

Research Article

Improved Hybrid Model for Predicting Concrete Crack Openings Based on Chaos Theory

Yao Xu ¹, Yaoying Huang ², Xiaofeng Xu ², and Fang Xiao ¹

¹State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin,
China Institute of Water Resources and Hydropower Research, Beijing 100038, China

²College of Hydraulic and Environmental Engineering, China Three Gorges University, 443002 Yichang, China

Correspondence should be addressed to Yaoying Huang; huangyaoying@sohu.com

Received 18 November 2021; Revised 28 December 2021; Accepted 5 January 2022; Published 31 January 2022

Academic Editor: Abílio De Jesus

Copyright © 2022 Yao Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conventional statistical models provide inaccurate predictions of concrete crack openings because they do not consider the nonlinear temperature response and the residual characteristics of concrete. To address this problem, this study introduces a nonlinear temperature factor and develops an improved statistical model of crack openings. The chaotic characteristics of residual time series of the improved statistical model are analyzed based on chaos theory and phase-space reconstruction theory. These theories are integrated with back-propagation (BP) artificial neural networks and genetic algorithms (GAs) to establish a GA-BP neural network model for predicting residuals. Finally, a hybrid model is developed for predicting the concrete crack opening behavior. The predictions of the conventional statistical model, the statistical model considering nonlinear temperature component, and the hybrid model are compared using the case study on the crack openings of a regulating sluice. The results show that the proposed hybrid model in this study for predicting concrete crack openings is significantly more accurate than the conventional statistical model and the statistical model considering nonlinear temperature component.

1. Introduction

Cracks are a common structural defect in hydraulic concrete. The operational behavior of macroscopic cracks in hydraulic concrete structures is directly related to the safety of the structure and is an important indicator of structural stability. To determine the operational behavior of concrete crack openings, the momentum misplacement is often monitored in real time by installing a crack meter group. Given the conditions of practical projects, concrete cracks are affected by various complex factors, including water pressure, temperature, and time. Hence, it is vital to establish an accurate mathematical model to aid in monitoring concrete crack openings and to quantitatively analyze, evaluate, and predict their behavior.

Concrete cracks fall into the category of concrete deformation. At present, mathematical models based on conventional statistical models of concrete deformations are widely used in engineering monitoring and parameter

inversion [1]. Such models usually include temperature component, hydraulic component, and aging component. Experiments demonstrate that temperature is crucial to control concrete deformation [2, 3] and that variations in temperature are an important factor in determining the opening and closing of concrete cracks [4]. The temperature component is thus considerably sensitive to the temperature variations [1]. Although the hydraulic and aging components of conventional statistical deformation models generally fit well to experimental results, the temperature component is underfitted because conventional statistical models [5, 6] generally consider only the linear response of the system to temperature but neglect the complex nonlinear relationship between the actual affected variables and environmental factors in practical engineering. Previous studies show that variations in temperature are related nonlinearly to concrete deformation, and including such nonlinearities in prediction models should provide more accurate results than conventional models [7, 8].

To further improve the accuracy of statistical models, chaos theory has recently been harnessed to analyze and predict the residual series of statistical models, effectively establishing a hybrid model that provides more accurate predictions than the statistical models. For instance, Wei et al. [9] established a back-propagation autoregressive integrated moving-average model based on the wavelet analysis that directly predicts chaotic residual time series. The nonlinear method predicts the residual, which is generated by a phase-space reconstruction, with the result being a hybrid model based on a conventional statistical model [10, 11]. The results show that hybrid models based on residual prediction are more accurate than conventional models. Bao and Wu [10] improved the mathematical model for monitoring concrete cracks by combining chaos theory with a conventional statistical model to establish a hybrid model of concrete crack openings. The results indicate that the stress field at the crack tip generates singularities when exposed to water pressure and temperature variations, which irreversibly deforms the concrete at the crack tip. Hence, crack openings also involve an irrecoverable part caused by nonlinear factors, suggesting that conventional statistical models cannot predict the deformation behavior as a function of water pressure and temperature. This part of the deformation contains significant information about the crack evolution process, which is usually described in terms of a residual series, that is, a chaotic residual time series. Hence, it is necessary to analyze the residual time series of statistical models of concrete crack openings.

At present, there are some reports on the prediction methods of time series [12]. The studies show that methods based on phase-space reconstruction are commonly used to predict chaotic time series. The essence of this approach is to transform the extension of time series into the interpolation of phase space. Based on the Takens embedding theorem [13], the optimized delay time and minimum embedding dimension are determined according to the actual time series. The one-dimensional (1D) time series can then be reconstructed into a phase space with the same topological meaning as the motive power system. A reasonable selection of the delay time and embedding dimension then allows reconstruction of the phase space with satisfactory adaptability, thereby producing the internal nonlinear mapping. By exploiting the properties of the Lyapunov exponent, this approach can recognize and make short-term predictions of chaotic time series [14, 15]. Currently, the methods for determining optimized delay time and minimum embedding dimension depend on the amount of data they are fed, the amount of computation, the anti-noise performance, and the objectivity with which the parameters are selected. The methods available for selecting optimized delay time typically include the autocorrelation coefficient method [16], the mutual information method [17], and the average displacement (AD) method [18], and the methods used to calculate the minimum embedding dimension typically include the geometric invariant method [14], the pseudo-neighbor method [19], and the Cao method [20].

Casdagli [21] proposed the basic principle for predicting chaotic time series, which was used to develop prediction methods involving artificial neural networks [9, 22, 23]. BP neural networks are suitable for nonlinear mapping, making them good candidates for predicting chaotic time series. However, single neural networks converge slowly, resulting in local minima and overfitting. Ding et al. [24] proposed GA-BP neural networks and demonstrated that genetic algorithms (GAs) provide obvious improvements over BP neural networks. Hence, phase-space reconstruction combined with GA-BP neural networks is expected to accurately predict chaotic time series.

In this study, the conventional statistical model is firstly improved by considering the nonlinear temperature factor. Further, the improved statistical model is combined with the GA-BP neural network model using chaos theory and phase-space reconstruction theory, to develop a hybrid model for predicting concrete crack openings. Finally, a detailed case study on the crack openings of a regulating sluice is presented to illustrate the application of the hybrid model.

2. Improved Statistical Model for Concrete Crack Openings

Hydraulic concrete structures with macroscopic cracks are affected by water pressure, temperature, and time. It is illustrated in Figure 1.

Consider a hydraulic concrete structure with K measuring points near a macroscopic crack. A conventional statistical model describing concrete crack openings (denoted as model I) is as follows:

$$\begin{aligned}\delta(t) &= \delta_T(t) + \delta_H(t) + \delta_\theta(t) + \delta_\varepsilon(t) \\ &= a_0 + \sum_{i=1}^K a_i T_i + \sum_{i=1}^4 b_i H^i + c_1 \theta + c_2 \ln \theta + \delta_\varepsilon(t),\end{aligned}\quad (1)$$

where $\delta(t)$ is the measured crack opening, $\delta_T(t)$ is the temperature component, $\delta_H(t)$ is the hydraulic component, $\delta_\theta(t)$ is the aging component, $\delta_\varepsilon(t)$ is the residual, T_i is the measured temperature of the i th measuring point, K is the number of temperature-measuring points, H is the water depth, $\theta = t/100$, t is time, and a_0 , a_i , b_i , c_1 , and c_2 are all regression coefficients.

Wang et al. [5] show that the temperature component in equation (1) does not consider a nonlinear temperature response. Since the concrete openings as a function of temperature is a complex nonlinear process, ignoring this nonlinearity will lead to underfitting. In contrast, including the environmental variables leading to nonlinearity alleviates this underfitting. Hence, we propose an improved statistical model of concrete crack openings, which is called statistical model considering nonlinear temperature component (denoted as model II) that accounts for the nonlinear temperature response [6–8]:

$$\begin{aligned}\delta(t) &= \delta_T(t) + \delta_H(t) + \delta_\theta(t) + \delta_\varepsilon(t) \\ &= a_0 + \sum_{p=1}^L \sum_{i=1}^K a_{ip} T_i^p + \sum_{i=1}^4 b_i H^i + c_1 \theta + c_2 \ln \theta + \delta_\varepsilon(t),\end{aligned}\quad (2)$$

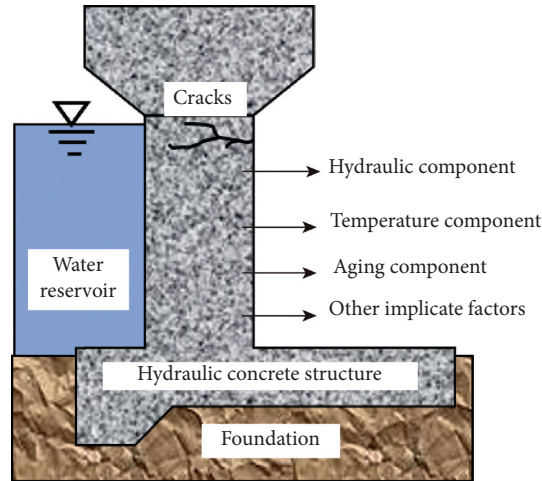


FIGURE 1: Influencing factors of macroscopic concrete cracks.

where L represents the order of the temperature polynomial reflecting the influence of the nonlinear temperature response (L is generally a positive integer between 1 and 4 and is determined by fitting to measured values); a_{ip} is the regression coefficient corresponding to the p th power of the measured temperature at the i th measurement point, and the other symbols have the same meaning as in equation (1). The regression coefficients of the statistical model are determined using regression analysis or optimization algorithms. The residual time series $\delta_e(t)$ is obtained by separating the measured data of the concrete crack openings.

3. Prediction of Residual Based on Chaos Theory

Under the action of water pressure and temperature variation, the stress field at the crack tip manifests singularity, which makes the concrete at the crack tip produce irreversible deformation. That is, the crack opening contains an irreversible part. Since the aging component cannot reflect the irreversible deformation accurately because of the limitations of its simple functions, the irreversible deformation is retained in the residual. In particular, when the crack opening suddenly changes due to overloading or drastic temperature variation, the hydraulic component, temperature component, and aging component of statistical model can hardly reflect this mutation of crack opening. In other words, the residual sequence contains a lot of information about the evolution of cracks [10] and some studies showed [9, 14] that this residual sequence contains chaotic characteristics. Therefore, chaos theory is used to analyze and predict the residual.

3.1. Phase-Space Reconstruction of Residual Sequence. Previous studies [10, 13] have shown that a 1D residual time series $x(t) = \{x(t_i) | i = 1, 2, 3, \dots, N\}$ can be reconstructed into the following form in m -dimensional phase space R^m :

$$\begin{bmatrix} x(t_1) & x(t_2) & \cdots & x(t_M) \\ x(t_1 + \tau) & x(t_2 + \tau) & \cdots & x(t_M + \tau) \\ \vdots & \vdots & & \vdots \\ x(t_1 + (m-1)\tau) & x(t_2 + (m-1)\tau) & \cdots & x(t_M + (m-1)\tau) \end{bmatrix} \quad (3)$$

where $M = N - (m-1)\tau$ is the number of phase points, $m > 0$ is the integer embedding dimension, and τ is the delay time, which is usually a positive integer multiple of the sampling interval of the residual time series $x(t)$.

The first-phase point in equation (3) can be expressed as follows:

$$X(t_i) = [x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)]^T \quad (i = 1, 2, \dots, n). \quad (4)$$

According to equations (3) and (4), the key of phase-space reconstruction lies in the determination of the delay time τ and the embedding dimension m .

3.1.1. Determination of Delay Time. Selecting the optimized delay time t_0 usually involves the autocorrelation coefficient method [16], the mutual information method [17], or the AD method [18]. The autocorrelation coefficient method is not suitable for nonlinear systems [25] and imposes a subjective selection of the method to calculate the descent coefficient. Although the mutual information method overcomes the shortcomings of the autocorrelation coefficient method, its calculation is relatively complicated. Hence, this study uses the AD method [18] to select the optimized delay time; it is an efficient calculation and avoids information redundancy and the complete irrelevance of delay coordinates caused by improper delay time. Under the 2-norm, the average displacement $S_2(m, \tau)$ is defined as follows:

$$S_2(m, \tau) \triangleq \frac{1}{M} \sum_{i=1}^M \sqrt{\sum_{l=1}^{m-1} [x(t_i + l\tau) - x(t_i)]^2}. \quad (5)$$

When given different embedding dimensions (≥ 2), $S_2(m, \tau)$ tends to be stable with increasing τ . Rosenstein et al. [18] suggested that optimized delay time can be determined according to the $S_2(m, \tau)$ vs. τ graph under different embedding dimension m ; that is, when the slope of $S_2(m, \tau)$ vs. τ decays to 40% of the initial slope, the corresponding τ is the optimized delay time. Research shows that this method is reliable for small, noisy datasets and that the computation time is short.

3.1.2. Determination of Embedding Dimension. Methods to calculate the minimum embedding dimension m_e include the geometric invariant method, the pseudo-neighbor method, and the Cao method. However, the geometric invariant method requires a noise-free time series, which is barely possible in practical applications. Cao [20] proposed the pseudo-neighbor point method, but this method is very subjective when selecting the decision threshold. They also proposed a practical method (called the ‘‘Cao method’’) for determining the minimum embedding dimension. This method not only is robust against noisy data but also reduces the interference of subjective factors and adverse effects caused by slight reductions in data size. For this reason, this study adopts the Cao method to calculate the minimum embedding dimension.

When an optimized time delay is given, the following parameters are defined in m -dimensional phase space:

$a(i, m)$ is defined as follows:

$$a(i, m) = \frac{\|X_i(m+1) - X_{n(i,m)}(m+1)\|_\infty}{\|X_i(m) - X_{n(i,m)}(m)\|_\infty} \quad (i = 1, 2, \dots, M), \quad (6)$$

where $X_i(m) = [x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)]^T$ ($i = 1, 2, \dots, M$), $\|X_i(m+1) - X_{n(i,m)}(m+1)\|_\infty$ is the distance defined under the infinite norm, and $X_{n(i,m)}(m)$ is the nearest point to the phase point $X_i(m)$, which means the phase point at the minimum Euclidean distance from $X_i(m)$.

The mean value of $E(m)$ of $a(i, m)$ is as follows:

$$E(m) = \frac{1}{M} \sum_{i=1}^M a(i, m). \quad (7)$$

When the embedding dimension is incremented from m to $m+1$, $E1(m)$ is as follows:

$$E1(m) = \frac{E(m+1)}{E(m)}. \quad (8)$$

Cao et al. [20] report that, given a fixed-point attractor in the residual time series, $E1(m)$ would stop changing once the embedding dimension m exceeds a certain value m_0 . In this case, $m_0 + 1$ is the minimum embedding dimension, which may be determined by a graphical analysis of $E1(m)$ vs. m .

In actual calculations, it is difficult to determine whether $E1(m)$ increases slowly or remains constant as m increases. For a group of random residual series, $E1(m)$ increases with

increasing m and converges for a deterministic residual series $E1(m)$. Hence, we must introduce $E2(m)$ to distinguish random residual series from deterministic residual series. $E2(m)$ is given as follows:

$$E2(m) = \frac{E^*(m+1)}{E^*(m)}, \quad (9)$$

where

$$E^*(m) = \frac{1}{M} \sum_{i=1}^M |x(t_i + m\tau) - x(t_{n(i,m)} + m\tau)|. \quad (10)$$

For random residual series, $E2(m) = 1$ for any m ; for deterministic residual series, $E2(m)$ is a nonconstant function of m ; that is, some m must exist so that $E2(m) \neq 1$. Therefore, Cao et al. [20] suggested that $E1(m)$ and $E2(m)$ be calculated at the same time to determine the minimum embedding dimension and distinguish between random residual series and deterministic residual series. For notational convenience, $E1(m)$ and $E2(m)$ are denoted as $E1$ and $E2$, respectively.

3.1.3. Identification of Chaotic Characteristics. The maximum Lyapunov exponent λ_1 is of important significance for determining the chaotic characteristics contained in residual time series. For phase space, $\lambda_1 > 0$ indicates that the system is chaotic. Rosenstein et al. [26] analyzed common methods of calculating the maximum Lyapunov exponent. For problems of unreliable calculation, intensive computation, and difficult-to-implement existing methods for small datasets, a practical method (called the ‘‘Rosenstein method’’ and also known as the ‘‘small-dataset method’’) was proposed for small datasets. The advantage of this method is that it is robust against the reconstruction dimension, time delay, and noise, thereby producing a more accurate maximum Lyapunov exponent. According to the theory of chaotic dynamics, the reciprocal λ_1^{-1} of the maximum Lyapunov exponent is the maximum prediction time and can be used as a reliability indicator of a short-term prediction [27].

When adopting the Rosenstein method to calculate the maximum Lyapunov exponent, $d_j(k)$ is used in the phase space reconstructed by the optimized delay time and minimum embedding dimension to express the Euclidean distance between phase point j , $X(t_j)$, and its initial nearest point $X(t_j)$ (the meaning is the same as the nearest point in (5)) at moment $k\Delta t$:

$$d_j(k) = \left\| X(t_j + k\Delta t) - X(t_j) \right\|, \quad (11)$$

where Δt is the sampling interval and is usually taken as $\Delta t = 1$ for calculational convenience.

The calculation result for $d_j(k)$ is expressed in matrix form as follows:

$$\begin{bmatrix} \|X(t_1) - X(t_1^\wedge)\| & \|X(t_2) - X(t_2^\wedge)\| & \dots & \|X(t_j) - X(t_j^\wedge)\| & \dots \\ \|X(t_1 + \Delta t) - X(t_1^\wedge + \Delta t)\| & \|X(t_2 + \Delta t) - X(t_2^\wedge + \Delta t)\| & \dots & \|X(t_j + \Delta t) - X(t_j^\wedge + \Delta t)\| & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \|X(t_1 + k\Delta t) - X(t_1^\wedge + k\Delta t)\| & \|X(t_2) - X(t_2^\wedge + k\Delta t)\| & \dots & \|X(t_j + k\Delta t) - X(t_j^\wedge + k\Delta t)\| & \dots \\ \vdots & \vdots & \dots & \vdots & \dots \end{bmatrix} \quad (12)$$

$d_j(k)$ is expressed as [26].

$$d_j(k) \approx C_j e^{\lambda_1(k\Delta t)}. \quad (13)$$

Taking the logarithm of both sides of equation (13) gives the following:

$$\text{Ind}_j(k) \approx \ln C_j + (k\Delta t)\lambda_1, \quad (14)$$

where C_j is the initial bifurcation and its value determined for j is constant. $y(k)$ is defined as follows:

$$y(k) = \frac{1}{\Delta t} \langle \text{Ind}_j(k) \rangle, \quad (15)$$

where $\langle \ln d_j(k) \rangle$ is the average over all j , that is, the average of the natural logarithm of the nonzero elements in row k of the matrix in equation (11).

In this case, we take the part of $y(k)$ that is close to linear with respect to k according to equations (13) and (14) and use the least-squares method to obtain a linear fit. The slope of the line is λ_1 , that is, the maximum Lyapunov exponent.

3.2. Predicted Value of Residual Based on GA-BP Neural Network. In the phase-space reconstruction of the residual time series, the “trajectory” in the reconstructed phase space is “dynamically equivalent” to the original system due to the concept of differential homeomorphism. Therefore, there is a unique mapping relation:

$$X(t + P_t) = F(X(t)), \quad (16)$$

where P_t is the prediction time, and the mapping relation F can be calculated by approximating all phase points in phase space. In this study, the nonlinear mapping relation is constructed using artificial neural networks with the whole domain method. The GA is a heuristic random evolutionary search algorithm, which has global search ability and can solve the error function without gradient information. Previous studies demonstrated that the GA-BP neural networks established by GAs provide better fits to nonlinear functions and more accurate short-term predictions of typical chaotic time series, even when compared with the predictions of BP neural networks [23, 28]. Therefore, this study uses GAs to optimize BP neural networks and thereby improves simple BP neural networks to attain the global maximum.

Setting the fitness function is quite important with GAs. In this study, we use the fitness function [29] as follows:

$$\text{Fitness} = 1000 \times \frac{R^2}{\text{RMSE}}, \quad (17)$$

which shows that a large R^2 and small root-mean-square error (RMSE) indicate a better fit. The optimal fitness is attained when the fitness becomes stable as a function of the evolutionary generations.

The equations for R^2 and the RMSE are as follows:

$$R^2 = \frac{\sum_{i=1}^{N_t} (\hat{\delta} - \bar{\delta})^2}{\left[\sum_{i=1}^{N_t} (\delta_t - \hat{\delta})^2 + \sum_{i=1}^{N_t} (\hat{\delta} - \bar{\delta}_t)^2 \right] \text{RMSE} = \left[\frac{1}{N_t} \sum_{i=1}^{N_t} (\delta_t - \hat{\delta})^2 \right]^{\frac{1}{2}}}, \quad (18)$$

where N_t is the number of samples in the test set, δ_t is the measured value, $\hat{\delta}$ is the predicted value, and $\bar{\delta}_t$ is the average measured value.

In GA-BP neural networks, the input vector is as follows:

$$X(m) = [x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)]^T, \quad (19)$$

and the output vector is as follows:

$$X(t + P_t) = [x(t + P_t), x(t + P_t - \tau), \dots, x(t + P_t - (m - 1)\tau)]^T. \quad (20)$$

The separated residual time series is trained to optimize the neural network model, which gives the weight and threshold of the GA-BP neural network model. The residual time series is then predicted.

3.3. Hybrid Model for Predicting Concrete Crack Opening. Model II and the GA-BP neural network (15) are integrated to establish the improved hybrid model for predicting concrete crack openings (denoted as “hybrid model”). Mathematically, the model is expressed as follows:

$$\begin{aligned} \delta(t + P_t) &= \delta_T(t + P_t) + \delta_H(t + P_t) \\ &+ \delta_\theta(t + P_t) + x(t + P_t), \end{aligned} \quad (21)$$

where $\delta_T(t + P_t)$, $\delta_H(t + P_t)$, and $\delta_\theta(t + P_t)$ predict the temperature, hydraulic, and aging components of P_t , respectively; $x(t + P_t)$ is the predicted value of the residual based on GA-BP neural network.

Figure 2 presents a flow chart of the procedure used by the hybrid model for predicting concrete crack openings.

4. Case Study of Regulating Sluice with Cracks

4.1. Background of Regulating Sluice with Cracks

4.1.1. Crack Distribution. Hubei Province has undertaken a large water control project designed for power generation and shipping, irrigation, aquaculture, and tourism. The project consists of a regulating sluice located in the main river with an earth-rockfill dam in the main riverbed, an earth-rockfill dam in the Gucheng Section, a powerhouse in the old river, a ship lock, a concrete gravity dam, and an earth-rockfill dam on both sides of the embankment.

On-site inspection has revealed three cracks at the top of the upstream traffic bridge of the gate storehouse on the left bank of the sluice (see Figure 3). Two of these cracks run upstream and downstream and stop at the lowest part of the breast wall. Meanwhile, a long-term calcium precipitation trace appears on the vertical surface of the retaining wall upstream of the gate storehouse and extends to the floor of the upstream wall of the gate storehouse, which is located between the two penetrating cracks at the top of the sluice and threatens the normal operation of gate storehouse.

4.1.2. Monitoring by Crack Meter. To diagnose the evolution of the concrete crack in the regulating sluice, nine sets of bidirectional vibrating string crack meter groups were installed to monitor the openings and misplacement momentum (see Figure 4). Mkj-1 and Mkj-2 were installed on the upstream side of the sluice, Mkj-3–Mkj-5 were installed near the top of the sluice, Mkj-6 and Mkj-7 were installed near the upstream surface of the sluice, and Mkj-8 and Mkj-9 were installed on the top of the sluice near the downstream section. From June 12 to 19, 2020, nine sets of bidirectional crack meters were installed and debugged. Figure 5 shows the measured crack openings and temperature lines of crack meter Mkj-3.

An obvious negative correlation appears between the crack opening at the sluice and the temperature. When the temperature increases, the crack closes and vice versa. Next, we use the measured crack opening in the hybrid model for concrete crack openings.

4.2. Model II of Concrete Crack Openings in Sluice

4.2.1. Determination of L in Model II. The various measuring points may give different results due to the nonlinear temperature response. To determine L in model II according to the principle mentioned above, we use nine temperature measurements by the crack meter. Using equation (2), we

establish model II for concrete crack openings for $L = 1, 2, 3, 4, 5$. The model regression coefficient corresponding to different results for L is obtained by regression analysis. R -square and RMSE of different improved models were calculated. The results are shown in Figure 6.

When L increases from 1 to 3, both R^2 and RMSE improve significantly; when $L \geq 3$, the two parameters remain stable. From the point of view of statistical models, this shows that temperature exerts a nonlinear effect on the concrete crack opening. Therefore, when model II with nonlinear temperature effects is included, $R^2 = 0.891$ and $\text{RMSE} = 0.04571$. Table 1 lists the regression coefficients obtained through regression analysis. Using these, model II for concrete crack openings is as follows.

$$\delta(t) = a_0 + \sum_{p=1}^3 \sum_{i=1}^9 a_{ip} T_i^p + \sum_{i=1}^4 b_i H^i + c_1 \theta + c_2 \ln \theta + \delta_\varepsilon(t). \quad (22)$$

4.2.2. Comparison with Other Models. Figure 7 compares model II for crack openings (22) with model I (1).

Model I considers only the linear response to temperature, as highlighted by the ellipses in Figure 7(a). Note that the fit near the extreme values is not ideal. In Figure 7(b), the nonlinear response to temperature is considered in model II, which makes the fit closer to the measured trajectory. This is seen in the decrease in the residual error of model II (Figure 7(c)).

Notably, limitations of the statistical models shown in Figures 7(b) and 7(c) still result in periods where the fit is poor, which will affect the accuracy of the models' predictions to some extent. Hence, the prediction methods based on the statistical model may still be improved. For this reason, we apply chaos theory to further analyze the residual, which leads to the more accurate model.

4.3. Phase-Space Reconstruction of Residual Series

4.3.1. Determination of Embedding Dimension and Delay Time. The residual time series from model II for crack openings is obtained by separating the measured values. From the residual time series, 1825 samples from 22:00 on June 23, 2020, to 7:00 on February 11, 2021, are selected as the training dataset, and 24 samples from 10:00 on February 11, 2021, to 7:00 on February 14, 2021, are selected as a posteriori prediction dataset. Figure 8 shows $S_2(m, \tau)$ as a function of delay time τ constructed using the AD plotting method, which is based on phase-space reconstruction theory. Figure 9 plots $E1$ and $E2$ (determined by the Cao method) as a function of embedding dimension m .

When the embedding dimension is constant, the slope of $S_2(m, \tau)$ vs. τ stabilizes as τ increases. The delay time is optimal when the slope of $S_2(m, \tau)$ vs. τ decreases to 40% of its initial value, which gives $\tau = 2$ as the optimal delay time. For a given optimal delay time, $E1$ stabilizes with increasing embedding dimension, and $E2 \neq 0$. Therefore, when $E1$ stabilizes, the corresponding embedding dimension (i.e., the minimum embedding dimension) is eight.

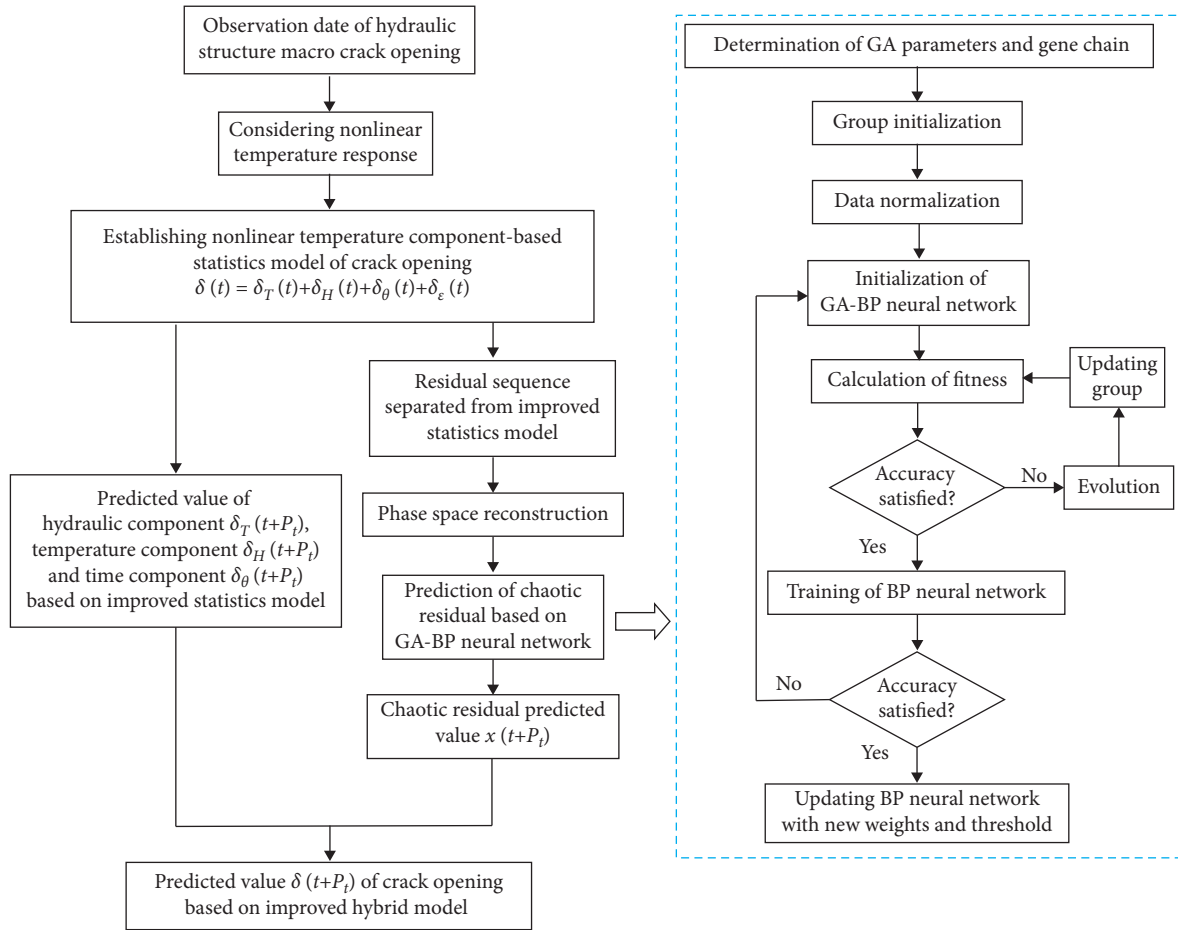


FIGURE 2: Flow chart of the hybrid model for predicting concrete crack openings.

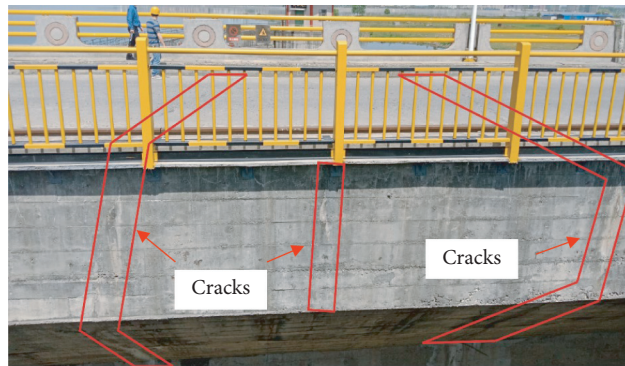


FIGURE 3: Main cracks in traffic bridge of sluice.

4.3.2. *Phase-Space Reconstruction of Residual Time Series and Determination of Prediction Step.* Given the embedding dimension and delay time, phase-space reconstruction is applied to the 1D residual time series via equation (3), as per the principle mentioned above. The maximum Lyapunov exponent is calculated via the Rosenstein method, indicating that the system contains chaotic components. The reciprocal of the system is then calculated, giving a maximum

prediction step of 40. Therefore, the posterior step is 24 (i.e., 24 samples at 7:00 on February 11, 2021, and 7:00 on February 14, 2021), and the duration is three days.

4.4. *Hybrid Model Based on GA-BP Neural Networks*

4.4.1. *Training to Optimize GA-BP Neural Networks.* As shown in Figure 1, we use 1825 samples from 22 on June 23,

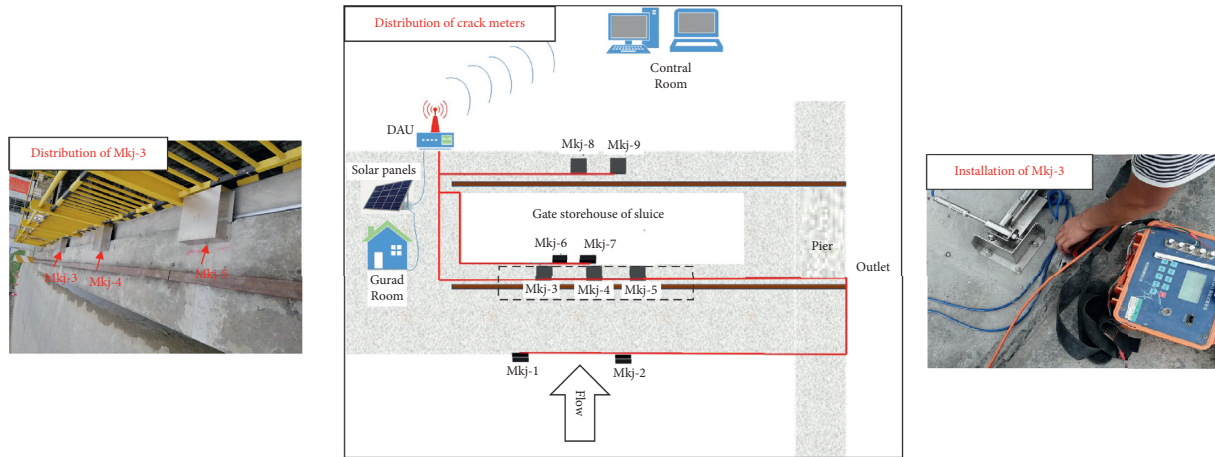


FIGURE 4: Layout of bidirectional crack meters.

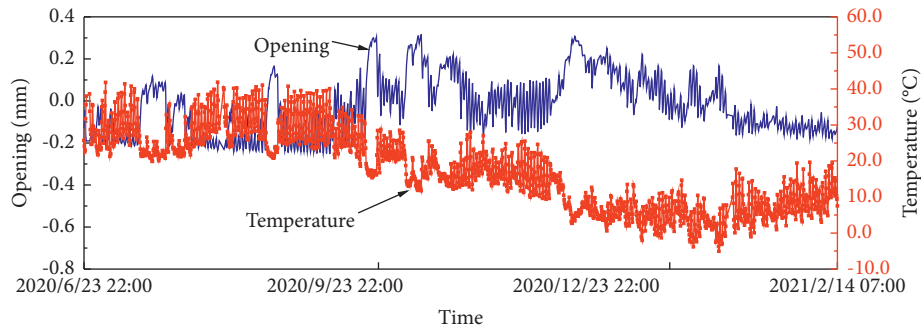


FIGURE 5: Concrete crack opening (blue) and temperature (red).

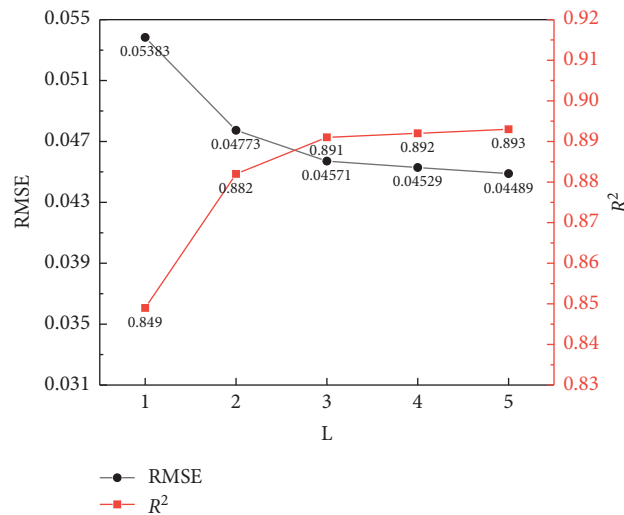


FIGURE 6: RMSE and R^2 as a function of L in model II.

2020, to 7:00 on February 11, 2021, to train the GA-BP neural networks. The structure of the BP neural network is 8-12-8, the GA's maximum number of generations is 30, the population is 40, the crossover probability is 0.8, and the variation probability is 0.1. Figure 10 shows the resulting fit.

When the fitness stabilizes (i.e., for >5 generations) and satisfies the precision requirements of neural networks, the optimized BP neural network is created using the weights and threshold parameters output at this time to train and predict the reconstructed residual time series.

TABLE 1: Regression coefficients of model II for concrete crack openings.

Regression coefficient	Equation (22)
a_{11}	-1.661×10^{-1}
a_{21}	9.235×10^{-2}
a_{31}	1.701×10^{-2}
a_{41}	0
a_{51}	0
a_{61}	7.707×10^{-2}
a_{71}	2.446×10^{-2}
a_{81}	9.186×10^{-2}
a_{91}	-1.097×10^{-1}
a_{12}	2.297×10^{-3}
a_{22}	-1.320×10^{-3}
a_{32}	0
a_{42}	0
a_{52}	1.158×10^{-3}
a_{62}	-2.587×10^{-3}
a_{72}	0
a_{82}	-3.704×10^{-3}
a_{92}	2.673×10^{-3}
a_{13}	0
a_{23}	0
a_{33}	0
a_{43}	-8.494×10^{-6}
a_{53}	-1.737×10^{-5}
a_{63}	1.863×10^{-5}
a_{73}	0
a_{83}	4.518×10^{-5}
a_{93}	-2.476×10^{-5}
b_1	-6.393×10^{-2}
b_2	-2.685
b_3	0
b_4	2.094×10^{-1}
c_1	2.537×10^{-1}
c_2	7.648×10^{-2}
a_0	2.931×10^{-1}

4.4.2. *Hybrid Model.* Based on the length of the posteriori predicted dataset of residual time series, the prediction is made starting at 10:00 on February 11, 2021, and the prediction step is 24 (three days). The residual time series predicts that $x(t+1)$, $x(t+2)$, \dots , $x(t+24)$ are obtained, and the hybrid model is as follows:

$$\delta(t + P_t) = a_0 + \sum_{k=p}^3 \sum_{i=1}^9 a_{ip} T_i^p + \sum_{i=1}^4 b_i H^i + c_1 \theta + c_2 \ln \theta + x(t + P_t), \quad (23)$$

where $\delta(t + P_t)$ is the predicted concrete crack opening at time P_t , where $P_t = 1, 2, \dots, 24$.

4.4.3. *Prediction Accuracy.* Mean square error (MSE), mean absolute error (MAE), and normalized mean square error

(NRMSE) are introduced to evaluate the prediction accuracy of the hybrid model. These are given by:

$$\begin{aligned} \text{MSE} &= \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\delta_t - \hat{\delta} \right)^2, \\ \text{MAE} &= \frac{1}{N_t} \sum_{i=1}^{N_t} \left| \delta_t - \hat{\delta} \right|, \\ \text{NRMSE} &= \frac{1}{\sigma} \left[\frac{1}{N_t} \sum_{i=1}^{N_t} \left(\delta_t - \hat{\delta} \right)^2 \right]^{1/2}, \end{aligned} \quad (24)$$

where σ is the standard deviation of the test set, N_t is the number of samples in the test set, δ_t is the measured concrete crack opening, and $\hat{\delta}$ is the predicted concrete crack opening. The closer the three evaluation indexes are to zero, the more accurate is the prediction.

According to the posteriori prediction of the residual time-series dataset, equation (23) gives the prediction of the hybrid model for concrete crack openings starting from 10:00 on February 11, 2021, with a prediction step of 24 (three days). Figure 11 compares this prediction with those of model I and model II. Table 2 gives the evaluation indicators for different models.

Unlike the values predicted by the conventional statistical model, those predicted by the statistical model considering nonlinear temperature component, which considers the nonlinear response to temperature, are closer to the measured values, although the improvement is minimal near the extreme values. However, the difference between the predicted values and the measured values near the extrema is significantly reduced when using the hybrid model. Compared with model I, the MSE, MAE, and NRMSE of the hybrid model decreased by 87.0%, 68.5%, and 63.9%, respectively. Compared with model II, the MSE, MAE, and NRMSE of the hybrid model decreased by 73.4%, 51.4%, and 48.4%, respectively.

In summary, a positive Lyapunov exponent indicates that concrete crack openings have chaotic components, so the hybrid model makes significantly improved predictions of concrete crack openings.

4.5. *Dynamic Prediction Based on Hybrid Model of Concrete Crack Openings.* Early warning is a dynamic process in real engineering because concrete crack openings evolve over time. Given new monitoring data, the parameters of the model must be updated and the dynamic prediction must be redone. To verify the reliability of the hybrid model, ten consecutive dynamic predictions were made, each with a time step of 24 (three days), and MSE, MAE, and NRMSE of model I, model II, and the hybrid model were statistically analyzed. Based on these three metrics, Table 2 gives the

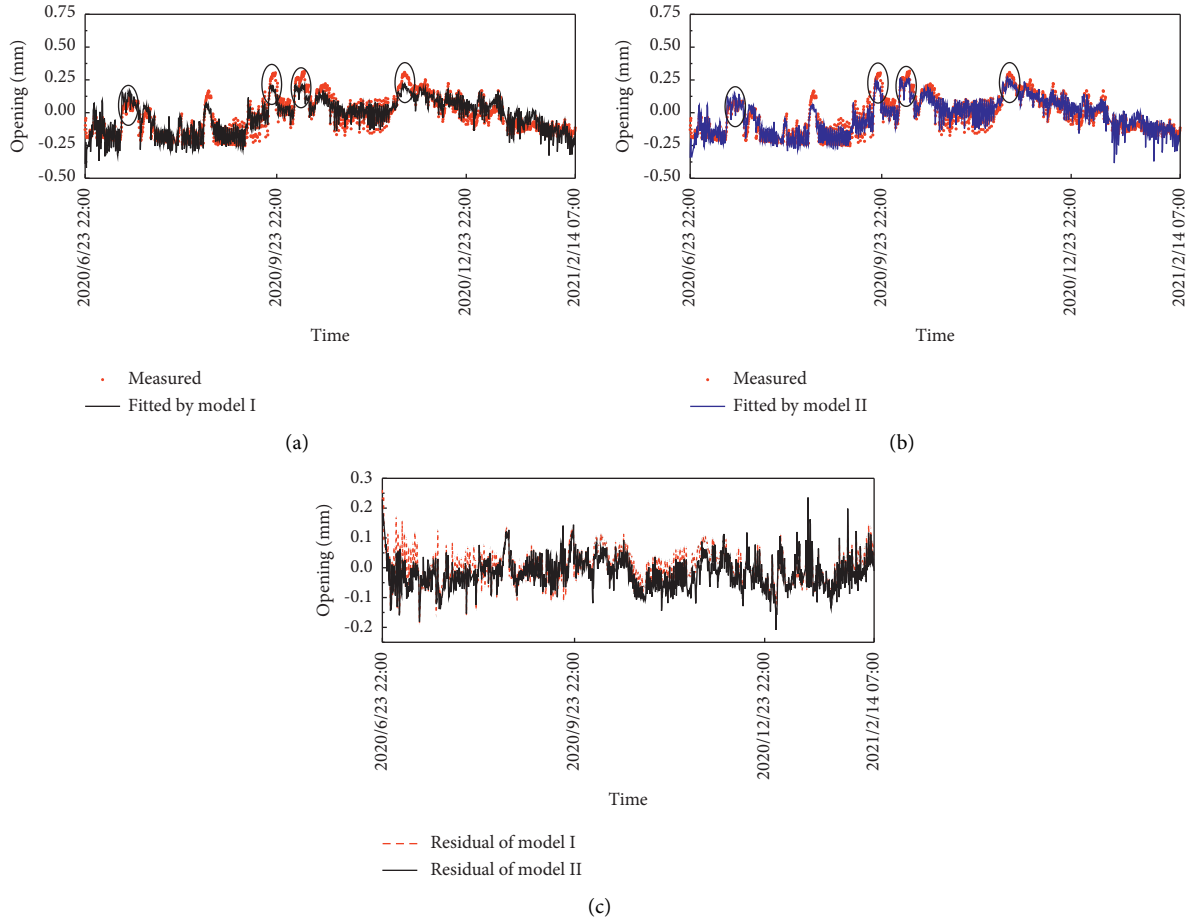


FIGURE 7: Fitting effect of model I and model II. (a) Model I fitting. (b) Model II fitting. (c) Residuals of model I and model II.

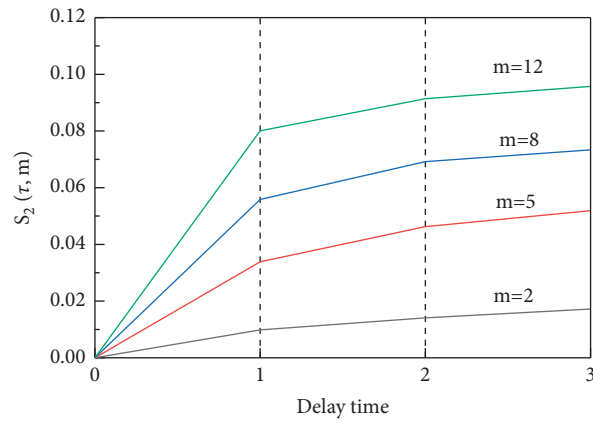


FIGURE 8: $S_2(m, \tau)$ as a function of delay time τ for different embedding dimensions.

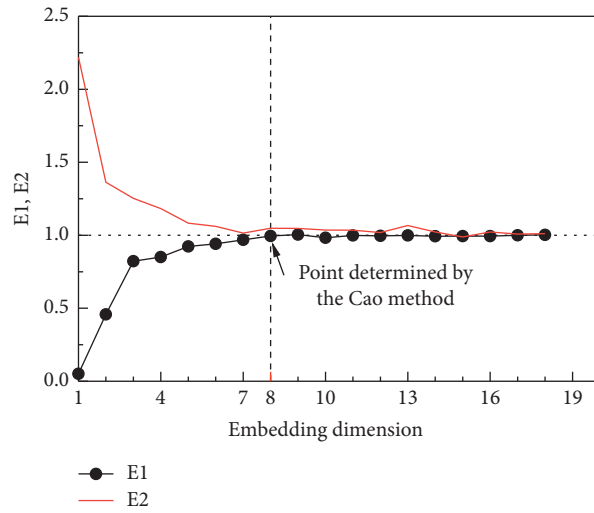


FIGURE 9: $E1$ and $E2$ as a function of embedding dimension m .

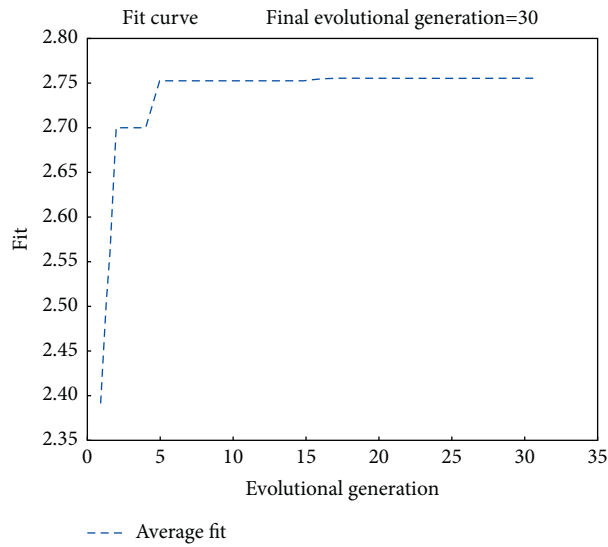


FIGURE 10: Fitness curve for training to optimize GA-BP neural networks.

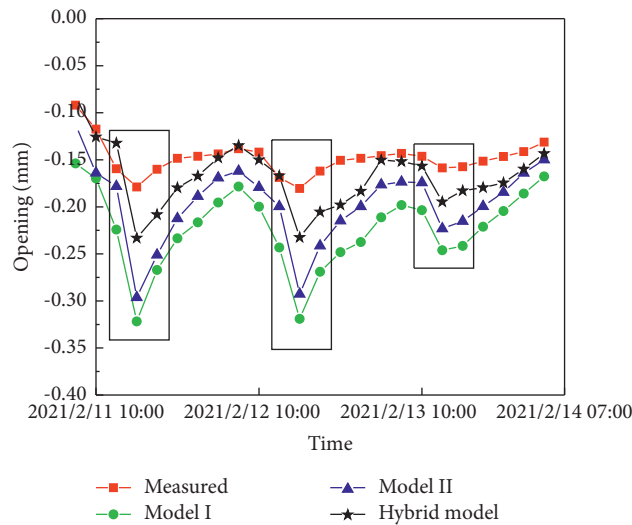


FIGURE 11: Evolution of predicted and measured values from the different models.

TABLE 2: Metrics for evaluating the prediction accuracy of the different models.

	MSE	MAE	NRMSE
Model I ($M1$)	6.371×10^{-3}	7.4935×10^{-2}	1.462×10^{-3}
Model II ($M2$)	3.118×10^{-3}	4.8492×10^{-2}	1.023×10^{-3}
Hybrid model ($M3$)	8.30×10^{-4}	2.3575×10^{-2}	5.28×10^{-4}
$(M2 - M3)/M2$	73.4%	51.4%	48.4%
$(M1 - M3)/M1$	87.0%	68.5%	63.9%

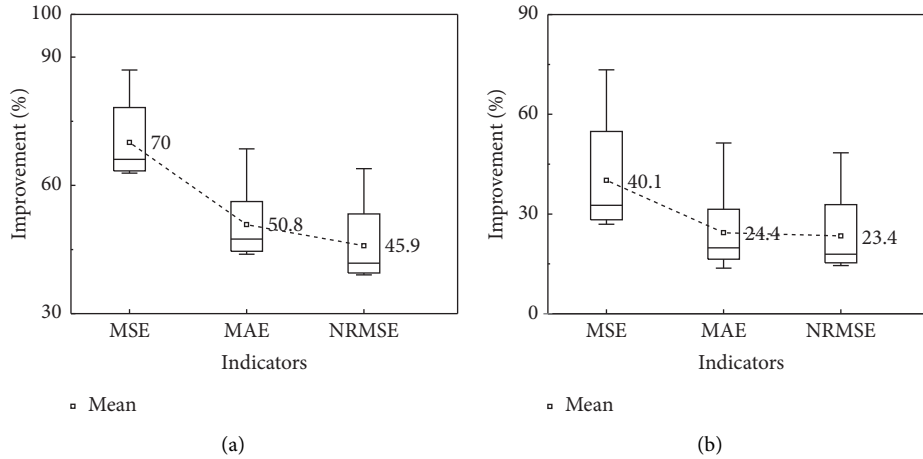


FIGURE 12: Dynamic prediction of hybrid model of concrete crack openings compared with that of statistical models. (a) Improvement due to hybrid model relative to model I. (b) Improvement due to hybrid model relative to model II.

calculated improvement for the hybrid model relative to model I and model II, and Figure 12 shows the box plots for a comparative analysis.

For the hybrid model, the three evaluation metrics are significantly improved compared with the two statistical models. The improvement for the hybrid model relative to model I is 45.9%–70%, and the maximum improvement in the MSE exceeds 80%. The average improvement for the hybrid model relative to model II is 23.4%–40.1%, and the maximum improvement in the MSE exceeds 70%. This indicates that the hybrid model provides more accurate predictions of concrete crack openings than does model II.

5. Conclusions

In this study, the conventional statistical model for concrete crack opening is improved by considering the nonlinear temperature factor. Using chaos theory and phase-space reconstruction theory, a hybrid model that combines the improved statistical model with a GA-BP neural network residual model is finally developed to predict concrete crack openings in hydraulic structures.

It is found that concrete crack openings are affected by the nonlinear temperature response. However, the conventional statistical model cannot accurately reflect the temperature component of the system. The improved statistical model makes a significant improvement in this area by introducing the second- and third-order temperature components.

The residual series of the statistical model considering nonlinear temperature component contains useful information of crack evolution; the presence of a chaotic component is demonstrated by calculating the maximum Lyapunov exponent. The phase-space reconstruction of the residual series is combined with a GA-BP neural network and the statistical model considering nonlinear temperature component to establish a hybrid model to predict the residuals. The predictions of the hybrid model are more accurate than those of the statistical model.

These results of the case study show that the hybrid model produces more reliable short-term dynamic predictions of concrete crack openings, making it of practical use for generating early warnings of concrete crack openings.

Data Availability

The data in this article were calculated and analyzed by the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was financially supported by the National Key Research and Development Program of China (No. 2018YFC0407103), the Open Research Fund of State Key Laboratory of Simulation and Regulation of Water Cycle in

River Basin (China Institute of Water Resources and Hydropower Research) under Grant No. IWHR-SKL-201917, and the IWHR Research and Development Support Program (No. SM0145B022021).

Supplementary Materials

Supplementary 1. Concise description for Figure 5. Figure 5: concrete crack opening (blue) and temperature (red). Column A is time. Column B is concrete crack opening. Column C is measured temperature. Supplementary 2. Concise description for Figure 6. Figure 6: RMSE and R^2 as a function of L in model II. Column A is the order of the temperature polynomial (denoted as L). Column B is RMSE. Column C is R^2 . Supplementary 3. Concise description for Figure 7. Figure 7: fitting effect of model I and model II. Column A is time. Column B is measured value. Column C is fitted value by model I. Column D is residuals of model I. Column E is fitted value by model II. Column F is residuals of model II. Supplementary 4. Concise description for Figure 8. Figure 8: $S_2(m, \tau)$ as a function of delay time τ for different embedding dimensions. Column A is delay time. Column B is $S_2(m, \tau)$ with $m = 2$. Column C is $S_2(m, \tau)$ with $m = 5$. Column D is $S_2(m, \tau)$ with $m = 8$. Column E is $S_2(m, \tau)$ with $m = 12$. Supplementary 5. Concise description for Figure 9. Figure 9: $E1$ and $E2$ are function of embedding dimension m . Column A is embedding dimension m . Column B is $E1$. Column C is $E2$. Supplementary 6. Concise description for Figure 10. Figure 10: fitness curve for training to optimize GA-BP neural networks. Column A is evolutionary generation. Column B is fit effect value. Supplementary 7. Concise description for Figure 11. Figure 11: evolution of predicted and measured values from the different models. Column A is time. Column B is measured values. Column C is predicted values by model I. Column D is predicted values by model II. Column E is predicted values by hybrid model. Supplementary 8. Concise description for Figure 12. Figure 12: dynamic prediction of hybrid model of concrete crack openings compared with that of statistical models. Column A is times of dynamic prediction. Column B is improvement of MSE due to hybrid model relative to model I. Column C is improvement of MAE due to hybrid model relative to model I. Column D is improvement of NRMSE due to hybrid model relative to model I. Column E is improvement of MSE due to hybrid model relative to model II. Column F is improvement of MAE due to hybrid model relative to model II. Column G is improvement of NRMSE due to hybrid model relative to model II. (*Supplementary Materials*)

References

- [1] Z. Wu, *Safety Monitoring Theories and its Application of Hydraulic Structures*, Higher Education Press, Beijing, China, 2003.
- [2] Y. Huang, "Optimization of temperature-control measures for concrete structures: a case study of the sluice project," *Advances in Civil Engineering*, vol. 2018, Article ID 4823130, 8 pages, 2018.
- [3] Y. Huang, L. Xiao, and T. Bao, "Transformation from measured strains to viscoelastic stresses considering temperature history for concrete dams," *KSCCE Journal of Civil Engineering*, vol. 23, no. 4, pp. 1725–1732, 2019.
- [4] J. Ouyang, X. Chen, Z. Huangfu, C. Lu, D. Huang, and Y. Li, "Application of distributed temperature sensing for cracking control of mass concrete," *Construction and Building Materials*, vol. 197, pp. 778–791, 2019.
- [5] J. Wang, W. Yang, and D. Zheng, "Reason of poor fitting in regression analysis for dam safety monitoring data," *Engineering Journal of Wuhan University*, vol. 36, no. 6, pp. 9–12, 2003.
- [6] J. Shen and D. Zheng, "Improved temperature component-based spatio-temporal distribution model for arch dam deformation," *Water Resources and Hydropower Engineering*, vol. 49, no. 5, pp. 38–43, 2018.
- [7] B. Xu, B. Liu, D. Zheng, L. Chen, and C. Wu, "Analysis method of thermal dam deformation," *Science China Technological Sciences*, vol. 55, no. 6, pp. 1765–1772, 2012.
- [8] J. Qiu, J. Liu, S. Tian, and J. Wu, "Spatial function monitoring model of dam deformation," *Yellow River*, vol. 39, no. 1, pp. 116–120, 2017.
- [9] B. Wei, D. Yuan, H. Li, and Z. Xu, "Combination forecast model for concrete dam displacement considering residual correction," *Structural Health Monitoring*, vol. 18, no. 1, pp. 232–244, 2019.
- [10] T. Bao and Z. Wu, "A hybrid model for prediction of concrete crack opening," *Water Resources and Hydropower Engineering*, vol. 36, no. 11, pp. 52–55, 2005.
- [11] H. Su, Z. Wen, Z. Chen, and S. Tian, "Dam safety prediction model considering chaotic characteristics in prototype monitoring data series," *Structural Health Monitoring*, vol. 15, no. 6, pp. 629–639, 2016.
- [12] N. D. Bokde, Z. M. Yaseen, and G. B. Andersen, "ForecastTB—an R package as a test-bench for time series forecasting-application of wind speed and solar radiation modeling," *Energies*, vol. 13, no. 10, p. 2578, 2020.
- [13] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics*, vol. 898, pp. 366–381, 1981.
- [14] B. Wei, D. Yuan, Z. Xu, and L. Li, "Modified hybrid forecast model considering chaotic residual errors for dam deformation," *Structural Control and Health Monitoring*, vol. 25, no. 1, 2018.
- [15] M. Alfaro, G. Fuertes, M. Vargas, J. Sepúlveda, and M. Veloso-Poblete, "Forecast of chaotic series in a horizon superior to the inverse of the maximum Lyapunov exponent," *Complexity*, vol. 2018, Article ID 1452683, 9 pages, 2018.
- [16] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, vol. 65, no. 4, pp. 1331–1392, 1993.
- [17] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [18] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D: Nonlinear Phenomena*, vol. 73, no. 1-2, pp. 82–98, 1994.
- [19] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, no. 6, pp. 3403–3411, 1992.
- [20] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D*, vol. 110, no. 1, pp. 43–50, 1997.

- [21] M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D: Nonlinear Phenomena*, vol. 35, no. 3, pp. 335–356, 1989.
- [22] J. W. Woolley, P. K. Agarwal, and J. Baker, "Modeling and prediction of chaotic systems with artificial neural networks," *International Journal for Numerical Methods in Fluids*, vol. 63, no. 8, pp. 989–1004, 2009.
- [23] B. Wang, B. Chen, G. Wang et al., "Back propagation (BP) neural network prediction and chaotic characteristics analysis of free falling liquid film fluctuation on corrugated plate wall," *Annals of Nuclear Energy*, vol. 148, Article ID 107711, 2020.
- [24] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artificial Intelligence Review*, vol. 36, no. 2, pp. 153–162, 2011.
- [25] S. Zhang Shu-Qing, J. Jia Jian, M. Gao Min, and X. Han Xu, "Study on the parameters determination for reconstructing phase-space in chaos time series," *Acta Physica Sinica*, vol. 59, no. 3, pp. 1576–1582, 2010.
- [26] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.
- [27] C. Liu, "Discussion on a new rule to find largest Lyapunov exponent by saturated embedded dimension," *Chinese Journal of Rock Mechanics and Engineering*, vol. 24, no. 22, p. 4088, 2005.
- [28] J. Huang, H. Luo, H. Wang, and B. Long, "Prediction of time sequence based on GA-BP neural net," *Journal of University of Electronic Science and Technology of China*, vol. 38, no. 5, pp. 687–692, 2009.
- [29] H. Xu, C. Chen, H. Zheng et al., "AGA-SVR-based selection of feature subsets and optimization of parameter in regional soil salinization monitoring," *International Journal of Remote Sensing*, vol. 41, no. 12, pp. 4470–4495, 2020.