Original article

Genetic programming approach for predicting service life of tunnel structures subject to chloride-induced corrosion

Wei Gao *, Xin Chen, Dongliang Chen

Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, PR China

HIGHLIGHTS

- GP is used to predict service life of tunnel structure subject to chloride-induced corrosion.
- This new method can construct an explicit expression of the prediction model.
- This new prediction model can take into account 17 main corrosion factors.
- The performance of the new model is compared with that of artificial neural network model.
- The effects of two main controlling parameters are analyzed detailed.

GRAPHICAL ABSTRACT

A new method for predicting the service life of tunnel structures subject to chloride-induced corrosion using data from real engineering examples and genetic programming (GP) is proposed. As a data-driven method, the new approach can construct explicit expressions of the prediction model. The new method was verified by comparing it with the chloride-ion diffusion model considering eight corrosion influence factors. Moreover, 25 datasets collected from tunnel engineering examples were used to construct the new prediction model considering 17 corrosion influence factors belonged to just one classification of engineering corrosion factors. In addition, the performance of the new model was verified through a comparative study with an artificial neural network (ANN) model which is frequently used in chloride-induced corrosion prediction for reinforced concrete structures. The comparison revealed that both the computational result and efficiency of the GP method were significantly better than those of the ANN model. Finally, to comprehensively analyze the new prediction model, the effects of the two main controlling parameters (population size and sample size) were analyzed. The results indicated that as
Introduction

For complicated underground environments and concealed engineering, it is important to study the durability of tunnel structures. For a typical reinforced concrete structure, their durability is generally analyzed via methods used to study reinforced concrete structures. Generally, for a reinforced concrete structure, chloride-induced corrosion is harmful, and the cost required for its repair is high. Thus, in this study, the chloride-induced corrosion of a tunnel structure was investigated. When structural damage due to reinforcement corrosion becomes evident, deterioration occurs at a late stage, and it may be too late to take preventive or protective measures. Therefore, predicting the service life of tunnel structures according to the extent of damage due to corrosion is important [1].

According to the corrosion mechanisms of reinforced concrete structures [2], the initiation period, which corresponds to the process of chloride ions penetrating concrete, is the main corrosion stage. Thus, the period of the initial stage has been regularly used to specify the service life of reinforced concrete structures. The initiation period is the time taken for chloride-induced corrosion from the surface to be observed at the concrete cover depth [3]. In traditional methods [4,5], by using the chloride-induced corrosion depth computed via an empirical deterioration model, the service life of a reinforced concrete structure can be obtained. Many studies on the chloride-induced corrosion of reinforced concrete structures have employed the empirical deterioration model based on the chloride-ion diffusion method, e.g., the large research projects dealing with chloride-induced corrosion called “HETEK (High quality chloride-ion diffusion method, e.g., the large research projects dealing with chloride-induced corrosion of tunnels called “DARTS (Durable with chloride-induced corrosion is harmful, and the cost required for its repair is high. Thus, in this study, the chloride-induced corrosion of a tunnel structure was investigated. When structural damage due to reinforcement corrosion becomes evident, deterioration occurs at a late stage, and it may be too late to take preventive or protective measures. Therefore, predicting the service life of tunnel structures according to the extent of damage due to corrosion is important [1].

According to the corrosion mechanisms of reinforced concrete structures [2], the initiation period, which corresponds to the process of chloride ions penetrating concrete, is the main corrosion stage. Thus, the period of the initial stage has been regularly used to specify the service life of reinforced concrete structures. The initiation period is the time taken for chloride-induced corrosion from the surface to be observed at the concrete cover depth [3]. In traditional methods [4,5], by using the chloride-induced corrosion depth computed via an empirical deterioration model, the service life of a reinforced concrete structure can be obtained. Many studies on the chloride-induced corrosion of reinforced concrete structures have employed the empirical deterioration model based on the chloride-ion diffusion method, e.g., the large research projects dealing with chloride-induced corrosion called “HETEK (High quality chloride-ion diffusion method, e.g., the large research projects dealing with chloride-induced corrosion of tunnels called “DARTS (Durable and Reliable Tunnel Structures)” [6] and “CHLOROTECH (Resistance of Concrete to Chloride Ingress – From Laboratory Tests to In-field Performance)” [7], the computer program for predicting the service life of reinforced concrete called “Life-365” [8], and other studies dealing with chloride penetration for large reinforced concrete structures [9–14]. Moreover, the chloride-ion diffusion method has been employed in studies on the chloride-induced corrosion of tunnels, e.g., the large research projects dealing with chloride-induced corrosion of tunnels called “DARTS (Durable and Reliable Tunnel Structures)” [15] and “DuraCrete (Durability design of conCrete structures)” [16]. The empirical deterioration model is generally constructed by using the chloride-ion diffusion theory. For this type of model, some parameters must be determined according to experimental results. Moreover, to construct the model, assumptions must be made. To determine the parameters simply and to use them easily, for this type of model, the considered factors must be certain and few. Because the empirical deterioration model is constructed under ideal experimental conditions, this type of method cannot consider complicated engineering uncertainty factors [17]. Thus, to predict the chloride-induced corrosion of reinforced concrete structures, artificial intelligence methods based on data analyses were applied.

First, the chloride permeability of concrete was predicted using artificial neural networks (ANNs) and other methods (decision tree and hybrid methods). For example, in the study of Boğa et al. [18], with data from laboratory tests, the chloride permeability of concrete containing ground-granulated blast furnace slag and calcium nitrite-based corrosion inhibitors was predicted using a four-layered feedforward neural network considering four factors: the cure type, curing period, blast furnace slag ratio, and corrosion inhibitor ratio. Using laboratory test data, Ghafoori et al. [19] predicted the chloride permeability of self-consolidating concrete through a four-layered feedforward neural network. Six factors were considered: the cementitious material content, water-to-cementitious material ratio, coarse-aggregate content, fine-aggregate content, amount of air–entraining admixture, and high-range water reducer. In the study of Inthata et al. [20], by using the datasets of laboratory tests, the chloride permeability of concrete containing ground pozzolans was predicted by a four-layered feedforward neural network considering six factors: the water-to-binder ratio, percent replacement, testing age, pozzolans type, aggregate-to-cement ratio, and actual compressive strength. Yasarer and Najjar [21] used datasets from laboratory tests to predict the chloride permeability of Kansas concrete mixes via a three-layered feedforward neural network considering six factors: the oven dry weight, saturated surface dry weight, weight in water, curing time, specific gravity, and percentage of water absorbed. In another study, in contrast to the foregoing studies using ANNs, the chloride permeability of concrete was predicted via the decision-tree method considering three factors: the water content, cement content, and high-calcium fly ash content [22]. Moreover, a hybrid method using support vector regression and particle swarm optimization was applied to data samples from laboratory tests for predicting the chloride permeability of concrete containing meta-kaoilin; the model considered six factors: the test time, coarse-aggregate, fine-aggregate, meta-kaoilin, water, and cement contents [23].

Second, because the concentration of chloride ions and chloride diffusion coefficient significantly influence the chloride permeability of concrete, ANNs have been used to determine these factors as well; datasets from laboratory tests have also been used for this purpose. For example, Peng et al. [24] predicted the concentration of chloride ions of concrete using a three-layer cascade-correlation neural network that considered five factors: the contents of cement, fly ash, and microsilica; the volume of the calcium nitrite solution; and the test time. Moreover, a three-layered feedforward neural network [25] and a single adaptive linear network [26] were used to determine the chloride diffusion coefficient of concrete considering factors such as the water-to-binder ratio, unit content of cement, amount of concrete mix, and environmental conditions.

Third, more complicated chloride-induced corrosion problems have been analyzed through artificial intelligence methods based on the analysis of laboratory experimental data. For example, four-layered feedforward neural networks were used to determine the relationship between the chloride diffusion coefficient and the concrete mix design in carbonated and noncarbonated concretes. This study considered factors such as the water-to-binder ratio, the silica fume content, and experimental data. Thus, the influence of carbonation on the chloride penetration in concrete was investigated [27]. Furthermore, considering eight factors (the contents of ordinary Portland cement, fly ash, silica fume, water, superplasticizer, coarse aggregates, and fine aggregates, as well as the water-to-binder ratio), an ANN model was developed for predicting the durability of high-performance concrete expressed in terms of the total charge passed through the concrete over a 6-h period [28].

However, the aforementioned studies only analyzed the chloride-induced corrosion of concrete. ANNs have also been used to predict the service life of concrete. In a study by Hodhod and Ahmed [29], using experimental test data and results obtained from the error function solution to Fick’s second law of diffusion, a three-layer ANN whose inputs included values of the concrete
cover depth, apparent chloride diffusion coefficient, chloride threshold, and surface chloride concentration was applied to study the corrosion initiation time of slag concrete exposed to chlorides.

An analysis of the previous studies indicates that for the extremely complicated corrosion environments, the underground structures cannot be investigated using artificial intelligence methods. The ANN is the most commonly used artificial intelligence method, and only datasets from laboratory tests have been applied. However, the chloride-induced corrosion model constructed by an ANN is a network model that is described by an implicit expression. Because the ANN model is not an explicit expression and cannot be described by a mathematical function, to use this model, it must be trained beforehand by the data samples. Thus, the ANN model cannot be used directly, and this approach cannot be easily popularized. In this study, according to data from real engineering examples, genetic programming (GP) was used to predict the service life of tunnel structures subject to chloride-induced corrosion considering 17 engineering corrosion influence factors belonged to just one classification of engineering factors. An explicit expression for the service life of tunnel structures was obtained, which is a definite mathematical function. Thus, a new service-life model for tunnel structures using GP is proposed that can be easily popularized in real tunnel engineering. This study is a novel effort to predict the service life of tunnel structures subject to chloride-induced corrosion, and can be applied to more complicated chloride-induced corrosion with the consideration of more engineering corrosion influence factors.

Chloride corrosion mechanism and chloride corrosion factors of tunnel structures

For the service environment of tunnel structures, sources of chloride ions are widely available [30]. The sources include permeated highway deicing salt, natural chloride salt from salt lakes or saline-alkali soil, corrosive products from industrial activities, and salt from cement adulterating agents. In particular, for undersea tunnels, the main source is salt from seawater. Chloride ions in the tunnel environment can enter the tunnel structure in different manners, including diffusion, convection, penetration, adsorption, and electromigration. For different service environments of tunnel structures, the transmission modes of chloride ions in the tunnel structure are different, and transmission may occur via only one mode or a combination of several modes. The transmission methods of chloride ions for different service environments of tunnel structures are presented in Table 1.

As chloride ions entering a tunnel structure accumulate and attain the critical concentration, on the surface of reinforcing steel bars, the passive film of the reinforcing steel is destroyed owing to the depassivation ability of the chloride ions. Thereafter, the steel rusts. Because of the expansive force caused by the rusting of the steel, the lining structure of the tunnel cracks and breaks, and the service life of the tunnel structure is reduced. During the chloride corrosion process, chloride ions are one type of catalyst that can stimulate and accelerate the rusting and expansion of the reinforcing steel in a tunnel structure. According to the aforementioned analysis on chloride corrosion of tunnel structures, the chloride ions accumulating on the structure surface was the first step of chloride corrosion which was mainly one physical process and affected by the external environments of the tunnel engineering. The second step was the chloride entering the structure and transmission in the structure which was the main process of the chloride corrosion and was the physical and chemical process determined by the characteristics of the chloride ions and the structure cement. The last step was the structure destruction which was the chemical and mechanical process. This process was mainly determined by the mechanical characteristics of the tunnel structure, which were related with the loading and structure features. The chloride corrosion mechanism for tunnel structures is shown in Fig. 1.

As a typical reinforced concrete structure, the tunnel structure has the same basic chloride corrosion mechanism as other reinforced concrete structures. To avoid reiteration and consider the length restriction, the chloride corrosion mechanism of the tunnel structure is simply introduced here. The detailed chloride corrosion mechanism of reinforced concrete structures can be found in the literatures. According to the aforementioned chloride corrosion mechanism and previous studies [31], the main factors that affect the chloride corrosion of tunnel structures can be divided into four groups. The first group includes the external corrosion environment factors of the tunnel engineering, such as the average annual temperature, average annual relative humidity, and design speed of the tunnel, which correspond the temperature environment, humidity environment, and air flow environment. The second group includes the factors related with the characteristics of the chloride ions, such as the critical chloride-ion ratio, surface chloride-ion ratio, initial chloride-ion ratio, chloride diffusion coefficient, and chloride-ion binding capability. The third group includes the factors related with the characteristics of the structure cement, such as the water-cement ratio for the tunnel structure, thickness of the concrete cover for the structure, depth of the convective region, stray current intensity, and degradation coefficient of the lining structure. The last group includes factors which determine the mechanical characteristics of the tunnel structure, such as the rock mass grading of the tunnel, burial depth of the tunnel, lining thickness of the tunnel, and clear width of the tunnel inner diameter. To see clearly, the main factors that affect the chloride corrosion of tunnel structures are presented in Table 2.

### Construction of service-life prediction model for tunnel structures using GP

GP is an evolutionary computation algorithm that was first proposed by the American scholar Koza in the 1990s [32]. With GP, a
mathematical model to describe engineering phenomena can be constructed. Then, GP can be used as a suitable pattern-recognition method in many engineering fields [33]. Therefore, in this study, GP was used to construct a new service-life model for tunnel structures subject to chloride-induced corrosion. The computation process of this new method is as follows.

(1) Selection of the controlling parameters

To apply GP, certain controlling parameters must be determined by the user. These parameters include the reproduction, crossover, and mutation probabilities, the number of individuals, the termination error, and the maximum number of evolutionary generations. The aforementioned parameters affect the performance of GP. If an unfavorable choice of GP input parameters occurs, the convergent speed will become low and it may be not convergent. However, studies on the selection of the parameters indicate that they can be determined according to past experience and testing.

(2) Generation of the initial population

The initial population is generated according to the data samples. In GP, the individual describes one function form, which includes one function set $F = \{ f_1, f_2, f_3, \ldots, f_n \}$ (number is $n$), and one terminator set $M = \{ a_1, a_2, a_3, \ldots, a_m \}$ (number is $m$). The function set includes arithmetic symbols, mathematical functions, Boolean operations, condition expressions, and iterated functions. The terminator set includes variables, constants, and even function relationships. One individual was represented by a layering structure tree. As an example, a layering structure tree describing one function was shown in Fig. 2.

In this study, an individual represents one service-life prediction model for a tunnel structure subject to chloride-induced corrosion described by a mathematical function.

(3) Fitness computation

The fitness function of an individual was generally a type of transformation of its objective function. The fitness value computed by the fitness function was the criterion used to evaluate the quality of an individual. In this study, the fitness value is the error between the real service life and the computed value via

![Fig. 2. Layering structure tree describing one function](image)

the prediction model for the tunnel structure, which is described as follows:

$$F = \frac{1}{1 + \sum_{i=1}^{m} (T_{ri} - T_{ci})^2}$$

where $T_{ri}$ and $T_{ci}$ are the values of the real and computed service lives of the tunnel structure, respectively, whose units are years, $n$ is the number of samples.

(4) Evolutionary operators

a) Reproduction operator

Based on a reproduction probability determined by the user, one individual in the current generation was randomly selected to create a new individual for the next generation. This operation mimicked natural selection in the evolutionary process. The degree of selection reservation was determined by the reproduction probability. In this study, one service life prediction model is randomly selected to be used in the next iteration.

b) Crossover operator

There were two steps for this operator. First, two individuals in the current generation were randomly selected according to their selection probability. In this study, the generally used roulette-wheel selection method was used, using which the selection probability of one individual was described as follows:

$$P = \frac{f_i}{\sum_{i=1}^{M} f_i}$$

where $f_i$ is the fitness of individual $i$, and $M$ is the number of individuals in the group.

Second, based on the crossover probability determined by the user, two new individuals were created by exchanging certain parts of two former individuals. This operation mimicked the genetic recombination of the evolutionary process, and its occurrence level was determined by the crossover probability. In this study, certain parts of two service life prediction models were exchanged to create two new models to be used in the subsequent iteration.

c) Mutation operator

There were two steps for this operator as well. First, one individual in the current generation was randomly selected according to its selection probability. Second, based on the mutation probability determined by the user, a new individual was created by changing certain parts of the former individual. This operation

| Factors                                      | Symbol | Unit  |
|----------------------------------------------|--------|-------|
| The average annual temperature               | $T$    | °C    |
| The average annual relative humidity         | $RH$   | No    |
| The water-cement ratio for tunnel structure  | $W/C$  | No    |
| The thickness of concrete cover for structure | $c$    | Mm    |
| The clear width of tunnel inner diameter     | $W$    | M     |
| The depth of the convective region           | $x$    | Mm    |
| The stray current intensity                  | $S$    | mA    |
| The chloride ion binding capability          | $r$    | No    |
| The degradation coefficient of lining structure | $K$  | No    |
| The critical chloride ion ratio              | $C_r$  | % by mass of concrete |
| The chloride ion ratio of lining surface     | $C_s$  | % by mass of concrete |
| The initial chloride ion ratio               | $C_i$  | % by mass of concrete |
| The chloride diffusion coefficient           | $D_c$  | m²/s  |
| The rock mass grading of the tunnel          | $Q$    | No    |
| The burial depth of the tunnel               | $D_b$  | M     |
| The lining thickness of the tunnel           | $L$    | cm    |
| The design speed of the tunnel               | $V$    | km/h  |

Table 2
Main factors affecting the chloride corrosion of tunnel structures.
GP was used for engineering applications, its performance was verified by comparison with theoretical models. Here, for simplicity, the three chloride-ion ratios \( C_i, C_c, \) and \( C_s \) are represented by one synthesized coefficient \( \beta \), which is expressed as follows:

\[
\beta = \frac{C_c - C_i}{C_i - C_s}
\]

(3)

where \( \beta \) is the synthesized coefficient of chloride-ion ratios.

The complicated chloride-ion diffusion model can consider eight factors: the average annual temperature \( (T) \), average annual relative humidity \( (RH) \), water-to-cement ratio of the structure \( (W/C) \), thickness of the concrete cover for the structure \( (c) \), load presented by the burial depth \( (D_b) \), chloride-ion binding capability \( (r) \), degradation coefficient of the structure \( (K) \), and synthesized coefficient of the chloride-ion content \( (\beta) \). GP was used to predict the chloride corrosion life considering the eight influencing factors to verify the performance of the method under complicated conditions. Thus, the function set was \( F = \{+, -, \times, \div, \text{power, ln, exp, sin, cos} \} \), and the terminator set was \( M = \{T, RH, W/C, c, x, r, K, \beta\} \). For verification, 20 samples considering the eight influencing factors randomly determined by the authors beforehand, whose service lives were determined via the traditional chloride ion diffusion model, were used, as shown in Table 3.

To obtain the service lives of the data samples in Table 3, the chloride-ion diffusion model, including the modification of the diffusion coefficient and considering several factors, was used in this study [34–38]. The model is defined as follows:

\[
L^{1-a} = \left[ \frac{C}{2\sqrt{[R horror formula]} f_0} \right]^{2} (4)
\]

where \( L \) is the service life; \( a \) is the influence coefficient of the concrete age, which can be described as \( 3(0.55 - W/C) \); and \( f \) is the influence coefficient of the load, which can be determined as follows. If \( D_b > 500 \), it is 1.1; if \( 500 > D_b > 200 \), it is 1.0; if \( 200 > D_b > 50 \), it is 0.9; and if \( 50 > D_b \), it is 0.8. \( D_b \) and \( t_0 \) are the benchmark content and time, respectively, and are given as \( D_b = 10^{-12.06 + 2.4 W/C} \) (m/s) and \( t_0 = 28 \) (days); \( f_0 \) is the influence coefficient of the relative humidity, which can be expressed as \( \left[ \frac{1 + \left( RH_c - 0.75 \right)}{1 - 0.75} \right]^{-1} \); \( RH_c = 0.75 \) is the critical relative humidity; \( f_r \) is the influence coefficient of the temperature, which is expressed as \( f_r = 15C \); \( T_0 = 145 \) (°C) is the benchmark temperature; and \( q \) is the cement activation constant, which is related to W/C and given as \( q = 3474.28 - 159.28 W/C - 0.49 \). The other parameters can be found in the foregoing sections.

The theoretical model is a complex chloride-ion corrosion model that considers eight factors and employs the basic chloride-ion diffusion principle. Thus, this model can only analyze the basic chloride-ion corrosion process and cannot be used in real engineering applications. The objective of this study was to verify the feasibility of GP for constructing a complex chloride-ion corrosion model. The lives computed using Eq. (4), which are presented in Table 3, were regarded as the real lives. According to experience and tests, the main parameters of GP were determined as follows: the number of individuals was 100, the reproduction probability was 0.1, the crossover probability was 0.9, the mutation probability was 0.05, and the maximum number of evolutionary generations was 100.

According to research experience, for the artificial intelligence method, the training samples were selected from the whole set of data samples randomly, and their number was generally 70%–80% of the total number of samples. The remaining samples were the testing samples. Because the data samples in Table 3 were randomly selected, for simplicity, the first 15 samples (called training samples) in Table 3 were used to construct the prediction model,
Table 3
Data samples used to construct the prediction model considering the eight factors.

| Number | T     | RH   | W/C  | c    | Db   | r   | K   | β   | Ta   |
|--------|-------|------|------|------|------|-----|-----|-----|------|
| 1      | 22.00 | 0.85 | 0.34 | 40   | 654  | 0.88| 10.0| 0.8 | 53.92 |
| 2      | 21.50 | 0.82 | 0.35 | 60   | 763  | 0.90| 9.0 | 0.5 | 67.88 |
| 3      | 20.30 | 0.79 | 0.34 | 50   | 952  | 0.78| 14.0| 0.7 | 62.87 |
| 4      | 18.00 | 0.78 | 0.33 | 60   | 550  | 0.77| 14.0| 0.7 | 113.55|
| 5      | 12.10 | 0.75 | 0.29 | 65   | 753  | 0.89| 12.5| 0.6 | 72.71 |
| 6      | 11.20 | 0.49 | 0.25 | 45   | 98   | 0.20| 24.0| 0.4 | 73.04 |
| 7      | 15.30 | 0.75 | 0.36 | 25   | 632  | 0.30| 18.0| 0.8 | 81.95 |
| 8      | 18.90 | 0.70 | 0.34 | 25   | 356  | 0.32| 18.5| 0.8 | 94.04 |
| 9      | 17.70 | 0.78 | 0.35 | 30   | 993  | 0.89| 15.5| 0.8 | 86.86 |
| 10     | 13.50 | 0.55 | 0.28 | 35   | 340  | 0.25| 25.0| 0.5 | 89.05 |
| 11     | 22.18 | 0.79 | 0.40 | 50   | 55   | 0.40| 19.0| 0.7 | 109.84|
| 12     | 18.50 | 0.82 | 0.40 | 52   | 63   | 0.57| 18.0| 0.7 | 105.08|
| 13     | 17.50 | 0.85 | 0.38 | 50   | 44   | 0.44| 19.0| 0.8 | 111.61|
| 14     | 24.15 | 0.84 | 0.38 | 50   | 320  | 0.70| 16.0| 0.7 | 84.86 |
| 15     | 9.80  | 0.70 | 0.30 | 53   | 55   | 0.37| 18.0| 0.5 | 48.16 |
| 16     | 10.30 | 0.89 | 0.38 | 70   | 65   | 0.35| 20.0| 0.6 | 92.21 |
| 17     | 11.00 | 0.84 | 0.35 | 65   | 452  | 0.60| 15.0| 0.6 | 33.34 |
| 18     | 23.20 | 0.82 | 0.38 | 65   | 551  | 0.79| 19.0| 0.6 | 59.96 |
| 19     | 11.50 | 0.88 | 0.36 | 50   | 650  | 0.70| 16.0| 0.7 | 46.84 |
| 20     | 22.00 | 0.85 | 0.39 | 55   | 85   | 0.47| 18.0| 0.7 | 109.43|

![Fig. 4. Evolutionary process of the prediction model considering the eight factors.](image)

and the last five (called testing samples) were used to verify the model. Using data from the training samples in Table 3, the GP model was trained. The evolutionary processes of the best and average fitness were shown in Fig. 4.

As shown in Fig. 4, the best and average fitness increased rapidly and stabilized as the evolutionary generation increased. Therefore, the search process of GP was satisfactory.

After the search process of GP was completed, the best individual was selected as the final result. Then, the suitable prediction model considering eight influence factors was obtained as follows:

\[
t = 0.00033 \times [(101.15\beta - 4.113) \times (2.01 + \sin2.001c) \times (4x \times (101.15\beta - 2)) + 5.5 \times T \times K - \sin10r \times K \times |\sin(101.15\beta - 2) + 4Db| \\
\times (100R + 335.01 \times \frac{W}{C} + 0.035)]
\]

(5)

where \( t \) is the life, and the other parameters can be found in the foregoing sections.

Using the above prediction model, the computed lives were compared with the supposed real lives (theoretically calculated by Eq. (4)) for the training samples, as shown in Table 4.

From Table 4, the computed results correlated well with the supposed real lives, and the average relative error was 0.13. Thus, for the training samples, the computational performance of the prediction method using GP was suitable.

To verify the prediction method using GP, the lives of the five testing samples were computed. Comparisons of the computed lives and supposed real lives of the testing samples are shown in Fig. 5.

As shown in Fig. 5, for the five testing samples, the lives computed using the prediction method agreed well with the supposed real lives. That is, using the prediction method employing GP, a suitable prediction of the service life considering complicated chloride corrosion conditions was obtained. Thus, GP can be used to predict the service life of tunnel structures considering complicated chloride corrosion conditions.

**Engineering application of new chloride corrosion life prediction method**

**Prediction of chloride corrosion life for real tunnel structures**

According to the foregoing verification study, the new method can be employed to accurately predict the chloride corrosion life of tunnel structures using only data samples. Thus, the new method was used to predict the service life of a real tunnel structure subject to chloride-induced corrosion. To predict the service life of a real tunnel structure subject to chloride-induced corrosion using GP, typical engineering examples were collected as data samples. In this study, according to the main influence factors listed in Table 2, 25 tunnel engineering examples involving a chloride corrosion environment collected from the related literatures were applied, as summarized in Table 5.

Based on the engineering examples in Table 5, a chloride corrosion life prediction model considering all 17 main influence factors (\( T, RH, W/C, c, x, S, r, K, C_c, C_i, D_0, D_w, W, L, Q, \) and \( V \)) was constructed. Here, the function set was \( F = \{+, -, \times, \cdot, \div, \text{pow}, \ln, \exp, \sin, \cos\} \), and the terminator set contained 17 influence factors. The first 20 samples (called training samples) were used to construct the prediction model. The last 5 samples (called testing samples) were used for prediction. Based on experience and tests, the main parameters of GP were determined as follows: the number of individuals was 300, reproduction probability was 0.1, crossover probability was 0.9, mutation probability was 0.05, \( \varepsilon \) was \( 10^{-3} \), and the maximum number of evolutionary generations was 200.

Data were collected from references supplied as Suppl. Materials.
Comparison for training samples by the prediction model considering the eight factors.

| Number | Real life | Computing life | Absolute error | Relative error |
|--------|-----------|----------------|----------------|----------------|
| 1      | 53.92     | 64.59          | 10.67          | 0.1980         |
| 2      | 67.88     | 79.59          | 11.71          | 0.1725         |
| 3      | 62.87     | 70.49          | 7.62           | 0.1212         |
| 4      | 113.55    | 117.69         | 4.14           | 0.0364         |
| 5      | 72.71     | 64.97          | 7.74           | 0.1065         |
| 6      | 73.04     | 61.70          | 11.34          | 0.1552         |
| 7      | 81.95     | 95.65          | 13.70          | 0.1672         |
| 8      | 94.04     | 103.90         | 9.87           | 0.1049         |
| 9      | 86.86     | 77.09          | 9.77           | 0.1125         |
| 10     | 89.05     | 98.40          | 9.36           | 0.1051         |
| 11     | 109.84    | 121.04         | 11.20          | 0.1020         |
| 12     | 105.08    | 94.68          | 10.40          | 0.0990         |
| 13     | 111.61    | 123.96         | 12.35          | 0.1106         |
| 14     | 84.86     | 93.09          | 8.23           | 0.0970         |
| 15     | 48.16     | 37.72          | 10.44          | 0.2169         |
| Average| —         | —              | 9.50           | 0.13           |

As indicated by Table 6, the computed results obtained using the new prediction model agreed well with the real lives, and the average relative error was small (0.0366). Therefore, the computation performance of the proposed chloride corrosion life prediction model was satisfactory.

Moreover, to verify the prediction performance of the model, the computed results for the five testing samples and their corresponding real lives are shown in Fig. 7.

As indicated by Fig. 7, the prediction results of the proposed model agreed well with the real lives.

According to the foregoing analyses, GP can be employed to predict the chloride corrosion life of tunnel structures using only engineering samples, and the prediction ability of this method is satisfactory.

Comparison study of prediction method via ANN

Because the ANN has been commonly used for the prediction of the chloride-induced corrosion of concrete in previous studies and is a data-based method that is similar to GP, to obtain a fair comparison, an ANN was used to predict the chloride corrosion life for a tunnel structure in this study. To validate the performance of the proposed GP method, the prediction results obtained using an ANN were compared with those of the GP method.

Through experience and testing, the construction of the ANN was selected as 17-27-1, which describes the number of neurons in the input, hidden, and output layers, respectively. The ANN was trained using the modified backpropagation (BP) algorithm. Thus, the ANN was a three-layer feedforward BP network. Its computation parameters were as follows: the iterating step and inertia parameter were 0.21 and 0.91, respectively; the termination error and maximum iteration number were set as 10⁻⁴ and 1000, respectively. Although these parameters affect the performance of the ANN, according to previous studies on their selection, they are generally determined via experience and testing.

For comparison, similar to the aforementioned studies for the data samples, the first 20 samples were selected as training samples and were used to construct the prediction model. The last five samples were the testing samples used to verify the model. The training samples were used to construct the ANN model, and the termination error for the ANN model was 0.0041. To compare the computational effect between the two methods (ANN and GP), the average relative errors of the training and testing samples for the two methods were calculated, as shown in Table 7. For a fair comparison, the computational efficiency of the two methods

Fig. 5. Comparison results for the testing samples by the prediction model considering the eight factors.

Using data from the training samples in Table 5, the GP algorithm was trained. The evolutionary processes of the best fitness and average fitness were shown in Fig. 6.

As indicated by Fig. 6, the evolutionary processes of the best fitness and average fitness approached stability as the evolutionary generation increased. Thus, the stability of the evolutionary process for the life prediction method considering 17 influence factors was satisfactory.

After the search process of GP was over, the best individual was selected as the final result. Then, the chloride corrosion life prediction model considering 17 influence factors could be obtained as follows:

\[
t = 54.0878 \times \left( \frac{\sin(X \cdot 1.8) - e^{0.094} - e^{2.25} - \sin \left( \frac{W \cdot 5.9}{9.1} + \cos \left( \frac{21.31}{17.24} \right) \right)}{\eta} \right) + 54.5352
\]

where

\[
\eta = e^{0.08} + e^{0.25} + e^{0.9} - \ln(0.49 - 0.49 \sin \left( \frac{W \cdot 0.32}{0.25} + \cos \left( \frac{X \cdot 24.2}{0.0586} \right) + \sin \left( \frac{3.7}{17.24} \right) - \ln \left( \frac{3.7}{17.24} \right) + \frac{0.15}{0.02} \right) + 0.9321 + 0.0952}
\]

\[
t \text{ is life, and the other parameters can be found in the foregoing sections.}
\]

To verify the computational performance of the new chloride corrosion life prediction model, the computed and real lives of the tunnel structures for the training samples in Table 5 were compared, as shown in Table 6.
Tunnel engineering examples in a chloride corrosion environment. The proposed method was better than that of the ANN method. Therefore, the computation effect of the GP method, the computation errors for the training and testing samples were approximately 2.2 and 93 times shorter, respectively, than those for the ANN method. Thus, the computation time of the GP method was significantly shorter than that of the ANN method; i.e., the computational efficiency of the GP method was far higher than that of the ANN method.

Furthermore, because the prediction model based on the GP method using the training samples employed an explicit mathematical expression, when a data sample was substituted into this function, the result could be computed directly. The computation process of the explicit mathematical expression was very easy and fast. Consequently, the computation speed of the prediction model using the GP method for the testing samples was very fast. However, the prediction model using the ANN method is a network model that cannot be described by mathematical expressions.

Table 5
Tunnel engineering examples in a chloride corrosion environment.

| Number | T (℃) | RH | W/C | c/mm | W/m | x/mm | S/mm | r | K | C |
|--------|-------|-----|-----|------|-----|------|------|---|---|---|
| 1      | 22.00 | 0.85 | 0.34 | 40   | 15.00 | 5.4  | 0    | 0.88 | 10.0 | 0.150 |
| 2      | 23.00 | 0.78 | 0.35 | 64   | 15.60 | 11.5 | 0    | 0.70 | 14.0 | 0.180 |
| 3      | 22.60 | 0.79 | 0.32 | 61   | 10.18 | 12.3 | 0    | 0.78 | 8.5  | 0.280 |
| 4      | 21.50 | 0.82 | 0.35 | 60   | 14.50 | 4.1  | 30   | 0.90 | 9.0  | 0.210 |
| 5      | 20.30 | 0.79 | 0.34 | 50   | 11.25 | 3.2  | 0    | 0.78 | 14.0 | 0.670 |
| 6      | 18.00 | 0.78 | 0.33 | 60   | 11.50 | 7.7  | 0    | 0.77 | 14.0 | 0.280 |
| 7      | 12.10 | 0.75 | 0.29 | 65   | 13.50 | 13.5 | 0    | 0.89 | 12.5 | 0.230 |
| 8      | 11.20 | 0.49 | 0.25 | 45   | 10.50 | 1.8  | 0    | 0.20 | 24.0 | 0.170 |
| 9      | 23.20 | 0.83 | 0.34 | 60   | 6.50  | 9.8  | 30   | 0.89 | 8.0  | 0.150 |

The table indicates that the average relative errors of the two methods were lower than 10%. Thus, the two methods can be used to predict the chloride corrosion life of tunnel structures using only engineering data samples. Their computed results were all suitable. Moreover, for the two methods, the average relative errors of the training samples were always lower than those of the testing samples. However, for the training and testing samples, the average relative errors of the GP method (3.66% and 6.7%, respectively) were smaller than those of the ANN method (4.41% and 9.23%, respectively)—particularly for the testing samples. For the GP method, the computation errors for the training and testing samples were approximately 20% and 38% smaller, respectively, than those for the ANN method. Therefore, the computation effect of the proposed method was better than that of the ANN method.

The computation time of the ANN method for the training samples was approximately 5.4 min, and that for the testing samples was approximately 0.94 min. The computation time of the GP method for the training samples was approximately 2.42 min, and that for the testing samples was approximately 0.01 min. For the GP method, the computation times for the training and testing samples were approximately 2.2 and 93 times shorter, respectively, than those for the ANN method. Thus, the computation time of the GP method was significantly shorter than that of the ANN method; i.e., the computational efficiency of the GP method was far higher than that of the ANN method.

Note: in Table 5, Tri is the value of the real service lives of the tunnel structure, whose unit is year.
Thus, the computation speed of the ANN model was significantly lower than that of the GP method.

Considering the computational effect and efficiency, the prediction model using the GP method was better than the ANN model for determining the suitable service life of tunnel structures subject to chloride-induced corrosion. Moreover, the prediction model using the GP method was an explicit mathematical expression. Thus, the new method using GP can be very simply and easily applied.

Discussion

As a typical evolutionary algorithm, there were certain controlling parameters to be determined for GP. These parameters significantly affected the performance of GP. To provide guidance for the determination of the parameters in this study, the effects of two main parameters (population size and sample size) on the performance of the chloride corrosion life prediction model constructed in the foregoing section were analyzed in detail. The sample size is an important parameter for collecting suitable data, which requires a considerable amount of effort in real engineering applications. Thus, it is important to determine the suitable sample size for a data-based method. Moreover, for GP, the population size is an important parameter that controls the computation time and results. For complicated engineering problems, the computation time is an important influence factor. Therefore, in this study, these two important parameters were analyzed.

Population size

As the population size increases, the computational accuracy increases, while the computational efficiency decreases; i.e., if the population size is small, the algorithm may converge to a local extreme very quickly, and the computed result is poor. In contrast,
if the population size is large, the computational expense increases significantly, and the computational efficiency is very low. Thus, the effects of the population size on the computational accuracy and efficiency were comprehensively analyzed, as follows.

(1) Effect on the computing accuracy

With different population sizes (20, 40, 80, 120, 180, 240, and 480), the average computation errors for the training and testing samples are presented in Table 8.

As indicated by Table 8, for the training and testing samples, the influence laws were similar. As the population size increased, the computation error decreased, and the extent of the decrease slowed. Moreover, when the population size was small, the relative errors of the training and testing samples were different, and their difference was large; i.e., when the population size was small, the difference between the approach and prediction errors was large. In contrast, when the population size was large, the difference between the approach and prediction errors was small; i.e., the errors were similar. Therefore, for a large population size, the relative errors of the training and testing samples were similar; i.e., the approach and prediction errors of the new prediction model using GP were similar. According to the above analysis, with regard to the computation error, a larger population size yields better performance of the GP prediction model.

To analyze the effect of the population size on the computation error more clearly, the relationship between the absolute errors of the training and testing samples and the population size was examined, as shown in Fig. 8.

The figure indicates that as the population size increased, its effect on the computation error decreased significantly, according to a power function. Therefore, the method for improving the computation error by increasing the population size had an application range and was only suitable for a small population size; for a large population size, increasing it further was not effective.

(2) Effect on the computing efficiency

To evaluate the computational efficiency of the chloride corrosion life prediction model with different population sizes (20, 40, 80, 120, 180, 240, and 480) more thoroughly, the effect of the number of objective-function evaluations during the evolutionary process, which is denoted by NOF and represents the computational efficiency for optimization algorithms [39], was studied.

The relationship between the NOF and the population size is shown in Fig. 9.

The figure indicates that as the population size increased, the NOF increased rapidly. The two parameters had an exponential relationship; i.e., as the population size increased, its effect on the NOF increased. Moreover, when the population size was larger than approximately 300, the NOF increased sharply. Furthermore, Fig. 9 shows that for a small population size, the effect of the population size on the computational efficiency was small. However, as the population size increased, its effect on the computational efficiency increased significantly. The computational efficiency of the new prediction model using GP decreased considerably with the increase in the population size. There was an optimal value for the population size; approximately 300. Therefore, considering only the computational efficiency, a smaller population size yields better performance of the prediction model using GP.

According to the comprehensive analysis of the computational accuracy and efficiency, in this study, the optimal population size was determined to be 300.

Sample size

As the sample size increased, the computed results improved. However, it was very difficult to collect suitable engineering examples. In particular, for complicated chloride corrosion problems of

| Population size | Training samples | Testing samples |
|-----------------|------------------|-----------------|
|                 | Absolute error   | Relative error  | Absolute error | Relative error |
| 20              | 15.44            | 0.2744          | 14.61          | 0.2974         |
| 40              | 8.73             | 0.1973          | 9.12           | 0.1944         |
| 80              | 2.85             | 0.0996          | 3.16           | 0.1104         |
| 120             | 1.37             | 0.0745          | 2.62           | 0.0803         |
| 180             | 1.69             | 0.0272          | 1.36           | 0.0501         |
| 240             | 0.84             | 0.0201          | 0.93           | 0.0219         |
| 480             | 0.63             | 0.0197          | 0.88           | 0.0193         |
tunnel structures, it was almost impossible to collect several engineering samples. Therefore, for the chloride corrosion life prediction method using GP, it was crucial to determine the minimal sample size. Because the sample size significantly affected the computational accuracy and only slightly affected the computational efficiency, only its effect on the computational accuracy was analyzed in this study.

For the training and testing samples, the average computation errors for different sample sizes (5, 10, 15, and 20) are presented in Table 9.

As indicated by Table 9, the influence laws were similar for the training and testing samples. As the sample size increased, the computation error decreased, and the extent of the decrease declined. In addition, the computation errors of the training and testing samples were all different, regardless of the sample size. The analysis results indicate that for a small sample size, the influence of the sample size on the computational accuracy was notable; however, for a large sample size, the influence of the sample size decreased. Thus, a larger sample size yielded better performance of the prediction model using GP. However, when the sample size increased to a certain value, the method for improving the computational accuracy by increasing the sample size was not effective.

To analyze the effect of the sample size on the computational accuracy in more detail, the relationship of the absolute errors of the training and testing samples with the sample size is shown in Fig. 10.

The figure shows that for a small sample size (<15), as the sample size increased, the computation error decreased rapidly; i.e., the computational accuracy improved rapidly. When the sample size was in the range of 15–20, the computation error decreased gradually. Moreover, when the sample size was approximately 20, the absolute errors for the two types of samples changed gradually and slightly. The extreme points for the two functions were approximately 20. Thus, for a sample size of 20, the computational accuracy improved very slowly or did not improve. Therefore, for this study, the optimal sample size was 20.

### Table 9

| Sample size | Training samples |          | Testing samples |          |
|-------------|------------------|----------|-----------------|----------|
|             | Absolute error   | Relative error | Absolute error | Relative error |
| S           |                  |           |                 |           |
| 5           | 20.03            | 0.4134   | 19.32           | 0.4427   |
| 10          | 15.72            | 0.2611   | 15.27           | 0.2753   |
| 15          | 9.94             | 0.1782   | 11.13           | 0.1634   |
| 20          | 8.05             | 0.1516   | 9.78            | 0.1179   |

### Conclusions

Using data collected from real engineering examples of 25 tunnels subject to chloride-induced corrosion, a method for predicting the service life of tunnel structures via the data-based method of GP was proposed. The explicit expression of the service-life prediction model for these tunnel structures can be obtained. For 20 randomly selected data samples, using the chloride-ion diffusion model considering eight corrosion influence factors, the effectiveness of the proposed method was verified. Finally, using data from tunnel engineering examples, a prediction model considering 17 corrosion influence factors for a real engineering environment using GP was constructed. Its computational effect was verified with data from other tunnel engineering examples. Moreover, the suitable performance of the proposed prediction model was verified by a comparative study with an ANN model. The results indicated that both the computational effect and efficiency of the GP method were significantly better than those of the ANN method. Moreover, the prediction model using GP consisted of an explicit mathematical expression, whereas that using the ANN method was a network model. Thus, the GP model could be applied more easily. Therefore, GP is a suitable method for predicting the service life of tunnel structures subject to chloride-induced corrosion.

To comprehensively analyze the proposed method, the effects of two main controlling parameters (population size and sample size) on the performance of the prediction model considering 17 corrosion influence factors were analyzed in detail. The results indicated that as the population size increased, the computation error decreased, and the extent of the decrease slowed. As the population size increased, its effect on the computation error decreased significantly. Furthermore, as the population size increased, the computational efficiency decreased sharply. When the population size was larger than approximately 300, the computational efficiency decreased sharply. According to a comprehensive analysis of the computational accuracy and efficiency of the new prediction model, the optimal population size was determined to be 300. Moreover, as the sample size increased, the computation error decreased, and the extent of the decrease declined. When the sample size was approximately 20, the absolute error for the data samples changed gradually and remained almost unchanged. Thus, the optimal sample size was 20 for the new prediction model.

### Conflict of interest

The authors have declared no conflict of interest.

### Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

### Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2019.07.001.
