

Research Article

Safety Monitoring Model of a Super-High Concrete Dam by Using RBF Neural Network Coupled with Kernel Principal Component Analysis

Siyu Chen ^{1,2,3} Chongshi Gu ^{1,2,3} Chaoning Lin,¹ Erfeng Zhao ^{1,2,3} and Jintao Song^{1,2,3}

¹College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China

²National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, Nanjing 210098, China

³State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Chongshi Gu; csgu@hhu.edu.cn

Received 27 March 2018; Revised 4 July 2018; Accepted 1 August 2018; Published 30 August 2018

Academic Editor: Andras Szekrenyes

Copyright © 2018 Siyu Chen 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.

Effective deformation monitoring is vital for the structural safety of super-high concrete dams. The radial displacement of the dam body is an important index of dam deformation, which is mainly influenced by reservoir water level, temperature effect, and time effect. In general, the safety monitoring models of dams are built on the basis of statistical models. The temperature effect of dam safety monitoring models is interpreted using approximate functions or the temperature values of a few points of measurement. However, this technique confers difficulty in representing the nonlinear features of the temperature effect on super-high concrete dams. In this study, a safety monitoring model of super-high concrete dams is established through the radial basis neural network (RBF-NN) and kernel principal component analysis (KPCA). The RBF-NN with strong nonlinear fitting capacity is utilized as the framework of the model, and KPCA with different kernels is adopted to extract the temperature variables of the dam temperature dataset. The model is applied to a super-high arch dam in China, and results show that the Hybrid-KPCA-RBF-NN model has high fitting and prediction precision and thus has practical application value.

1. Introduction

More than 90 dams with height greater than 200 m are built worldwide (dams in construction are included), more than 60% of which are concrete dams [1]. The operation state of these super-high concrete dams is complicated due to the ambient temperature, water pressure, and concrete mechanical and geomechanical factors. The collapse of these dams may pose a serious threat to the downstream areas. The deformation of the dam is an important indicator that can intuitively reflect the operation property of the dam [2]. Daily monitoring and analysis of the data obtained from the monitoring system is important, which can help guarantee the safe operation of the dams.

Deterministic and statistical models are extensively used in dam safety monitoring [2, 3]. Deterministic models use numerical calculation techniques to predict the dam and

foundation responses under the environmental load [2, 4]. The deterministic models are usually utilized during dam construction and storage periods. The advantage of the deterministic models is its interpretation of the dam behavior in combination with the concepts of mechanics and materials. However, the implementation of a deterministic model is cumbersome and time-consuming, which requires substantial modeling effort and computational work. Meanwhile, the statistical models are dam safety monitoring models and built on the basis of measured data (water pressure, temperature, and dam effect) and some regression methods, including multivariate linear regression [5, 6], stepwise regression [7, 8], partial least squares regression [9, 10], and kernel function partial least squares regressions [11]. Other than the deformation monitoring, statistical models have been extensively applied in other aspects of dam safety monitoring, such as seepage and crack [12–14].

In recent years, artificial neural network is regarded as a powerful machine learning tool [15], which enriches conventional models and opens a new way for the safety monitoring of concrete dams. Mata applied artificial neural networks to evaluate the behavior of a concrete gravity dam [6]. Ching-Yun Kao adopted artificial neural networks to monitor the long-term static deformation data of Fei-Tsui Arch Dam [16]. Zhu [17] and Wang [18] applied backpropagation neural network (BP-NN) to construct dam deformation monitoring model and used intelligent algorithm to optimize the parameters of the network. Kang [19] utilized extreme learning machine (ELM) to predict dam deformation and achieved better prediction performance than that of BP-NN. The proper selection of the input variables has significant influence on the performance of the neural network model for dam safety monitoring. In general, the input variables of the neural network involve hydrostatic pressure, temperature, and irreversible components (time effect), which are calculated on the basis of the hydrostatic, seasonal, and time (HST) statistical models [20] or hydrostatic, thermal, and time (HTT) statistical models [5]. In HST models, temperature effect is considered to be a simple harmonic function to describe the variations of air and water temperature. Different from HST models, in HTT models, the measured temperatures are usually analyzed to construct temperature component. The advantages of the use of measured temperatures of the concrete dam have been presented by several authors [5, 21, 22]. However, a large number of thermometers are embedded in high concrete dams, and selecting and calculating the input variables concerning the temperature of the dam are challenging. In the previous studies, principal component analysis (PCA) was utilized to address this limitation, for its performance on dimensionality reduction, and overcoming multicollinearity [5, 23]. However, the temperature data collected from the thermometer embedded in dams have nonlinear features [24]. PCA is a linear technique and has the difficulty in reducing large amounts of temperature variables to a few new uncorrelated variables while simultaneously minimizing the loss of nonlinear information. For these reasons, the residual error of the models may, therefore, remain large in some cases.

Presently, some super-high dams have been equipped with a large number of monitoring instruments and complete safety monitoring system, which makes utilization of huge amount of monitoring data and construction of digital-driven safety monitoring models with high accuracy possible. The objective of this study is to develop a precise safety monitoring model for predicting the displacement of super-high concrete dam. The radial basis function neural network (RBF-NN) is adopted to construct the model framework. A new methodology that calculates input variables concerning the temperature effect will also be introduced. The methodology proposed, on the basis of kernel principal component analysis (KPCA) and modified artificial fish-swarm algorithm (MAFSA), helps to reduce the dimensionality of temperature measurements and minimize the loss of nonlinear information while simultaneously improving the prediction accuracy of the model.

2. Statistical Model for Horizontal Displacements Monitoring of Concrete Dams

The displacement of concrete dams at an arbitrary point can be divided into horizontal and vertical displacements. According to present research results, the horizontal displacement of the dam is mainly affected by water pressure, ambient temperature, and time, which can be quantitatively interpreted and approximated using the following [25].

$$\delta = \delta_H(t) + \delta_T(t) + \delta_\theta(t) + \varepsilon, \quad (1)$$

where $\delta_H(t)$ denotes the hydrostatic pressure variable, $\delta_T(t)$ denotes the temperature variable, $\delta_\theta(t)$ denotes the time effect variable, and ε denotes the residual error.

Hydrostatic pressure variable can be described using a polynomial function depending on reservoir water levels as follows:

$$\delta_H(t) = \sum_i^N a_i H^i, \quad (2)$$

where H is the height of the reservoir. The value of N is considered to be related to the type of the dams. Quartic polynomial ($N = 4$) is adopted to describe the hydrostatic pressure variable of arch dams, whereas the cubic polynomial ($N = 3$) is adopted to describe the hydrostatic pressure variable of gravity dams.

Time effect variable, caused by creep and plastic deformation of dam concrete and bedrock, reflects the irreversible deformation of the dam toward a certain direction over time. According to present research results, time effect variable can be described by a polynomial function consisting of linear, exponential, logarithmic, and hyperbolic functions, as follows:

$$\delta_\theta(t) = \sum_{i=1}^4 c_i F_i \quad (3)$$

$$\begin{aligned} F_1 &= \theta, \\ F_2 &= 1 - e^{-\theta}, \\ F_3 &= \ln(\theta + 1), \\ F_4 &= \frac{\theta}{(\theta + 1)}, \end{aligned} \quad (4)$$

where θ is the time calculation parameters related to the time of observation date t and time of the initial date t_0 , which can be expressed as $\theta = (t - t_0)/100$.

Temperature variable is related to ambient temperature and the internal temperature field of the dam. In addition, the selection and calculation of temperature variable depend on the integrity and continuity of the temperature monitoring data in the period. When the temperature monitoring data

are not continuous, inadequate, or unavailable, the temperature variable can be expressed by the combination of harmonic function as follows:

$$\delta_{T1}(t) = \sum_{i=1}^2 (b_{1i} \sin(id) + b_{2i} \cos(id)), \quad (5)$$

where $d = 2\pi t/365$ and t is the number of days from the observation date to the beginning of the monitoring sequence. On the basis of the double angle formula, (5) can be transformed to other forms as follows:

$$\begin{aligned} \delta_{T1}(t) = & b_1 \sin(d) + b_2 \cos(d) + b_3 \sin(d) \cos(d) \\ & + b_4 \sin^2(d). \end{aligned} \quad (6)$$

The approach above shows that the temperature effect is assumed to vary by either 1-year or 6-month periodicity and can be expressed by the sum of sinusoidal functions. When the temperature monitoring data is continuous and adequate, the temperature variable can be expressed as follows:

$$\delta_{T2}(t) = \sum_{i=1}^2 b_i T_i, \quad (7)$$

where T_i is the measurement data obtained from the thermometers buried in concrete dam body. The thermometer measurement data T_i near the displacement point to be analyzed are usually utilized to calculate the temperature variable.

3. Extracting the Kernel Principal Component of the Thermometer Data

In this section, a new methodology for nonlinear feature extraction of temperature effect is proposed. We utilize different novel kernels rather than radial basis function to construct KPCA for better performance on feature extraction and dimensionality reduction of KPCA. Determination of kernel parameters can be considered to be an optimization problem; hence, MAFSA is adopted to address this task.

3.1. KPCA. PCA is a powerful tool that can convert a set of data of possibly correlated variables into a set of linearly uncorrelated features called principal components (PCs) by adopting the orthogonal transformation. However, this statistical procedure is a linear technique, which is not suitable to discover and capture nonlinear features among the original dataset. KPCA [26] is an extension of PCA by using kernel methods, which makes the PCA perform in Hilbert space. The KPCA is illustrated as follows.

In high-dimension feature space $\mathbf{R}^n \rightarrow \Omega$, the matrix $X \rightarrow \Phi(x)$ is constructed by samples $x_i (i = 1, 2, \dots, m)$.

We assume that $\sum_{i=1}^m \Phi(x) = 0$, so the covariance matrix \mathbf{C}^Ω in the feature space Ω can be expressed as follows:

$$\mathbf{C}^\Omega = \frac{1}{m} \sum_{j=1}^m \Phi(x_j) \Phi(x_j)^T. \quad (8)$$

The eigenvalue λ and eigenvector \mathbf{v} can be obtained by solving the following equation:

$$\lambda \mathbf{v} = \mathbf{C}^\Omega \mathbf{v}. \quad (9)$$

Equation (9) is equivalent to the following equation:

$$\lambda (\Phi(x_k) \cdot \mathbf{v}) = \Phi(x_k) \cdot \mathbf{C}^\Omega \mathbf{v}, \quad k = 1, 2, \dots, m. \quad (10)$$

The solutions \mathbf{v} lie in the span of $\Phi(x_1), \Phi(x_2), \dots, \Phi(x_m)$, but $\alpha_i (i = 1, 2, \dots, m)$ must exist to satisfy the equation as follows:

$$\mathbf{v} = \sum_{i=1}^m \alpha_i \Phi(x_i). \quad (11)$$

Let $\mathbf{K}(x_i, x_j) = \Phi(x_i) \Phi(x_j)$. By combining (8), (9), and (11), we can get

$$m\lambda \boldsymbol{\alpha} = \mathbf{K} \boldsymbol{\alpha}, \quad (12)$$

where $\boldsymbol{\alpha}$ denotes the column vector with $\alpha_i (i = 1, 2, \dots, m)$. By solving (12), we can obtain the vector α_i^k corresponding to λ_k .

To normalize the eigenvectors $\boldsymbol{\alpha}$, we require that

$$1 = \mathbf{v}^k \cdot \mathbf{v}^k = \sum_{i=1}^m \alpha_i \alpha_j \mathbf{K} = \lambda_k (\boldsymbol{\alpha}^k \cdot \boldsymbol{\alpha}^k). \quad (13)$$

The kernel PCs can be calculated as follows:

$$x_k = (\mathbf{v}^k \cdot \Phi(x)) = \sum_{i=1}^m \alpha_i^k \mathbf{K}(x_i, x_j). \quad (14)$$

The number of kernel PCs are determined by the following.

$$\frac{\sum_{i=1}^l \lambda_k}{\sum_{i=1}^m \lambda_k} > 90\% \quad (15)$$

The derivation above is performed on the basis of the assumption that $\sum_{i=1}^m \Phi(x) = 0$. In general, this assumption is difficult to establish, which indicates that x is not guaranteed to be centered in the feature space Ω and needs to be centralized. Let \mathbf{L}_m be a square matrix of order m and $(\mathbf{L}_m)_{ij} = 1/m$ and \mathbf{K} can be centralized and presented as follows:

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{K} \mathbf{L}_m - \mathbf{L}_m \mathbf{K} - \mathbf{K} \mathbf{L}_m \mathbf{K} \quad (16)$$

3.2. Selecting of Kernel Function. The functions that satisfy the Mercer's theorem are kernel functions [27]. Kernel functions are mainly divided into two categories, namely, local and global kernel functions. Interpolation and extrapolation abilities are the key indexes to determine the performance of KPCA, which is affected by the selection of kernel functions and its parameters. Local kernel function has better interpolation ability, whereas global kernel function shows better extrapolation capability. According to this feature, combining local and global kernel functions by using mixtures is a feasible solution to obtain kernels that have good interpolation and extrapolation capabilities.

Furthermore, radial basis function kernel is a typical local kernel. However, the exponential calculation of radial basis function kernel is computationally inefficient. The K-type kernel [28] has been proposed to be an alternative to radial basis function kernel due to its good performance on calculation and interpolation ability, which is expressed as follows:

$$\bar{\mathbf{K}}(x_i, x_j) = \frac{m}{(1 + k^2 \|x_i - x_j\|^2)}, \quad (17)$$

where $m > 0, k > 0$ and can represent the width of the K-type kernel.

The polynomial kernel, a typical global kernel, is expressed as follows:

$$\bar{\mathbf{K}}(x_i, x_j) = [a(x_i, x_j) + 1]^d, \quad (18)$$

where a is the free parameter and d is the order of the polynomial kernel.

Linear and multiplicative combinations are typical method of mixing kernels [29], which can be expressed as follows:

$$\begin{aligned} \bar{\mathbf{K}}(x_i, x_j) &= \mathbf{K}_1(x_i, x_j) + \mathbf{K}_2(x_i, x_j) \\ \bar{\mathbf{K}}(x_i, x_j) &= \alpha \mathbf{K}(x_i, x_j), \end{aligned} \quad (19)$$

where $\alpha > 0$ and $\bar{\mathbf{K}}$ and \mathbf{K} denote the kernel functions. Thus, a hybrid kernel containing K-type kernel and polynomial kernel can be combined as follows:

$$\begin{aligned} \bar{\mathbf{K}} &= \gamma \mathbf{K}_k + (1 - \gamma) \mathbf{K}_{poly} \\ \bar{\mathbf{K}}(x_i, x_j) &= \gamma \left(\frac{m}{(1 + k^2 \|x_i - x_j\|^2)} \right) \\ &\quad + (1 - \gamma) \left([a(x_i, x_j) + 1]^d \right), \end{aligned} \quad (20)$$

where k, m, a, d , and γ are the parameters of the kernel.

3.3. MAFSA. Artificial fish-swarm algorithm (AFSA) is an effective swarm intelligent optimization algorithm [30], which is on the basis of the four behaviors of a natural fish, namely, preying, swarming, following, and random behaviors. In AFSA, each artificial fish allows mutual information communications and independently performs the four behaviors proposed until the global optimal state is obtained. We suppose there are N artificial fish and the position of any artificial fish is X . The food consistence at position X is denoted by $Y = f(X)$ (objective function), where X is the vector in n -dimensional space and n is the number of parameters to be optimized on the basis of practical problems. Let $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ and $X_j = (x_{j1}, x_{j2}, \dots, x_{jn})$, which are the current and next positions of an artificial fish, respectively. Hence, the straight distance between X_i and X_j can be expressed as $D_{ij} = \|X_i - X_j\|$. $STEP$ and $VISUAL$

are defined as the moving step and perception range of each artificial fish, respectively. MAFSA is an improvement of AFSA, which modifies the moving step and perception range in basic AFSA. The basic idea of MAFSA will be described briefly as follows.

(1) *Preying Behavior.* In general, fish gather to the position with higher food concentration. If $Y_i < Y_j$, then the artificial fish will move a $STEP$ from X_i to X_j , which can be expressed as follows:

$$X_j = X_i + STEP \cdot \frac{X_j - X_i}{D_{ij}} \times rand(0, 1). \quad (21)$$

If $Y_i > Y_j$, then the artificial fish will randomly select the next position X_j and determine whether the forward condition is satisfied. Num denotes the number of maximum iteration of the artificial fish. The reattempt times must be less than TRY_Num , or the artificial fish will execute a random behavior.

(2) *Swarming Behavior.* During the moving process of the fish swarm, each artificial fish moves toward the center position X_c of its neighbor artificial fish ($D_{ic} < VISUAL$). The number of the neighbor artificial fish is n_f . δ denotes the crowded degree of the artificial fish swamp. If $Y_c/n_f > \delta Y_i$, then the artificial fish will move a $STEP$ from X_i to X_c , which can be expressed as follows:

$$X_j = X_i + STEP \cdot \frac{X_c - X_i}{D_{ic}} \times rand(0, 1). \quad (22)$$

If $Y_c/n_f < \delta Y_i$, which indicates that the food consistence at center position X_c is low and swarm is relatively crowded, then the artificial fish will perform preying behavior.

(3) *Following Behavior.* In the fish swarm, when an artificial fish finds the area with high food consistence at position X_b , it will quickly be followed by the neighbor artificial fish ($D_{ib} < VISUAL$). n_f denotes the number of the neighbor artificial fish. If $Y_b/n_f > \delta Y_i$, then the artificial fish will move a $STEP$ from X_i to X_b , which can be expressed as follows:

$$X_j = X_i + STEP \cdot \frac{X_b - X_i}{D_{ib}} \times rand(0, 1). \quad (23)$$

If $Y_b/n_f < \delta Y_i$, then the artificial fish will perform preying behavior.

(4) *Random Behavior.* A random behavior is the default of the preying behavior, which indicates that the artificial fish will opt a position randomly in its visual scope and move toward this position. The expression is determined as follows:

$$X_j = X_i + VISUAL \cdot rand(0, 1). \quad (24)$$

(5) *Improvement of the Step and Visual.* In basic AFSA, the moving step and the perception range of the artificial fish are fixed. Relatively large values of $STEP$ and $VISUAL$ make most artificial fish gather to the global optimal solution quickly,

but this may result in the premature convergence. On the contrary, the local search capability of the artificial fish can be enhanced by adopting relatively small values of $STEP$ and $VISUAL$. Consequently, the artificial fish may fall into the local optimum and miss the global optimal solution. To achieve the faster convergence rate and keep the balance between the global and local searches, a variable parameter $\beta(t)$ is introduced to improve the values of the step and visual, which can be presented as follows:

$$\beta(t) = \exp\left(-23 \times \left(\frac{t}{NUM}\right)^k\right)$$

$$VISUAL_t = \beta(t) \times VISUAL + VISUAL_{\min} \quad (25)$$

$$STEP_t = \beta(t) \times STEP + STEP_{\min},$$

where k denotes the attenuation parameters. If the value of k is small, then $\beta(t)$ will significantly decrease at the initial iteration stage. t is the current iteration number. $VISUAL_{\min}$ and $STEP_{\min}$ are the minimum of the perception range and step, respectively.

3.4. Parameter Optimization of Kernel Function on the Basis of MAFSA. The values of the kernel parameters directly affect the performance of KPCA. Therefore, we utilize MAFSA to search the optimal solution to the parameters of different kernels. Here, we consider the contribution rate of the first kernel PC PC_1 as objective function. The purpose is to represent the original temperature data using another set of uncorrelated variables with minimum numbers while simultaneously minimizing the loss of information. The basic procedures of these optimization problems are described as follows.

Step 1. Set the parameters of the artificial fish, namely, N , $VISUAL$, $VISUAL_{\min}$, δ , $STEP$, $STEP_{\min}$, TRY_Num , and Num . Initialize the values of the kernel parameters of each artificial fish. Take the polynomial kernel as an example; express the initial position as $X_i = (a_1, d_1)$, where a_1 and d_1 represent the random values in the searching range.

Step 2. Use KPCA to analyze the temperature dataset obtained from the dam body and calculate the contribution rate of the first kernel PC PC_1 , which can be expressed as $PC_1 = \lambda_{k1} / \sum_{i=1}^m \lambda_k$, where λ_{k1} denotes the largest eigenvalue of all eigenvalues and m denotes the number of the eigenvalues.

Step 3. Perform preying, swarming, following, and random behaviors of the MAFSA.

Step 4. Record the optimal values of the objective function in the bulletin.

Step 5. If terminal conditions of MAFSA are satisfied, then output the optimal values of the kernels and calculate the eigenvector α^{k1} . The temperature variables can be expressed as $T \cdot \alpha^{k1}$, where T denotes temperature dataset. Otherwise, the MAFSA is initiated to proceed to another iteration in Step 2. The number of total iterations is Num .

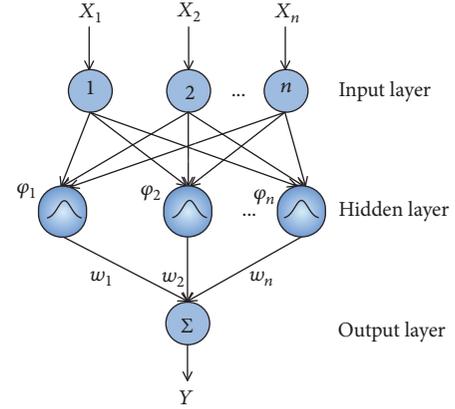


FIGURE 1: Topology of RBF neural network.

4. Dam Safety Monitoring Model Using RBF-NN

4.1. RBF Neural Network. RBF-NN is one of the most popular neural networks and extensively used in several fields, especially in classification and regression. The RBF-NN is a three-layer neural network [31], which includes input, hidden, and output layers. Similar to other feed-forward neural networks, RBF-NN is composed of large quantities of interconnected artificial neurons. Every neuron in the radial basis function network is completely connected with each neuron of the next layer. The topology of RBF-NN is shown in Figure 1. In comparison with the BP-NN, RBF-NN can approach the nonlinear continuous function with arbitrary precision and avoid falling into the local minimum point effectively.

RBF-NN contains the following important features.

(1) The input domain contains the center vector denoted by C_{jp} which is also called cluster center. C_{jp} is stored as a weight factor from the input layer to the hidden layer. Euclidean distance, typically represented as $ED_j = \|x_p - C_{jp}\|$, is used to measure the distance of the input sample x_p from the cluster center C_{jp} , where $j = 1, 2, \dots, M$, $p = 1, 2, \dots, P$, M is the total number of data centers, and P is the total number of samples. The clustering center C_{jp} is determined by K-means clustering algorithm.

(2) Gaussian function is typically selected as the hidden layer transfer function, which is expressed as follows:

$$\varphi(r)_j = \exp\left(-\frac{ED_j^2}{2\sigma^2}\right), \quad (26)$$

where σ is the spread constant of the radial basis function, which is denoted by $\sigma = (ED_j)_{\max} / \sqrt{M}$. $(ED_j)_{\max}$ is the maximum Euclidean distance among selected centers.

(3) The output of the RBF-NN is represented by the following equation:

$$y_i = \sum_{j=1}^N w_{ij} \varphi(r)_j, \quad i = 1, 2, \dots, N, \quad (27)$$

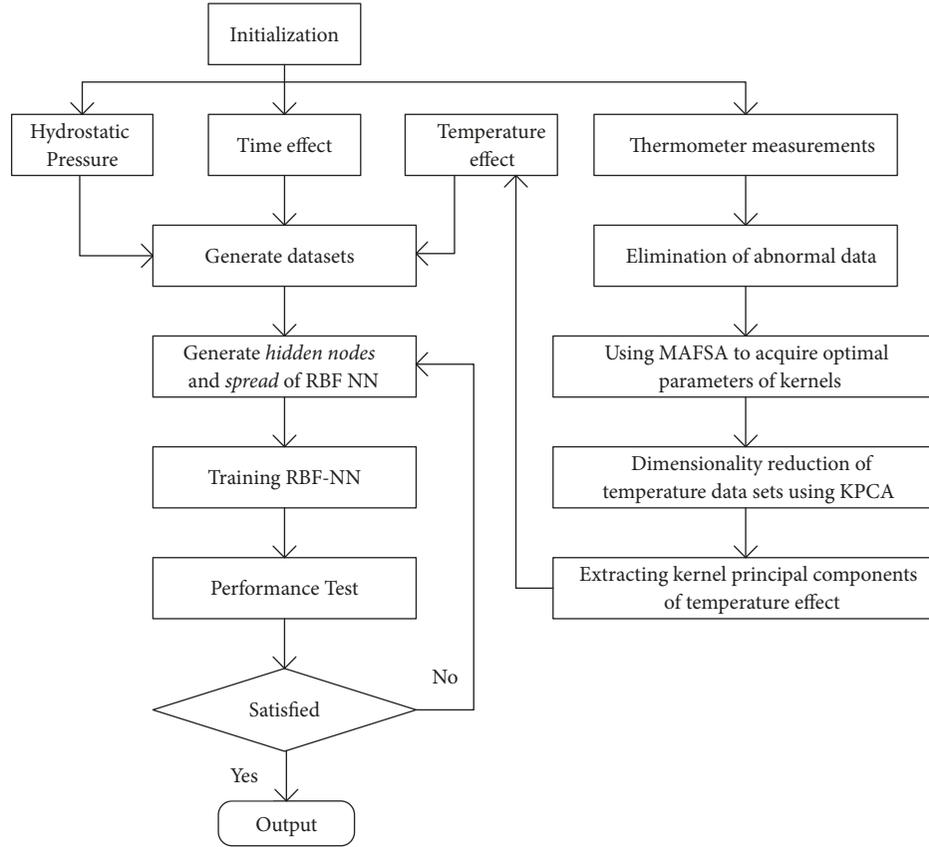


FIGURE 2: Flowchart of the developed model for dam safety monitoring.

where w_{ij} is the weight of the hidden layer to the output layer w_{ij} , w_{ij} is determined by gradient descent algorithm [32], and N is the total number of hidden layer nodes.

4.2. Procedure of the Model. RBF-NN is adopted to construct the dam displacement prediction model. Hydrostatic pressure, time effect, and kernel PC of temperature are selected as the input variables of the RBF-NN model. The output of the model is the radial displacement of the single or multiple measuring points of the dam. In the structure of RBF-NN, two important tuning parameters are used, namely, the maximum number of hidden layers and the spread. The performance and the prediction accuracy of the developed model depend on the values of these two parameters. In this paper, trial and error is utilized to evaluate the optimum values of these two parameters to obtain a precise and reliable performance for the model. The developed model is implemented on the MATLAB platform, and the main steps are as follows.

Step 1. Collect data from concrete dam monitoring system.

Step 2. Build the training and testing sets of the RBF-NN model establishment.

Step 3. Set the training error threshold. Carry out the trial and error approach by changing the values of hidden nodes

and the spread and evaluating the overall root mean square error (RMSE) of the model.

Step 4. Describe the process of the trial and error approach by using diagram, and record the best values of the hidden nodes and the spread.

Step 5. Establish the model using the best parameters obtained in Step 4 and evaluate the results with several performance criteria.

The flowchart of the model is shown in Figure 2. The performance criteria in Step 5 include the decision coefficient (R^2), root mean square error (RMSE), average absolute error (MAE), and maximum absolute error (E), which are expressed as follows:

$$R^2 = \frac{[\sum_{i=1}^N (y_S(i) - \bar{y}_S)(y(i) - \bar{y})]^2}{\sum_{i=1}^N (y_S(i) - \bar{y}_S)^2 \sum_{i=1}^N (y(i) - \bar{y})^2} \quad (28)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_S(i) - y(i))^2} \quad (29)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_S(i) - y(i)| \quad (30)$$

$$E = \max |y_S(i) - y(i)| \quad (31)$$

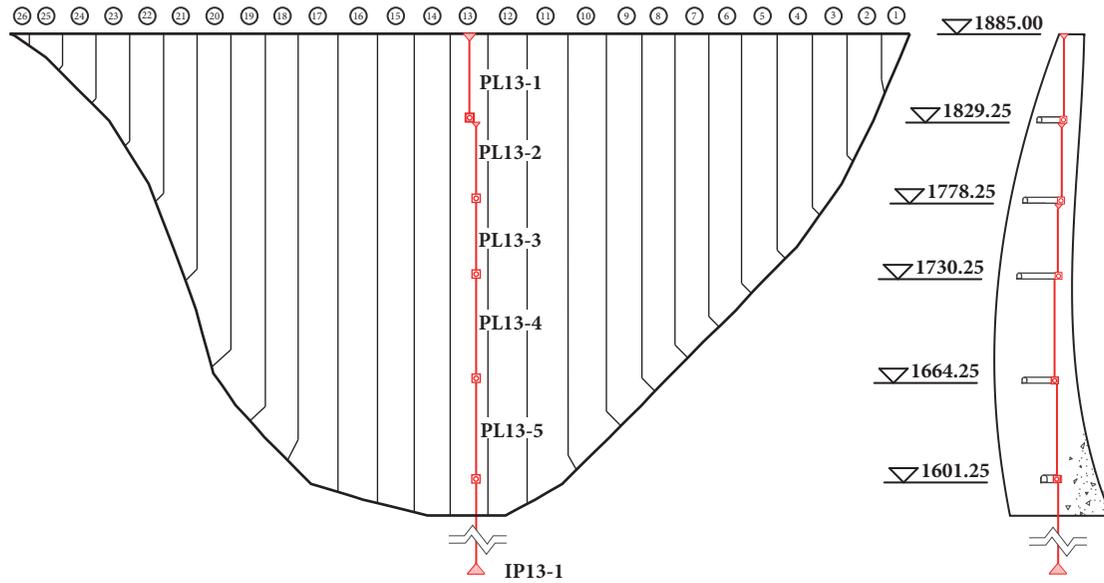


FIGURE 3: Distribution of pendulums and thermometers in Dam Section 13#.

where $y_s(i)$ and $y(i)$ represent the simulated and measured values of the model, respectively ($i = 1, 2, \dots, N$); \bar{y}_s and \bar{y} denote the simulated and measured average values, respectively; and N denotes the number of observations. Larger R^2 and smaller $RMSE$, MAE , and E indicate that the model has better performance.

5. Application Example

5.1. Project Profile. A double-curvature super-high concrete arch dam is located downstream on the Yalong River in Sichuan Province, China. The maximum dam height above ground level is 305 m and the elevation of the dam crest is 1885 m. The dam consists of 26 dam sections, with a dam crest length of 552 m. The width of the dam crest and bottom is 16 m and 63 m, respectively. The storage capacity of the reservoir is 7.76 billion m^3 . The normal reservoir level is about 1880 m above sea level. The dam construction began in 2005 and was completed in 2014, with its main purposes as hydroelectric power generation and flood control.

The dam automatic monitoring system is composed of several types of instruments, including water level gauges, pendulums, thermometers, strain gauges, piezometers, and osmometers. In this paper, radial displacement data acquired from the PL13-2 direct pendulum is studied. As is shown in Figure 3, the #13 dam section is located in the middle of the valley, and PL13-2 direct pendulum is equipped in the upper part of the #13 dam section where the radial displacement range is relatively large. Furthermore, the monitoring of data is continuous, which is helpful for modeling and analysis. The dataset for modeling is established from July 2013 to November 2016, which accumulated a total of 175 groups of data.

5.2. Temperature Variables. The thermometers installed in #13 dam section are approximately distributed within the

range of 1620–1881 m in elevation. The temperature data collected from 45 thermometers, where the data are continuous, is studied and analyzed. To test the proposed KPCA's performance on feature extraction and dimensionality reduction of input temperature dataset, the comparison with PCA is carried out.

5.2.1. Temperature Variables Calculated by PCA. By using PCA to analyze the temperature data matrix, five PCs are obtained. Figure 4 depicts the five PCs of input temperature dataset, and the proportion of the total PCs is illustrated in Figure 5. These five PCs, which do not consider approximately 75% of the information contained in the original temperature dataset, are considered to represent the temperature effect and selected as the input variables of the model. The PC1, which has the maximum contribution rate among 45 principle components, explains only 49% of the information contained in the original temperature dataset.

5.2.2. Temperature Variables Calculated by KPCA. K-type, polynomial, and proposed hybrid kernels are adopted to construct KPCA. The MAFSA is used to search the optimal parameters of each kernel. Table 1 shows the initial parameters of MAFSA. Figure 6 and Table 2 show the optimization results of the kernels. Figure 7 shows the temperature PCs calculated by K-type KPCA, Polynomial KPCA, and the proposed Hybrid-KPCA. On the basis of the results, we can find that the optimization result is stable within 20 iterations, and the KPCA constructed with the three kernels optimized by MAFSA has good performance on feature extraction. The contribution rate of the first PC calculated by proposed KPCA all exceeds 85%, which is significantly larger than that in PCA.

5.3. Model Establishment. As described in Sections 2 and 3, the hydrostatic pressure, temperature, and time effect variables are determined as independent variables, whereas the

TABLE 1: Parameters of MAFSA.

| Description | Formula | Values |
|--|----------------|--------|
| Number of artificial fish | N | 20 |
| Vision distance of artificial fish | $VISUAL$ | 1.0 |
| Minimum vision distance of artificial fish | $VISUAL_{min}$ | 0.1 |
| Crowded degree | δ | 0.618 |
| Step of artificial fish | $STEP$ | 0.5 |
| Minimum step of artificial fish | $STEP_{min}$ | 0.1 |
| Number of preying behaviors | TRY_Num | 100 |
| Number of total iterations | Num | 20 |

TABLE 2: Optimization results of kernels calculated by MAFSA.

| Kernel | Optimal parameters of kernel | Maximum contribution rate of the first kernel PC PC_1 (%) |
|-------------------|--|---|
| K-type kernel | $k = 0.100; m = 3.590$ | 85.5 |
| Polynomial kernel | $a = 1.326; d = 0.987$ | 99.7 |
| Hybrid kernel | $k = 0.100; m = 1.000; a = 1.140; d = 0.998; \gamma = 0.802$ | 99.1 |

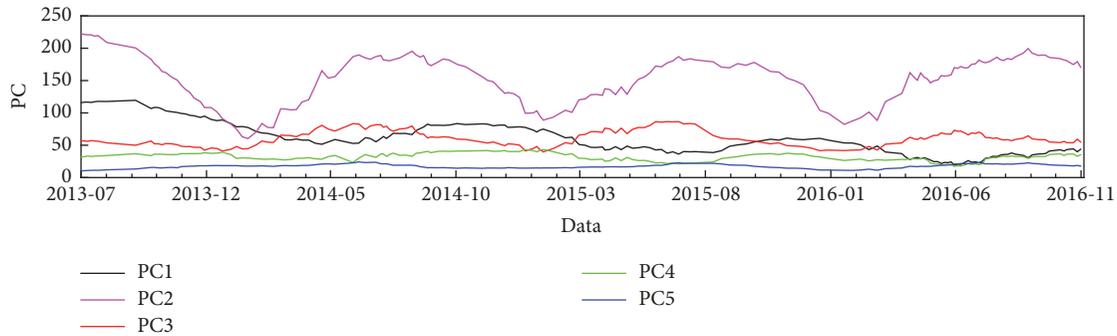


FIGURE 4: Temperature component calculated by PCA.

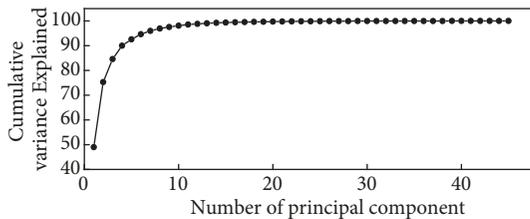


FIGURE 5: Proportion of the total of variables calculated by PCA.

measured radial displacements from 13-2 point are selected as dependent variables. Hydrostatic pressure variables contain four factors, namely, H , H^2 , H^3 , and H^4 . Meanwhile, the time effect variables have four factors, namely, θ , $1 - e^{-\theta}$, $\ln(\theta + 1)$, and $\theta/(\theta + 1)$. Figures 8 and 9 show the measured upstream water level and measured radial displacements from 13-2 point, respectively. In Figure 9, negative values denote the radial displacements toward the upstream. The RBF-NN is used as a framework to build the model. Therefore, the dependent variables are selected as inputs, and independent variables are selected as output of the RBF-NN.

The first 160 groups are utilized for training, and the remaining 15 samples are adopted as testing set. Here, we

establish four safety monitoring models on the basis of RBF-NN that involve the same hydrostatic pressure, time effect, and different temperature variables. Model 1 contains a total of thirteen input variables, of which the temperature variables are calculated using PCA and involve five factors. Models 2, 3, and 4 each contain nine input variables, and their temperature variables are calculated by utilizing K-type KPCA, Polynomial KPCA, and the proposed Hybrid-KPCA, respectively.

For these four safety monitoring models on the basis of RBF-NN, the training objective of the mean square error is set to 10^{-4} . The values of spread and hidden nodes are determined using trial and error. The value of spread is selected in the interval $[1, 3]$ with a minimum step of 0.01, whereas the value of hidden layer nodes is selected in the interval $[10, 40]$ with a minimum step of 1. Figure 10 shows the effects of the values of spread and hidden nodes on the performance of these four models, and the optimum values are summarized in Table 3. In Figure 10, the errors of models 3 and 4 are relatively smaller than the others. Moreover, the error variation range of model 4 is the smallest.

5.4. Performance Evaluations and Discussions. To graphically and statistically evaluate the performance of the proposed

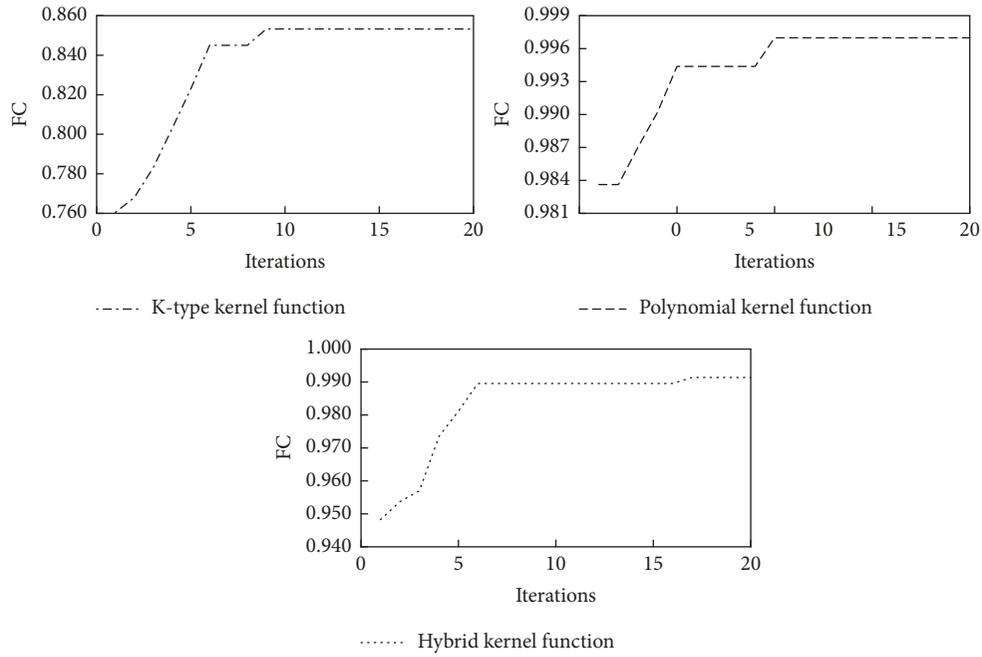


FIGURE 6: PC1 optimization process using MAFSA.

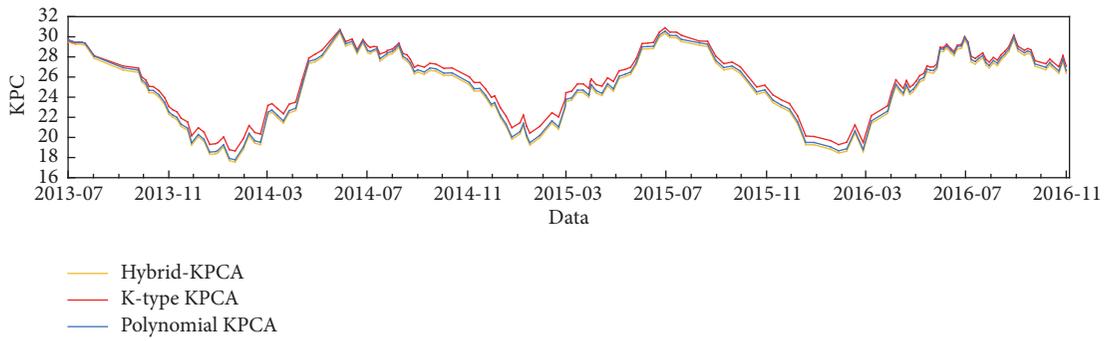


FIGURE 7: Temperature component calculated by KPCA.

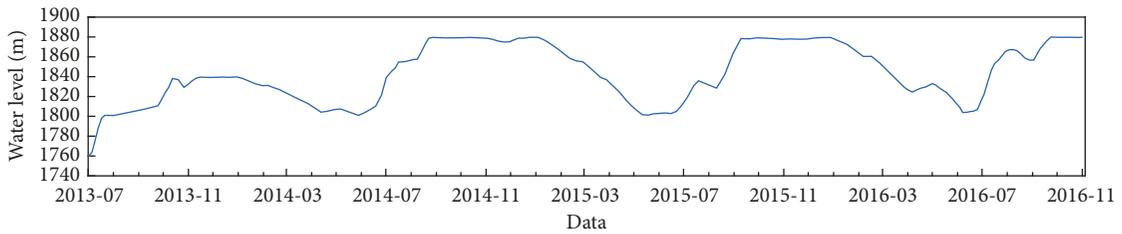


FIGURE 8: Measured upstream water level (H).

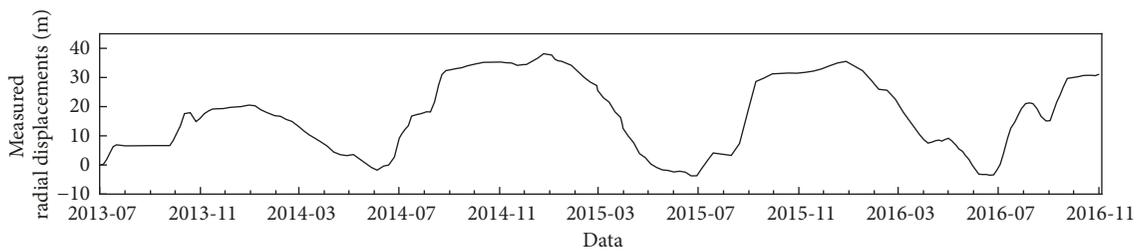


FIGURE 9: Measured radial displacements.

TABLE 3: Parameters of RBF-NN.

| Model | Input layer | Spread | Hidden nodes | Minimum RMSE |
|-------------------------------------|-------------|--------|--------------|--------------|
| Model 1: PCA and RBF-NN | 13 | 2.40 | 28 | 1.134 |
| Model 2: K-type KPCA and RBF-NN | 9 | 1.11 | 36 | 1.138 |
| Model 3: Polynomial KPCA and RBF-NN | 9 | 1.13 | 36 | 0.853 |
| Model 4: Hybrid-KPCA and RBF-NN | 9 | 1.33 | 36 | 0.819 |

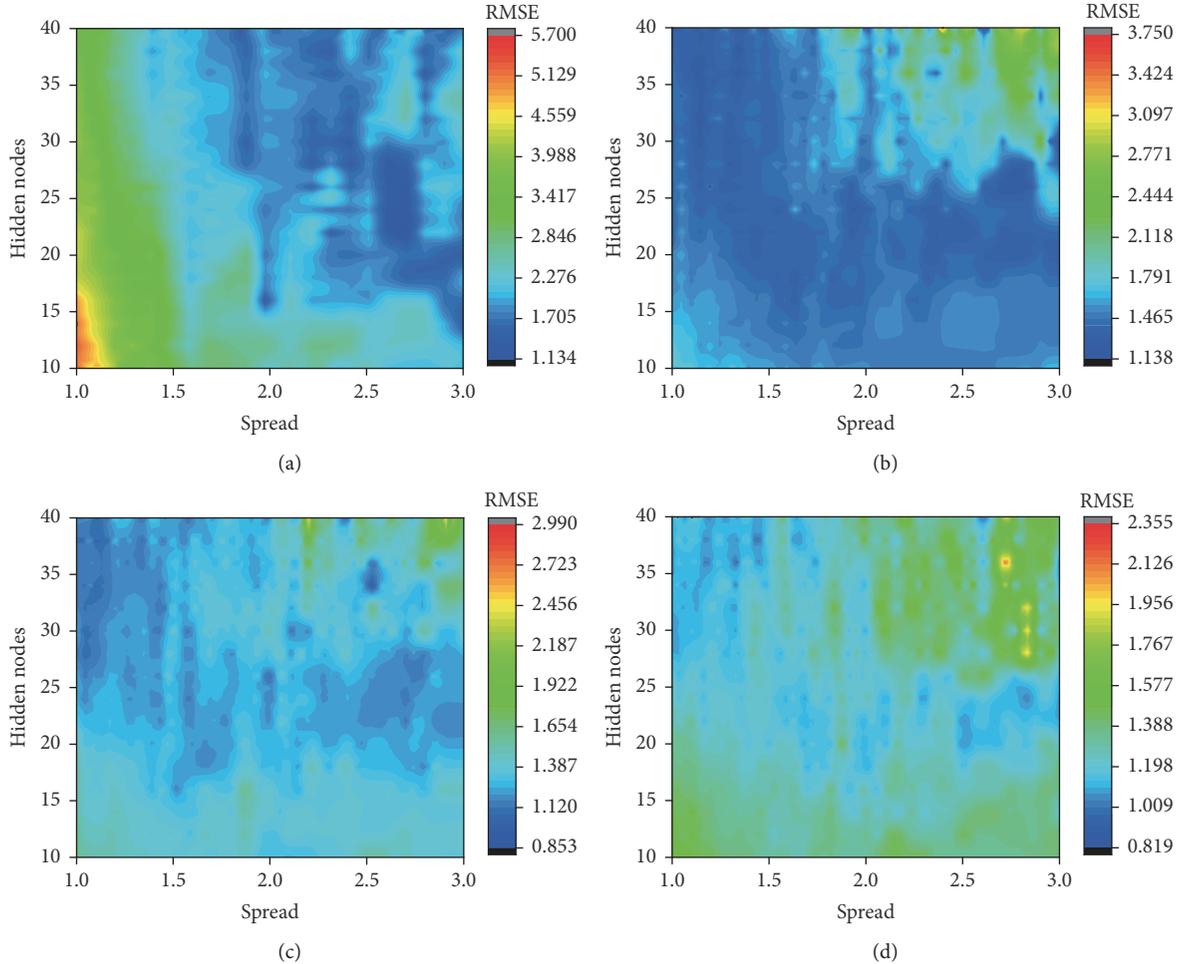


FIGURE 10: Effects of the values of spread and number of hidden nodes to the models performance: (a) RBF-NN model coupled with PCA; (b) RBF-NN model coupled with K-type KPCA; (c) RBF-NN model coupled with Polynomial KPCA; (d) RBF-NN model coupled with Hybrid-KPCA.

models, comparisons among the measured, fitted, and predicted values are firstly carried out by plotting from Figures 11–14. Then, the values of performance criteria for the four models are listed in Table 4, and the best results are shown in boldface. The obtained evaluation conclusions can be summarized as follows.

(1) These four models' decision coefficients (R^2) are all larger than 0.99. In addition, the four models are all capable of fitting and predicting the trend of radial displacements accurately due to the good performance of RBF-NN.

(2) The predicted deviations of Polynomial KPCA-RBF-NN model and Hybrid-KPCA-RBF-NN model are relatively smaller than PCA-RBF-NN model and K-type KPCA-RBF-NN model. In Table 4, we can find that K-type KPCA-RBF model obtains the smallest absolute error (E), average

root mean square error ($RMSE$), and average absolute error (MAE). Although the fitting performance of K-type KPCA-RBF-NN model is not the best, its predictive performance is better than PCA-RBF-NN model.

(3) The prediction results of PCA-RBF-NN model are more unstable than the other models, which may be due to the loss of nonlinear feature in the temperature variables by using PCA.

6. Conclusion

On the basis of the traditional safety monitoring statistical model's establishment theory, RBF-NN is combined with other methods, such as kernel principle component analysis (KPCA) and modified artificial fish-swarm algorithm

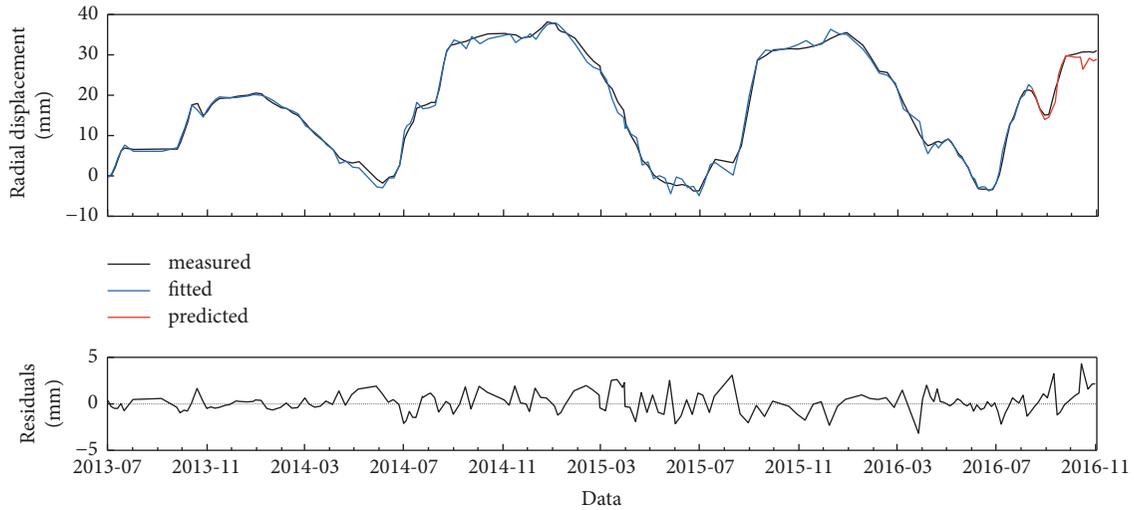


FIGURE 11: Performance of the RBF-NN model coupled with PCA (model 1).

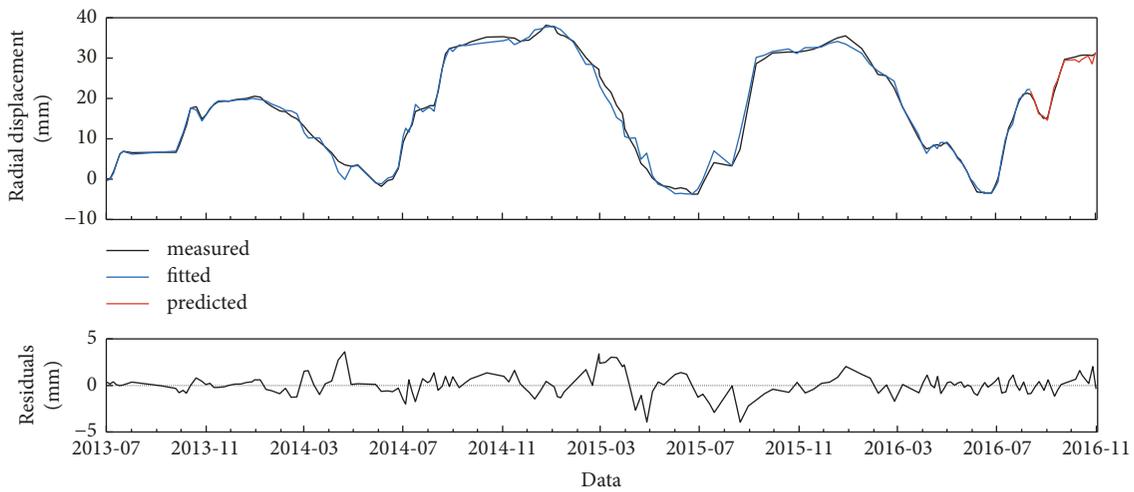


FIGURE 12: Performance of the RBF-NN model coupled with K-type KPCA (model 2).

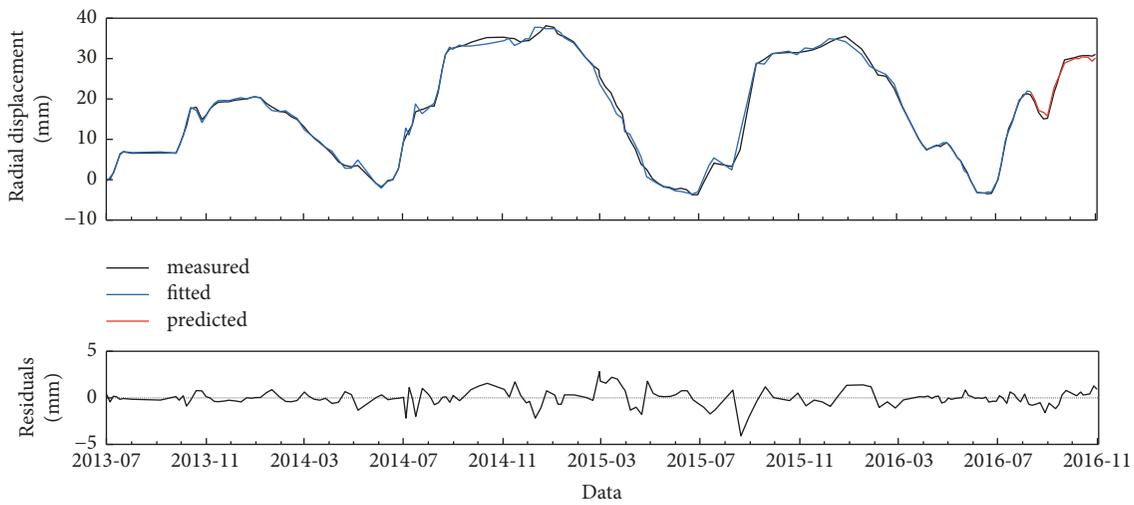


FIGURE 13: Performance of the RBF-NN model coupled with Polynomial KPCA (model 3).

TABLE 4: Comparison of prediction error for the models.

| Model | R^2 | | | RMSE | | | MAE | | | E | |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Training | Testing | Average | Training | Testing | Average | Training | Testing | Average | Training | Testing |
| 1 | 0.993 | 0.937 | 0.992 | 1.054 | 1.780 | 1.134 | 0.799 | 1.382 | 0.849 | 3.167 | 4.306 |
| 2 | 0.991 | 0.982 | 0.992 | 1.159 | 0.884 | 1.138 | 0.817 | 0.699 | 0.807 | 3.962 | 2.035 |
| 3 | 0.995 | 0.994 | 0.995 | 0.856 | 0.819 | 0.853 | 0.585 | 0.730 | 0.598 | 4.098 | 1.281 |
| 4 | 0.996 | 0.985 | 0.996 | 0.813 | 0.885 | 0.819 | 0.578 | 0.694 | 0.588 | 2.937 | 0.969 |

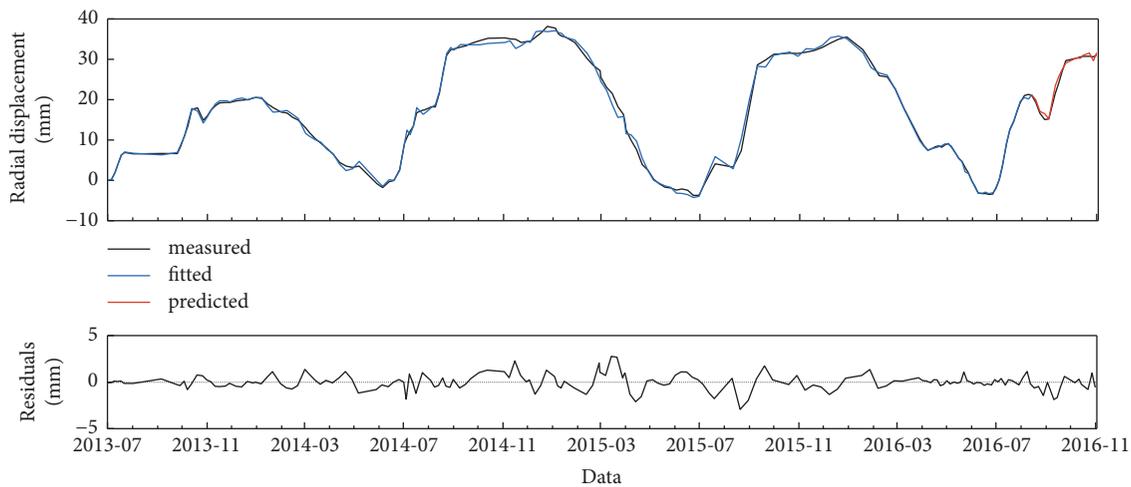


FIGURE 14: Performance of the RBF-NN model coupled with Hybrid-KPCA (model 4).

(MAFSA), to build the precise monitoring model of dam deformation. The temperature variables selection methods of super-high concrete dam safety monitoring model are discussed. The key problems on kernel principle component analysis, namely, the kernel function selection and the kernel parameter optimization, are investigated. The monitoring models of a super-high concrete dam in Southwestern China are built using the proposed methods in this paper and other methods.

(1) The KPCA, with three different kernels optimized by MAFSA, are applied to construct input temperature variables of super-high concrete dam for the purpose of capturing the nonlinear features and minimizing the information loss from temperature dataset. The results assert that the proposed optimized KPCA methods have better performance in dimensionality reduction of temperature dataset than PCA.

(2) Four safety monitoring models that their temperature variables are calculated by utilizing PCA, K-type KPCA, Polynomial KPCA, and the proposed Hybrid-KPCA are used on the basis of RBF-NN. The comparison of four different RBF-NN based monitoring models is carried out and results indicate that RBF-NN and the proposed Hybrid-KPCA are a good combination that has the best performance among the models. The validity and accuracy of the model also have been verified by various statistical and graphical presentations.

(3) In the future, the proposed model can be extended to the online safety monitoring for the displacement and other structural behaviors of super-high concrete dams.

Data Availability

The data used to support the findings of this study have not been made available.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the research funds provided by National Key R&D Program of China (2016YFC0401601, 2017YFC0804607, and 2018YFC0407104), National Natural Science Foundation of China (Grant nos. 51739003, 51479054, 51779086, 51379068, and 51579083), Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (YS11001), Jiangsu Natural Science Foundation (Grant no. BK20160872), Special Project Funded of National Key Laboratory (20145027612 and 20165042112), and Key R&D Program of Guangxi (AB17195074).

References

- [1] *World Register of Dams*, <http://www.icold-cigb.org>.
- [2] Z. Wu, *Safety Monitoring, Theory and Its Application of Hydraulic Structures*, Higher Education Press, Beijing, China, 2003.

- [3] B. Stojanovic, M. Milivojevic, M. Ivanovic, N. Milivojevic, and D. Divac, "Adaptive system for dam behavior modeling based on linear regression and genetic algorithms," *Advances in Engineering Software*, vol. 65, pp. 182–190, 2013.
- [4] P. L ger and S. Seydou, "Seasonal thermal displacements of gravity dams located in northern regions," *Journal of Performance of Constructed Facilities*, vol. 23, no. 3, pp. 166–174, 2009.
- [5] J. Mata, A. Tavares de Castro, and J. S  da Costa, "Constructing statistical models for arch dam deformation," *Structural Control and Health Monitoring*, vol. 21, no. 3, pp. 423–427, 2014.
- [6] J. Mata, "Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models," *Engineering Structures*, vol. 33, no. 3, pp. 903–910, 2011.
- [7] C. Shen and Z. Wu, "Statistical model for uplift on concrete dam foundations," *Journal of Hohai University*, vol. 4, pp. 261–262, 1989.
- [8] W. W. Shen and J. M. Ren, "Multiple stepwise regression analysis crack open degree data in gravity dam," *Applied Mechanics and Materials*, vol. 477–478, pp. 888–891, 2014.
- [9] C. S. Gu, B. O. Li, G. L. Xu, and H. Yu, "Back analysis of mechanical parameters of roller compacted concrete dam," *Science China Technological Sciences*, vol. 53, no. 3, pp. 848–853, 2010.
- [10] B. Li, C. Gu, Z. Li, and Z. Zhang, "Monitoring model for dam seepage based on partial least-squares regression and partial least square support vector machine," *Journal of Hydraulic Engineering*, vol. 39, no. 12, pp. 1390–1400, 2008.
- [11] Hao Huang, Bo Chen, and Chungao Liu, "Safety Monitoring of a Super-High Dam Using Optimal Kernel Partial Least Squares," *Mathematical Problems in Engineering*, vol. 2015, Article ID 571594, 13 pages, 2015.
- [12] J. Qiu, D. Zheng, and K. Zhu, "Seepage Monitoring Models Study of Earth-Rock Dams Influenced by Rainstorms," *Mathematical Problems in Engineering*, vol. 2016, Article ID 1656738, 11 pages, 2016.
- [13] T. Bao, D. Qin, X. Zhou, and G. Wu, "Abnormality monitoring model of cracks in concrete dams," *Science China Technological Sciences*, vol. 54, no. 7, pp. 1914–1922, 2011.
- [14] Y. Xiang, S.-Y. Fu, K. Zhu, H. Yuan, and Z.-Y. Fang, "Seepage safety monitoring model for an earth rock dam under influence of high-impact typhoons based on particle swarm optimization algorithm," *Water Science and Engineering*, vol. 10, no. 1, pp. 70–77, 2017.
- [15] G. W. Housner, L. A. Bergman, T. K. Caughey et al., "Structural control: past, present, and future," *Journal of Engineering Mechanics*, vol. 123, no. 9, pp. 897–971, 1997.
- [16] C. Y. Kao and C. H. Loh, "Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches," *Structural Control and Health Monitoring*, vol. 20, no. 3, pp. 282–303, 2013.
- [17] Y. Zhu, C. I. Gu, E. Zhao, J. Song, and Z. Guo, "Structural Safety Monitoring of High Arch Dam Using Improved ABC-BP Model," *Mathematical Problems in Engineering*, vol. 2016, Article ID 6858697, 9 pages, 2016.
- [18] X. Wang, K. Yang, and C. Shen, "Study on MPGA-BP of Gravity Dam Deformation Prediction," *Mathematical Problems in Engineering*, vol. 2017, Article ID 2586107, 13 pages, 2017.
- [19] F. Kang, J. Liu, J. Li, and S. Li, "Concrete dam deformation prediction model for health monitoring based on extreme learning machine," *Structural Control and Health Monitoring*, vol. 24, no. 10, 2017.
- [20] D. Tonini, "Observed behavior of several Italian arch dams," *Journal of the Power Division*, vol. 82, no. 6, p. 26, 1956.
- [21] R. Ardito, G. Maier, and G. Massalongo, "Diagnostic analysis of concrete dams based on seasonal hydrostatic loading," *Engineering Structures*, vol. 30, no. 11, pp. 3176–3185, 2008.
- [22] P. L ger and M. Leclerc, "Hydrostatic, temperature, time-displacement model for concrete dams," *Journal of Engineering Mechanics*, vol. 133, no. 3, pp. 267–277, 2007.
- [23] H. Yu, Z. R. Wu, T. F. Bao, and L. Zhang, "Multivariate analysis in dam monitoring data with PCA," *Science China Technological Sciences*, vol. 53, no. 4, pp. 1088–1097, 2010.
- [24] Z. Wu, C. Gu, H. Su, and B. Chen, "Review and prospect of calculation analysis methods in hydro-structure engineering," *Journal of Hohai University*, vol. 43, no. 5, pp. 395–405, 2015.
- [25] C. Gu and Z. Wu, *Safety Monitoring of Dams and Dam Foundations-Theories & Methods and Their Application*, Hohai University Press, Nanjing, China, 2006.
- [26] B. Sch olkopf, A. Smola, and K.-R. M ller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [27] N. Deng and Y. Tian, *New Method in Data Mining Support Vector Machine*, Science Press, Beijing, China, 2004.
- [28] C. Sun, "Support vector machine based K-type kernel function," *Journal of Huaihai Institute of Technology*, vol. 15, no. 4, pp. 4–7, 2006.
- [29] G. F. Smits and E. M. Jordaan, "Improved SVM regression using mixtures of kernels," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '02)*, vol. 3, pp. 2785–2790, May 2002.
- [30] X. Li, Z. Shao, and J. Qian, "An optimizing method based on autonomous animals: Fish-Swarm Algorithm," *Systems Engineering - Theory & Practice*, vol. 11, no. 22, pp. 32–38, 2002.
- [31] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 2014.
- [32] S. S. Panda, D. Chakraborty, and S. K. Pal, "Flank wear prediction in drilling using back propagation neural network and radial basis function network," *Applied Soft Computing*, vol. 8, no. 2, pp. 858–871, 2008.

