

Research Article

A Nonlinear Method for Component Separation of Dam Effect Quantities Using Kernel Partial Least Squares and Pseudosamples

Wenzhong Yin,^{1,2,3} Erfeng Zhao ^{1,2,3} Chongshi Gu ^{1,2,3} Hao Huang,^{1,2,3} and Yang Yang⁴

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

²State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

³National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, Nanjing 210098, China

⁴Dam Safety Management Department, Nanjing Hydraulic Research Institute, Nanjing 210029, China

Correspondence should be addressed to Erfeng Zhao; zhaofeng@hhu.edu.cn

Received 8 May 2019; Accepted 18 November 2019; Published 23 December 2019

Academic Editor: Venu G. M. Annamdas

Copyright © 2019 Wenzhong Yin 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.

Existing component separation methods fail to consider the complex nonlinear relationship between dam effect quantities and environmental variables. In this study, a novel nonlinear component separation method for the effect quantities is proposed by combining kernel partial least squares (KPLS) and pseudosamples. By this method, a nonlinear monitoring model is established based on KPLS, and the complicated nonlinear relationship between the effect quantities and environmental variables can be determined accurately through the model. Furthermore, special pseudosamples are constructed to separate independent components and coupling influence components of environmental factors from the KPLS model. These methods have been applied into a super-high arch dam, and the separated displacement components conform to the general deformation law. The presented results indicate that it is more reliable than traditional multiple linear regression models.

1. Introduction

Dams provide substantial comprehensive benefits, including power generation, flood control, and irrigation. Dam safety significantly affects the security of personal property and ecological environment in the surrounding areas. Based on prototype monitoring data, dam safety monitoring models can be useful to the supervision and management of long-term operation to ensure dam safety operation [1–3]. Most models predict dam effect quantities (displacement, seepage pressure, and stress) and compare with the observation to identify possible abnormalities. In addition, some models separate the components of influence factors (hydrostatic pressure, temperature, and time effect) of dam effect quantities and physically interpret these components, especially the component of the time effect, to appraise the practical service status during long-term operation.

The most popular dam monitoring model is the multiple linear regression (MLR) model [1, 2, 4–11]. A common feature of the MLR model is that dam measured behavior is

regarded as a linear combination of its influence factors, and model coefficients are fitted by ordinary least squares. However, two key drawbacks of the MLR model are multicollinearity and linearity. The multicollinearity refers to the complete or approximate linear correlation that exists between the influence variables. For example, the multicollinearity may exist between the reservoir level and temperature given seasonal operations of the reservoir, among different terms of the polynomial function to describe the effect of hydrostatic pressure [12] or among influence factors and their lagged variables when considering delayed effects. Linearity refers to the effects of different influence factors that are considered independent and satisfy the superposition principles. However, nonlinear interactions between influence factors are measured. For example, the reservoir level affects the thermal response of the dam because of different air and water temperatures [10, 13, 14]. In addition, the loads of hydrostatic pressure and temperature increase the creep of concrete dams, which is a kind of time effect. The multicollinearity and linearity may lead to poor

prediction accuracy and misinterpretation of dam behavior given the inaccurate or incorrect separation of the influence components [2, 12, 15, 16]. Considering the multicollinearity and linearity of the MLR model, many theories and methods are used to establish improved models. On the one hand, stepwise regression method [17], principal component analysis (PCA) [18], partial least squares (PLS) [19], and panel data theory [20] are used to address the multicollinearity. On the other hand, machine learning is also used to establish nonlinear monitoring models that can capture complex interactions among influence factors, such as neural network (NN) models [16, 21–24], and support vector regression models [18, 19, 25–27]. These nonlinear models own better prediction accuracy than the MLR model, but the components of the influence factors of dam behavior are difficult to separate and physically interpret using these nonlinear models due to the lack of explicit expressions for each component. Mata [16] used the ceteris paribus analysis to interpret the effects of hydrostatic pressure and temperature on dam displacement in the NN model, but the key time effect was ignored.

Kernel PLS (KPLS) [28] is a new nonlinear PLS to address nonlinear problems. In KPLS, the original input data are nonlinearly transformed into a high-dimensional space via a kernel function, and then a linear PLS model is created in the high-dimensional space. The linear relationship obtained by PLS in the high-dimensional space corresponds to the nonlinear relationship in the original input space. KPLS not only retains all of the advantages of PLS, but also has strong nonlinear mapping ability. Thus, KPLS can overcome the drawbacks of multicollinearity and linearity of the MLR model. Huang et al. [29] used KPLS to establish a safety monitoring model of super-high dams to indicate that the KPLS model had better prediction accuracy than MLR and PLS models. In order to address the black box problem in kernel function of KPLS, Postma et al. [30] obtained a column of response trace of the model through constructing a column of special pseudosamples, thus realizing a visual display of the nonlinear relation between the original input and the output of the model and disclosing important input variables. The outcome provides a basis for further simplification of models.

In this study, a nonlinear component separation method for dam effect quantities is proposed by combining KPLS and pseudosamples. The remainder of this paper is organized as follows. Sections 2 and 3 describe the method of KPLS and the principles of constructing the pseudosamples, respectively. Section 4 introduces the nonlinear method for combining the KPLS and the pseudosamples to separate the components of dam effect quantities. Section 5 provides an application to verify the feasibility and rationality of the proposed method.

2. KPLS Method

Assume that there are m independent variables $\{x_1, x_2, \dots, x_m\}$ and p dependent variables $\{y_1, y_2, \dots, y_p\}$, and the number of training samples is N , which constitute independent variables matrix $\mathbf{X}_{N \times m}$ and dependent variables matrix $\mathbf{Y}_{N \times p}$. The PLS algorithm first obtains A mutually unrelated implicit variables $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A\}$ and $\{\mathbf{u}_1, \mathbf{u}_2,$

$\dots, \mathbf{u}_A\}$ from the independent and dependent variables by linear mapping, respectively, and then establishes a linear relationship between $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A\}$ and $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_A\}$. However, the linear PLS cannot simulate a system with complicated nonlinear relationships between independent variables and dependent variables. KPLS transforms the original nonlinear data into linear in high-dimensional feature space through kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ ($k(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$) and then constructs the linear PLS model in high-dimensional feature space, as shown in Figure 1.

The KPLS regression model is described as follows:

$$\mathbf{Y} = \Phi \mathbf{B} + \varepsilon, \quad (1)$$

where \mathbf{B} is a $(c \times p)$ matrix of the regression coefficients and ε is a $(N \times p)$ matrix of residuals.

By introducing a kernel function, it is possible to avoid calculating a specific nonlinear mapping and dot product operations in the feature space. The Gram kernel matrix $\mathbf{K}_{N \times N}$ ($\mathbf{K} = \Phi \Phi^T$) is composed of the cross inner product of all mapping vectors. The KPLS algorithm [31] is expressed as follows (Algorithm 1):

Based on the KPLS algorithm, the score matrix $\mathbf{T}_{N \times A} = [\mathbf{t}_1, \dots, \mathbf{t}_A]$ of the matrix $\Phi_{N \times c}$ and $\mathbf{U}_{N \times A} = [\mathbf{u}_1, \dots, \mathbf{u}_A]$ of the matrix $\mathbf{Y}_{N \times p}$ can be determined. Both two matrixes are orthogonal in accordance with columns. The expression for calculating the regression coefficient matrix \mathbf{B} and the fitted values is as follows:

$$\mathbf{B} = \Phi^T \mathbf{U} (\mathbf{T}^T \mathbf{K} \mathbf{U})^{-1} \mathbf{T}^T \mathbf{Y}, \quad (2)$$

$$\hat{\mathbf{Y}} = \Phi \mathbf{B} = \mathbf{K} \mathbf{U} (\mathbf{T}^T \mathbf{K} \mathbf{U})^{-1} \mathbf{T}^T \mathbf{Y}. \quad (3)$$

The predicted values of output can be calculated using

$$\hat{y} = \mathbf{B} \varphi(\mathbf{x}) = \mathbf{Y}^T \mathbf{T} [\mathbf{U} (\mathbf{T}^T \mathbf{K} \mathbf{U})^{-1}]^T \mathbf{k}(\mathbf{x}). \quad (4)$$

It should be pointed out that before the KPLS operation, the data needs to be centralized in the high-dimensional feature space. This process can be accomplished by replacing \mathbf{K} and $\mathbf{k}(\mathbf{x})$ by $\bar{\mathbf{K}}$ and $\bar{\mathbf{k}}(\mathbf{x})$ in the following equations, correspondingly:

$$\bar{\mathbf{K}} = (\mathbf{I} - \mathbf{E}) \mathbf{K} (\mathbf{I} - \mathbf{E}), \quad (5)$$

$$\bar{\mathbf{k}}(\mathbf{x}) = (\mathbf{I} - \mathbf{E}) (\mathbf{k}(\mathbf{x}) - \mathbf{K} \mathbf{e}), \quad (6)$$

where $\mathbf{I} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$; $\mathbf{E} = 1/N \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$; and $\mathbf{e} = 1/N [1, 1, \dots, 1]^T$.

3. Pseudosamples

Pseudosamples were first used in the sensitivity analysis of uncertainty factors. Pseudosamples that change with a single factor are constructed to estimate the influence degree on the objective indicator, thus generating major influence factors. Postma et al. [30] constructed a column of special pseudosamples and obtained a column of response traces of the

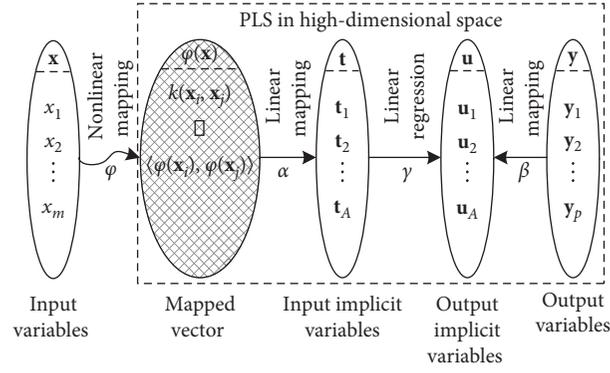


FIGURE 1: KPLS simulation of a complicated nonlinear system.

- (1) Set $i = 1$, $\mathbf{K}_1 = \mathbf{K}$, $\mathbf{Y}_1 = \mathbf{Y}$
- (2) Initialize \mathbf{u}_i randomly. \mathbf{u}_i is equal to the column with the maximum variance in \mathbf{Y}_i .
- (3) $\mathbf{t}_i = \mathbf{K}_i \mathbf{u}_i / \|\mathbf{K}_i \mathbf{u}_i\|$
- (4) $\mathbf{c}_i = \mathbf{Y}_i^T \mathbf{t}_i$
- (5) $\mathbf{u}_i = \mathbf{Y}_i \mathbf{c}_i / \|\mathbf{Y}_i \mathbf{c}_i\|$
- (6) If \mathbf{t}_i converges, turn to step (7); otherwise, return to step (3)
- (7) $\mathbf{K}_{i+1} = (\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T) \mathbf{K}_i (\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T)$, $\mathbf{Y}_{i+1} = \mathbf{Y}_i - \mathbf{t}_i \mathbf{t}_i^T \mathbf{Y}_i$
- (8) Save the data in the matrices: $\mathbf{T} \leftarrow \mathbf{t}_i$, $\mathbf{U} \leftarrow \mathbf{u}_i$
- (9) Set $i = i + 1$. If $i > A$ (A is the number of selected implicit variables), the loop ends; otherwise, return to (2)

ALGORITHM 1

model to realize a visual display of the nonlinear relationship between inputs and outputs of the system. The special pseudosamples take all weights into one input variable, whereas the rest were set to 0. Assuming a KPLS model with m input variables, a column of pseudosamples for the first input variables can be constructed as follows:

$$\begin{bmatrix} z_1, 0, \dots, 0 \\ z_2, 0, \dots, 0 \\ \vdots \\ z_k, 0, \dots, 0 \end{bmatrix}_{k \times m}, \quad (7)$$

where z_1, z_2, \dots, z_k is the equivalent interval in the value range of input z .

$$z_i = z_{\min} + \frac{(z_{\max} - z_{\min})}{(k-1) \times (i-1)}, \quad i = 1, 2, \dots, k, \quad (8)$$

where z_{\max} and z_{\min} are the maximum and minimum values of the input variables z and k is the number of pseudosamples.

The value of k is positively related to the nonlinearity of the system. This column of pseudosamples has been used as the new inputs of the KPLS model. k response values of the outputs can be determined using equation (4). Assuming that the response values of a output variable (w) are w_1, w_2, \dots, w_k , a scattering point diagram can be drawn in accordance with z_1, z_2, \dots, z_k and w_1, w_2, \dots, w_k , which is a trace line. It offers a visual display of the nonlinear relationship between the input variables z and the output variables w . Similarly, the trace lines of all inputs can be

obtained. Through comparing these trace lines, major inputs that influence outputs can be determined.

4. Nonlinear Method for Component Separation of Dam Effect Quantities

A complex nonlinear relationship between dam effect quantities and environmental variables is observed due to nonlinear factors, including nonlinear constitutive relation of dam materials, transverse joints of dam body, and contact nonlinear of dam foundation fault. To separate the real components of the effect quantities, this study initially establishes a nonlinear monitoring model using KPLS and then separates the influence component of various environmental factors from the KPLS model through constructing special pseudosamples.

The construction of the dam safety monitoring model discloses the determined relationship between the effect quantities and environmental variables using the prototype monitoring data based on mathematical, mechanical, and information science methods. Then, the effect quantities can be predicted and separated. The dam safety monitoring system is depicted in Figure 2.

The nonlinear expansion of environmental factors is performed with a nonlinear mapping (φ) to realize the nonlinear relationship with the effect quantities in a high-dimensional feature space. The effect quantities and the environmental variables contain noise, while some unusable and redundant components for interpreting the effect quantities are added through the nonlinear expansion. Therefore, the effective

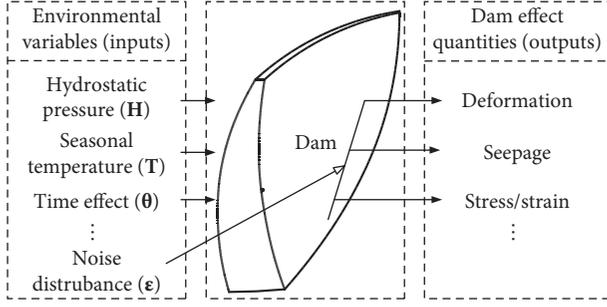
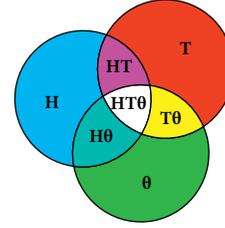


FIGURE 2: Dam safety monitoring system.

components \mathbf{t}_i and \mathbf{u}_i corresponding to the environmental variables and the effect quantities are, respectively, obtained by the linear mapping of α and β , and the linear relationship between \mathbf{t}_i and \mathbf{u}_i is established. Moreover, the components extracted by PLS are extracted together through the input and output variables. Substantially, this process is called weighted PCA. Therefore, the effective components of the environmental variables and the effect quantities can interpret dam service conditions well. Moreover, as a whole structure, the dam has a strong correlation between the effect quantities, thereby making the multidependent variable KPLS modeling more advantageous than the single-dependent variable KPLS modeling, which can save modeling time and eliminate data noise and redundancy effectively [28]. Therefore, this study uses KPLS to establish a nonlinear model for multiple effect quantities. The prediction model is described in equation (4).

Based on the established model, the practical samples are directly reconstructed into a type of special pseudosamples as the input of the model. This study aims to separate the influence components of various environmental variables from the KPLS model. The environmental factors are assumed to affect the dam behavior in the model and can be divided into three groups, namely, hydrostatic pressure, temperature, and time effect. To obtain the corresponding hydrostatic pressure component, the values of temperature and time effect factor in the actual sample can be directly set to 0. The output value can be obtained by constructing a column of pseudosamples and input the pseudosamples into the KPLS model (equation (4)). After deducting the initial value, the output value is nearly the desired hydrostatic pressure component. Similarly, the temperature and time effect components can be obtained. The three components obtained are only the independent components of hydrostatic pressure, temperature, and time effect on dam behavior. In addition, couplings among the effects of the three environmental factors occur, as shown in Figure 3. For example, the time effect deformation of a concrete dam can be intensified by increasing the water level. Changes in water level may also affect the boundary conditions of the temperature field in the concrete dam, further affecting the effects of temperature. Therefore, it is necessary to analyze the coupling components to understand the operation condition of the dam well and make the monitoring safety operation of the dam beneficial. The corresponding coupling components can also be separated from the KPLS model based on the pseudosamples. The corresponding method is presented in Table 1, and its proof is provided in Appendix.



- H** — Hydrostatic pressure-independent component
- T** — Temperature-independent component
- θ** — Time effect-independent component
- Hθ** — Hydrostatic pressure-time effect coupling component
- Tθ** — Temperature-time effect coupling component
- HT** — Hydrostatic pressure-temperature coupling component
- HTθ** — Hydrostatic pressure-temperature-time effect coupling component

FIGURE 3: Schematic of influence components of dam environmental factors.

TABLE 1: Component separation method of the KPLS model based on pseudosamples.

Component	Pseudosamples	Output	Calculation
H	$[\mathbf{H}, \mathbf{0}, \mathbf{0}]_{N \times m}$	$f_k(\mathbf{H}, \mathbf{0}, \mathbf{0})$	$\underline{f}_{kH}(\mathbf{H})$
T	$[\mathbf{0}, \mathbf{T}, \mathbf{0}]_{N \times m}$	$f_k(\mathbf{0}, \mathbf{T}, \mathbf{0})$	$\underline{f}_{kT}(\mathbf{T})$
θ	$[\mathbf{0}, \mathbf{0}, \mathbf{θ}]_{N \times m}$	$f_k(\mathbf{0}, \mathbf{0}, \mathbf{θ})$	$\underline{f}_{kθ}(\mathbf{θ})$
Hθ	$[\mathbf{H}, \mathbf{0}, \mathbf{θ}]_{N \times m}$	$f_k(\mathbf{H}, \mathbf{0}, \mathbf{θ})$	$\underline{f}_{kHθ}(\mathbf{Hθ})$
Tθ	$[\mathbf{0}, \mathbf{T}, \mathbf{θ}]_{N \times m}$	$f_k(\mathbf{0}, \mathbf{T}, \mathbf{θ})$	$\underline{f}_{kTθ}(\mathbf{Tθ})$
HT	$[\mathbf{H}, \mathbf{T}, \mathbf{0}]_{N \times m}$	$f_k(\mathbf{H}, \mathbf{T}, \mathbf{0})$	$\underline{f}_{kHT}(\mathbf{HT})$
HTθ	$[\mathbf{H}, \mathbf{T}, \mathbf{θ}]_{N \times m}$	$f_k(\mathbf{H}, \mathbf{T}, \mathbf{θ})$	$\underline{f}_{kHTθ}(\mathbf{HTθ})$

Notes: **H**, **T**, and **θ** correspond to the actual sample matrices of hydrostatic pressure, temperature, and time effect in the KPLS model. **0** denotes the corresponding reset zero matrix, and the underline indicates the model output of pseudosamples deducted by the initial value (dam effect quantities are relative values only).

From the abovementioned analysis, a nonlinear method for component separation of dam effect quantities can be established by combining KPLS and pseudosamples. The entire process of this method is demonstrated in Figure 4. The specific steps are presented as follows:

- (1) Several correlation effect quantities, which can reflect dam behavior (e.g., deformation, seepage, or strain) are selected as the output of the KPLS model.
- (2) Similar to the MLR model, the factors of environmental variables that affect the dam behavior are selected as the input of the KPLS model in accordance with the physical mechanical analysis.
- (3) A period of normal history for the monitoring data of the environmental variables and effect quantities is selected. Using the monitoring data, the KPLS model is constructed through the method discussed in Section 2.
- (4) In accordance with the method described in the third part of this paper, pseudosamples with environmental factors that affect the abovementioned dam behavior are constructed.

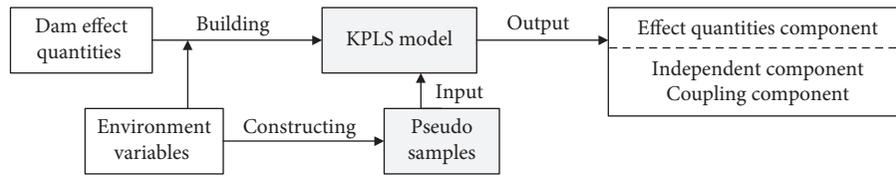


FIGURE 4: Entire flow of the nonlinear method for component separation of dam effect quantities.

- (5) By inputting the constructed pseudosamples into the KPLS model established in (3), the influence components of each environmental factor can be separated.

5. Validation and Application

In order to verify the proposed method, a case study based on an arch dam is carried out. This arch dam is a concrete double-curvature arch dam, with a maximum height of 294.5 m. The normal water level and total capacity of the reservoir are 1245 m and $149.14 \times 10^8 \text{ m}^3$, respectively. Positive and reversed vertical lines are set at 9 dam segments (4#, 9#, 15#, 19#, 22#, 25#, 29#, 35#, and 41#) to monitor horizontal deformation of the dam body and foundation (Figure 5).

5.1. Modeling. In this study, the radial deformation of the 22# segment is analyzed. In the 22# segment, radial displacements at five positive vertical measuring points and 1 reversed vertical measuring point are analyzed, which are recorded as y_0, y_1, y_2, y_3, y_4 , and y_5 from the dam foundation to dam crest. With reference to the factor selection method of the statistical model, the environmental variables in the KPLS model are chosen as $\{H, H^2, H^3, H^4, \sin(2\pi t/365), \cos(2\pi t/365), \theta, \ln \theta\}$.

The first four items are hydraulic pressure factors, the two middle items are temperature factors, and the last two items are time factors. H is water head and t is the accumulative days during the monitoring period $\theta = t/100$. On this basis, a multieffect KPLS model is constructed based on the monitoring data from July 1, 2010, to December 17, 2012. The prediction accuracy of the constructed KPLS model is verified by the monitoring data from December 18, 2012, to December 31, 2012. The training samples and the test samples are shown in Figure 6. During the same period, the measured values of the reservoir water level are shown in Figure 6.

The mean square error (MSE) is applied to compare the fitting and forecast accuracy of the two models. The expression of the MSE is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (9)$$

where n is the number of training samples or test samples; y_i is the i th observed value; and \hat{y}_i is the i th model fitting value or forecast value.

It can be seen from Figure 7 that the mean square error of the KPLS model is smaller than that of the MLR model in both the fitting period and the prediction period. The average fitting and prediction mean square error are 0.074 and

0.090, respectively, and the accuracy is about 1 and 2 orders of magnitude higher than the MLR model. The high accuracy of this model provides a strong guarantee for the next component separation.

5.2. Component Separation. In accordance with the method presented in Table 1, the corresponding pseudosamples are constructed and input into the KPLS model. The independent and coupling influence components of hydrostatic pressure, temperature, and time effect can be obtained through calculation. The corresponding component hydrographs of y_5 radial displacement on the crest of a crown cantilever are exhibited in Figures 8 and 9. The results indicate that each environmental variable has dominant and independent influences on the arch dam, and the coupling influence is significant. Moreover, the variation in the coupling component is complex, thereby confirming that a complex nonlinear relationship exists between the effect quantities and the environmental variables.

To verify the rationality of the aforementioned nonlinear separation results, the independent influence components are compared with the components from the MLR model. The comparison results are illustrated in Figures 10–12.

- (1) The variation in the two models is the same and satisfies the general deformation law of the arch dam. That is, the reservoir water level rises, the upstream water thrust increases, the dam deforms downstream, the displacement of the hydrostatic pressure component to the downstream increases, and vice versa. When the temperature decreases, dam concrete shrinks, the dam deforms downstream, and the displacement of the temperature component to downstream also increases.

At the beginning of the modeling period, the dam has just been poured, and the arching is completed. The water level of the reservoir is low, and the dam has a creep deformation upstream given the influence of the dam's overhang weight. In the later period, with the rise in the reservoir water level, the upstream water thrust increases, and the dam turns to creep deformation downstream. The time effect component shifts to the upstream and then to the downstream with time. In addition, the modeling period belongs to the initial stage of dam impounding operation. The time effect deformation induced by water load develops rapidly, and the displacement of the time effect component to the downstream owns a linear growth trend in the later period.

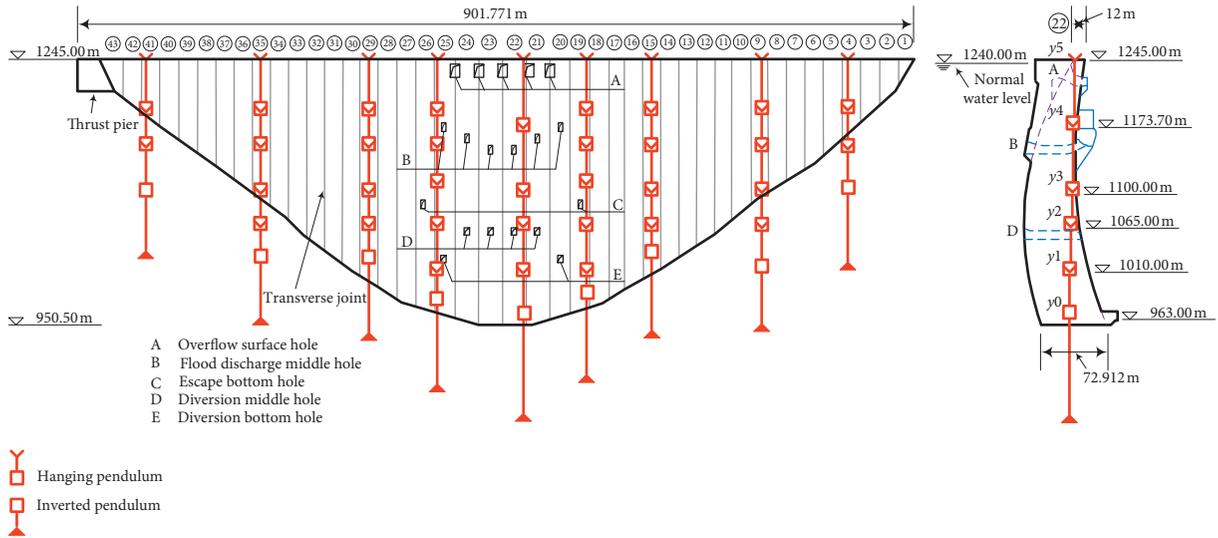


FIGURE 5: Distribution map of the measuring points.

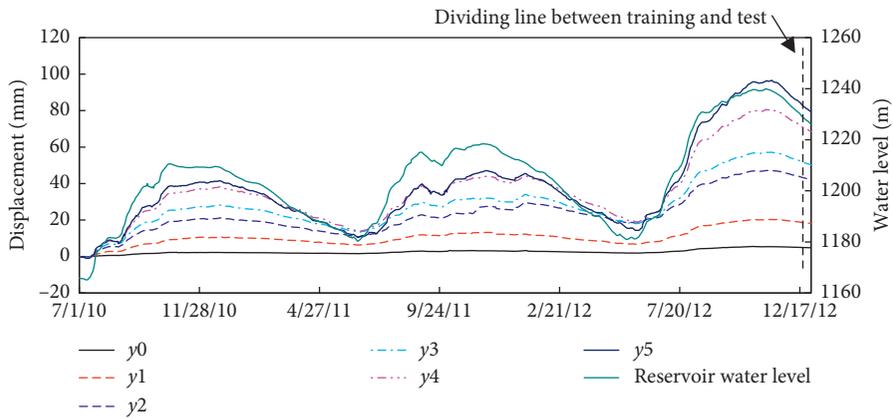


FIGURE 6: Time evolution of measuring point displacement and reservoir water level.

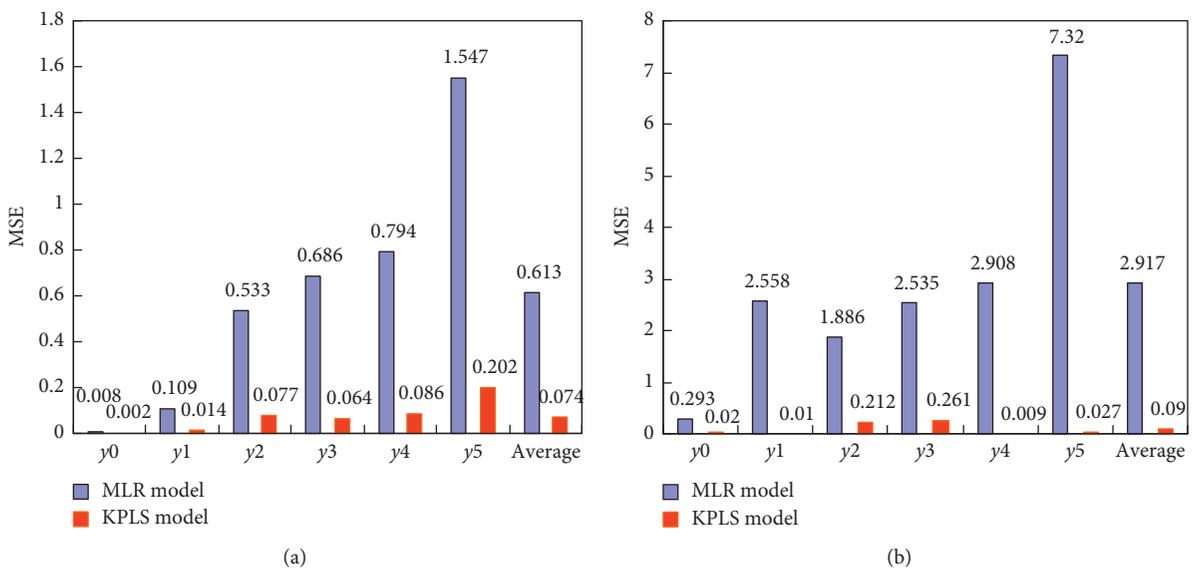


FIGURE 7: Fitting and forecast MSE of the MLR model and KPLS model. (a) Fitting. (b) Forecast.

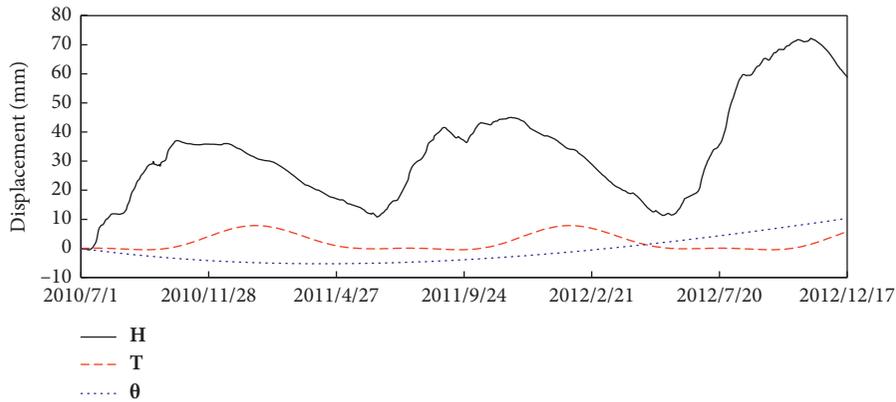


FIGURE 8: Independent component hydrographs of the KPLS model for radial displacement.

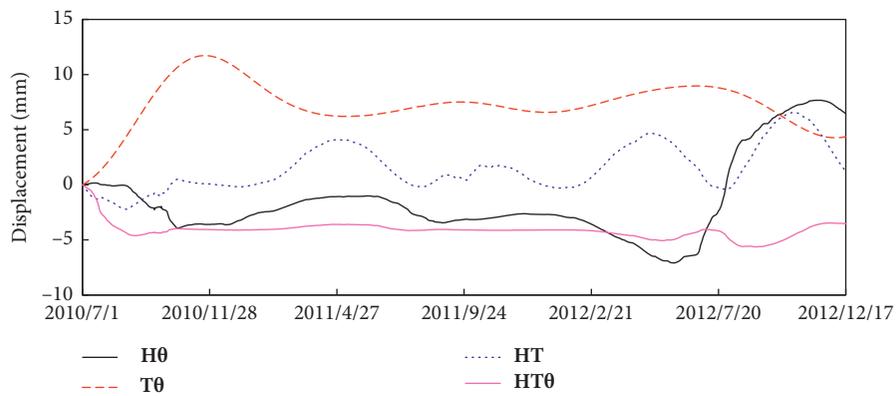


FIGURE 9: Coupling component hydrographs of the KPLS model for radial displacement.

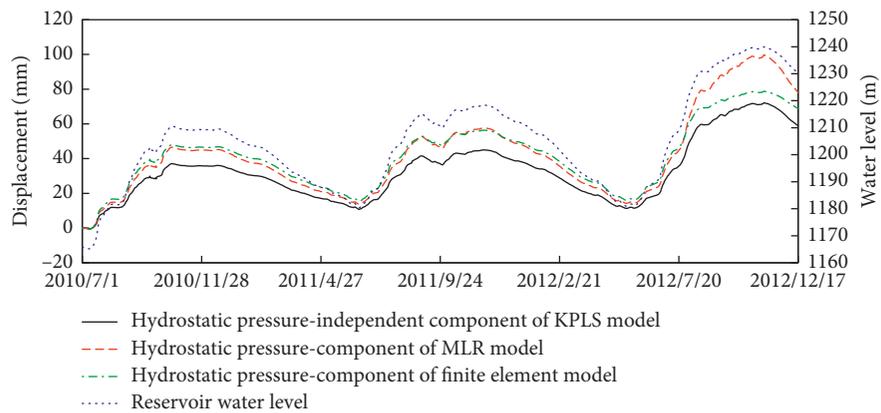


FIGURE 10: Comparison of the hydrostatic pressure components of models for radial displacement.

(2) The independent component of the hydrostatic pressure is relatively smaller in the KPLS model than in the MLR model, and the difference between the two models is evident when the water storage is close to the normal water storage level. To explore their rationality, the arch dam is subjected to an elastic finite element calculation analysis. The finite element calculation model is

depicted in Figure 13. The centerline of the arch dam is used as the reference, and the right and the left bank directions are selected as 800 m. Taking the origin of the dam crest as the reference, 400 m is taken upstream and 600 m is taken downstream. The length selected below the base is nearly the height of the dam, which is approximately 653 m elevated. The slope of the rock mass

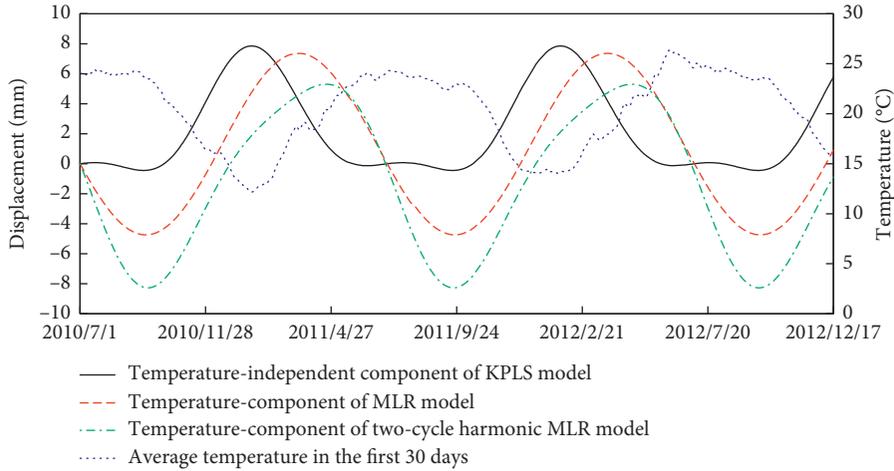


FIGURE 11: Comparison of the temperature components of models for radial displacement.

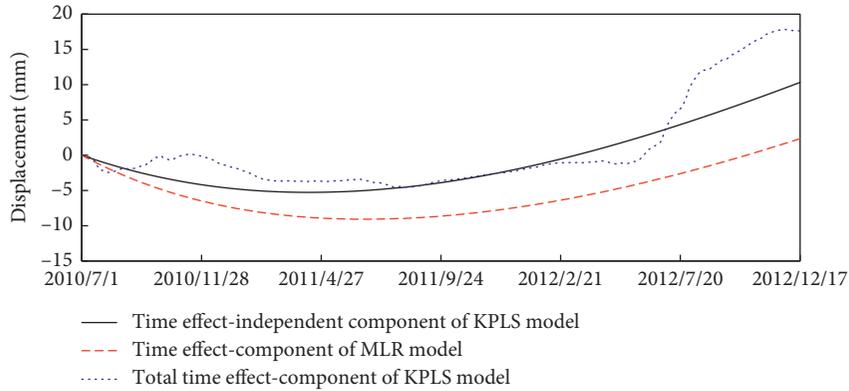


FIGURE 12: Comparison of time effect of models for radial displacement.

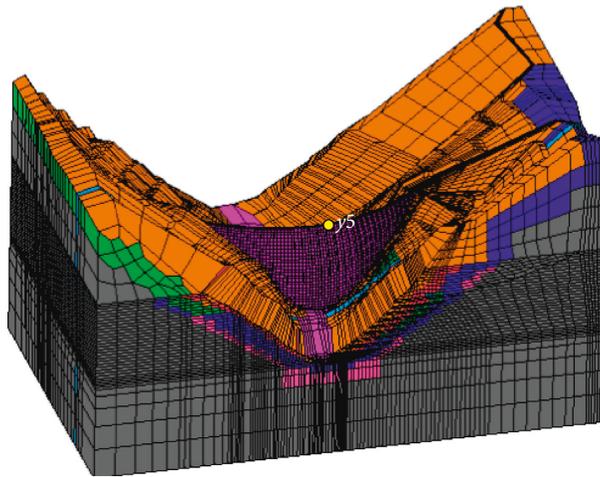


FIGURE 13: 3D finite element model of the dam.

above the elevation of the dam crest is cut to the natural boundary elevation. The whole model consists of 122008 units and 137301 nodes, and the dam body includes 10018 units and 18747 nodes.

The finite element method (FEM) is used to calculate the displacement of the dam under different water depths, and the expression of the relationship between the radial displacement of the y_5 of the arch crest and the reservoir water level can be obtained as follows:

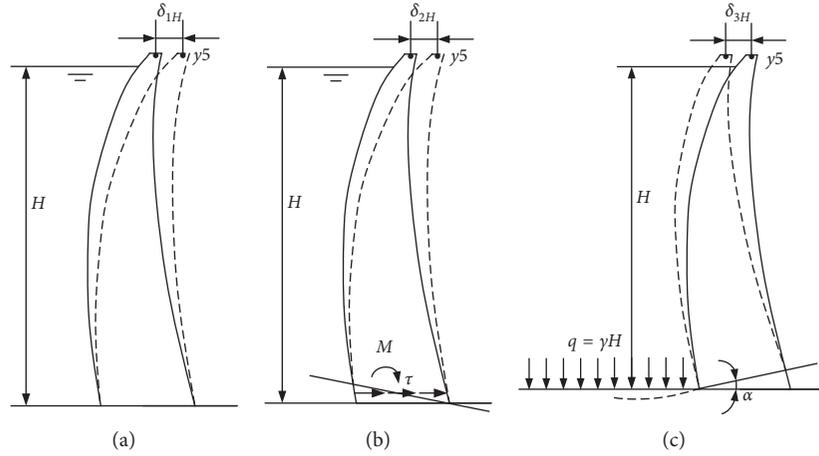


FIGURE 14: Three components (i.e., (a) δ_{1H} , (b) δ_{2H} , and (c) δ_{3H}) of the hydrostatic pressure component.

$$\begin{aligned} \delta_H = & -0.1054\mathbf{H} + 0.0022\mathbf{H}^2 + 7.4297 \times 10^{-6}\mathbf{H}^3 \\ & - 2.0882 \times 10^{-8}\mathbf{H}^4. \end{aligned} \quad (10)$$

In accordance with the actual change in the water level in the modeling period, the hydrostatic pressure component is calculated using equation (10), and the result is exhibited in Figure 10. The FEM and the MLR model hydrostatic pressure components are close in the first two water storage processes, and the difference between the two models is evident when the third water storage is close to the normal level. However, the FEM hydrostatic pressure component is larger than the KPLS model, and the difference between the three water storage processes is minimal. When the third water storage is close to the normal level, the difference in the hydrostatic pressure component between the KPLS model and FEM model is smaller than the difference in the hydrostatic pressure component between the MLR model and the FEM model.

The mechanical analysis shows that the hydrostatic pressure component of the arch dam deformation consists of three parts, as displayed in Figure 14: (1) the internal force generated by the reservoir hydrostatic pressure on the dam causes the dam displacement (δ_{1H}); (2) the internal force generated on the rock base surface causes the dam displacement (δ_{2H}); and (3) the weight of the reservoir water acts on the reservoir basin, thus causing the dam displacement (δ_{3H}) due to the rotation of the reservoir basin. Considering the limited extension of the FEM model to the upstream, the influence of the deformation of the reservoir basin on the displacement of the dam cannot be fully considered, thus the rotational displacement of the parts of the dam upstream is neglected. In accordance with this limitation,

the hydrostatic pressure component of the abovementioned finite element is corrected; that is, the hydrostatic pressure component downstream is reduced. The FEM hydrostatic pressure component in Figure 10 is translated downward, which is close to the hydrostatic pressure independent component of the KPLS model. In addition to the elastic deformation of the dam and foundation, the hydrostatic pressure can also cause creep deformation and some irreversible plastic deformations. The later three deformations are related to the hydrostatic pressure and time effect and belong to the coupling component of hydrostatic pressure and time effect. The MLR model constructed in this study cannot consider the coupling effect of hydrostatic pressure and time effect. Therefore, the separated hydrostatic pressure component is inevitably affected by hydrostatic pressure-time effect coupling component, which is inconsistent with the FEM hydrostatic pressure component. When the dam initially approaches the normal water level, the influence of hydrostatic pressure-time effect coupling is prominent. Based on the abovementioned analysis, the hydrostatic pressure can be reasonably considered an independent component of the KPLS model.

- (3) The temperature-independent component of the KPLS model is basically displaced downstream, whereas the MLR model temperature component has upstream and downstream displacement. At the beginning of the modeling period, the dam has just been poured, and the arching is completed. The water level of the reservoir has been stored to 1166 m, which is approximately three-fourths of the dam height. At this time, the concrete hydration heat of the dam has been completed, and the temperature of the dam is mainly affected by air and reservoir water temperatures. The starting time of this modeling was on July 1, 2010, which is

in the high-temperature period of the reservoir area. Therefore, compared with the initial moment, the deformation temperature component must be displaced substantially downstream and reach the maximum in the low-temperature period. Given that the reservoir water temperature lags behind air temperature change and dam concrete heat conduction time effect, the temperature component of dam deformation has a hysteresis effect relative to air temperature, and the lag time is generally nearly 1 month. Thus, the temperature component is compared with the average temperature in the first 30 days in the reservoir area (Figure 11). Under the premise of the same environmental variable factors, the temperature-independent component of the KPLS model is negatively correlated with the average temperature during the first 30 days. In addition, considering that the MLR model only uses an annual cycle harmonic temperature factor, a deficiency may exist. In this study, a semiperiodic harmonic temperature factor ($\{\sin(4\pi t/365), \cos(4\pi t/365)\}$) is added on the basis of the abovementioned modeling factor, and a two-period harmonic MLR model is constructed. The temperature component separated by the MLR model is shown in Figure 10. The displacement to the upstream is large, and the correlation with the average temperature in the first 30 days is poor. This result is considered unreasonable. In summary, the temperature-independent component of the KPLS model separation is regarded as reasonable.

- (4) The time effect-independent component of the KPLS model has a small upstream displacement caused by the overhanging self weight of the dam in the early stage of modeling. When the reservoir water level is close to the normal water storage level, the time effect-independent component changes to a positive downstream displacement. However, the time effect component of the MLR model has a large displacement to the upstream in the early stage. It remains a negative upstream displacement when the reservoir water level reaches the normal water storage level in the later stage. At the beginning of the modeling period, the reservoir water level has been stored to approximately three-fourths of the dam height. At this time, the upstream creep deformation caused by the overhanging self weight of the dam is slightly larger than the downstream creep deformation caused by the reservoir water pressure. Moreover, the reservoir water level still rises, the superposition of the two deformations does not cause excessive upstream time effect displacement, and engineering experience shows that the time effect deformation is typically a positive downstream displacement when the reservoir is stored to the normal level. In summary, the time effect-independent component separated by the KPLS model is considered reasonable.

In addition, the time effect deformation includes the unrecoverable creep and plastic deformation of the dam and foundation, and their recoverable creep deformation. However, the time effect component obtained by separating the time effect factors is only the average trend amount of the abovementioned time effect deformation. The hydrostatic pressure-time effect coupling component, the temperature-time effect coupling component, and the hydrostatic pressure-temperature-time effect coupling component separated by the KPLS model are superimposed on the time effect-independent component to obtain the total time effect component of the KPLS model, as illustrated in Figure 12. The total time effect component of the KPLS model fluctuates with the hydrostatic pressure and the temperature load in the vicinity of its independent time effect component, thereby reflecting the irreversible and recoverable time effect deformation. When the dam's third impoundment is close to the normal level for the first time, the total time effect component of the KPLS model has a large downstream displacement increment. This result is due to the fact that the arch dam is subjected to the huge hydrostatic pressure that corresponds to the normal level for the first time. Moreover, structural and stress adjustments occur, and the dam body and foundation have produced numerous irreversible plastic deformations. In addition, the total time effect component tends to stabilize, which is a normal phenomenon.

6. Conclusion

In this study, a nonlinear method for the component separation of dam effect quantities is proposed by combining KPLS and pseudosamples. The method initially uses KPLS to establish a nonlinear analysis model for dam safety monitoring. The model can determine the complex nonlinear relationship between dam effect quantities and environmental variables, avoid the influence of noise and environmental factor multicollinearities, and fully utilize the correlation information between multiple effect quantities to eliminate noise effectively. The model has high fitting and prediction accuracy, thus providing a strong guarantee for subsequent component separation. On the basis of the proposed KPLS model, the independent and coupling influence components of each environmental factor are separated by constructing special pseudosamples. The application indicates that the separation result of the nonlinear method is consistent with the MLR model. This result is in accordance with the general deformation law, and the separation result is more credible in the nonlinear method than in the MLR model.

The pseudosample component separation method designed in this study is also applicable to other dam nonlinear monitoring models, such as NN models and SVM models. However, the coupling influence components of the environmental factors separated in this study are complex, and further efforts should be put on physical causes of the coupling influence components and thus understand the working state of the dam well and ensure its safety.

Appendix

Mathematical Proof of KPLS Model Separation Method Based on Pseudosamples

Suppose a KPLS model of dam safety monitoring analysis contains p dam effect quantities $\mathbf{y} = \{y_1, y_2, \dots, y_p\}$ and m environmental variables $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$. Besides, environmental variables can be divided into hydrostatic pressure $\mathbf{H} = \{H_1, H_2, \dots, H_{m1}\}$, temperature $\mathbf{T} = \{T_1, T_2, \dots, T_{m2}\}$, and time effect $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_{m3}\}$, $m = m1 + m2 + m3$. Therefore, this KPLS model can be expressed as follows:

$$\begin{cases} y_1 = f_1(H_1, H_2, \dots, H_{m1}, T_1, T_2, \dots, T_{m2}, \theta_1, \theta_2, \dots, \theta_{m3}) + \varepsilon_1, \\ y_2 = f_2(H_1, H_2, \dots, H_{m1}, T_1, T_2, \dots, T_{m2}, \theta_1, \theta_2, \dots, \theta_{m3}) + \varepsilon_2, \\ \vdots \\ y_p = f_p(H_1, H_2, \dots, H_{m1}, T_1, T_2, \dots, T_{m2}, \theta_1, \theta_2, \dots, \theta_{m3}) + \varepsilon_p, \end{cases} \quad (\text{A.1})$$

where f_1, f_2, \dots, f_p are functions of p effect quantities about all environmental variables. $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ are the corresponding residuals. Any function f_k ($1 \leq k \leq p$) can be approximated as follows after the Taylor series expansion:

$$\begin{aligned} f_k(\mathbf{X}) \approx & b_0 + f_{kH}(\mathbf{H}) + f_{kT}(\mathbf{T}) + f_{k\theta}(\boldsymbol{\theta}) + f_{kH\theta}(\mathbf{H}\boldsymbol{\theta}) \\ & + f_{kT\theta}(\mathbf{T}\boldsymbol{\theta}) + f_{kHT}(\mathbf{HT}) + f_{kHT\theta}(\mathbf{HT}\boldsymbol{\theta}), \end{aligned} \quad (\text{A.2})$$

where b_0 is a constant term. $f_{kH}(\mathbf{H})$, $f_{kT}(\mathbf{T})$, and $f_{k\theta}(\boldsymbol{\theta})$ are polynomial functions that contain hydrostatic pressure (\mathbf{H}), temperature (\mathbf{T}), and time effect ($\boldsymbol{\theta}$) only, without constant items. They represent independent effects of three environmental variables. $f_{kH\theta}(\mathbf{H}\boldsymbol{\theta})$, $f_{kT\theta}(\mathbf{T}\boldsymbol{\theta})$, $f_{kHT}(\mathbf{HT})$, and $f_{kHT\theta}(\mathbf{HT}\boldsymbol{\theta})$ are polynomial functions that only contain cross terms of hydrostatic pressure (\mathbf{H}), temperature (\mathbf{T}), and time effect ($\boldsymbol{\theta}$) only and have no constant items. They represent coupling effects of three environmental variables.

The constructed pseudosamples in Table 1 are input into equation (A.2), thus obtaining the following equation:

$$\begin{cases} f_k(\mathbf{H}, \mathbf{0}, \mathbf{0}) \approx b_0 + f_{kH}(\mathbf{H}), \\ f_k(\mathbf{0}, \mathbf{T}, \mathbf{0}) \approx b_0 + f_{kT}(\mathbf{T}), \\ f_k(\mathbf{0}, \mathbf{0}, \boldsymbol{\theta}) \approx b_0 + f_{k\theta}(\boldsymbol{\theta}), \\ f_k(\mathbf{H}, \mathbf{0}, \boldsymbol{\theta}) \approx b_0 + f_{kH}(\mathbf{H}) + f_{k\theta}(\boldsymbol{\theta}) + f_{kH\theta}(\mathbf{H}\boldsymbol{\theta}), \\ f_k(\mathbf{0}, \mathbf{T}, \boldsymbol{\theta}) \approx b_0 + f_{kT}(\mathbf{T}) + f_{k\theta}(\boldsymbol{\theta}) + f_{kT\theta}(\mathbf{T}\boldsymbol{\theta}), \\ f_k(\mathbf{H}, \mathbf{T}, \mathbf{0}) \approx b_0 + f_{kH}(\mathbf{H}) + f_{kT}(\mathbf{T}) + f_{kHT}(\mathbf{HT}), \\ f_k(\mathbf{H}, \mathbf{T}, \boldsymbol{\theta}) \approx b_0 + f_{kH}(\mathbf{H}) + f_{kT}(\mathbf{T}) + f_{k\theta}(\boldsymbol{\theta}) \\ \quad + f_{kH\theta}(\mathbf{H}\boldsymbol{\theta}) + f_{kT\theta}(\mathbf{T}\boldsymbol{\theta}) + f_{kHT}(\mathbf{HT}) + f_{kHT\theta}(\mathbf{HT}\boldsymbol{\theta}). \end{cases} \quad (\text{A.3})$$

Next, the constant item b_0 can be eliminated by deducting the initial value of the model output. Based on

equation (A.3), the approximate solution of dam effect quantity components can be gained, as shown in the following equation:

$$\begin{cases} \underline{f}_{kH}(\mathbf{H}) \approx \underline{f}_k(\mathbf{H}, \mathbf{0}, \mathbf{0}), \\ \underline{f}_{kT}(\mathbf{T}) \approx \underline{f}_k(\mathbf{0}, \mathbf{T}, \mathbf{0}), \\ \underline{f}_{k\theta}(\boldsymbol{\theta}) \approx \underline{f}_k(\mathbf{0}, \mathbf{0}, \boldsymbol{\theta}), \\ \underline{f}_{kH\theta}(\mathbf{H}\boldsymbol{\theta}) \approx \underline{f}_k(\mathbf{H}, \mathbf{0}, \boldsymbol{\theta}) - \underline{f}_k(\mathbf{H}, \mathbf{0}, \mathbf{0}) - \underline{f}_k(\mathbf{0}, \mathbf{0}, \boldsymbol{\theta}), \\ \underline{f}_{kT\theta}(\mathbf{T}\boldsymbol{\theta}) \approx \underline{f}_k(\mathbf{0}, \mathbf{T}, \boldsymbol{\theta}) - \underline{f}_k(\mathbf{0}, \mathbf{T}, \mathbf{0}) - \underline{f}_k(\mathbf{0}, \mathbf{0}, \boldsymbol{\theta}), \\ \underline{f}_{kHT}(\mathbf{HT}) \approx \underline{f}_k(\mathbf{H}, \mathbf{T}, \mathbf{0}) - \underline{f}_k(\mathbf{H}, \mathbf{0}, \mathbf{0}) - \underline{f}_k(\mathbf{0}, \mathbf{T}, \mathbf{0}), \\ \underline{f}_{kHT\theta}(\mathbf{HT}\boldsymbol{\theta}) \approx \underline{f}_k(\mathbf{H}, \mathbf{T}, \boldsymbol{\theta}) - \underline{f}_k(\mathbf{H}, \mathbf{0}, \boldsymbol{\theta}) - \underline{f}_k(\mathbf{0}, \mathbf{T}, \boldsymbol{\theta}) \\ \quad - \underline{f}_k(\mathbf{H}, \mathbf{T}, \mathbf{0}) + \underline{f}_k(\mathbf{H}, \mathbf{0}, \mathbf{0}) + \underline{f}_k(\mathbf{0}, \mathbf{T}, \mathbf{0}) + \underline{f}_k(\mathbf{0}, \mathbf{0}, \boldsymbol{\theta}), \end{cases} \quad (\text{A.4})$$

where the underline denotes the deducted initial value (components of dam effect quantities are only relative values).

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant nos. 51779086, 51739003, and 51609149) and the National Foundation for Studying Abroad of China (grant no. 201806715053).

References

- [1] Z. R. Wu, *Safety Monitoring Theory & its Application of Hydraulic Structure*, Higher Education Press, Beijing, China, 2003.
- [2] F. Salazar, R. Morán, M. Á. Toledo, and E. Oñate, "Data-based models for the prediction of dam behaviour: a review and some methodological considerations," *Archives of Computational Methods in Engineering*, vol. 24, no. 1, pp. 1–21, 2017, in English.
- [3] Z. Shi, C. Gu, B. Chen et al., "Calculation methods for the permeability coefficient of concrete face rockfill dam with cracks," *Advances in Civil Engineering*, vol. 2019, Article ID 6571092, 13 pages, 2019.
- [4] D. Tonini, "Observed behaviour of several Italian arch dams," *Journal of the Power Division*, vol. 82, no. 6, pp. 1–26, 1956.
- [5] M. Rocha, J. Serafim, and A. Silveira, "A quantitative method for the interpretation of the results of the observation of dams," in *Proceedings of the 6th Congress on Large Dams*, New York, NY, USA, 1958.
- [6] L. E. Chouinard, D. W. Bennett, and N. Feknous, "Statistical analysis of monitoring data for concrete arch dams," *Journal of Performance of Constructed Facilities*, vol. 9, no. 4, pp. 286–301, 1995.

- [7] P. Léger and M. Leclerc, "Hydrostatic, temperature, time-displacement model for concrete dams," *Journal of Engineering Mechanics*, vol. 133, no. 3, pp. 267–277, 2007, in English.
- [8] 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, in English.
- [9] J. Mata, A. T. de Castro, and J. S. da Costa, "Constructing statistical models for arch dam deformation," *Structural Control & Health Monitoring*, vol. 21, no. 3, pp. 423–437, 2014, in English.
- [10] M. Tatin, M. Briffaut, F. Dufour, A. Simon, and J. P. Fabre, "Statistical modelling of thermal displacements for concrete dams: influence of water temperature profile and dam thickness profile," *Engineering Structures*, vol. 165, pp. 63–75, 2018, in English.
- [11] J. Hu and S. H. Wu, "Statistical modeling for deformation analysis of concrete arch dams with influential horizontal cracks," *Structural Health Monitoring-An International Journal*, vol. 18, no. 2, pp. 546–562, 2019, in English.
- [12] F. Amberg, "Interpretative models for concrete dam displacements," in *Proceedings of the 23th ICOLD Congress*, vol. q91-R43, Brasilia, Brazil, May 2009.
- [13] M. Tatin, M. Briffaut, F. Dufour, A. Simon, and J.-P. Fabre, "Thermal displacements of concrete dams: accounting for water temperature in statistical models," *Engineering Structures*, vol. 91, pp. 26–39, 2015, in English.
- [14] F. Salazar and M. Á. Toledo, "Discussion on "thermal displacements of concrete dams: accounting for water temperature in statistical models"," *Engineering Structures*, vol. 171, pp. 1071–1072, 2018, in English.
- [15] Y. Q. Chen, "Research on importance of dam safety influencing factors and quantitative separation method," Master thesis, Hohai University, Nanjing, China, 2009.
- [16] 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.
- [17] J. Y. Chen and X. Lin, *Processing Method of Observation Data*, Shanghai Jiaotong University Press, Shanghai, China, 1985.
- [18] X. D. Chen, C. S. Gu, and H. N. Chen, "Early warning of dam seepage with cooperation between principal component analysis and least squares wavelet support vector machine," *Fresenius Environmental Bulletin*, vol. 22, pp. 500–507, 2013, in English.
- [19] B. Li, C.-S. Gu, Z.-L. Li, and Z.-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. 1390–1394, 2008.
- [20] C. Shao, C. Gu, M. Yang, Y. Xu, and H. Su, "A novel model of dam displacement based on panel data," *Structural Control & Health Monitoring*, vol. 25, no. 1, p. e2037, 2018, in English.
- [21] Y.-S. Kim and B.-T. Kim, "Prediction of relative crest settlement of concrete-faced rockfill dams analyzed using an artificial neural network model," *Computers and Geotechnics*, vol. 35, no. 3, pp. 313–322, 2008, in English.
- [22] T. Bao, C. Gu, Z. Wu, and H. Yu, "Analysis of crack stability in concrete dams with chaos optimized neural network," in *Proceedings of the Sixth International Conference on Natural Computation*, Yantai, China, August 2010.
- [23] D. Santillan, J. Fraile-Ardanuy, and M. A. Toledo, "Dam seepage analysis based on artificial neural networks: the hysteresis phenomenon," in *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, August 2013.
- [24] V. Ranković, A. Novaković, N. Grujović, D. Divac, and N. Milivojević, "Predicting piezometric water level in dams via artificial neural networks," *Neural Computing & Applications*, vol. 24, no. 5, pp. 1115–1121, 2014, in English.
- [25] J. Annan and L. Bing, "Nonlinear time series prediction model for dam seepage flow based on PSO-SVM," *Journal of Hydraulic Engineering*, vol. 37, no. 3, pp. 331–335, 2006, in Chinese.
- [26] W. Xinzhou, F. A. N. Qian, Xuchengquan, and L. I. Zhao, "Dam deformation prediction based on wavelet transform and support vector machine," *Geomatics and Information Science of Wuhan University*, vol. 33, no. 5, p. 469, 2008, in Chinese.
- [27] V. Ranković, N. Grujović, D. Divac, and N. Milivojević, "Development of support vector regression identification model for prediction of dam structural behaviour," *Structural Safety*, vol. 48, pp. 33–39, 2014, in English.
- [28] R. Rosipal and L. J. Trejo, "Kernel partial least square regression in reproducing kernel hilbert SPace," *Journal of Machine Learning Research*, vol. 2, pp. 97–123, 2001.
- [29] H. Huang, B. Chen, and C. G. 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, in English.
- [30] G. J. Postma, P. W. T. Krooshof, and L. M. C. Buydens, "Opening the kernel of kernel partial least squares and support vector machines," *Analytica Chimica Acta*, vol. 705, no. 1-2, pp. 123–134, 2011, in English.
- [31] J. L. Godoy, D. A. Zumoffen, J. R. Vega, and J. L. Marchetti, "New contributions to non-linear process monitoring through kernel partial least squares," *Chemometrics And Intelligent Laboratory Systems*, vol. 135, pp. 76–89, 2014, in English.

