soils.*

1. Introduction

Shallow foundations are one of the most common structures for transferring loads to the near-surface ground. Generally, there are two main criteria that must be considered in the design process of shallow foundations, including the soil-bearing capacity and foundation settlement. However, the design of shallow foundations on cohesionless soils is generally more controlled by settlement than bearing capacity [1]. Thus, accurate settlement prediction is of paramount importance. The settlement of shallow foundations on cohesionless soils usually occurs because of the following two main reasons: (i) soil compressibility due to induced applied stresses and consequent rearrangement of soil particles and (ii) lateral deformation of the foundation subsoil because of the tendency of soil to move away from underneath the foundation.

** Corresponding author: hshahnazari@iust.ac.ir*
1 School of Civil Engineering, Iran University of Science and Technology, P.O. Box 16765-163, Narmak, Tehran, Iran
2 Department of Civil Engineering, Curtin University, Perth Western Australia 6845, Australia

Poulus listed the following major causes of shallow foundation settlement [2]:

(i) Static loads imposed by the weight of structures.

(ii) Dynamic loads produced by machinery, earthquakes, moving loads, etc.

(iii) Changes in the moisture content of soils due to various natural phenomena, such as seasonal fluctuation in the water table.

(iv) Further effects of nearby construction projects that mainly result from adjacent excavation, pile driving and dewatering.

In this paper, the first case will be addressed. Predicting the settlement of a shallow foundation on cohesionless soils due to static loads is considered to be a compound geotechnical problem because it involves some uncertainties. The stress-strain history of subsoil, compressibility potential of soil, real distribution of applied stresses, difficulty in obtaining undisturbed samples of cohesionless soil, intricacy in forecasting the real magnitude of the imposed loads and heterogeneous nature of the soils are the main factors of the uncertainties.

The estimation of shallow foundation settlement on cohesionless soils has been a topic of research that attracted many investigators and several estimation methods have been proposed over the years. The most widely used methods include those proposed by Terzaghi and Peck [3], Meyerhof [4], Bazaraa [5], Schmertmann [1], Schultze and Sherif [6], Schmertmann et al. [7] and Burland and Burbidge [8]. Comprehensive studies of existing methods (e.g., Jorden [9]; Jeyapalan and Boehm [10]; Gifford et al. [11]; Tan and Duncan [12]; Wahls [13]; Sivakugan et al. [14]) indicate inconsistencies in the settlements predicted by these methods; therefore, more accurate and consistent methods are still needed. For instance, a comparison between the observed and predicted settlements according to the Terzaghi and Peck method for 79 cases is shown in Figure 1 (Sivakugan et al. [14]). This figure demonstrates that this method involves a safety factor of approximately 2.2 for most of the cases, which leads to a conservative design.

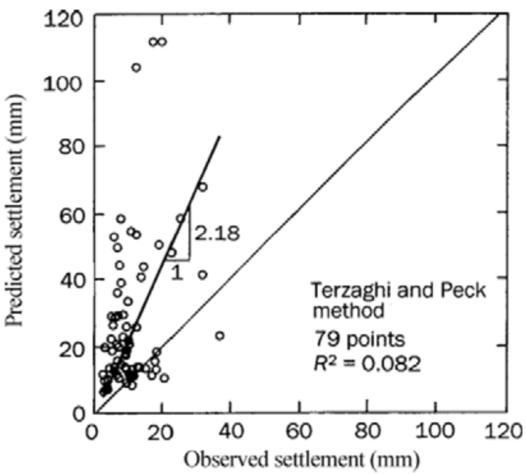

Fig. 1 Comparison of predicted with observed settlements for 79 foundations based on Terzaghi and Peck method (Sivakugan et al., 1998)

More recently, the feasibility of using soft computing to more accurately and consistently predict the settlement of shallow foundations on cohesionless soils has been investigated. For instance, Shahin et al. [15] applied artificial neural networks (ANNs), which were shown to have great potential in predicting the complex, nonlinear behavior of shallow footings on sandy soils. In addition, Rezania and Javadi [16] proposed the use of genetic programming (GP), and Samui and Sitharam [17] suggested the least squares support vector machines (LSSVM).

The current paper represents an attempt to obtain an accurate settlement prediction of shallow foundations on cohesionless soils using a new evolutionary-based approach, i.e., evolutionary polynomial regression (EPR). In addition, two new GP-based modeling techniques including the classical GP model and gene expression programming (GEP) are developed to compare with the new EPR-based formula. A reliable database gathered from different case histories was utilized to develop the models. The settlements predicted from the proposed formulations were compared with those previously obtained from the soft-computing techniques. A comparative parametric study confirmed the robustness of the proposed numerical correlations.

2. Evolutionary Polynomial Regression (EPR)

Soft computing techniques have been developed rapidly during recent years. They have been applied to different civil engineering complicated problems such as predicting the behavior of plastic pipes embedded in reinforced sand [18], predicting the creep effects in masonry structures [19, 20], determining the deviatoric stress of calcareous sands [21], and predicting the pavement condition index (PCI) [22].

Recenetly, Giustolisi and Savic [23] developed a novel data-driven method, i.e., evolutionary polynomial regression (EPR), based on evolutionary computing that combines the best features of conventional numerical regression techniques with genetic programming (GP) and symbolic regression techniques. One of the most useful applications of EPR is finding the best model to fit observed data (e.g., fitting a line or curve through a set of points). Evaluation of liquefaction potential based on cone penetration test (CPT) results [24], assessment of earthquake-induced soil liquefaction and lateral displacement [25] and prediction of total sediment load of rivers [26] are some instances of EPR applications in civil engineering.

A physical system with an output y is based on a set of input variables $\boldsymbol{X}$ and parameters $\boldsymbol{\theta}$ can be formulated mathematically as follows:

$$y = F(\mathrm{X}, \theta) \tag{1}$$

Where F is a function of n-dimensional space in which n is the number of inputs. Artificial neural networks (ANNs) and genetic programming (GP) are two well-known methods that try to reconstruct the function F using an input/output dataset. GP generates a population of terms for F, coded in tree-based structures of variable size, and explores the best format for F based on a fitness function. On the other hand, ANNs focus on mapping F rather than finding a reasonable structure for it. GP considers the functional relationships between the variables $\boldsymbol{X}$. However, ANNs function on a lower level of knowledge of the functional relationships between $\boldsymbol{X}$ [23].

Both methods (i.e., ANNs and GP) are powerful techniques with great potential in modeling the nonlinear complex problems that are difficult to model with conventional methods, but they have their inherent negative aspects. The main drawbacks are as follows [23, 24]:

(i) GP uses an evolutionary approach to determine a mathematical form of F, but the parameter values (i.e., vector $\boldsymbol{\theta}$) are generated as nonadjustable constants, referred to as ephemeral random constants. Therefore, the constants do not necessarily represent optimal values, and good structures of F may be missed in the modeling process because of unsuitable constants.

(ii) Bloating may occur within GP processes. Bloating causes an excessive growth of the GP-based expressions without any effective advance in its overall performance. Therefore, bloating can cause the evolutionary process of GP to be relatively inefficient.

(iii) The main drawback of the ANNs is the difficulty in determining the network structure due to its great complexity because it represents knowledge in terms of a weight matrix together and bias terms, which are not accessible or easy to work with.

(iv) The risk of becoming stuck in local minima for the training algorithm is the other considerable disadvantage of the ANN approach.

Compared with the other data-driven and classical regression techniques, EPR is a hybrid data-driven technique that attempts to overcome some of the abovementioned shortcomings. The most beneficial features of EPR include:

(i) Needs a small number of constants to be estimated, which helps avoid over-fitting problems, especially for small data sets.

(ii) Uses linear parameter estimation, ensuring that a unique solution is found when the inverse problem is well conditioned.

(iii) Can feature automatic model construction, which avoids the need to preselect the functional form and the number of parameters in the target model.

(iv) Can perform both linear and nonlinear analyses in a single iteration.

EPR can be expressed in the following general mathematical form [23]:

$$y = \sum_{j=1}^{n} F(\mathrm{X}, f(\mathrm{X}), a_j) + a_0 \quad (2)$$

Where y is the estimated vector of the process outputs; F is the function generated by the process; $\mathbf{X}$ is a matrix of input variables; f is a function defined by the user; a_j is a constant value; and n is the number of terms of the target expression.

EPR has two major stages in the process of constructing a symbolic regression model. The first stage is the identification of the main structure, and the second is the parameter estimation. EPR first uses the standard genetic algorithm (GA) strategy to search for the best form of the model structure (i.e., functional form of the main expression). In the second stage, EPR performs a least-square regression to find the adjustable parameters. In other words, it estimates function parameters based on a linear optimization technique. As stated by Giustolisi and Savic [23], the two abovementioned steps represent the simplest relation between the symbolic essence and numerical regressive nature of EPR (i.e., functional form and parameters, respectively). More explanations of the technique can be found in Giustolisi and Savic [23, 27] and Rezania et al. [28].

3. Development of EPR-Based Settlement Model

3.1. Model inputs and outputs

The recognition of the effective parameters in the settlement value of a shallow foundation on cohesionless soils has a direct impact on the precision of a predictive numerical model and strongly supports the technical acceptance of the model. In most important traditional methods, the foundation width (B), foundation net applied pressure (q), foundation embedment ratio (i.e., ratio of embedment dapth to foundation width, D_f/B) and soil compressibility within the depth of influence of the foundation are the main parameters affecting the foundation settlement [15, 29]. Also, the effect of foundation geometry is an important parameter that can be considered by introducing the ratio of the length to width (L/B) into the model development parameters. Burland and Burbidge [8] stated that there are two other parameters (i.e., the thickness of soil layer beneath the foundation and depth of groundwater) that have minor degrees of importance. Due to the lack of sufficient data regarding the thickness of soil layer in the available database, this parameter was not considered in the current study. It was also assumed that the effect of the water table has already been reflected in the measured SPT blow count, as proposed by Meyerhof [4]. In addition, the length-to-width ratio (L/B) of circular footings is considered to be equal to unity. It should be noted that more information about field condition parameters reflecting ground and testing conditions can surely help researchers to propose more reliable models in future.

The database used for the EPR model development includes 189 individual case histories; 5 cases were reported by Bazaraa [5], 22 cases by Burbidge [30], 125 cases by Burland and Burbidge [8], one case by Picornell and Del Monte [31], 30 cases by Wahls [13], 2 cases by Maugeri et al. [32] and 4 cases by Briaud and Gibbens [33]. The database covers a wide range of cohesionless soil and foundation properties.

3.2. Data division

As a common practice, the data were divided into two subsets; training (i.e., calibration) and testing (i.e., validation). The training subset is mainly used to generate the model, and the testing subset is used to examine the performance of the constructed model. In this study, 152 cases (80%) were used for the training process and the other 37 cases (20%) are used to confirm the validity of the trained EPR-based model.

Studies by Shahin et al. [34] and Tokar and Johnson [35] revealed that the way the data are divided has a noteworthy impact on the obtained results. In this study, the statistically consistent method described in detail by Shahin et al. [34] was used. The data were divided in such a way that the main statistical parameters of the training and testing subsets (i.e., mean, maximum, minimum, and standard deviation) became as close to each other as possible and accordingly represent similar statistical populations. The statistical properties of the training and testing subsets are summarized in Table 1.

Table 1 Statistical characteristics of training and testing subsets

Parameter	Dimension	Subset	Statistical characteristics			
			Mean	Standard Deviation	Maximum	Minimum
B	m	Training	8.6	10.2	60.0	0.8
		Testing	9.4	10.1	41.2	0.9
q_{net}	kPa	Training	186.9	125.6	697.0	18.3
		Testing	188.0	114.6	575.0	33.0
SPT-N	-	Training	24.6	13.4	60.0	4.0
		Testing	24.3	14.2	55.0	4.0
L/B	-	Training	2.2	1.8	10.6	1.0
		Testing	2.1	1.8	8.1	1.0
D_f/B	-	Training	0.5	0.6	3.4	0.0
		Testing	0.6	0.6	3.0	0.0
$S_{measured}$	mm	Training	20.5	27.0	121.0	0.6
		Testing	20.4	25.2	120.0	1.3

3.3. Initial settings and model optimization

In this study, the EPR toolbox version 2.SA was used to model the shallow foundation settlement in granular soils. Giustolisi and Savic [23] developed this software based on the homonymous modeling methodology using a hybrid evolutionary paradigm. In the development process of an EPR-based model, some initial settings and a set of constraints can be devised to optimize the constructed model by considering some properties such as equation length, set of allowable functions, number of terms, set of allowable exponents, and number of generations. In this study, based on the results of different analyses, the maximum number of generations and number of terms were set equal to 14 and 10, respectively. For simplicity with respect to the evolved models, the set of allowable exponents was ±4, ±3.5, ±3, …, ±0.5 and 0. Other initial parameters were set to their recommended default values in the EPR toolbox. Complete descriptions about the effective initial parameters can be found in Giustolisi and Savic [23, 27].

After model development, statistical analyses were performed to measure the model performance. The mathematical definitions of the statistical criteria used include the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) (see Table 2).

Table 2 Statistical criteria used for evaluation of models

Performance index	Mathematical Definition
Coefficient of determination (R^2)	$R^2 = 1 - \dfrac{\sum_{i=1}^{n}(y_m - y_p)^2}{(y_m - \overline{y_m})^2}$
Root mean square error (RMSE)	$\mathrm{RMSE} = \sqrt{\dfrac{\sum_{i=1}^{n}(y_m - y_p)^2}{n}}$
Mean absolute error (MAE)	$\mathrm{MAE} = \dfrac{\sum_{i=1}^{n}\lvert y_m - y_p \rvert}{n}$

3.4. EPR-based settlement model

To obtain simple and straightforward formulae, numerous attempts with various initial settings were executed and performance of the statistical analyses was evaluated. The best model was selected according to the best statistical properties and model simplicity. Equation (3) is the optimum obtained EPR-based formula after simplification.

$$S = \frac{7.2q}{N^2} - \frac{190}{NL} + 3.4\frac{B\sqrt{q} - 10\sqrt{BD_f}}{N} - 0.3\sqrt{Bq} + 16 \tag{3}$$

Where S is the predicted settlement (mm), B is the foundation width (m), L is the foundation length (m), D_f is the foundation embedment depth (m), q is the foundation net applied pressure (kPa) and N is the average SPT below count. Equation (3) describes the settlement as a function of the most effective factors obtained from the main properties of the soil-foundation system. It should be noted that the proposed formula is only valid for the ranges shown in Table 1.

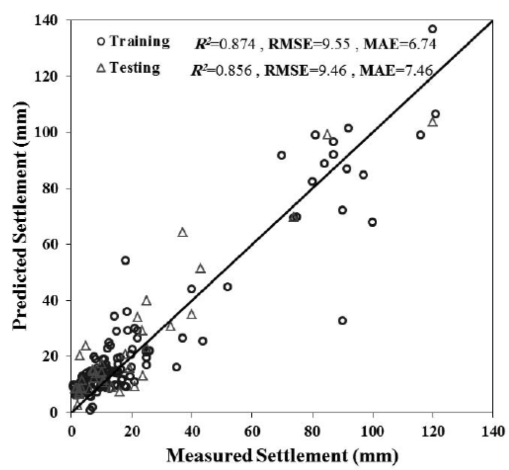

Fig. 2 Comparison of measured versus predicted settlements obtained from the EPR-based model

Figure 2 shows the measured versus predicted values obtained from Equation (3), and values of R^2, RMSE and MAE are also illustrated in the figure. It can be seen that the relatively high performance and accuracy of the EPR-based model in predicting the case histories of the measured settlements is clearly illustrated.

4. Development of GP- and GEP-Based Settlement Models

4.1. Overview on genetic programming

The main advantage of GP-based methods over simple regression and other soft computing techniques is their ability to generate predictive formulae without making any assumption about the basic form of the existing relationship [36]. The Darwinian natural selection principle is the basis of GP-based algorithms, which was used by Koza [37] to develop the classical GP method in the late 1980s. The classical GP method is also called the tree-based GP [37] or the traditional GP method [36].

In this study, two new GP-based models were developed to predict the foundation settlement of cohesionless soils using classical GP and GEP. GP is sufficiently well known among geotechnical researchers, and several problems have been solved using this technique in recent years (e.g., Baziar et al. [36]; Kermani et al. [38]; Rezania and Javadi [16]; Teodorescu [39]). Full description of the GP technique is beyond the scope of this paper and can be found in many publications (e.g., Koza [37]; Banzhaf et al. [40]).

Gene expression programming (GEP) is a branch of GP that was first invented by Ferreira [41]. In fact, this method is a natural development of genetic algorithms (GAs) and genetic programming (GP). Similar to the classical GP, there are different configuration parameters in GEP that can be used to optimize the target model and minimize the error levels: function set, terminal set, fitness function, control parameters, and termination condition. The function set contains the basic mathematical operators and Boolean logical functions (or any other user-defined function), while the terminal set contains the numerical constants. The fitness function is mainly defined to determine the fitness of each individual; this function is based on the minimization of the error. Control parameters and termination condition apply the necessary specifications to control the process of modeling.

The main difference between the classical GP and GEP lies in the representation of the models. In GEP, the models are created with a fixed length of character strings and are further described as computer solutions in tree-based structures that are named expression trees (ETs) [42]. On the other hand, in the classical GP, the created models are represented in tree-based structures and expressed in a functional programming language [37, 42]. Detailed description of the GEP technique can be found in Ferreira [41, 43], Gandomi et al. [42], Mollahasani et al. [44] and Alavi and Gandomi [45].

4.2. Model inputs and outputs

The input and output parameters of the GP- and GEP-based models were the same as those used to develop the EPR-based model, and the same data division was also used.

4.3. Initial settings and model optimization

In this study, a new GP code named GPTIPS was utilized to develop a predictive model based on GP [46]. In GPTIPS, it is possible to create a set of restrictions to avoid bloating. Bloating is the excessive growth of a model without any considerable improvement in the fitness value. An effective technique called lexicographic tournament selection was used in the evolutionary process of model development by GPTIPS to control the bloating of the model [47]. Table 3 shows the range of initial parameters used in the GP realizations, and other initial parameters were considered to be equal to their recommended default values according to Searson [48]. It should be noted that in this study the fitness function used for the classical GP-based modeling was the root mean square error (RMSE).

Table 3 Range of initially defined parameters in GP

Parameter	Range
Population size	20–10000
Number of generations	100–5000
Maximum number of nodes	10–100
Maximum depth of trees	3–50
Numerical constants lower bound	–10
Numerical constants upper bound	+10
Function set	+ , – , × , ÷

For development of the GEP-based model, GeneXproTools [49] was used. Similar to GP modeling, RMSE was used as the fitness function. Other functions such as the mean absolute error (MAE) and mean square error (MSE) were also tried but the best results were obtained with the RMSE. Table 4 illustrates the range of initial parameters used in GEP modeling. The other parameters of GEP-based modeling were set to their default values according to the GEPSOFT [49].

Table 4 Range of initially defined parameters in GEP

Parameter	Settings
Number of generations	100–100000
Number of chromosomes	10–50
Number of genes	1–5
Number of constants per gene	1–8
Numerical constants lower bound	–20
Numerical constants upper bound	+20
Function set	+ , – , × , ÷ , power , ln(x)

4.4. GP and GEP model formulations

In an attempt to obtain the optimal GP- and GEP-based models, numerous realizations were performed with different initial settings and the best models were selected according to their R^2, RMSE and MAE. Ultimately, the following models were found to be optimal and their predictive abilities are shown in Figures 3 and 4. Equations (4) and (5) represent the developed GP- and GEP-based formulae, respectively.

$$S = \frac{2.5B\left(\frac{N}{B} - 1 + \frac{B+1}{D_f + 0.16B} + \frac{2B-N}{L} + \frac{q}{N}\right)}{N + \frac{D_f}{B}\left(B - \frac{L}{B}\right) + \frac{B}{N}} + 4 \quad (4)$$

$$S = \frac{(q + B^2)}{N + \frac{D_f}{B}\left(3N - 3.7\frac{L}{B}\right)} + 5.6 \quad (5)$$

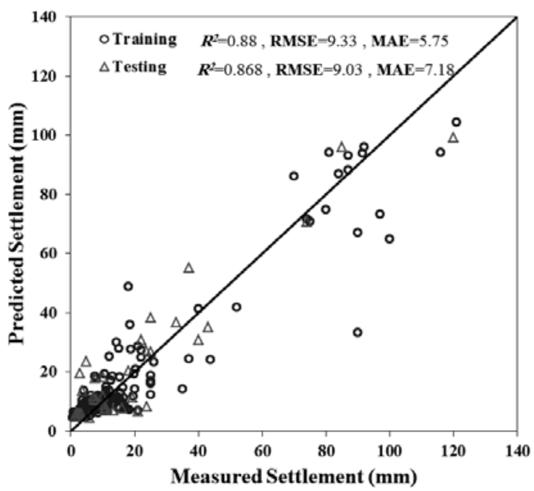

Fig. 3 Comparison of settlements measured and predicted by the GP-based model

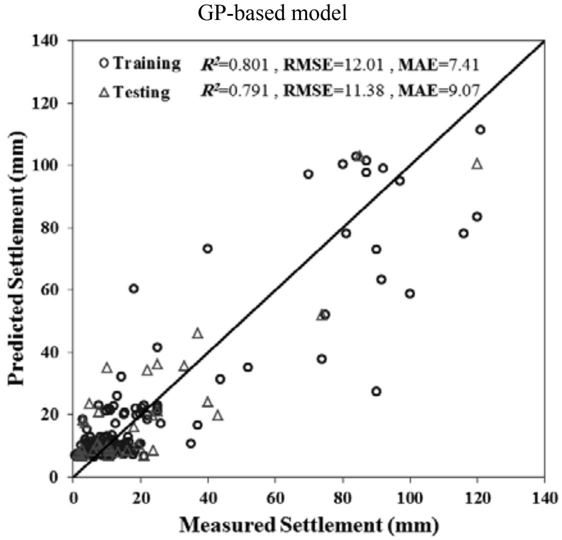

Fig. 4 Comparison of settlements measured and predicted by the GEP-based model

By comparing the formulations in equations (3–5), it can be seen that the GEP model (i.e., equation 5) is simpler than both the EPR model (i.e., equation (3)) and GEP model (i.e., equation 4). However, it can also be seen from Figures 2–4 that the performance of the GEP model is worse with less R^2 and higher RMSE and MAE, and much more scatter around the line of equality between the measured and predicted settlements. It should be noted that the proposed formulae are only valid for the ranges shown in Table 1.

5. Comparison of EPR, GP and GEP Models With Other Available Models

In this section, the developed EPR, GP and GEP settlement prediction models are compared with other available soft computing models, including the ANN model developed by Shahin et al. [15, 29] and the GP model developed by Rezania and Javadi [16]. It should be noted that the comparison between the soft computing settlement prediction models and traditional methods such as those of Meyerhof [4], Schultze and Sherif [6] and Schmertmann et al. [7] was investigated in previous studies by Shahin et al. [15] and Rezania and Javadi [16].

5.1. Statistical analyses

The concurrent consideration of statistical criteria including R^2, RMSE, and MAE data is a reasonable method for comparing the models. Table 5 shows the statistical performance of the models developed in this work and the ANN- and GP-based models developed in previous studies. It can be seen that the EPR- and GP-based models are the most precise models in predicting the settlement of shallow foundations on cohesionless soils. It can also be seen that the accuracy of the ANN model proposed by Shahin et al. [29] and the GP model developed by Rezania and Javadi [16] is high but not as high as the EPR and GP models of the current study. On the other hand, the statistical results show that the performance of the GEP model is considerably lower than that of all other models. Overall, the statistical results indicate that the application of EPR and GP methods provides a more potential improvement over the previously developed models. Considering these findings, it can be concluded that the new EPR- and GP-based models are the most robust models.

Table 5 Statistical performances of the proposed models and the previously published ones

Ref.	Method	R^2	RMSE	MAE
Shahin et al. (2002a)	ANN	0.851	10.25	7.14
Rezania and Javadi (2007)	GP	0.826	11.07	6.77
Current study	EPR	0.871	9.53	6.88
Current study	GP	0.878	9.27	6.03
Current study	GEP	0.799	11.89	7.73

5.2. Parametric study

For further verification of the predictive ability of the models developed in this study and for more comparison with previously developed models by Shahin et al. [29] and Rezania and Javadi [16], a parametric study was conducted to investigate the effect of the model inputs on predicted settlements. The parametric study verifies whether or not the behavior of the predictive models matches that observed in the experimental investigation. Therefore, the impact of varying each input parameter was studied while the other input parameters were maintained constant at their mean values according to the available database. Figures 5-9 illustrate the results of this parametric study in which the sensitivity of the model output (i.e., settlement) to changes in each input parameter is illustrated.

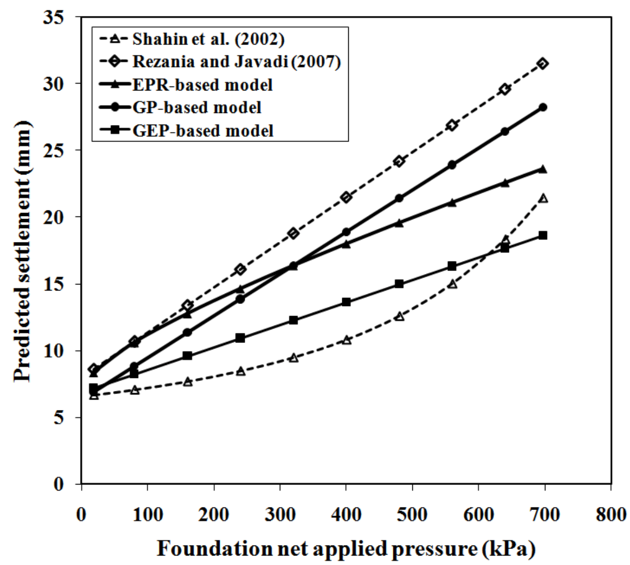

Fig. 6 Results of parametric study: effect of foundation net applied pressure

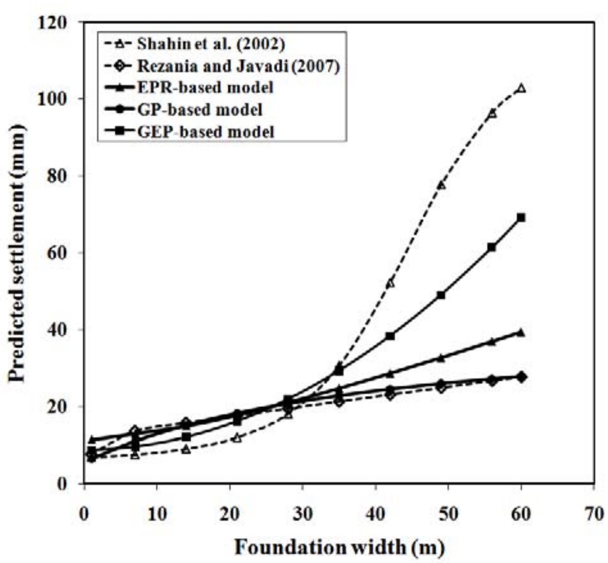

Fig. 5 Results of parametric study: effect of foundation width

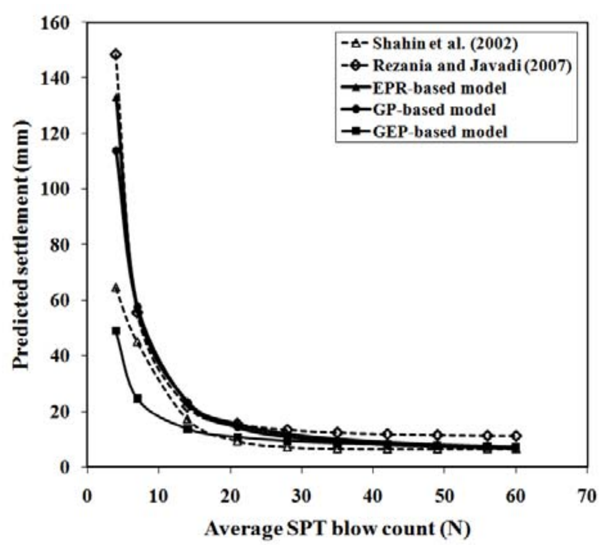

Fig. 7 Results of parametric study: effect of average SPT blow count

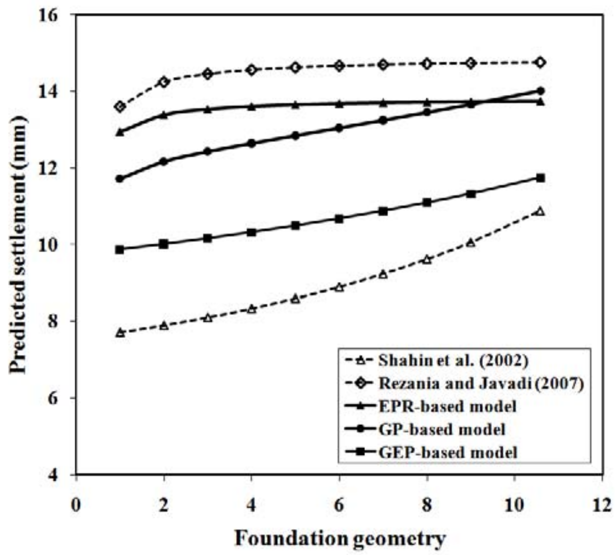

Fig. 8 Results of parametric study: effect of foundation geometry (the ratio of length to width, L/B)

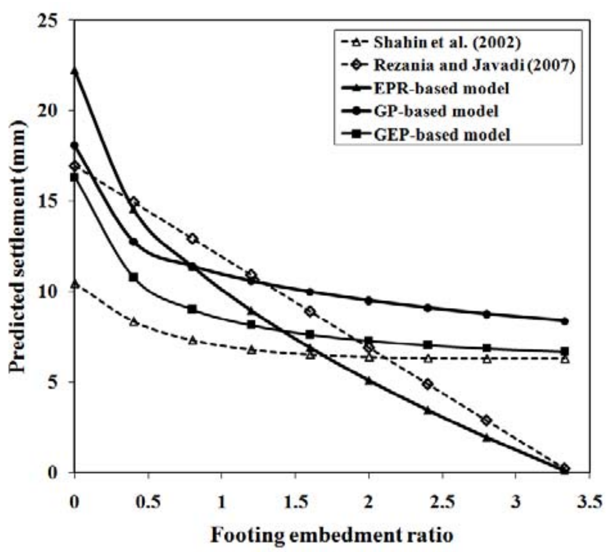

Fig. 9 Results of parametric study: effect of foundation embedment ratio

It can be seen from Figures 5-9 that all models were able to capture the underlying physical meaning of settlement problem and experimental investigation. For example, it was found in all models that the foundation settlement increases with increasing the foundation width, foundation geometry and net applied pressure, as one would expect. On the other hand, the settlement decreases with increasing the foundation embedment ratio and average SPT blow count, also as expected. It should be noted that despite the ability of all models to capture the appropriate trends of settlement behavior, the values of settlement obtained from each model under similar conditions show some differences. For example, the settlement obtained from the ANN-based model developed by Shahin et al. [29] for foundation widths larger than 40 m is considerably different from that of the other models. In addition, the settlement predicted by the GEP-based model for an average SPT-N below 20 is also different from that of the other models. Based on the parametric study, the overall statement that can be made regarding settlement of shallow foundations on granular soils is that all input parameters are effective, but the weights of their effectiveness are different. For example, for soils with SPT-N values smaller than 20, the average SPT blow count is the most effective parameter. Consequently, measuring the actual SPT-N value for loose and relatively medium dense sandy soil is essential. In addition, the foundation width and net applied pressure affect settlement considerably.

6. Summary and Conclusions

In this study, three powerful soft computing approaches including evolutionary polynomial regression (EPR), genetic programming (GP) and gene expression programming (GEP) were employed to assess the complex behavior of shallow foundation settlement on cohesionless soils. The models were trained and tested using a databank of field measurements. In addition, a comprehensive parametric study was performed for further verification of the models. Finally, the newly generated models were compared with the most precise previous ones reported in the technical literature. Based on the results of this study, the following conclusions can be made:

1. The proposed evolutionary-based models using EPR, GP and GEP techniques are able to accurately predict shallow foundation settlement on cohesionless soils. With respect to their precision and simplicity, the developed models can be successfully applied in practical projects.
2. The comprehensive parametric study insured the proper performance of the new models considering the ability of the models to match the underlying physical meaning of geotechnical aspects of settlement problem. The results were also consistent with the findings of previous studies.
3. The average SPT below count of subsoil, foundation width and net applied pressure are the most significant input parameters that affect the foundation settlement.
4. The comparison between the newly developed EPR, GP and GEP models and the current most accurate ANNs and GP models available in the literature showed that the newly developed EPR and GP models outperform the previously developed ANNs and GP models.
5. The feasibility of the EPR, GP and GEP approaches in finding solutions for highly nonlinear problems such as settlement of shallow foundations on granular soils was clearly illustrated in the current study, thus, these techniques can also be used to solve other complex engineering problems. Moreover, it is possible to further improve the predictive ability of the models developed by the proposed techniques by re-training the models with new cases when more data are made available. In the current topic, further information about field condition parameters reflecting ground and loading

conditions (e.g., thickness of soil layer, level of water table, etc.) can surely help researchers to propose more reliable models for more precise prediction of shallow foundations settlements on cohesionless soils.

References

[1] Schmertmann J.H. Static cone to compute static settlement over sand, Journal of the Soil Mechanics and Foundations Division, ASCE, No. 3, Vol. 96 1970, pp. 1011-1043.

[2] Poulos H.G. Settlement of isolated foundations, Proceedings of the General Session of the Symposium Held at The University of New South Wales, Kensington, NSW, July 14-18, 1975, pp. 181-212.

[3] Terzaghi K, Peck R.B, Soil mechanics in engineering practice, John Wiley, New York. 1948.

[4] Meyerhof G.G. Shallow foundations, Journal of the Soil Mechanics and Foundations Division, ASCE, 1965, Vol. 91(SM2), pp. 21–31.

[5] Bazaraa A.R.S.S. Use of the standard penetration test for estimating settlements of shallow foundations on sand, PhD thesis, University of Illinois, Urbana, 1967.

[6] Schultze E, Sherif G. Prediction of settlements from evaluated settlement observations for sand, Proceedings of 8th International Conference on Soil Mechanics and Foundation Engineering, 1973, Vol. 1, pp. 225-230.

[7] Schmertmann J.H, Hartman J.P, Brown P.B. Improved strain influence factor diagrams, Journal of the Geotechnical Engineering Division, 1978, Vol. 104 (GT8), pp. 1131-1135.

[8] Burland J.B, Burbidge M.C. Settlement of foundations on sand and gravel, Proceeding of the Institution of Civil Engineers, London, 1985, 78-Part 1, pp. 1325-1381.

[9] Jorden E.E. Settlement in sand-Methods of calculating and factors affecting, Ground Engineering, 1977, No. 1, Vol. 10, pp. 30-37.

[10] Jeyapalan J.K, Boehm R. Procedures for predicting settlements in sands, Proceedings of Settlement of Shallow Foundations on Cohesionless Soils: Design and Performance, Geotechnical Special Publication, ASCE, Seattle, Washington, 1986, No. 5, pp. 1-22.

[11] Gifford D.G. Wheeler J.R, Kraemer S.R, McKown A.F. Spread footings for highway bridges, Final Report, FHWA/RD-86/185, Federal Highway Administration, Washington, DC, 1987.

[12] Tan C.K, Duncan J.M. Settlement of footings on sands-Accuracy and reliability, Proceedings of Geotechnical Engineering Congress, Boulde, Colorado, Geotechnical Special publication, 1991, No. 27, pp. 446-455.

[13] Wahls H.E. Settlement analysis for shallow foundations on sand, Proceedings of the 3rd International Geotechnical Engineering Conference, Cairo, Egypt, 1997, pp. 7-28.

[14] Sivakugan N, Eckersley J, Li H. Settlement predictions using neural networks, Australian Civil Engineering Transactions, 1998, Vol. CE40, pp. 49-52.

[15] Shahin M.A, Maier H.R, Jaksa M.B. Predicting settlement of shallow foundations using neural networks, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 2002, No. 9, Vol. 128, pp. 785–793.

[16] Rezania M, Javadi A.A. A new genetic programming model for predicting settlement of shallow foundations, Canadian Geotechnical Journal, 2007, No. 12, Vol. 44, pp. 1462-1473.

[17] Samui P, Sitharam T.G. Least-square support vector machine applied to settlement of shallow foundations on cohesionless soils, International Journal for Numerical and Analytical Methods in Geomechanics, 2008, No. 17, Vol. 32, pp. 2033-2043.

[18] Moghaddas Tafreshi S.N, Tavakoli Mehrjardi G. The use of neural network to predict the behavior of small plastic pipes embedded in reinforced sand and surface settlement under repeated-load, Engineering Applications of Artificial Intelligence, 2008, Vol. 21 No. 6, pp. 883-894.

[19] Reda Taha M.M, Noureldin A, El-Sheimy N, Shrive N.G. Artificial neural networks to predict creep with an example application to structural masonry, Canadian Journal of Civil Engineering, 2003, Vol. 30 No. 3, pp. 523-532.

[20] Abed M.M, El-Shafie A, Osman S.A.B. Creep predicting model in masonry structure utilizing dynamic neural network, Journal of Computer Science, 2010, No. 5, Vol. 6, pp. 597-605.

[21] Shahnazari H, Tutunchian M.A, Rezvani R, Valizadeh F. Evolutionary-based approaches for determining the deviatoric stress of calcareous sands, Computers and Geosciences, 2013, Vol. 50, pp. 84-94.

[22] Shahnazari H, Tutunchian M.A, Mashayekhi M, Amini A.A. Application of soft computing for prediction of pavement condition index, Journal of Transportation Engineering, ASCE, 2012, Vol. 138 No. 12, pp. 1495-1506.

[23] Giustolisi O, Savic D.A. A symbolic data-driven technique based on evolutionary polynomial regression, Journal of Hydroinformatics, 2006, No. 3, Vol. 8 pp. 207-222.

[24] Rezania M, Javadi A.A, Giustolisi O. Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression, Computers and Geotechnics, 2010, Nos. 1-2, Vol. 37, pp. 82–92.

[25] Rezania M, Faramarzi A, Javadi A.A. An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement, Engineering Applications of Artificial Intelligence, 2011, No. 1, Vol. 24, pp. 142–153.

[26] Ghani N.A.A, Shahin M.A, Nikraz H.R. Use of evolutionary polynomial regression (EPR) for prediction of total sediment load of Malaysian rivers, International Journal of Engineering, 2012, No. 5, Vol. 6, pp. 262–277.

[27] Giustolisi O, Savic D.A. A novel strategy to perform genetic programming: Evolutionary polynomial regression, 6th International Conference on Hydroinformatics, Singapore, Liong, Phoon & Babovic (eds), World Scientific Publishing Company, 2004, Vol. 1, pp. 787-794.

[28] Rezania M, Javadi A.A, Giustolisi O. An evolutionary-based data mining technique for assessment of civil engineering systems, Journal of Engineering Computations, 2008, No. 6, Vol. 25, pp. 500–517.

[29] Shahin M.A, Jaksa M.B, Maier H.R. Artificial neural network-based settlement prediction formula for shallow foundations on granular soils, Australian Geomechanics, 2002, No. 4, Vol. 37, pp. 45-52.

[30] Burbidge M.C. A case study review of settlements on granular soil, MSc. thesis, Imperial College of Science

and Technology, University of London, London, 1982
[31] Picornell M, Del Monte E. Prediction of settlement of cohesive granular soils, Proceedings of Measured Performance of Shallow Foundations, Nashville, Tennessee, Edited by M. Picornell, Geotechnical Special Publication, 1988, No. 15, ASCE, New York, pp. 55–72.
[32] Maugeri M, Castelli F, Massimino M.R, Verona G. Observed and computed settlements of two shallow foundations on sand, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 1998, No. 7, Vol. 124, pp. 595-605.
[33] Briaud J.L, Gibbens R. Behavior of five large spread footings in sand, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 1999, No.9, Vol. 125, pp. 787–796.
[34] Shahin M.A, Maier H.B, Jaksa M.B. Data division for developing neural networks applied to geotechnical engineering, Journal of Computing in Civil Engineering, ASCE, 2004, No. 2, Vol. 18, pp. 105-114.
[35] Tokar S.A, Johnson P.A. Rainfall-runoff modeling using artificial neural networks, Journal of Hydrologic Engineering, ASCE, 1999, No. 3, Vol. 4, pp. 232–239.
[36] Baziar M.H, Jafarian Y, Shanazari H, Movahed V, Amin Tutunchian M. Prediction of strain energy-based liquefaction resistance of sand–silt mixtures: an evolutionary approach, Computers and Geosciences, 2011, No. 11, Vol. 37, pp. 1883-1893.
[37] Koza J. Genetic programming: on the programming of computers by means of natural selection, Cambridge (MA): MIT Press, 1992, 813 p.
[38] Kermani E, Jafarian Y, Baziar M.H. New predictive models for the v_{max}/a_{max} ratio of strong ground motions using genetic programming, International Journal of Civil Engineering, 2009, No. 4, Vol. 7, pp. 236-247.
[39] Teodorescu L. Gene Expression Programming Approach to event selection in High Energy Physics, IEEE Transaction on Nuclear Science, 2006, Vol. 53, pp. 2221-2227.
[40] Banzhaf W, Nordin P, Keller R.E, Francone F.D. Genetic programming: An introduction: On the automatic evolution of computer programs and its application, Morgan Kaufmann Publishers, Inc, San Francisco, CA, USA, 1998.
[41] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems, Complex Systems, 2001, No. 2, Vol. 13, pp. 87–129.
[42] Gandomi A.H, Alavi A.H, Mirzahosseini R, Moghadas Nejad F. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures, Journal of Materials in Civil Engineering, 2011, No.3, Vol. 23, pp. 248-263.
[43] Ferreira C. Gene expression programming: Mathematical modeling by an artificial intelligence, Second Edition, Springer-Verlag, Germany, 2006.
[44] Mollahasani A, Alavi A.H, Gandomi A.H. Empirical modeling of plate load test moduli of soil via gene expression programming, Computers and Geotechnics, 2011, No. 2, Vol. 38, pp. 281-286.
[45] Alavi A.H, Gandomi A.H. A robust data mining approach for formulation of geotechnical engineering systems, Engineering Computations, 2011, Vol. 28, No. 3, pp. 242-274.
[46] Searson D.P. GPTIPS: Genetic Programming & Symbolic Regression for MATLAB, available at: http://gptips.sourceforge.net, 2009 (accessed 11 February 2010).
[47] Luke S, Panait L. Lexicographic parsimony pressure, Proceedings of the Genetic and Evolutionary Computation Conference, London, Morgan Kaufmann, San Francisco, CA, 2002, pp. 829–836.
[48] Searson D.P. GPTIPS: Genetic Programming & Symbolic Regression for MATLAB, User Guide1.0, available at: http://gptips.sourceforge.net, 2009 (accessed 11 February 2010).
[49] GEPSOFT. GeneXproTools. Version 4.0, available at: http://www.gepsoft.com/, 2006.