Computers & Geosciences 37 (2011) 1860–1869

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression

A. Ahangar-Asr, A. Faramarzi, N. Mottaghifard, A.A. Javadi *

Computational Geomechanics Group, College of Engineering, Computing and Physical Sciences, University of Exeter, Exeter, UK

ARTICLE INFO

Article history:
Received 14 January 2011
Received in revised form
15 April 2011
Accepted 20 April 2011
Available online 13 May 2011

Keywords:
Optimum moisture content
Maximum dry density
Permeability
Evolutionary computing
Data mining

ABSTRACT

This paper presents a new approach, based on evolutionary polynomial regression (EPR), for prediction of permeability (*K*), maximum dry density (MDD), and optimum moisture content (OMC) as functions of some physical properties of soil. EPR is a data-driven method based on evolutionary computing aimed to search for polynomial structures representing a system. In this technique, a combination of the genetic algorithm (GA) and the least-squares method is used to find feasible structures and the appropriate parameters of those structures. EPR models are developed based on results from a series of classification, compaction, and permeability tests from the literature. The tests included standard Proctor tests, constant head permeability tests, and falling head permeability tests conducted on soils made of four components, bentonite, limestone dust, sand, and gravel, mixed in different proportions. The results of the EPR model predictions are compared with those of a neural network model, a correlation equation from the literature, and the experimental data. Comparison of the results shows that the proposed models are highly accurate and robust in predicting permeability and compaction characteristics of soils. Results from sensitivity analysis indicate that the models trained from experimental data have been able to capture many physical relationships between soil parameters. The proposed models are also able to represent the degree to which individual contributing parameters affect the maximum dry density, optimum moisture content, and permeability.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In the construction of many civil engineering structures such as road embankments, loose soil must be compacted to a desired density and water content. In other projects such as earth dams and compacted soil liners for containing contaminated solid and liquid wastes, the soil should be compacted for the density as well as the permeability requirements. The permeability of compacted soils very much depends on the compaction conditions. The required compaction is usually expressed in terms of degree of compaction (dry density) and water content of the soil. To achieve the required degree of compaction, the water content must be near its optimum value. Thus, both the maximum dry density and the optimum water content are essential parameters for design of compacted earthwork. Furthermore, for soil lining construction, the permeability of a compacted soil liner must be very low. Since permeability, maximum dry density, and optimum water content are normally determined from time-consuming laboratory tests, it is desirable to have prediction models capable of predicting compacted soil characteristics based on some easily measurable physical properties of soil.

Many research works have been conducted to relate permeability and compaction characteristics of soils to their physical properties. The physical properties used generally include plasticity characteristics (liquid limit, plastic limit, shrinkage limit, and plasticity index), specific gravity, and grain size distribution that are easily attainable from relatively straightforward laboratory tests. However, specific index properties used in various correlation equations differ considerably. Rowan and Graham (1948) used gradation, specific gravity, and shrinkage limit in their correlation equations. Davidson and Gardiner (1949) eliminated the specific gravity from the equations of Rowan and Graham (1948), but included a plasticity index. Turnbull (1948) related the optimum moisture content with gradation, while Jumikis (1946) correlated the optimum moisture content with liquid limit and plasticity index.

Ring et al. (1962) developed two sets of correlation equations, one for optimum moisture content and the other for maximum dry density. The physical properties used were liquid limit, plastic limit, plasticity index, D_{50}, content of particles finer than 0.001 mm, and fineness average (FA). The fineness average was determined as one-sixth of the summation of the percentages of soil mass finer than No. 10, No. 40, and No. 200 sieves. Liquid limit alone was correlated with both maximum dry density and

* Corresponding author.
E-mail addresses: aa375@ex.ac.uk (A. Ahangar-Asr), af269@ex.ac.uk (A. Faramarzi), n.mottaghifard@gmail.com (N. Mottaghifard), a.a.javadi@ex.ac.uk (A.A. Javadi).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.04.015

optimum moisture content by Ramiah et al. (1970) and Blotz et al. (1998). Linveh and Ishai (1978) also developed some relationships using specific gravity and liquid limit as input. Gupta and Larson (1979) presented a model for predicting packing density of soils from grain size distribution.

The permeability of a soil varies with many factors, such as soil density, water content, degree of saturation, void ratio, and soil structure. Available correlations between these factors and permeability include those of Carman (1937), Burmister (1954), Lambe (1951), Michaels and Lin (1954), Olson (1963), Mitchell et al. (1965), and Garcia-Bengochea et al. (1979). Various relationships between permeability and grain size distribution of soils have been reported. Hazen (1911) suggested that, for filter sands having relatively uniform particles, the permeability is directly proportional to the square of the effective grain size, D_{10}. Zunker (1930) developed a theoretical linear relationship, in full logarithmic scales, between the grain size and the permeability for spherical particles of uniform size. Taylor (1948) formulated a theoretical equation, based on the capillary tube model, for flow through porous media, relating the permeability with a representative grain size. Considering the effect of grain size on permeability, Burmister (1954) recommended that the type of grading (namely, the shape of gradation curve), range of grain size, and fineness (namely, D_{10}) must be taken into account. For a given type of gradation and grain size range, he found that the permeability can be better related with D_{50} than with D_{10}. Horn (1971) related the permeability with the mean grain size on the basis of Zunker's work in 1930. Chen et al. (1977) found that the permeability is strongly related with D_{50}, and Hauser (1978) related the permeability with the aggregate size. Based on the previous research works it can be concluded that the permeability is strongly dependent on grain size distribution. However, a general correlation equation between permeability and gradation applicable to a wide range of soils is not yet available. To develop such a relationship, the entire spectrum of grain size distribution must be considered. More importantly, the density or void ratio of the soil mass should also be considered.

Taking into account a much broader range of influencing factors, Wang and Huang (1984) developed regression equations for predicting maximum dry density, optimum water content, and permeability for two levels of compaction degree (90% and 95%). Najjar et al. (1996) used neuronets to determine the optimum moisture content and maximum dry density of soils. Sinha and Wang (2008) proposed models based on the artificial neural network (ANN) to predict permeability, maximum dry density, and optimum moisture content. Although neural networks are very effective in capturing and representing the behavior of engineering systems, they are also known to suffer from a number of shortcomings. One of the drawbacks of a neural network is that the optimum structure of ANN (e.g., number of hidden layers, number of neurons, and transfer functions) should be identified a priori. This is usually done through a trial and error procedure. The other major shortcoming is related to the black box nature of an ANN model and the fact that the relationship between input and output parameters of the system is described in terms of a weight matrix and biases that are not easily accessible to users understanding. In fact the black box nature and lack of interpretability have prevented ANNs from achieving their full potential in engineering applications.

This paper introduces a new approach to predict permeability (K), maximum dry density (MDD), and optimum moisture content (OMC) of soils using EPR. EPR is a new data mining technique that overcomes the shortcomings of ANNs by providing a structured and transparent model representing the behavior of the system. EPR models are developed to relate K, MDD, and OMC to physical properties of the soils. The results of EPR model predictions are compared with those of a neural network model, a correlation equation from the literature, and the experimental data. A parametric study is conducted to assess the level of contribution of each parameter in the developed models.

2. Evolutionary polynomial regression

EPR integrates numerical and symbolic regressions to perform evolutionary polynomial regression. The strategy uses polynomial structures to take advantage of their favorable mathematical properties. The key idea behind the EPR is to use evolutionary search for exponents of polynomial expressions by means of a genetic algorithm (GA) engine. This allows (i) easy computational implementation of the algorithm, (ii) efficient search for an explicit expression, and (iii) improved control of the complexity of the expression generated (Giustolisi and Savic, 2006). EPR is a data-driven method based on evolutionary computing, aimed to search for polynomial structures representing a system. A physical system, having an output y, dependent on a set of inputs $\mathbf{X}$ and parameters $\boldsymbol{\theta}$, can be mathematically formulated as

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

where F is a function in an m-dimensional space and m is the number of inputs. To avoid the problem of mathematical expressions growing rapidly in length with time in EPR the evolutionary procedure is conducted in the way that it searches for the

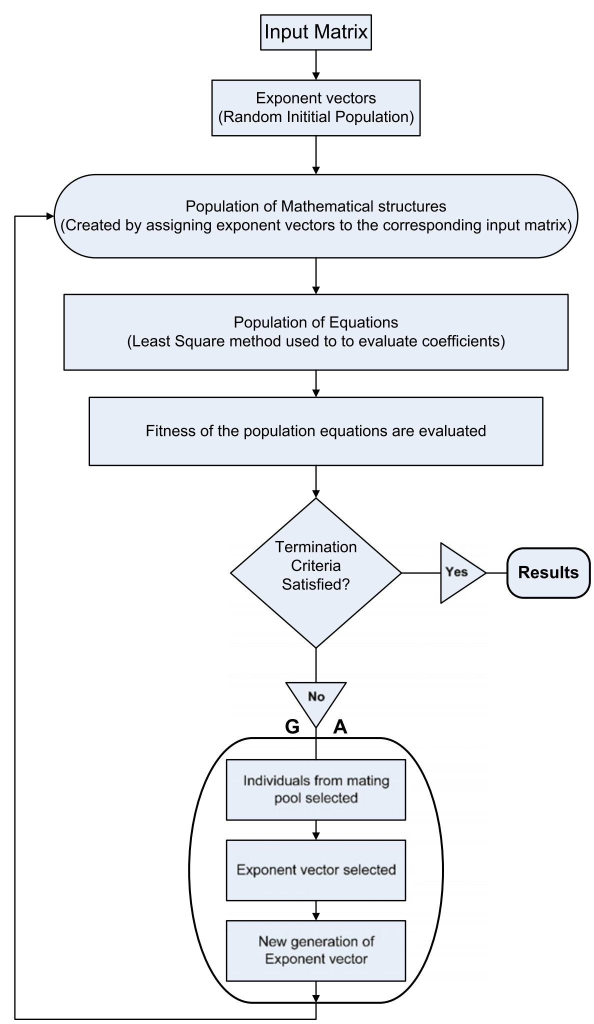


Fig. 1. Typical flow diagram for EPR procedure.

exponents of a polynomial function with a fixed maximum number of terms. During one execution it returns a number of expressions with increasing numbers of terms up to a limit set by the user to allow the optimum number of terms to be selected. The general form of expression used in EPR can be presented as (Giustolisi and Savic, 2006)

$$y = \sum_{j=1}^{m} F(\mathbf{X}, f(\mathbf{X}, a_j) + a_0 \tag{2}$$

where y is the estimated vector of output of the process; a_j is a constant; F is a function constructed by the process; $\mathbf{X}$ is the matrix of input variables; f is a function defined by the user; and m is the number of terms of the target expression. The first step in identification of the model structure is to transfer Eq. (2) into the vector form

$$\mathbf{Y}_{N\times1}(\boldsymbol{\theta},\mathbf{Z}) = [\,\mathbf{I}_{N\times1} \quad \mathbf{Z}^{j}_{N\times m}][\,a_0 \quad a_1 \quad \ldots \quad a_m]^{T} = \mathbf{Z}_{N\times d}\boldsymbol{\theta}^{T}_{d\times1} \tag{3}$$

where $\mathbf{Y}_{N\times1}(\boldsymbol{\theta},\mathbf{Z})$ is the least-squares estimate vector of the N target values; $\boldsymbol{\theta}_{1\times d}$ is the vector of $d=m+1$ parameters a_j and a_0 ($\boldsymbol{\theta}^T$ is the

Table 1
Actual gradation and physical properties of test soils.

Soil no.	Nominal gradation (%)				Actual gradation (%)				Specific gravity	Atterberg limits		Grain size		
	Clay	Silt	Sand	Gravel	Clay	Silt	Sand	Gravel		LL (%)	PL (%)	< No. 4	< No. 40	< No. 200
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	100	0	0	0	84	16	0	0	2.76	495	46	100	100	100
2	80	20	0	0	71	28	1	0	2.76	444	36	100	100	99
3	60	40	0	0	57	41	2	0	2.76	351	36	100	100	98
4	40	60	0	0	44	53	3	0	2.75	203	38	100	99	97
5	20	80	0	0	30	70	0	0	2.87	84	31	100	100	98
6	0	100	0	0	17	83	0	0	2.75	24	22	100	100	97
7	0	60	20	0	13	63	24	0	2.73	0	0	100	88	76
8	0	60	40	0	10	47	43	0	2.72	0	0	100	77	57
9	0	40	60	0	6	31	63	0	2.7	0	0	100	66	37
10	0	20	80	0	3	16	81	0	2.68	0	0	100	55	19
11	0	0	100	0	0	0	100	0	2.67	0	0	100	44	0
12	20	0	80	0	16	3	81	0	2.69	136	25	100	55	19
13	20	20	60	0	20	18	62	0	2.7	132	23	100	66	38
14	20	40	40	0	24	34	42	0	2.73	94	26	100	70	58
15	20	60	20	0	28	50	22	0	2.74	81	27	100	88	78
16	40	40	20	0	40	38	22	0	2.74	222	38	100	88	78
17	40	20	40	0	37	22	41	0	2.72	240	35	100	78	59
18	40	0	60	0	33	6	61	0	2.71	277	29	100	66	39
19	60	0	40	0	50	10	40	0	2.73	389	32	100	78	60
20	60	20	20	0	54	25	21	0	2.74	362	42	100	89	79
21	80	0	20	0	67	18	20	0	2.76	467	39	100	89	80
22	0	0	90	10	0	0	90	10	2.67	0	0	90	40	0
23	0	20	70	10	3	16	71	10	2.72	0	0	90	51	10
24	0	40	50	10	7	32	51	10	2.76	0	0	90	62	39
25	0	60	30	10	10	48	32	10	2.81	0	0	90	74	58
26	0	80	10	10	14	64	12	10	2.86	0	0	90	85	78
27	0	90	0	10	16	72	2	10	2.88	0	0	90	90	88
28	10	10	70	10	10	9	71	10	2.69	75	15	90	51	19
29	10	30	50	10	23	26	41	10	2.75	60	10	90	72	49
30	10	50	30	10	17	42	31	10	2.72	50	10	90	74	59
31	10	70	10	10	21	57	12	10	2.84	45	20	90	85	78
32	10	80	0	10	22	62	2	10	2.86	50	25	90	90	88
33	30	10	50	10	27	12	51	10	2.71	210	30	90	62	39
34	30	30	30	10	30	28	32	10	2.73	175	35	90	73	59
35	30	50	10	10	34	44	12	10	2.74	165	40	90	84	78
36	30	60	0	10	35.6	53	1.8	9.6	2.84	162	47	90	90	89
37	50	10	30	10	44	16	30	10	2.74	342	32	90	74	60
38	50	30	10	10	47	32	11	10	2.74	330	40	90	85	79
39	50	40	0	10	49.2	40	1.2	9.6	2.75	322	37	90	90	89
40	70	10	10	10	61	19	10	10	2.74	445	40	90	85	80
41	70	20	0	10	63	26	1	10	2.75	435	40	90	90	89
42	90	0	0	10	76	14	0	10	2.75	495	46	81	90	90
43	0	0	80	20	0	0	81	19	2.67	0	0	81	36	0
44	0	20	60	20	4	15	62	19	2.69	0	0	81	47	19
45	0	40	40	20	7.2	31.2	42.4	19	2.71	0	0	81	58	38
46	0	60	20	20	11	47	23	19	2.72	0	0	81	69	58
47	0	80	0	20	14	64	2	20	2.86	24	22	81	81	78
48	20	0	60	20	17	3	61	19	2.7	170	30	81	47	20
49	20	20	40	20	21	19	41	19	2.71	140	25	81	58	39
50	20	40	20	20	24	34	22	19	2.73	110	25	81	69	58
51	20	60	0	20	28	50	3	19	2.74	110	40	81	80	78
52	40	0	40	20	36	6	39	19	2.72	340	32	81	60	42
53	40	20	20	20	38	22	21	19	2.72	300	40	81	69	60
54	40	40	0	20	41	38	2	19	2.75	285	40	81	80	79
55	60	0	20	20	51	10	20	19	2.74	455	40	81	70	61
56	60	20	0	20	55	25	1	19	2.75	425	40	81	81	80
57	80	0	0	20	68	13	0	19	2.75	495	46	81	81	81

transposed vector); and $\mathbf{Z}_{N\times d}$ is a matrix formed by $\mathbf{I}$ (unitary vector) for bias a_0, and m vectors of variables $\mathbf{Z}^j$. For a fixed j, the variables $\mathbf{Z}^j$ are a product of the independent predictor vectors of inputs, $\mathbf{X} = <\mathbf{X}_{1\times 2}\ldots\ \mathbf{X}_k>$.

In general, EPR is a two-stage technique for constructing symbolic models. Initially, using a standard genetic algorithm, it searches for the best form of the function structure, i.e., a combination of vectors of independent inputs, $\mathbf{X}_{s=1:k}$, and secondly it performs a least-squares regression to find the adjustable parameters, $\boldsymbol{\theta}$, for each combination of inputs. In this way a global search algorithm is implemented for both the best set of input combinations and related exponents simultaneously, according to the user-defined cost function (Giustolisi and Savic, 2006). The adjustable parameters, a_j, are evaluated by means of the linear least-squares (LS) method based on minimization of the sum of squared errors (SSE) as the cost function. The SSE function, which is used to guide the search process toward the best fit model, is

$$\mathrm{SSE} = \frac{\sum_{i=1}^{N} (y_a - y_p)^2}{N} \tag{4}$$

where y_a is the target value in the training dataset and y_p is the model prediction. The global search for the best form of the EPR

Table 2
Compaction test data and gradations of test soils.

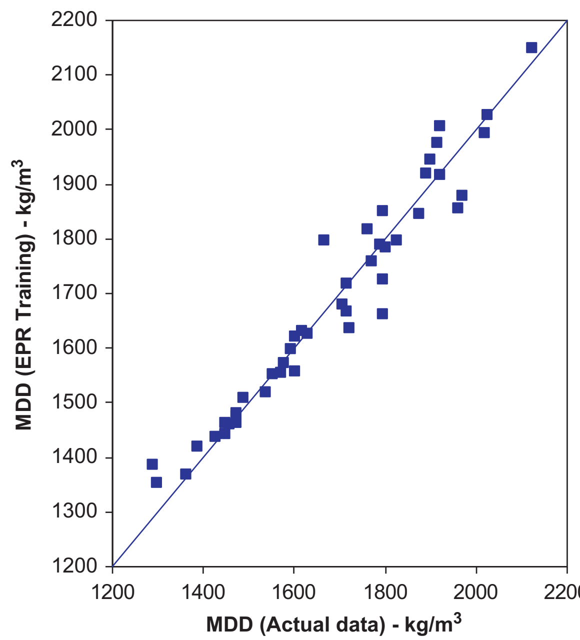
Soil no. (1)	W opt (%) (2)	γ_{dmax}(kg/m^3) (3)	D_{50} (10^{-4} mm) (4)	D_{10} (10^{-5} mm) (5)	Fineness modulus, Fm (6)	Uniformity coefficient (U) (7)	$F_{0.001}$ (%) (8)
1	28	1297	3.5	5.6	0.37	10.36	71
2	31	1289	6	5	0.621	22	58
3	29	1362	14	10	0.472	25	45
4	28	1450	29	13	1.123	37.69	33
5	28	1458	42	40	1.282	14	19
6	26	1490	54	140	1.52	4.79	5
7	21	1602	100	300	2.024	0.47	6
8	16	1714	1500	200	2.48	80	4
9	11	1762	3500	160	2.857	225	2
10	10.5	1898	3600	630	3.276	71.43	0
11	14	1826	4500	19,000	3.68	2.95	0
12	13	1874	3700	46	3.043	1043.48	13
13	10	1666	2300	40	2.861	850	15
14	20	1618	140	30	2.175	416.67	17
15	24	1546	70	25	1.741	44	19
16	28	1474	42	15	1.527	53.33	31
17	17	1602	85	12	1.924	708.33	31
18	13	1794	2200	11	2.37	290.91	28
19	15	1554	20	5.6	1.865	1517.86	42
20	27	1450	16	7	1.286	54.29	44
21	30	1386	7	5	1.035	24	57
22	13	1890	5500	18,000	3.949	4.17	0
23	12.5	2058	4100	600	3.521	96.67	0
24	12	1922	2300	270	3.097	140.74	2
25	16	1788	120	190	2.667	63.16	3
26	22	1706	70	160	2.218	5.94	4
27	24	1618	60	150	1.996	5.13	5
28	10	1914	4200	220	3.417	263.64	8
29	13.5	1962	2500	140	2.972	150	18
30	17.5	1794	120	100	2.538	128	10
31	23.5	1629	65	70	2.114	12.68	11
32	27.5	1578	57	60	1.887	12	12
33	16	1770	2300	18	2.745	2000	22
34	20	1722	120	23	2.323	565.22	24
35	25	1602	60	26	1.916	38.46	25
36	29	1525	41	30	1.666	20	25
37	20	1594	50	6	2.083	1666.67	36
38	28	1498	25	8	1.661	75	38
39	32.5	1450	21	11	1.44	38.18	39
40	27	1474	9.5	5.5	1.406	36.36	50
41	30	1426	9	5	1.2	34	52
42	24	1394	4.4	3.8	0.955	21.05	64
43	8.5	2026	6200	18,500	4.218	5.14	0
44	9.5	2122	4600	700	3.796	102.86	2
45	12	2018	3000	280	3.374	160.71	3
46	15	1794	190	190	2.952	78.95	5
47	22.5	1682	70	150	2.482	6.33	6
48	8	1922	4800	40	3.551	1825	14
49	12	1970	2800	35	3.129	1760	16
50	19.5	1738	140	44	2.707	295.45	18
51	26	1618	70	60	2.243	1.83	24
52	11.5	1802	2500	6.8	2.816	6176.47	31
53	20	1714	90	11	2.462	1181.82	31
54	24	1570	40	16	2.03	50	32
55	18	1618	20	6.5	2.212	523.08	43
56	18	1538	15	7.5	1.79	45.33	45
57	30	1474	6	4.4	1.556	28.41	57

equation is performed by means of a standard GA over the values in the user-defined vector of exponents. The GA operates based on Darwinian evolution, which begins with random creation of an initial population of solutions. Each parameter set in the population represents the individual's chromosomes. Each individual is assigned a fitness based on how well it performs in its environment. Through crossover and mutation operations, with the probabilities P_c and P_m, respectively, the next generation is created. Fit individuals are selected for mating, whereas weak individuals die off. The mated parents create a child (offspring) with a chromosome set, which is a mix of parents' chromosomes. In EPR integer GA coding with single point crossover is used to determine the location of the candidate exponents.

The EPR process stops when the termination criterion, which can be the maximum number of generations, the maximum number of terms in the target mathematical expression, or a

Table 3
Permeability test data.

Soil No. (1)	Permeability (10^{-7} cm/s) [k_{90}] (2)	Permeability (10^{-7} cm/s) [k_{95}] (3)	Void ratio (e_{90}) (4)	Void ratio (e_{95}) (5)
1	0.00052	0.00036	1.362	1.238
2	0.0005	0.00025	1.377	1.252
3	0.017	0.014	1.225	1.133
4	0.258	0.75	1.107	0.996
5	1	0.64	1.095	1.072
6	90	45	1.174	0.942
7	11	1.15	0.983	0.793
8	270	44	0.762	0.654
9	120	27	0.717	0.627
10	5000	8400	0.568	0.497
11	5211	48	0.653	0.566
12	0.013	0.003	0.594	0.512
13	0.35	0.18	0.817	0.722
14	1.7	1.3	0.912	0.811
15	0.51	0.5	0.969	0.065
16	0.041	0.029	1.064	13.956
17	0.0031	0.00215	0.904	0.804
18	0.001	0.00081	0.678	0.589
19	0.0038	0.0004	0.951	0.849
20	0.0014	0.00091	1.099	0.989
21	0.0007	0.0006	1.204	1.088
22	15,000	3200	0.49	0.412
23	800	100	0.468	0.39
24	200	60	0.594	0.511
25	75	20	0.747	0.655
26	350	42	0.862	0.764
27	24	7.5	0.977	0.873
28	10	2.5	0.561	0.479
29	550	25	0.566	0.575
30	40	9	0.733	0.595
31	50	36	0.936	0.834
32	19	12	1.013	0.907
33	0.5	0.04	0.7	0.611
34	0.52	0.42	0.761	0.688
35	5.2	1.2	0.9	0.8
36	6.3	1.7	1.068	0.959
37	0.016	0.0048	0.902	0.802
38	0.043	0.034	1.032	0.925
39	0.19	0.056	1.107	0.966
40	0.0034	0.0028	1.065	0.956
41	0.0038	0.0041	1.143	1.03
42	0.0095	0.0023	1.191	1.076
43	17,000	2000	0.39	0.317
44	4000	1700	0.408	0.334
45	2000	540	0.492	0.413
46	35	20	0.676	0.588
47	60	20	0.888	0.789
48	0.115	0.06	0.56	0.478
49	0.32	0.2	0.527	0.447
50	2.3	1.7	0.744	0.653
51	3	1.1	0.881	0.782
52	0.0037	0.0037	0.676	0.588
53	0.013	0.008	0.763	0.67
54	0.42	0.12	0.945	0.843
55	0.00115	0.00081	0.881	0.782
56	0.0175	0.015	0.987	0.882
57	0.00075	0.00058	0.057	0.942



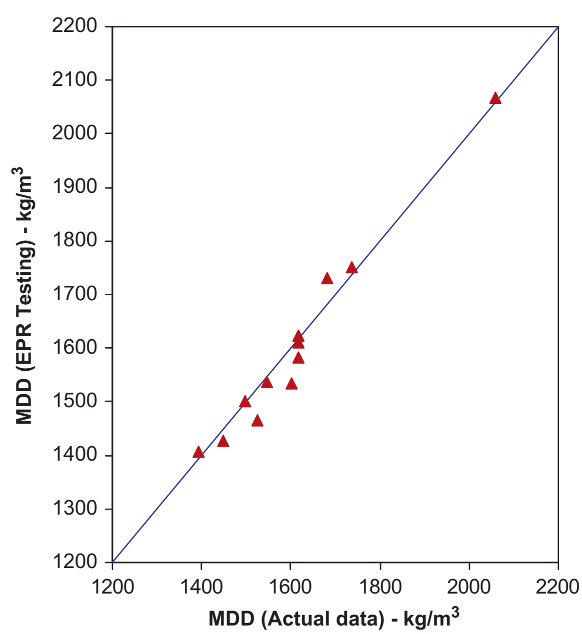


Fig. 2. Comparison between the predicted maximum dry density and the actual values.

Table 4
Coefficient of determination for predicted MDD values.

Model	COD values (%)
Evolutionary polynomial regression (EPR)	96 (for unseen testing data)
Artificial neural network (ANN)	98
Wang and Huang (1984)	95

particular allowable error, is satisfied. A typical flow diagram for the EPR procedure is illustrated in Fig. 1.

3. Database

Some experimental data from the literature (Sinha and Wang, 2008) are used to develop the EPR models. Table 1 includes the gradation properties of soils, Table 2 contains the compaction test data as well as some physical properties, and Table 3 summarizes the permeability test data. Data are from a soil made of four different major components—gravel, sand, limestone dust, and bentonite with different proportions. The bentonite contained Na-montmorillonite as the primary clay mineral. The limestone dust was a by-product of limestone quarry, which had a grain size ranging from 0.002 to 0.047 mm. The sand component was a well-graded fine aggregate used for making Portland cement concrete. Its grain size ranged from 0.074 to 4.76 mm. The gravel component was a coarse aggregate having a particle size range of 4.76–10.05 mm. All tests were conducted following the standard testing procedures stipulated in the ASTM Standard, e.g., ASTM D-422 for mechanical analysis, ASTM D-423 for liquid limit, and ASTM D-424 for plastic limit tests. The laboratory compaction tests were conducted using the standard Proctor compaction effort in accordance with the standard test procedures of ASTM D-558. The details of testing procedure and results of analysis have been presented by Sinha and Wang (2008).

4. EPR procedure

In the evolutionary process of building EPR models, a number of constraints can be implemented to control the output models in terms of the type of functions used, number of terms, range of exponents, number of generations, etc. In this process there is a potential to achieve different models for a particular problem, which enables the user to gain additional information for different scenarios (Rezania et al., 2008). Applying the EPR procedure, the evolutionary process starts from a constant mean of output

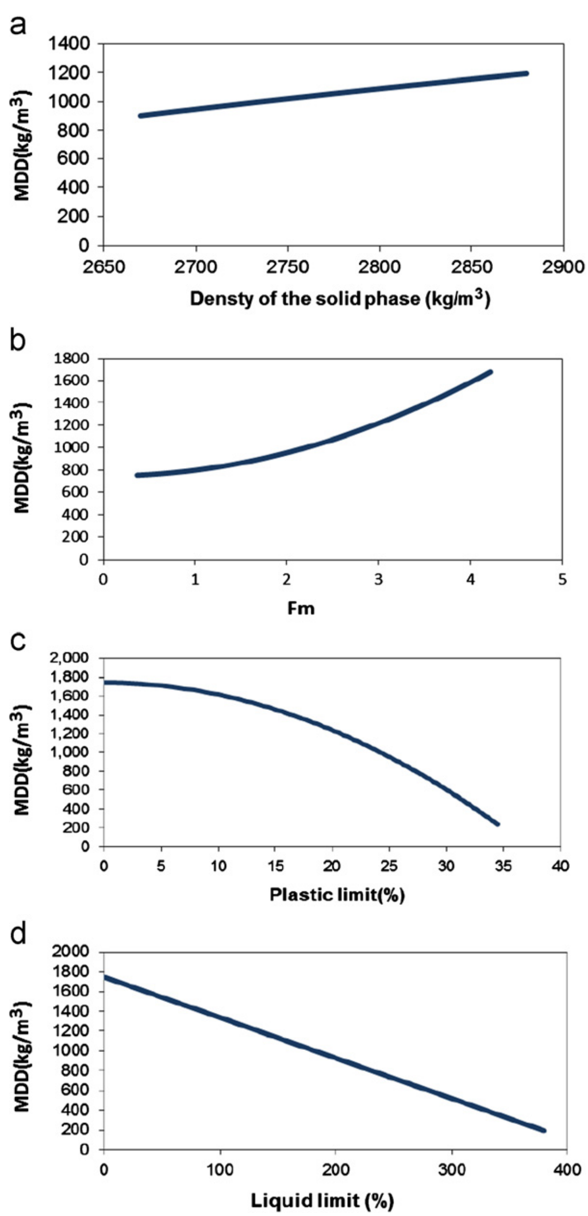


Fig. 3. Parametric study results of the maximum dry density against (a) density of the solid phase of the soil, (b) fineness modulus, (c) plastic limit, and (d) liquid limit.

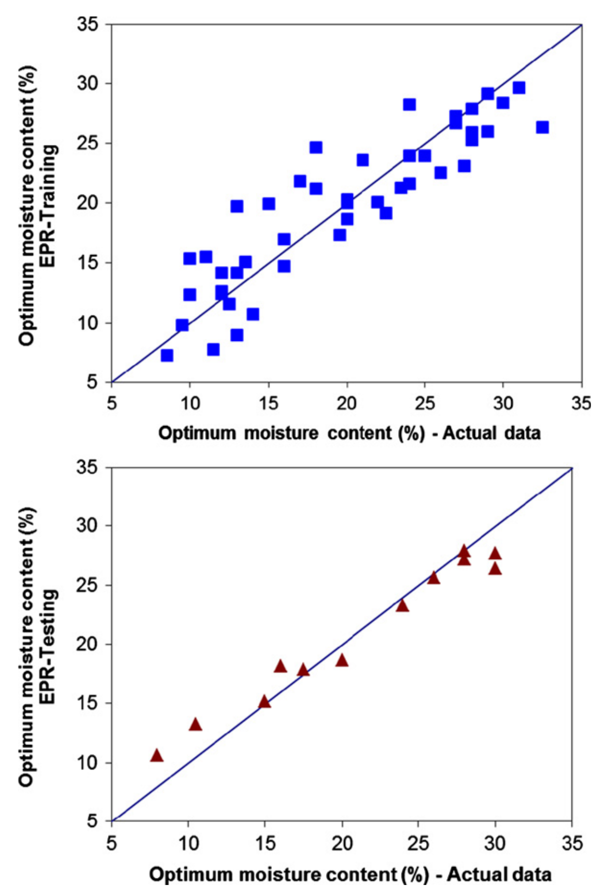


Fig. 4. Comparison between the predicted and the actual optimum moisture content values.

Table 5
Coefficient of determination for predicted OMC values.

Model	COD values (%)
Evolutionary polynomial regression (EPR)	94 (for unseen testing data)
Artificial neural network (ANN)	92
Wang and Huang (1984)	89

values. By increasing the number of evolutions it gradually picks up different participating parameters in order to form equations describing the relationship between the parameters of the system. Each proposed model is trained using the training data and tested using the testing data provided. The level of accuracy at each stage is evaluated based on the coefficient of determination (COD), i.e., the fitness function as

$$COD = 1 - \frac{\sum_N (Y_a - Y_p)^2}{\sum_N \left(Y_a - \frac{1}{N}\sum_N Y_a\right)^2} \tag{5}$$

where Y_a is the actual output value, Y_p is the EPR predicted value, and N is the number of data on which COD is computed. If the model fitness is not acceptable or the other termination criteria (in terms of maximum number of generations and maximum number of terms) are not satisfied, the current model goes through another evolution in order to obtain a new model.

5. Data preparation

The way in which the data are divided into training and validation sets has a significant effect on the results. In this study the dataset was divided into several random combinations of training and validation sets until a robust representation of the whole population was achieved for both training and validation sets. To select the most robust representation, a statistical analysis was performed on the input and output parameters of the randomly selected training and validation sets. The aim of the analysis was to ensure that the statistical properties of the data in each of the subsets were as close to each other as possible and therefore they represented the same statistical population (Rezania et al., 2008). After the analysis, the most statistically consistent combination was used for construction and validation of the EPR models. The parameters used in statistical analysis include the maximum, minimum, mean, and standard deviation. Testing sets of data were chosen in a way that all the parameters were in the range between the maximum and the minimum values in the whole dataset.

6. EPR model for maximum dry density (MDD)

Five input parameters were used for the EPR model for MDD including dry density of solid phase (γ_s) expressed in kg/m^3, fineness modulus (Fm), effective grain size (D_{10}) expressed in mm, plastic limit (PL) expressed in %, and liquid limit (LL) expressed in %. The only output was the maximum dry density. The EPR model developed for maximum dry density is

$$MDD = -\frac{1.38 \times 10^{12}(D_{10})}{\gamma_s^2} + 1.92 \times 10^5 (D_{10}) - 0.661 (D_{10}) PL^2 \times LL + 52.35 Fm^2 + 1385.07 \tag{6}$$

Fig. 2 shows a comparison between the results of the EPR model training and testing and the actual experimental data. Table 4 presents the values of the coefficient of determination (COD) for the models. The table shows that the EPR model performs well and represents a very accurate prediction for unseen cases of data.

The results of the parametric study are shown in Fig. 3. The results show that increasing the density of the soil grains increases the maximum dry density, as it is correctly predicted by the proposed model (Fig. 3a). Increasing the fineness modulus implies a coarser soil and results in increasing maximum dry density, which is also consistent with predictions of the model (Fig. 3b). Also, increasing the liquid limit and plastic limit implies finer soil and results in decrease in the maximum dry density values (Sridharan and Nagaraj, 2005) as shown in Fig. 3c and d.

7. EPR model for optimum moisture content (OMC)

Three input variables were used to develop the EPR model for OMC. These are the fineness modulus (Fm), coefficient of uniformity (U), and plastic limit (PL) expressed in percentage. The EPR model developed to predict the optimum moisture content is

$$OMC = \frac{9.47}{Fm^3 \times U} - \frac{3.57 \times 10^{-5} PL^3}{Fm^2} - \frac{4.55 \times 10^{-3} U}{Fm} + 1.72 \times 10^{-3} PL^2 - 6.36 Fm + 34.09 \tag{7}$$

The optimum moisture content of soils predicted using the EPR model is compared with the experimental data (Fig. 4). The values of coefficient of determination for the models are shown in Table 5. The results indicate excellent performance of the proposed EPR model. The results of the sensitivity analysis are shown in Fig. 5. Fig. 5a shows the variation of OMC with fineness modulus. As the fineness modulus increases (the soil grains get coarser) the specific surface of soil grains decreases, which in turn causes the optimum moisture content to decrease, as correctly captured by the model. A similar trend is reported by previous

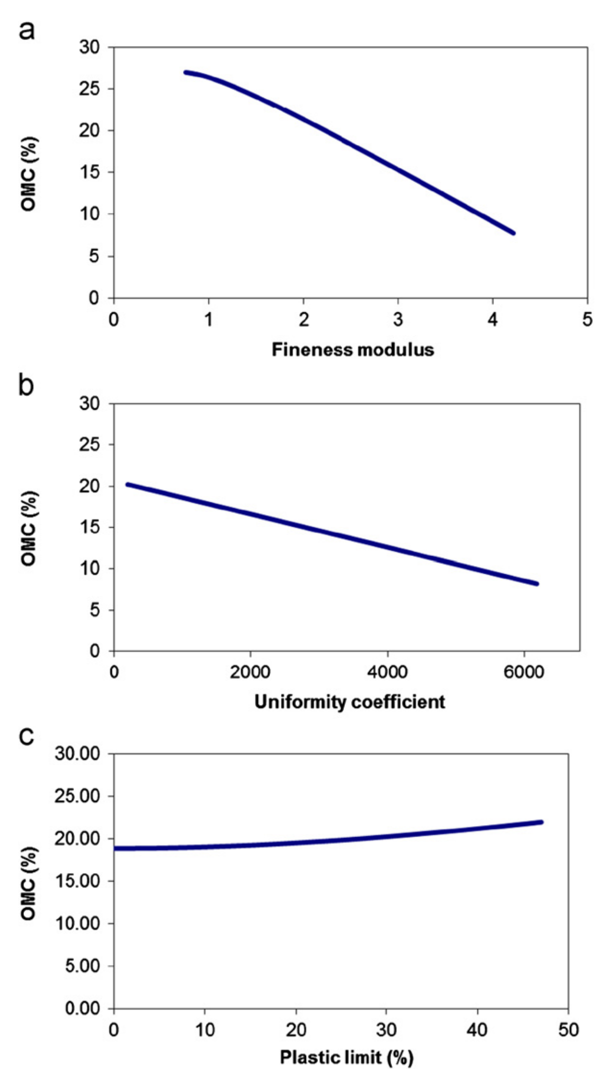


Fig. 5. Parametric study results of the optimum moisture content against (a) fineness modulus of the soil, (b) uniformity coefficient, and (c) plastic limit.

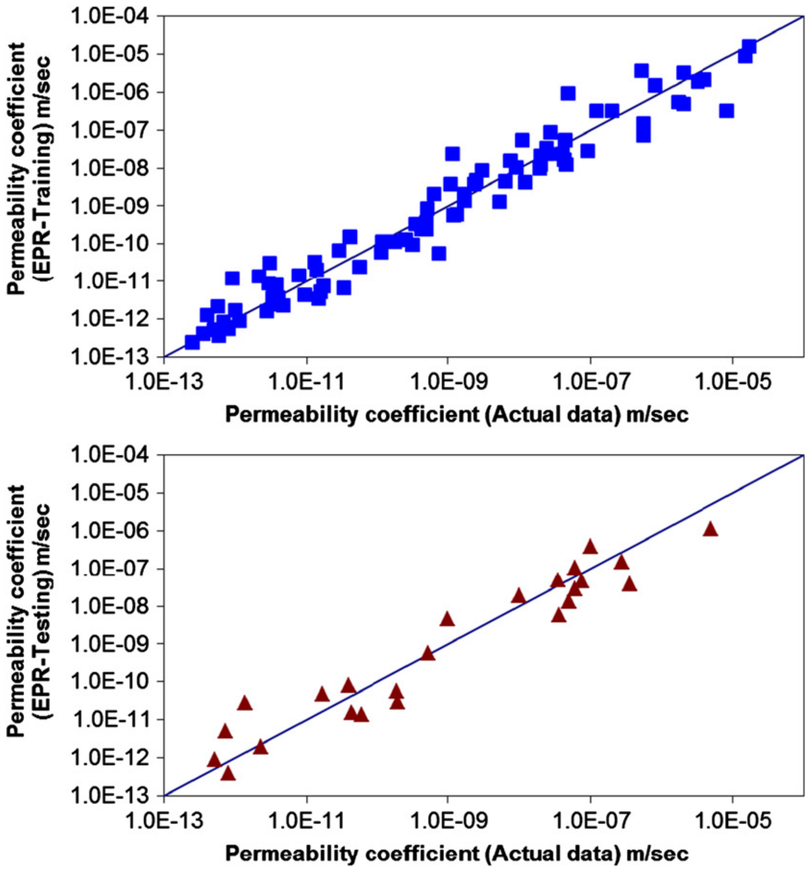


Fig. 6. Comparison between the predicted permeability coefficient and the actual values.

Table 6
Coefficient of determination for predicted K values.

Model	COD values (%)
Evolutionary polynomial regression (EPR)	92 (for unseen testing data)
Artificial neural network (ANN)	90
Wang and Huang (1984)	89

researchers such as Venkatarama and Gupta (2008). The effect of coefficient of uniformity on OMC is shown in Fig. 5b. The higher the coefficient of uniformity, the larger the range of particle sizes in the soil and hence the lower the optimum moisture content (Craig, 1998). Again, this trend is correctly predicted by the model. Fig. 5c shows that increasing plastic limit results in the increase in the optimum moisture content due to the increase in the specific surface of the soil grains. A similar trend of variation of optimum moisture content with plastic limit is reported by Sridharan and Nagaraj (2005). The results also show that optimum moisture content is greatly affected by the fineness modulus and the coefficient of uniformity and plastic limit appears to have less effect on the optimum moisture content of soil.

8. EPR model for coefficient of permeability (K)

Five input parameters were used for the EPR model for the coefficient of permeability (m/s) including degree of compaction (P) expressed in %, mean grain size (D_{50}) expressed in (mm), effective grain size (D_{10}) expressed in mm, plastic limit (PL) expressed in %, and the liquid limit (LL) expressed in %. The EPR model developed to predict the permeability coefficient is

$$\begin{aligned}\mathrm{Log}_{10}(K) = {} & \frac{600.25}{P}\left(1+\frac{4600(D_{50})}{P^2}\right)-\frac{121.96}{{D_{50}}^2\times 10^8}+\frac{2\times 10^{-6}\mathrm{PL}^2}{D_{50}} \\ & -\frac{3.42\mathrm{LL}-1983.12}{{D_{10}}^3\times 10^{15}}-\frac{1.53\times 10^3({D_{50}}^3)+47.2}{D_{10}\times 10^5} \\ & -5.09\times 10^{-4}\mathrm{PL}^2(1+0.774LL(D_{50}) \\ & -2.66\times 10^{-2}\mathrm{LL}\times \mathrm{PL}(D_{50}))-13.8 \qquad (8)\end{aligned}$$

The results of the developed EPR model are compared with the experimental data (Fig. 6) as well as with two other prediction models (Table 6). The EPR model provides a very good prediction of the coefficient of permeability of soils. Fig. 7 shows the results of the parametric study on the EPR permeability model. Fig. 7a shows that, as expected, by increasing the degree of compaction the volume of voids decreases, which causes the permeability of the soil to decline. As the effective grain size increases, the soil becomes coarser, and the permeability increases up to a point after which it increases at a very slow rate (Fig. 7b). Increasing plasticity index, which could imply greater fines content in the soil, also causes the permeability of the soil to decrease (Blotz et al., 1998) as correctly predicted by the proposed model (Fig. 7c). The results show that the plasticity index has the greatest effect on the permeability of the soil while the effects of degree of compaction and effective grain size are relatively moderate.

9. Summary and conclusion

The process of compaction is extensively employed in the construction of embankments and strengthening subgrades of

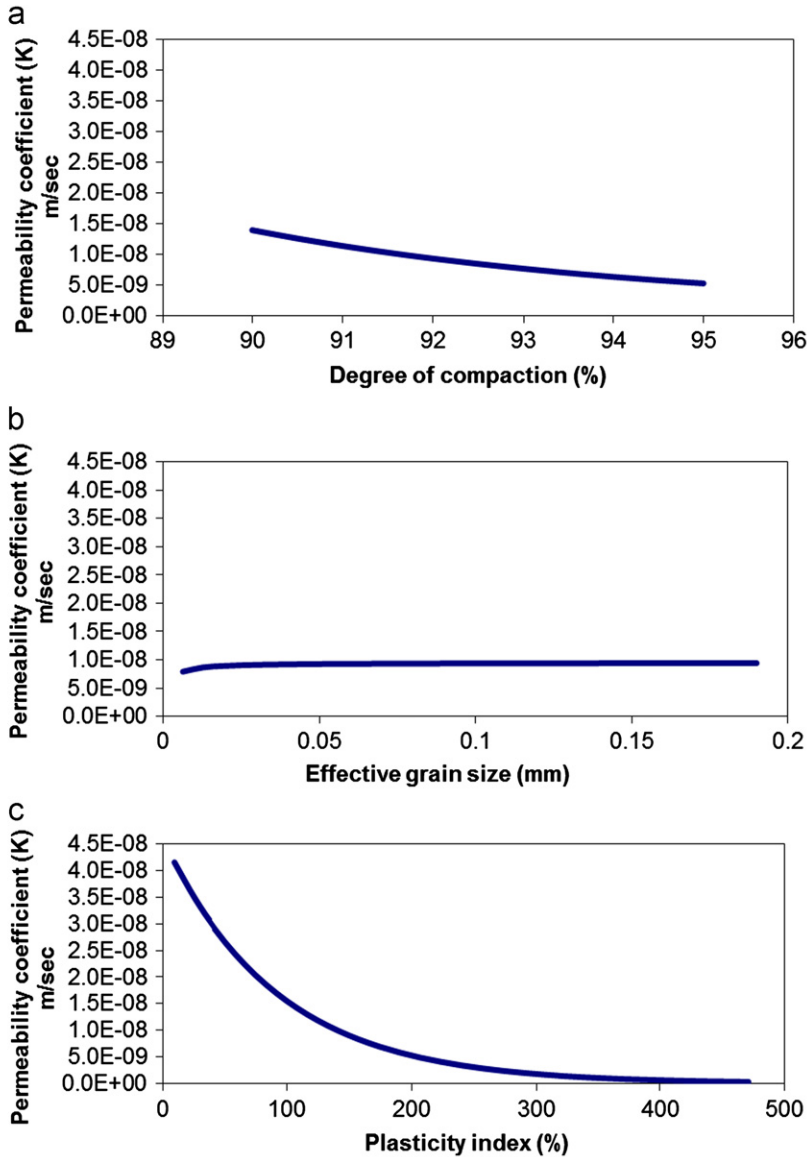


Fig. 7. Parametric study results of the permeability coefficient model against (a) soil compaction, (b) effective grain size, and (c) plasticity index.

roads and runways. In recent years the use of pattern recognition methods such as artificial neural network has been introduced as an alternative method for predicting compaction characteristics and permeability of soils. These methods have the advantages that they do not require any simplifying assumptions in developing the model. However, neural network based models also have some shortcomings. In this research a new approach was presented to describe the relationships between permeability and compaction characteristics, with some physical properties of soils.

Three separate EPR models were developed and validated using a database of experiments involving test data on compaction and permeability characteristics of a number of soils. The results of the model predictions were compared with the experimental data as well as results from other models including a neural network. A parametric study was conducted to evaluate the effects of different parameters on permeability and compaction characteristics of soils. Comparison of the results shows that the developed EPR models provide very accurate predictions. They can capture and represent various aspects of compaction and the permeability behavior of soils directly from the experimental data. The developed models present a structured and transparent representation of the system allowing a physical interpretation of the problem that gives the user an insight into the relationship between the soil compaction and the permeability behavior and various contributing physical properties. From a practical point of view, the EPR models presented in this paper are accurate and easy to use. In the EPR approach, no preprocessing of the data is required and there is no need for normalization or scaling of the data.

An interesting feature of EPR is in the possibility of obtaining more than one model for a complex phenomenon. Selecting an appropriate objective function, assuming preselected elements (based on engineering judgment), and working with dimensional information enable refinement of final models (Giustolisi and Savic, 2006).

The best models are chosen on the basis of their performance on a set of unseen data. This allows examining the generalization capabilities of the developed models. Thus, an unbiased performance indicator is obtained on the real capability of the models.

Another major advantage of the EPR approach is that as more data become available, the quality of the prediction can be easily improved by retraining the EPR model using the new data.

References

Blotz, L.R., Benson, C.H., Boutwell, G.P., 1998. Estimating optimum water content and maximum dry unit weight for compacted clays. Journal of Geotechnical and Geo-Environmental Engineering, ASCE 124 (9), 907–912.

Burmister, D.M., 1954. Principles of permeability testing of soils, symposium on permeability of soils. ASTM Special Technical Publication, vol. 163, pp. 3–26.

Carman, P.C., 1937. Fluid Flows Through Granular Beds, vol. 15. Transportation Institute Chemistry Engineers, London, England, pp. 150–166.

Chen, C.Y., Bullen, A.G.R., Elnaggar, H.A., 1977. Permeability and related principles of coal refuse. Transportation Research Record 640, 49–52.

Craig, R.F., 1998. Soil Mechanics, 6th ed. E&FN Spon.

Davidson, D.T., Gardiner, W.F., 1949. Calculation of standard proctor density and optimum moisture content from mechanical analysis, shrinkage factors, and plasticity index. In: Proceedings of the HRB 29, pp. 447–481.

Garcia-Bengochea, I., Lovell, C.W., Altschaeffl, A.G., 1979. Pore distribution and permeability of silty gays. Journal of Geotechnical Engineering Division, ASCE 105 (GT7), 839–855.

Giustolisi, O., Savic, D.A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. Journal of Hydroinformatics, IWA-IAHR Publishing, UK 8 (3), 207–222.

Gupta, S.C., Larson, W.E.I.I., 1979. A model for predicting packing density of soils using particles-size of distribution. Soil Science Society of America Journal 43 (4), 758–764.

Hauser, V.L., 1978. Seepage control by particle size selection. Transactions of the ASAE 21 (4), 691–695.

Hazen, A., 1911. Discussion of dams on sand foundations. Transactions of the ASCE 73, 199–203.

Horn, M.E., 1971. Estimating soil permeability rates. Journal of the Irrigation and Drainage Division, ASCE 97 (IR2), 263–274.

Jumikis, A.R., 1946. Geology and soils of the Newark (N.J.) metropolitan area. Journal of the Soil Mechanics and Foundations Division, ASCE 93 (SM2), 71–95.

Lambe, T.W., 1951. Soil Testing for Engineers. John Wiley & Sons Inc., NY.

Linveh, M., Ishai, I., 1978. Using indicative properties to predict the density-moisture relationship of soil. Transportation Research Record (60P), pp. 22–28.

Michaels, A.S., Lin, C.S., 1954. The permeability of kaolinite. Industrial & Engineering Chemistry Research 46, 1239–1246 (ACS Publications).

Mitchell, J.K., Hopper, D.R., Campanella, R.C., 1965. Permeability of compacted clay. Journal of the Soil Mechanics and Foundations Division, ASCE 91 (SM4), 41–65.

Najjar, Y.M., Basheer, I.A., Naouss, W.A., 1996. On the identification of compaction characteristics by neuronets. Computers and Geotechnics 18 (3), 167–187.

Olson, R.E., 1963. Effective stress theory of soil compaction. Journal of the Soil Mechanics and Foundations Division, ASCE 89 (SM2), 27–45.

Ramiah, B.K., Viswanath, V., Krishnamurthy, H.V., 1970. Interrelationship of compaction and index properties. In: Proceedings of the Second Southeast Asian Conference on Soil Engineering, Singapore, pp. 577–587.

Rezania, M., Javadi, A.A., Giustolisi, O., 2008. An evolutionary-based data mining technique for assessment of civil engineering systems. Journal of Engineering Computations 25 (6), 500–517.

Ring, G.W., Sallgerb, J.R., Collins, W.H., 1962. Correlation of compaction and classification test data. HRB Bulletin 325, 55–75.

Rowan, H.W., Graham, W.W., 1948. Proper compaction eliminates curing period in construction fills. Civil Engineering 18, 450–451.

Sinha, S.K., Wang, M.C., 2008. Artificial neural network prediction models for soil compaction and permeability. Geotechnical and Geological Engineering 26, 47–64.

Sridharan, A., Nagaraj, H.B., 2005. Plastic limit and compaction characteristics of fine grained soils. Ground Improvement 9 (1), 17–22.

Taylor, D.W., 1948. Fundamentals of Soil Mechanics. John Wiley & Sons Inc., NY.

Turnbull, J.M., 1948. Computation of the optimum moisture content in the moisture–density relationship of soils. In: Proceedings of the Second International Conference on Soil Mechanics and Foundation Engineering, Rotterdam, Holland, IV, pp. 256–262.

Venkatarama, B.V., Gupta, A., 2008. Influence of sand grading on the characteristics of mortars and soil–cement block masonry. Construction and Building Materials 22, 1614–1623.

Wang, M.C., Huang, C.C., 1984. Soil compaction and permeability prediction models. Journal of Environmental Engineering, ASCE 110 (6), 1063–1083.

Zunker, F., 1930. Das Verhalten des Bodens Zum Wasser. Handbuch der Bodenlehre 6, 66–220.