An evolutionary approach to modelling concrete degradation due to sulphuric acid attack

Alani, Amir and Faramarzi, Asaad (2014) An evolutionary approach to modelling concrete degradation due to sulphuric acid attack. Applied Soft Computing, 24. pp. 985-993. ISSN 1568-4946

http://dx.doi.org/10.1016/j.asoc.2014.08.044

An evolutionary approach to modelling concrete degradation due to sulphuric acid attack

Amir M. Alani [1], Asaad Faramarzi *

Department of Civil Engineering, School of Engineering, University of Greenwich, Central Avenue, Chatham Maritime, Kent ME4 4TB, United Kingdom

ARTICLE INFO

Keywords:
Evolutionary polynomial regression
Optimisation
Hybrid techniques
Sulphuric acid attack
Corrosion
Sewer pipes

ABSTRACT

Concrete corrosion due to sulphuric acid attack is known to be one of the main contributory factors for degradation of concrete sewer pipes. This article proposes to use a novel data mining technique, namely, evolutionary polynomial regression (EPR), to predict degradation of concrete subject to sulphuric acid attack. A comprehensive dataset from literature is collected to train and develop an EPR model for this purpose. The results show that the EPR model can successfully predict mass loss of concrete specimens exposed to sulphuric acid. Parametric studies show that the proposed model is capable of representing the degree to which individual contributing parameters can affect the degradation of concrete. The developed EPR model is compared with a model based on artificial neural network (ANN) and the advantageous of the EPR approach over ANN is highlighted. In addition, based on the developed EPR model and using an optimisation technique, the optimum concrete mixture to provide maximum resistance against sulphuric acid attack has been identified.

1. Introduction

Sewer systems are essential infrastructures that play a pivotal role in economy, prosperity, social well-being, quality of life and especially the health of a country. The nature of the wastewater and the propensity for anaerobic conditions in the buried pipes lead to complex chemical and biochemical transformations in the pipes, resulting in inevitable deterioration of pipe materials due to a variety of mechanisms such as hydrogen sulphide induced corrosion of concrete. The sewer networks have had to expand as a result of population growth and thus the extended hydraulic retention time of wastewater in the sewer pipes tends to create a suitable environment for sulphide production, leading to the corrosion of pipes [1]. In addition it is also believed that the widely projected climate change induced temperature rise will further accelerate corrosion. The pipe corrosion results in reduction of wall thickness, leading to collapse of the pipes and possibly the whole system, unless proactive intervention is carried out in a timely manner, based on an accurate prediction of their remaining safe life. The consequences of the collapses of sewers are socially, economically and environmentally devastating, causing enormous disruption of daily life, massive costs, and widespread pollution [1].

Concrete corrosion due to sulphuric acid attack is known to be one of the main contributory factors for degradation of concrete sewer pipes. Sulphate, which exists in wastewater, is reduced to sulphide by anaerobic bacteria. These bacteria are present in a thin slime layer on the submerged surface of the sewer pipe and the production of sulphide occurs in this slime layer. The generated sulphide escapes to the exposed sewer atmosphere where it is transformed to sulphuric acid by aerobic bacteria. The acid reacts with calcium hydroxide in the cementitious sewer pipe which forms gypsum and causes corrosion [2–4].

Pomeroy [3] proposed a model to predict the corrosion rate in cementitious sewer pipes.

$$c = 11.5\frac{k\phi_{sw}}{A} \tag{1}$$

In this equation, c is the average rate of corrosion of the material (mm/year), k is a factor representing the acid formation based on climate condition, ϕ_{sw} is the average flux of sulphide to the pipe wall (g/m^2 h) and A is the alkalinity of the pipe material.

* Corresponding author at: School of Civil Engineering, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. Tel.: +0121 4145146.
E-mail addresses: m.alani@gre.ac.uk (A.M. Alani), A.Faramarzi@bham.ac.uk (A. Faramarzi).
[1] Tel.: +44 1634 883293.

Eq. (1) shows that amongst pipe material characteristics, alkalinity (A) is the most influential factor in the corrosion of concrete sewer pipes. Many researchers have investigated the effect of acid attack on different mixtures and admixtures of concrete. Attiogbe and Rizkalla [5] evaluated the response of four different concrete mixtures including two different cement types (ASTM Type I and ASTM Type V) to accelerated acid attack. The concrete samples were immersed in sulphuric acid solutions with a pH of 1.0. This concentration of sulphuric acid was selected since it was a representative of what is expected in sewer pipes in the process of deterioration. After 70 days of immersion, the results of the experiment showed that the weight loss of concrete samples with cement Type V is slightly more than those samples created with cement Type I. It was concluded that in the long term, the sulphate resistant cement does not contribute to an improved resistance of concrete compared to ordinary Portland cement when they are subjected to sulphuric acid attack. Ehrich et al. [6] carried out biogenic and chemical sulphuric acid tests to monitor the corrosion of different cement mortars. They used ordinary and sulphate resistant Portland cement as well as calcium aluminate cement to produce different mortars. The biogenic tests were carried out using a simulation chamber where the temperature, humidity and amount of sulphide were monitored and controlled. For the chemical test, the mortar samples were immersed in PVC containers filled with sulphuric acid. The results of both chemical and biogenic tests showed that calcium aluminate cement mortars had greater resistance against both types of acid attacks. Monteny et al. [7] simulated chemical and biogenic sulphuric acid corrosion of different concrete compositions including ordinary and polymer cement concrete. For the biogenic tests, they put small concrete samples in a microbiological suspension containing bacteria, sulphur and nutrients which generated sulphuric acid in a biogenic manner. The chemical tests were performed using a rotating apparatus. Concrete samples were set up on an axis which was rotating in such a way that the concrete samples were only partially immersed in a solution of sulphuric acid with a pH of around 1.0. The results of both tests revealed that concrete mixtures with styrene-acrylic ester polymer showed a higher resistance compared to the concrete with high sulphate resistance cement. On the other hand the concrete mixtures with acrylic polymer and styrene butadiene polymer showed a lower strength than the high sulphate resistance concrete. De Belie et al. [8] presented the results of biogenic and chemical sulphuric acid tests carried out on different types of commercially produced concrete sewer pipes. They performed both types of tests on different mixtures of concrete including different aggregate and cement types. The results of both chemical and biogenic tests showed that the aggregate type had the largest effect on degradation of concrete samples. In addition, based on the results obtained from their studies, they proposed an equation to predict the degradation depth taking into account both alkalinity and water absorption of concrete (Eq. (2)).

$$C = \frac{c_1}{A} + c_2 W \tag{2}$$

where C is degradation depth after four cycles of the microbiological test (mm), A is alkalinity, W is water absorption (%) and c_1 and c_2 are the coefficients of the equation. Chang et al. [9] investigated the use of different aggregates and cements to improve the resistance of concrete subject to sulphuric acid attack. The concrete samples were produced with limestone, and siliceous aggregate, and Portland, binary and ternary cements. The water/cement ratio was kept constant (i.e. $W/C = 0.4$) for all the samples. The concrete specimens were immersed into a sulphuric acid solution with a pH between 1.27 and 1.35. The changes in weight and compression strength of samples were examined at different ages up to 168 days. It was shown that the use of limestone aggregates and ternary cement containing silica fume and fly ash will help to reduce the weight loss and reduction in compressive strength of concrete under sulphuric acid attack. Hewayde et al. [10] carried out an investigation on 78 different concrete mixtures including different cement types, different water/cement ratios and various admixtures subject to sulphuric acid attack. The concrete samples were immersed in sulphuric acid solutions with pH levels of 0.3, 0.6, and 1.0. The authors stated that the solution with a pH of 0.6 represents conditions with a high count of anaerobic bacteria that exist in the submerged surface of the sewer pipes, while the solution with a pH of 0.3 represents a supercritical condition that may occur in industrial sewer systems subject to high temperature and humidity. The experiment consisted of determining the compressive strength of samples at different ages and measuring the changes in weight at different pH values. Using the data collected from the tests, they developed two artificial neural network (ANN) models to predict the mass loss and compressive strength of concrete. They showed that the developed ANN models are capable of predicting both compressive strength and mass loss of concrete samples under exposure to sulphuric acid, providing the required parameters (i.e. the concrete contents) have been inputted. The studies presented above and many more in literature show that the constituents of concrete mix including admixtures play an important role in the alkalinity of concrete and consequently its vulnerability to sulphuric acid induced corrosion. However, insufficient work has been carried out in relation to the modelling and development of an explicit relationship to predict the deterioration of concretes with various mixtures subject to sulphuric acid. No doubt the development of such model(s) would help industry to evaluate and possibly improve the concrete mix design of their sewer pipes. In addition if the concrete mix design of existing pipes is known, water companies can carry out proactive intervention, based on the accurate predictions provided by such models.

The rapid development in computational software and hardware in recent decades has introduced several soft computing and data-driven approaches to modelling engineering problems. Although there are various data-driven techniques based on artificial intelligence, artificial neural network (ANN) and genetic programming (GP) are among the best known techniques that have been used to model civil and mechanical engineering problems. ANN uses models composed of many processing elements (neurons) connected by links of variable weights (parameters) to form black box representations of systems. ANNs are capable of dealing with a large amount of data and can learn complex model functions from examples, by training sets of input and output data [11,12]. ANNs have the ability to model complex, nonlinear processes without having to assume the form of the relationship between input and output variables [13,14]. However, ANN has shown to possess some drawbacks. A major disadvantage of ANN is the large complexity of the network structure; it represents the knowledge in terms of a weight matrix and biases which are not accessible to the user. ANN models, as a black box class of models, gives no information on how the input parameters affect the output(s). In addition, parameter estimation and over-fitting are other disadvantages of models constructed by ANN [15,16]. Genetic programming (GP) is another modelling approach that has been used to model engineering phenomena. GP is an evolutionary computing method that generates transparent and structured mathematical expressions to represent the system being studied. The most common type of GP method is symbolic regression, which was proposed by Koza [17]. This technique creates mathematical expressions to fit a set of data points using the evolutionary process of genetic programming. The genetic programming procedure mimics natural selection as the 'fitness' of the solutions in the population improves through successive generations [18,19]. However, GP also has some limitations. It is proven that GP is not very powerful in finding constants and, more importantly, that it tends to produce functions that grow in length over time [15].

In this article, using a dataset collected from literature and a novel hybrid data-driven technique that overcomes the shortcomings of ANN and GP, a model is developed to predict the degradation of concrete subject to sulphuric acid attack. This new data mining technique, called evolutionary polynomial regression (EPR), provides a structured, transparent and concise model representing the behaviour of the system. Description of EPR technique is provided in following sections. Then development of the model to predict the degradation of concrete subject to acid attack is presented. A parametric study is carried out for the proposed model in order to investigate the effect of changes in different input parameters on the output. In addition the developed EPR model is compared with a neural network model to show the advantageous of the proposed technique. Using the developed model and optimisation techniques, the optimum ingredients of concrete mixtures to resist against acid attack is determined.

2. Evolutionary polynomial regression

Evolutionary polynomial regression (EPR) is a new hybrid technique for creating true or pseudo-polynomial models from observed data by integrating the power of least square regression with the efficiency of genetic algorithm. A typical formulation of EPR can be expressed in the following equation [15]:

$$y = \sum_{j=1}^{m} F(\mathbf{X}, f(\mathbf{X}), a_j) + a_0 \tag{3}$$

In this equation, y is the estimated output of the system; a_j is a constant value; F is a function constructed by process; $\mathbf{X}$ is the matrix of input variables; f is a function defined by user; and m is the number of terms of expression excluding the bias term a_0. The general functional structure represented by $F(\mathbf{X}, f(\mathbf{X}), a_j)$ is constructed from elementary functions by EPR using genetic algorithm (GA). The function of GA is to select the useful input vectors from $\mathbf{X}$ to be combined together. The building blocks (elements) of the structure of F are defined by the user based on understanding of the physical process. While the selection of feasible structures to be combined is done through an evolutionary process, the parameters a_j are estimated by the least square method.

The first step to identify the structure of the model is to convert Eq. (3) into the following vector form [15,20]:

$$\mathbf{Y}_{N\times1}(\boldsymbol{\theta}, \mathbf{Z}) = \begin{bmatrix} \mathbf{I}_{N\times1} & \mathbf{Z}^j_{N\times m} \end{bmatrix} \times [a_0 \quad a_1 \cdots a_m]^T = \mathbf{Z}_{N\times d} \times \boldsymbol{\theta}^T_{d\times1} \tag{4}$$

where $\mathbf{Y}_{N\times1}(\boldsymbol{\theta}, \mathbf{Z})$ is the least square (LS) estimate vector of N target values; $\boldsymbol{\theta}_{d\times1}$ is the vector of $d=m+1$ parameters a_j and a_0 ($\boldsymbol{\theta}^T$ is the transposed vector); and $\mathbf{Z}_{N\times d}$ is a matrix formed by $\mathbf{I}$ (unity 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, $\mathbf{X} = \langle \mathbf{X}_1\, \mathbf{X}_2 \cdots \mathbf{X}_k \rangle$.

EPR starts from Eq. (4) and searches for the best structure, i.e. a combination of vectors of independent variables (inputs) $\mathbf{X}_{S=1:k}$. The matrix of input $\mathbf{X}$ is given as [15]:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{3k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_k \end{bmatrix} \tag{5}$$

where the kth column of $\mathbf{X}$ represents the candidate variable for the jth term of Eq. (4). Therefore the jth term of Eq. (4) can be written as:

$$\mathbf{Z}^j_{N\times1} = \left[(\mathbf{X}_1)^{ES(j,1)} \cdot (\mathbf{X}_2)^{ES(j,2)} \quad \cdots \quad (\mathbf{X}_k)^{ES(j,k)} \right] \tag{6}$$

where $\mathbf{Z}^j$ is the jth column vector in which its elements are products of candidate independent inputs and $\mathbf{ES}$ is a matrix of exponents. Therefore, the problem is to find the matrix $\mathbf{ES}_{k\times m}$ of exponents whose elements can be values within user-defined bounds. For example, if a vector of candidate exponents for inputs, $\mathbf{X}$, (chosen by user) is $\mathbf{EX} = [0, 1, 2]$ and number of terms (m) (excluding bias) is 4, and the number of independent variables (k) is 3, then the polynomial regression problem is to find a matrix of exponents $\mathbf{ES}_{4\times3}$ [15]. An example of such matrix can be as following:

$$\mathbf{ES} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix} \tag{7}$$

When this matrix is applied to Eq. (6) the following set of mathematical expression is obtained:

$$\begin{aligned} \mathbf{Z}_1 &= (\mathbf{X}_1)^0 \cdot (\mathbf{X}_2)^1 \cdot (\mathbf{X}_3)^2 = \mathbf{X}_2 \cdot \mathbf{X}_3^2 \\ \mathbf{Z}_2 &= (\mathbf{X}_1)^0 \cdot (\mathbf{X}_2)^1 \cdot (\mathbf{X}_3)^1 = \mathbf{X}_2 \cdot \mathbf{X}_3 \\ \mathbf{Z}_3 &= (\mathbf{X}_1)^1 \cdot (\mathbf{X}_2)^2 \cdot (\mathbf{X}_3)^0 = \mathbf{X}_1 \cdot \mathbf{X}_2^2 \\ \mathbf{Z}_4 &= (\mathbf{X}_1)^1 \cdot (\mathbf{X}_2)^1 \cdot (\mathbf{X}_3)^0 = \mathbf{X}_1 \cdot \mathbf{X}_2 \end{aligned} \tag{8}$$

Thus the expression of Eq. (4) is:

$$\begin{aligned} \mathbf{Y} &= a_0 + a_1 \cdot \mathbf{Z}_1 + a_2 \cdot \mathbf{Z}_2 + a_3 \cdot \mathbf{Z}_3 + a_4 \cdot \mathbf{Z}_4 \\ &= a_0 + a_1 \cdot \mathbf{X}_2 \cdot \mathbf{X}_3^2 + a_2 \cdot \mathbf{X}_2 \cdot \mathbf{X}_3 + a_3 \cdot \mathbf{X}_1 \cdot \mathbf{X}_2^2 + a_4 \cdot \mathbf{X}_1 \cdot \mathbf{X}_2 \end{aligned} \tag{9}$$

It should be noted that each row of $\mathbf{ES}$ determines the exponents of the candidate variable of the jth term in Eqs. (3) and (4). Each of the exponents in matrix $\mathbf{ES}$ corresponds to a value from user-defined vector $\mathbf{EX}$. This allows the transformation of the symbolic regression problem into finding the best $\mathbf{ES}$, i.e. the best structure of the EPR equation, e.g. in Eq. (9).

In addition to the above structure, EPR can construct non-polynomial mathematical expressions. It is possible to assume a function f, such as natural logarithm, hyperbolic tangent, hyperbolic secant and exponential and a structure among the following [15]:

$$\begin{aligned} &\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (\mathbf{X}_1)^{ES(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{ES(j,k)} \cdot f\left((\mathbf{X}_1)^{ES(j,k+1)}\right) \cdot \ldots \cdot f\left((\mathbf{X}_k)^{ES(j,2k)}\right) && \text{Case 1} \\ &\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot f\left((\mathbf{X}_1)^{ES(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{ES(j,k)}\right) && \text{Case 2} \\ &\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (\mathbf{X}_1)^{ES(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{ES(j,k)} \cdot f\left((\mathbf{X}_1)^{ES(j,k+1)} \cdot \ldots \cdot (\mathbf{X}_k)^{ES(j,2k)}\right) && \text{Case 3} \\ &\mathbf{Y} = g\left(a_0 + \sum_{j=1}^{m} a_j \cdot (\mathbf{X}_1)^{ES(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{ES(j,k)}\right) && \text{Case 4} \end{aligned} \tag{10}$$

An integer GA coding is used in EPR to determine the location of the candidate exponents of $\mathbf{EX}$ in the matrix $\mathbf{ES}$ [20,21]. For example, the positions in $\mathbf{EX}=[0,1,2]$ correspond to the following string for the matrix of Eq. (7) and the expression of Eq. (9):

$$\mathbf{EX} = [123, \quad 122, \quad 231, \quad 221] \tag{11}$$

It is clear that the presence of a zero in **EX** ensures the ability to exclude some of the inputs and/or input combinations from the regression equation.

The modelling process of EPR starts by evolving equations. As the number of evolutions increases, EPR gradually picks up the different contributing parameters to form equations representing the system being studied.

In order to provide the best symbolic model(s) of the system being studied to the users, EPR is facilitated with different objective functions to optimise. The original EPR methodology used only one objective (i.e., the accuracy of data fitting) to explore the space of solutions while penalising complex model structures using some penalisation strategies [15]. However the single-objective EPR methodology showed some shortcomings, and therefore the multi-objective genetic algorithm (MOGA) strategy has been added to EPR [22]. The multi-objective approach in EPR (MOGA-EPR) is designed to seek those model structures that on one hand satisfy the fitness and on the other hand controlling the structural complexity. In this approach the control of fitness and complexity are demanded to different singly acting objective functions. The objectives represented by the functions are mutually conflicting, and therefore their optimisation returns a trade-off surface of models [20–22]. MOGA-EPR tackles a multi-model strategy by varying the structural parsimony (i.e. the number of constant values in the equation) while working on the objective function used in Single-Objective EPR. Then, MOGA-EPR finds the set of symbolic expressions that perform well according to two (or more) conflicting criteria considered simultaneously, the level of agreement between simulated and observed measurements, and structural parsimony of the expressions obtained. The objective functions used are: (i) Maximisation of the fitness; (ii) Minimisation of the total number of inputs selected by the modelling strategy; (iii) Minimisation of the length of the model expression. A further advantage of MOGA-EPR is the increased pressure to achieve structural parsimony because a large number of a_j values or a large total number of inputs must be justified by the fitness of the model (note that the Pareto dominance criterion and the function are to be minimised). The introduced objective functions can be used in a two objective configuration or all together [20–22]. At least one objective function limits the complexity of the models while the other one control the fitness of the models. The multi-objective strategy returns a trade-off surface (or line) of complexity versus fitness which allows the user to achieve a lot of purposes of the modelling approach to the phenomenon studied [20–22]. In this study the multi-objective EPR is used to develop the EPR-based models. Further details of the EPR technique can be found in [15,20–22].

The accuracy of the developed models by EPR is measured at each stage using the coefficient of determination (CoD) [23]:

$$\mathrm{CoD} = 1 - \frac{\sum_N (Y_a - Y_p)^2}{\sum_N \left(Y_a - 1/N \sum_N Y_p\right)^2} \tag{12}$$

where Y_a is the actual input value; Y_p is the EPR predicted value and N is the number of data points on which the CoD is computed. If the model fitness is not acceptable or other termination criteria (e.g., maximum number of generation and maximum number of terms) are not satisfied, the current model should go through another evolution in order to obtain a new model [20]. A typical flow diagram for the EPR procedure is presented in Fig. 1.

The EPR algorithm has been implemented in MATLAB by “hydroinformatics” research group at the Technical University of Bari, Italy [20–24]. EPR has a friendly and easy-to-use interface and offers a wide range of options to control the complexity and structure of the models. EPR is proven to be capable of learning complex non-linear relationships from a set of data, and it has many desirable features for engineering applications. The EPR technique has

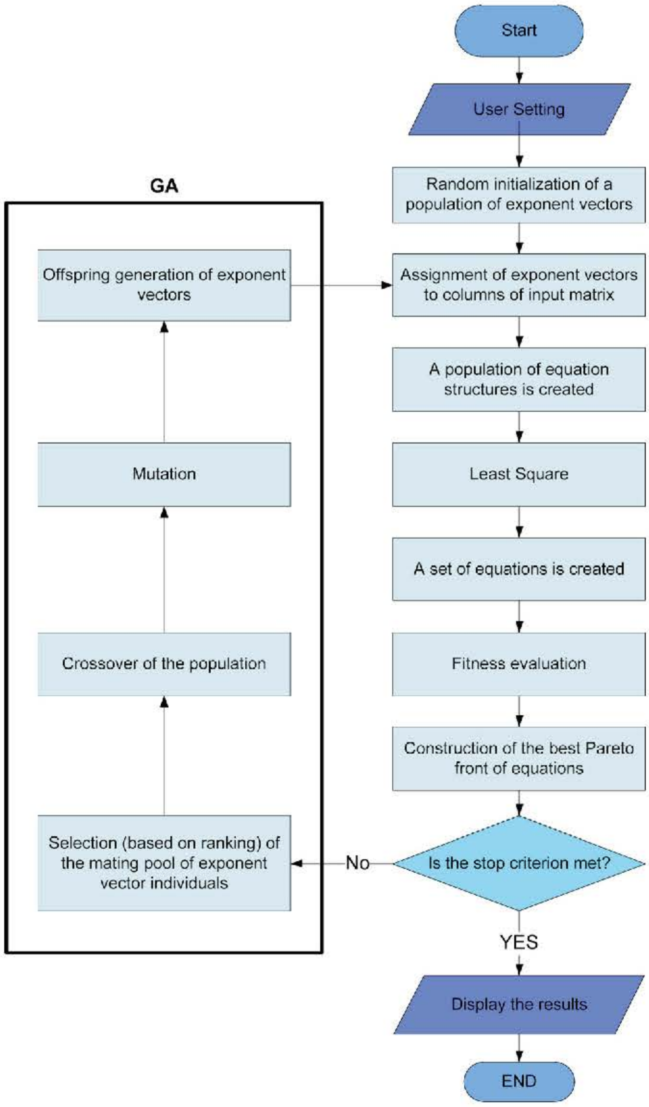


Fig. 1. Typical flow diagram for EPR procedure.

been successfully applied to modelling a wide range of complex engineering problems including modelling sewer failure [24], pipe break prediction [25], mechanical behaviour of rubber concrete [26], torsional strength of reinforced concrete beams [27] and many other applications in civil and mechanical engineering [28–30].

3. Development of models

3.1. Database

The database to train and develop EPR models is collected from a study by Hewayde [31]. Hewayde [31] carried out a set of experiments to evaluate the compressive strength and mass loss of different concrete mixtures under sulphuric acid attack. The experiment involved the preparation of several concrete cylinders with various mix design, followed by immersing them in sulphuric acid solutions with different pH values in order to measure the level of degradation. Degradation of samples was evaluated by means of measuring and recording the mass loss of concrete samples after immersion in acid solution. Two different cement types (ASTM Type I and ASTM Type V), siliceous fine and coarse aggregate and various admixtures including silica fume, metakaolin, geopolymer cement, organic corrosion inhibitor (OCI), Caltite, and Xypex were

Table 1
Input and output parameters of models I and II.

	Model I (Mass loss)													
	Inputs													Output
Parameter	Cement	Gravel	Sand	Water	SP[a]	Slag	Silica Fume	Meta.[b]	OCI	Caltite	Xypex	Geo[c]	pH	Mass Loss
Unit	kg/m^3	kg/m^3	kg/m^3	L/m^3	L/m^3	kg/m^3	kg/m^3	kg/m^3	L/m^3	L/m^3	kg/m^3	kg/m^3	–	(%)
Symbol	C	G	S	W	H	Sg	SF	M	OCI	Clt	X	Geo	pH	ML

[a] SP: superplasticizer.
[b] Meta: Metakaolin.
[c] Geo: geopolymer cement.

Table 2
Statistics of the training and testing data used to develop the EPR model.

Parameters	C	G	S	W	H	Sg	SF	M	OCI	Clt	X	Geo	pH	ML
Training data														
Minimum	140.0	745.0	798.0	109.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
Maximum	571.0	1009.0	926.0	202.0	2.8	150.5	64.5	64.5	7.0	35.0	13.1	215.0	1.0	70.0
Mean	352.7	870.1	869.1	148.3	1.1	46.9	4.8	5.8	0.5	2.7	0.9	20.8	0.5	0.2
Standard deviation	96.4	35.5	21.7	17.9	0.6	65.7	14.0	15.3	1.6	8.3	2.8	58.6	0.2	0.1
Testing data														
Minimum	182.0	851.0	829.0	120.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
Maximum	430.0	952.0	892.0	168.3	2.4	150.5	64.5	43.0	6.0	30.0	8.6	172.0	1.0	0.3
Mean	341.6	875.8	870.2	147.2	1.1	53.1	5.4	8.0	0.8	2.9	1.5	21.5	0.5	0.2
Standard deviation	83.7	18.3	11.0	10.3	0.5	67.8	14.9	14.7	1.9	8.9	3.3	55.1	0.2	0.1

used to prepare concrete specimens. The effect of using ASTM Type V cement in the mixtures was presented in terms of percentage of slag since Type V cement, is a blended cement made of 65% ordinary Portland cement (ASTM Type I) and 35% finely ground granulated blast furnace slag. The concrete samples had different values of water/cement ratio and aggregate contents as well as various percentages of superplasticizer and admixtures which made a very suitable collection of data to train and develop EPR models. Further details of the experiments are described in [10,31]. In this study all the above ingredients of concrete are considered as input parameters of the EPR model and percentage of mass loss as an indication of degradation as the output. Details of the all parameters, symbols and units used to develop the model are presented in Table 1.

3.2. *EPR procedure*

In order to ensure the validity and reliability of the developed models, before the EPR procedure starts, the data is divided into two independent training and validation sets. This is also a common approach in most of the data mining techniques based on artificial intelligence such as neural network and genetic programming [10–20]. The construction of the model takes place by adaptive learning over the training set and the performance of the constructed model is then appraised using the validation set. In order to select the most robust representation of the whole data for training and validation sets, a statistical analysis was carried out on the input and output parameters of several randomly selected sets of data. The purpose of the analysis is to ensure that the statistical properties of the data in each of the subsets were as close to each other as possible. After the analysis, the most statistically consistent combination was used for construction and validation of the EPR models. In addition the statistical analysis will help to keep the validation data in the range of the maximum and minimum values of the training data as generally the EPR technique (like other data-mining techniques) is stronger in interpolation than extrapolation over the data. Maximum, minimum, average and standard deviations are the parameters used to perform the analysis. The result of the statistical analysis is presented in Table 2.

Before the start of the EPR process the training data was shuffled to avoid any bias during the training process over a particular part of the data. Once the training and validation sets are chosen, the EPR process can start. To develop the EPR models, a number of settings can be adjusted to manage the constructed models in terms of the type of the functions, number of terms, range of exponents, etc. [20]. When the EPR starts, the modelling procedure commences by evolving equations. As the number of evolutions increases, EPR gradually learns and picks up the participating parameters in order to form equations. Each proposed model is trained using the training data and tested using the validation data. The level of accuracy at each stage is measured using the CoD (Eq. (12)). Several EPR runs were carried out and the analysis was repeated with various combinations and ranges of exponents, different functions and different numbers of terms in order to obtain the most suitable form for the model. The following setting returned the strongest set of models. Range of exponents: [0 ½ 1 2 3]; number of terms: 20; function type: no function; MOGA strategy: CoD vs. (% a_j). The EPR process with the setting outlined above completed in 4 min and 49 s on a

Table 3
A summary of EPR results for degradation model.

Model No.	No. participating parameters	Number of terms	CoD training (%)	CoD testing (%)
1	0	1	0.0	0.0
2	2	2	68.0	78.4
3	4	3	72.9	79.6
4	8	4	84.3	79.9
5	9	5	87.3	87.3
6	10	6	88.8	90.4
7	12	7	89.8	90.9
8	12	8	91.4	91.3
9	12	9	94.8	90.0
10	10	10	94.9	89.2
11	11	11	95.0	87.4
12	11	12	96.8	91.0
13	10	13	97.0	94.0
14	10	14	97.2	89.7
15	12	15	97.2	88.7
16	13	16	94.7	87.3
17	13	17	96.6	96.1
18	13	18	97.0	88.3
19	13	19	97.3	94.2
20	12	20	97.7	96.0

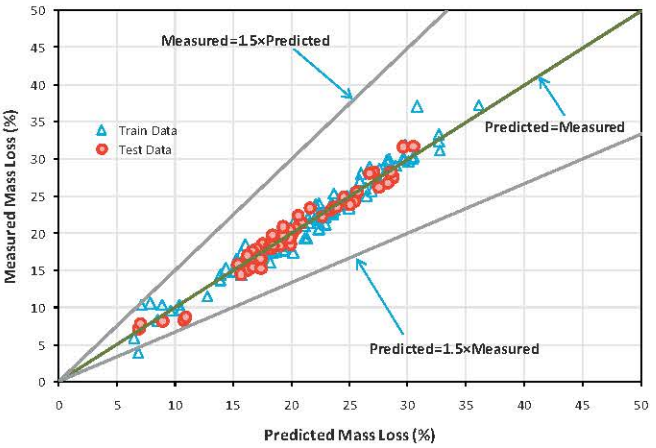


Fig. 2. Prediction results of model I for training and validation data.

personal computer with Intel® Core™ i7 processor with 2.2 GHz of speed and 4GB memory. As mentioned earlier the MOGA-EPR returns a trade-off curve of the model complexity versus accuracy which allows the user to select the most suitable model based on his judgement and knowledge of the problem. The results of the EPR process are presented in Table 3. The EPR models in this table are ranked based on the number of terms. It can be seen from this table that of the 20 equations constructed by EPR only relationship number 16, 17, 18, and 19 include all the participating parameters. Based on the simplicity of the models and the CoD values of both training and testing datasets model number 17 (Eq. (13)) is found to be the most robust models for predicting degradation of concrete.

$$\begin{aligned}\mathrm{ML} = {} & 1.5\times10^{-4}(\mathrm{Sg})^2 + 4.7\times10^{-7}(\mathrm{W})(\mathrm{SF})\sqrt{\mathrm{S}} - 2.2\times10^{-6}(\mathrm{W})\sqrt{(\mathrm{H})(\mathrm{Clt})(\mathrm{S})} + 1.6\times10^{-2}\sqrt{(\mathrm{G})(\mathrm{Sg})} \\ & -1.5\times10^{-7}(\mathrm{H})(\mathrm{Geo})\sqrt{(\mathrm{G})(\mathrm{pH})(\mathrm{W})(\mathrm{Sg})} + 2.8\times10^{-6}(\mathrm{X})\sqrt{(\mathrm{G})(\mathrm{S})(\mathrm{M})} - 1.3\times10^{-6}(\mathrm{G})(\mathrm{W})\sqrt{\mathrm{pH}} \\ & +1.9\times10^{-8}(\mathrm{G})^2\sqrt{(\mathrm{W})} - 9.4\times10^{-11}(\mathrm{G})^2(\mathrm{Sg})(\mathrm{S}) - 3.2\times10^{-4}\sqrt{(\mathrm{C})(\mathrm{S})} + 5.6\times10^{-8}(\mathrm{S})(\mathrm{pH})\sqrt{(\mathrm{C})(\mathrm{W})(\mathrm{Geo})} \\ & -7.2\times10^{-13}(\mathrm{G})^3(\mathrm{pH})^2\sqrt{(\mathrm{C})(\mathrm{W})} - 5.8\times10^{-6}(\mathrm{pH})^2(\mathrm{C})(H)\sqrt{(\mathrm{W})(\mathrm{SF})} - 5.2\times10^{-7}(\mathrm{C})(\mathrm{W})\sqrt{\mathrm{H}} \\ & +3.3\times10^{-5}(\mathrm{C})\sqrt{(\mathrm{G})} - 2.2\times10^{-15}(\mathrm{S})^3(\mathrm{H})^3(C)\sqrt{(\mathrm{G})(\mathrm{X})(\mathrm{pH})} - 1.2\times10^{-10}(\mathrm{C})^2\sqrt{(\mathrm{G})(\mathrm{S})(\mathrm{OCI})}\end{aligned} \tag{13}$$

The symbols used in Eq. (13) are described in Table 1. The predictions provided by this relationship for both training and validation data is illustrated in Fig. 2. From this figure and the CoD values presented in Table 3 it is evident that the EPR model performs well and represent a very accurate prediction for unseen cases of data.

3.3. *Parametric study*

A parametric study was carried out for further examination of the prediction capabilities of the proposed EPR model (i.e. Eq. (13)). The parametric study will help to assess the extent to which the EPR model represents the physical relationships between different parameters and the effects of different input parameters on the model output. All the input parameters except the one being examined were set to their mean values and the model predictions for different values of the parameter being studied were investigated. Each parameter was varied within the range of its maximum and minimum values. Fig. 3 shows the results of the parametric study conducted to investigate the effect of change in cement content and W/C ratio on the developed model. The results are presented for three different pH values (i.e. 0.3, 0.6 and 1.0). The results show that the mass loss of concrete subject to sulphuric acid attack escalates by increasing cement content or reduction in W/C ratio. Both of these behaviours are consistent with previous studies [10]. These results show that as the cement content of concrete increases, the sulphuric acid will expand its reaction with the cement which leads to further corrosion of the concrete. The sensitivity of the EPR model to one of the admixtures (OCI) is presented in Fig. 4. It is evident from this figure that as the amount of OCI increases the mass loss is reduced. This indicates that adding a limited amount of OCI as a partial replacement of cement will reduce the deterioration of concrete against sulphuric acid. In addition it can be observed that Eq. (13) has captured the effect of different values of pH and its effect on the degradation of the concrete. As expected Fig. 3 and 4 show that a lower value of pH, which represents a harsher acidic environment, cause further degradation in concrete. These predictions are in agreement with those reported in Hewayde [31]. It can be seen from the figures above that the developed EPR model was successful in capturing the sensitivity of mass loss to changes of different concrete mixture and admixture contents.

3.4. *Comparison with ANN model*

The results of the developed EPR model (Eq. (13)) is compared with other existing models to assess the performance of EPR and further validate reliability of the developed model. From literature the work carried out by Hewayde et al. [10] is the only study on prediction of concrete degradation as a result of sulphuric acid attack that includes all the concrete ingredients mentioned above. As explained before Hewayde et al. [10] developed an ANN model to predict the mass loss of the concrete samples immersed in sulphuric acid solutions. In this study the model developed by Hewayde et al. [10] is used as a reference to examine the performance of the developed EPR model. Hewayde et al. [10] did not report any CoD or R^2 values for their developed models. Therefore for a fair comparison a feed-forward back-propagation neural network was developed using the same training and testing datasets as those used in the development of the EPR model. The structure and architecture of the neural network was kept same as the one presented in Hewayde et al. [10]. The neural network model comprised of 13 elements in input layer representing the mixture ingredients, one hidden layer with 10 processing elements and one node in output layer representing the mass loss of concrete. The performance of EPR and accuracy of the EPR-based model is compared with the ANN model in terms of coefficient of determination (CoD), root mean square error (RMSE) and mean absolute error (MAE). These coefficients are defined in Eqs. (12), (14) and (15) respectively. The result of this comparison is presented in Table 4.

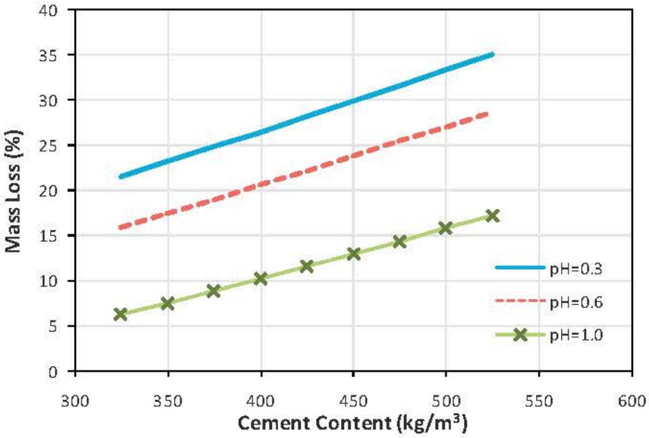
$$\mathrm{RMSE} = \sqrt{\frac{\sum_N (Y_a - Y_p)^2}{N}} \tag{14}$$

$$\mathrm{MAE} = \frac{\sum_N \left|Y_a - Y_p\right|}{N} \tag{15}$$

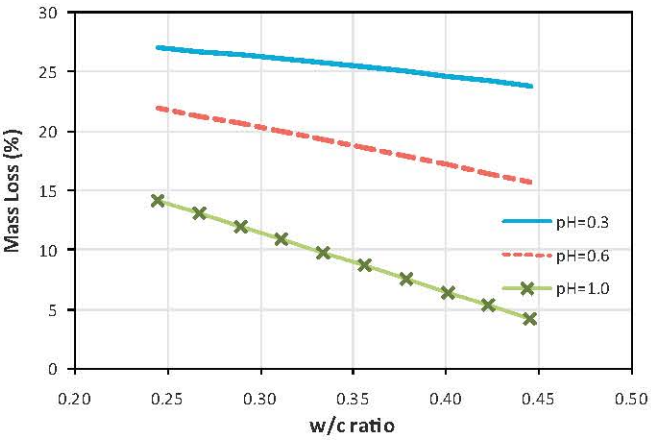
Table 4 shows that the EPR model has captured the underlying relationship between the parameters in different levels and has performed slightly better than the ANN model in all three criteria for both training and testing datasets. However apart from the small differences between these coefficients for EPR and ANN, the fact that the EPR models are transparent, concise, and practical mathematical equations, makes EPR approach more favourable compare

Table 4
Performance of EPR and ANN model in prediction of concrete degradation.

Data subset	*CoD*		*RMSE*		*MAE*	
	EPR	ANN	EPR	ANN	EPR	ANN
Training	96.61	94.28 ± 0.22	1.22	1.71 ± 0.06	0.68	0.78 ± 0.00
Validation	96.14	95.16 ± 0.48	1.08	1.21 ± 0.04	0.89	0.95 ± 0.01



(a)



(b)

Fig. 3. Changes in mass loss with (a) cement content (b) W/C ratio.

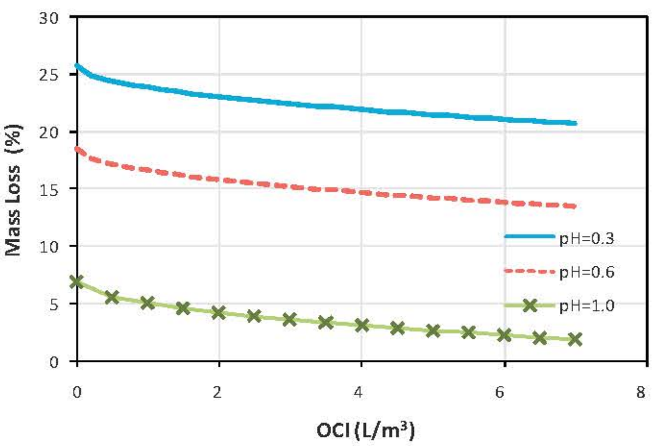


Fig. 4. Changes of mass loss with OCI.

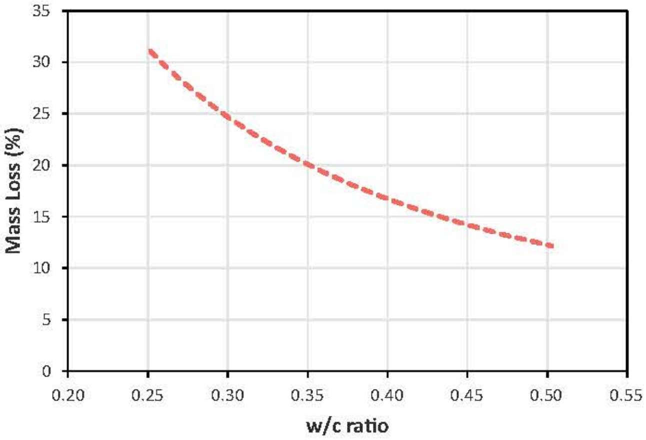


Fig. 5. Changes of mass loss versus W/C ratio in the customised model (Eq. (16)).

with ANN models which are made of complex black box of weight matrices and cannot be readily accessed by the user.

3.5. *Customised model*

As shown in previous sections, Eq. (13) is the general EPR model that includes all the mixture and admixture parameters and can accurately predict the degradation of concrete exposed to sulphuric acid. However it is also possible to use these models for the concretes that have been prepared with no admixtures or with only some of the admixtures. This can be done by adapting Eq. (13) when those admixture parameter(s) are equal to zero. The results of such evaluations lead to the generation of more concise and practical equations that include all the essential concrete ingredients. As an example, Eq. (13) is customised here for the case when no admixture is used, and pH value is equal to 0.6. The result of these adjustments is presented in Eq. (16).

$$\begin{aligned} ML = {} & -9.8 \times 10^{-7}(G)(W) + 1.9 \times 10^{-8}(G)^2\sqrt{(W)} - 3.2 \times 10^{-4} \\ & \sqrt{(C)(S)} - 2.6 \times 10^{-13}(G)^3\sqrt{(C)(W)} \\ & -5.2 \times 10^{-7}(C)(W)\sqrt{H} + 3.3 \times 10^{-5}C\sqrt{G} \end{aligned} \tag{16}$$

The customised Eq. (16) is a practical tool that can be used to evaluate the degree of deterioration of ordinary concretes exposed to sulphuric acid. The sensitivity analysis of Eq. (16) is examined for changes of W/C ratio which is known to be a key parameter in concrete mass loss due to sulphuric acid attack [32]. The result is shown in Fig. 5. It can be observed that Eq. (16) has successfully predicted the reduction in mass loss as the W/C ratio increases. This shows the reliability of the customised model in predicting concrete degradation.

4. Optimum mixture of concrete subject to sulphuric acid attack

From the results of the parametric study it is evident that different concrete ingredients may have different effects on the

Table 5
Optimum concrete mixture for minimum mass loss.

Parameter Unit	Cement (kg/m^3)	Gravel (kg/m^3)	Sand (kg/m^3)	Water (kg/m^3)	Superplasticizer (L/m^3)	W/C –	Mass loss (%)
Mix design	404.0	778.2	926.0	202.0	2.0	0.50	10.0
	447.0	745.0	926.0	201.1	1.5	0.45	11.2
	478.1	745.0	926.0	191.2	1.5	0.40	13.5
	513.9	745.0	926.0	179.9	1.5	0.35	16.3
	555.6	745.0	926.0	166.7	1.5	0.30	19.6
	571.0	794.4	926.0	142.8	1.5	0.25	24.4

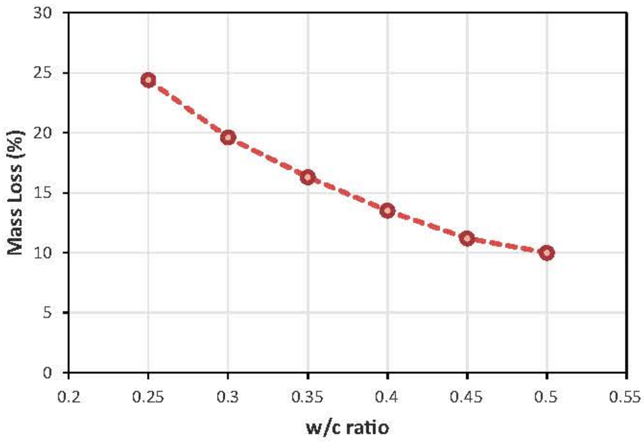


Fig. 6. The results of optimisation: minimum mass loss for different W/C ratios.

degradation of concrete. For example while increasing cement content will escalate the corrosion due to the mass loss, adding more water will help to reduce the concrete degradation. Therefore it is important to find a concrete mixture that can minimise the concrete degradation when it is exposed to sulphuric acid attack. In this section, using optimisation techniques and customised model (Eq. (13)), different optimum concrete mixtures to minimise degradation are obtained. Although only main concrete ingredients (i.e. cement, gravel, sand, water and superplasticizer) are optimised here, the technique can be extended to find both the optimum mixtures and admixtures in Eq. (13).

Eq. (16) was minimised using a nonlinear programming optimisation technique. Lower limits and upper limits of each variable in the equation were set based on the minimum and maximum values of those parameters in the dataset. A constraint was defined to ensure that the total volume of concrete is always equal to unit value during the optimisation process. This process was carried out several times for different values of W/C ratios. The results of this optimisation are presented in Table 5. From this table it can be concluded that the W/C ratio has a significant influence on the vulnerability of the concrete when encounter an acidic environment. This has also been reported by other researchers in previous studies [32]. The results show that it is possible to achieve a minimum 10% mass loss with a W/C value of 0.50 and the presented mix design. The relationship between W/C ratio and mass loss is also depicted in Fig. 6. While the W/C ratio is evidently a key role in the rate of degradation, the influence of other ingredients such as gravel and sand seems to be complex. This can be related to the nature of aggregate materials which are non-homogenous materials (unlike cement and water) as well as the effect of different types of aggregate which has different reaction in the vicinity of an acidic environment. Further investigation and experiments on various types of aggregate can help to understand its function in amount of the concrete degradation due to acid attack.

5. Summary and conclusions

Sulphuric acid attack is recognised as one of the main causes for concrete sewer pipe degradation. Degradation of sewer pipes results in reduction of pipe's wall thickness and the eventual breakdown of the system. The collapse of sewer systems can incur many financial and social problems.

In this article a new approach is presented for the prediction of degradation of concretes subject to sulphuric acid attack. Using a fairly comprehensive dataset from several acid attack experiments on various concrete mixtures and admixtures and a hybrid data mining technique (EPR), a model was developed and validated to predict the mass loss percentage of concrete when it is exposed to sulphuric acid. EPR integrates numerical and symbolic regression to perform evolutionary polynomial regression. The strategy uses polynomial structures to take advantage of their favourable mathematical properties. The developed EPR model presents 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 degradation and various contributing parameters.

The main feature of the EPR approach presented in this article is the possibility of getting more than one model for concrete degradation. The best model is chosen on the basis of simplicity and its performances on a test set of unseen data. For this purpose, the initial dataset is split into two subsets, (i) training and (ii) validation. The validation data set is not seen by EPR in the model construction phase and predictions provided by EPR models based on this data can be used as an unbiased performance indicator of generalisation capabilities of the proposed models. Another major advantage of the EPR approach is that, as more data becomes available, the quality of the prediction can be easily improved by retraining the EPR model using the new data.

A parametric study was conducted to evaluate the effect of the contributing parameters (i.e. concrete contents) on the predictions of the proposed EPR models. Combined effects of the parameters were also considered in the sensitivity analysis to investigate the interdependencies of parameters and their effect on the EPR predictions. The results show that the developed EPR models provide very accurate predictions for concrete degradation and are easy to use from a practical viewpoint. The results of the EPR model were compared with an ANN model and it was shown that the EPR model provided more accurate results on both training and validation datasets. In addition unlike ANN, EPR returns structured, transparent, concise and practical mathematical equations which allow the user to have a better understanding on the relationship between input and output parameters. Using the developed EPR models, a customised model was obtained in which it only includes the essential concrete contents (i.e. cement, gravel, sand, water and superplasticizer). The proposed EPR model was optimised in order to find the optimum concrete mixture that provides the maximum resistance against sulphuric acid attack. The results of the optimisation confirmed that, degradation or mass loss is highly dependent on water-cement ratio. When using the models developed by EPR or finding optimum solutions using the developed models, precautions should be taken as the models are only valid and reliable within the range of the data that has been used for training them. Any attempt to use these models outside the training range may lead to unreliable predictions and unexpected errors.

Acknowledgements

This research was funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC) grant number EP/1032150/1 (Assessing Current State of Buried Sewer Systems and Their Remaining Safe Life).

References

[1] A.M. Alani, A. Faramarzi, M. Mahmoodian, K.F. Tee, Prediction of sulphide build-up in filled sewer pipes, Environ. Technol. 35 (2014) 1721–1728.
[2] R.D. Pomeroy, J.D. Parkhurst, The forecasting of sulfide build-up rates in sewers, Prog. Water Technol. 9 (1977) 621–628.
[3] The problem of hydrogen sulphide in sewers, Clay Pipe Development Association (1976).
[4] A. Beeldens, D. Van Gemert, Biogenic sulphuric acid attack of concrete sewer pipes: a prediction of the corrosion rate, in: V.M. Malhotra (Ed.), Proceedings of the 5th CANMET/ACI International Conference on Recent Advances in Concrete Technology, Singapore; July 29–August 1, vol. SP-200, 2001, pp. 595–606.
[5] E.K. Attiogbe, S.H. Rizkalla, Response of concrete to sulfuric acid attack, ACI Mater. J. 84 (1988) 481–486.
[6] S. Ehrich, L. Helard, R. Letourneux, J. Willocq, E. Bock, Biogenic and chemical sulfuric acid corrosion of mortars, J. Mater. Civ. Eng. 11 (1999) 340–344.
[7] J. Montenya, N. De Beliea, E. Vinckeb, W. Verstraeteb, L. Taerwea, Chemical and microbiological tests to simulate sulfuric acid corrosion of polymer-modified concrete, Cem. Concr. Res. 31 (2001) 1359–1365.
[8] N. De Belie, J. Monteny, A. Beeldens, E. Vincke, D. Van Gemert, W. Verstraete, Experimental research and prediction of the effect of chemical and biogenic sulfuric acid on different types of commercially produced concrete sewer pipes, Cem. Concr. Res. 34 (2004) 2223–2236.
[9] Z. Chang, X. Song, R. Munna, M. Marosszeky, Using limestone aggregates and different cements for enhancing resistance of concrete to sulphuric acid attack, Cem. Concr. Res. 32 (2005) 1486–1494.
[10] E. Hewayde, M. Nehdi, E. Allouche, G. Nakhla, Neural network prediction of concrete degradation by sulphuric acid attack, Struct. Infrastruct. Eng. 3 (2007) 17–27.
[11] S. Pandey, D.A. Hindoliya, R. Mod, Artificial neural networks for predicting indoor temperature using roof passive cooling techniques in buildings in different climatic conditions, Appl. Soft Comput. 12 (2012) 1214–1226.
[12] S. Liu, J. Xu, J. Zhao, X. Xie, W. Zhang, An innovative method for dynamic update of initial water table in XXT model based on neural network technique, Appl. Soft Comput. 13 (2013) 4185–4193.
[13] J. Ghaboussi, J.H. Garrett, X. Wu, Knowledge-based modeling of material behavior with neural networks, J. Eng. Mech. 117 (1991) 132–153.
[14] J. Ghaboussi, D.A. Pecknold, M. Zhang, R.M. Haj-Ali, Autoprogressive training of neural network constitutive models, Int. J. Numer. Methods Eng. 42 (1998) 105–126.
[15] O. Giustolisi, D. Savic, A symbolic data-driven technique based on evolutionary polynomial regression, J. Hydroinf. 8 (2006) 207–222.
[16] O. Giustolisi, D. Laucelli, Increasing generalisation of input-output artificial neural networks in rainfall-runoff modelling, Hydrol. Sci. J. 50 (2005) 439–457.
[17] J.R. Koza, Genetic Programming On the Programming of Computers by Natural Selection, MIT Press, Cambridge, MA, 1992.
[18] V. Garg, V. Jothiprakash, Evaluation of reservoir sedimentation using data driven techniques, Appl. Soft Comput. 13 (2013) 3567–3581.
[19] M. Asadi, M. Eftekhari, M.H. Bagheripour, Evaluating the strength of intact rocks through genetic programming, Appl. Soft Comput. 11 (2011) 1932–1937.
[20] A. Doglioni, A Novel Hybrid Evolutionary Technique for Environmental Hydraulic Modelling, PhD Thesis, Technical University of Bari, Italy, 2004.
[21] A. Faramarzi, Intelligent computational solutions for constitutive modelling of materials in finite element analysis, PhD Thesis, University of Exeter, UK, 2011.
[22] O. Giustolisi, D.A. Savic, Advances in data-driven analyses and modelling using EPR-MOGA, J. Hydroinf. 11 (2009) 225–236.
[23] A. Ahangar-Asr, A. Faramarzi, A.A. Javadi, A new approach for prediction of the stability of soil and rock slopes, Eng. Comput. 7 (2010) 878–893.
[24] D. Savic, O. Giustolisi, L. Berardi, W. Shepherd, S. Djordjevic, A. Saul, Modelling sewer failure by evolutionary computing, Water Manag. 159 (2006) 111–118.
[25] Q. Xu, Q. Chen, W. Li, J. Ma, Pipe break prediction based on evolutionary data-driven methods with brief recorded data, Reliab. Eng. Syst. Safety 96 (2011) 942–948.
[26] A. Ahangar-Asr, A. Faramarzi, A.A. Javadi, O. Giustolisi, Modelling mechanical behaviour of rubber concrete using evolutionary polynomial regression, Eng. Comput. 28 (2011) 492–507.
[27] A. Fiore, L. Berardi, G. Carlo Marano, Predicting torsional strength of RC beams by using evolutionary polynomial regression, Adv. Eng. Softw. 47 (2012) 178–187.
[28] A. Faramarzi, A.A. Javadi, A.M. Alani, EPR-based material modelling of soils considering volume changes, Comput. Geosci. 48 (2012) 73–85.
[29] A. Faramarzi, A.A. Javadi, A. Ahangar-Asr, Numerical implementation of EPR-based material models in finite element analysis, Comput. Struct. 118 (2013) 100–108.
[30] A. Faramarzi, A.M. Alani, A.A. Javadi, An EPR-based self-learning approach to material modelling, Comput. Struct. 137 (2013) 67–71.
[31] E. Hewayde, Investigation on degradation of concrete sewer pipes by sulfuric acid attack, PhD Thesis, The University of Western Ontario, 2005.
[32] N.I. Fattuhi, B.P. Hughes, The performance of cement paste and concrete subjected to sulphuric acid attack, Cem. Concr. Res. 18 (1988) 545–553.