Research Article

Detection Model for Seepage Behavior of Earth Dams Based on Data Mining

Zhenxiang Jiang ¹ and Jinping He ^{1,2}

¹School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan 430072, China

²State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

Correspondence should be addressed to Jinping He; whuhjp@163.com

Received 3 September 2017; Revised 4 February 2018; Accepted 22 February 2018; Published 18 April 2018

Academic Editor: Bin Jiang

Copyright © 2018 Zhenxiang Jiang and Jinping He. 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.

Seepage behavior detecting is an important tool for ensuring the safety of earth dams. However, traditional seepage behavior detection methods have used insufficient monitoring data and have mainly focused on single-point measures and local seepage behavior. The seepage behavior of dams is not quantitatively detected based on the monitoring data with multiple measuring points. Therefore, this study uses data mining techniques to analyze the monitoring data and overcome the above-mentioned shortcomings. The massive seepage monitoring data with multiple points are used as the research object. The key information on seepage behavior is extracted using principal component analysis. The correlation between seepage behavior and upstream water level is described as mutual information. A detection model for overall seepage behavior is established. Result shows that the model can completely extract the seepage monitoring data with multiple points and quantitatively detect the overall seepage behavior of earth dams. The proposed method can provide a new and reasonable means of quantitatively detecting the overall seepage behavior of earth dams.

1. Introduction

Seepage is an important factor that affects the safety of earth dams. Based on the statistical data of the International Commission on Large Dams, approximately 52.2% of earth dam crashes are caused by seepage damage [1].

The upstream water level under normal service conditions causes earth dams to form a stable seepage field in the dam body and foundation, thereby indicating a stable seepage behavior. However, excessive seepage gradient, excessive seepage pressure, and other abnormal seepage phenomena may occur in earth dams due to the construction defects and material aging. These phenomena can cause seepage damage, increase in instability of a dam's slope, and lead to dam breakage.

However, dam safety can be controlled. Safety detecting provides the basis for dam safety control. Several osmometers are typically placed along the key points in the dam to detect seepage behavior. The measured value of osmometers fluctuates within a reasonable range when the seepage behavior of

earth dam is normal. The measured value will exhibit sudden changes or trends when the seepage behavior of earth dam is abnormal. Therefore, analyzing the data of these osmometers should be conducted to detect the seepage behavior of earth dams. Mathematical and mechanical methods are used to analyze the data and detect seepage behavior. Appropriate measures are reinforced when an abnormal seepage phenomenon is detected to effectively reduce the risk of dam breakage and provide technical assurance on the service safety of earth dams.

Currently, the detection methods for the seepage behavior of earth dams are divided into three types. The first type uses a statistical regression method for analyzing the monitoring data. The factors that influence the seepage behavior of earth dams are summarized as water level, rainfall, temperature, and aging. A seepage monitoring model for a single point is established. Abnormal seepage behavior is detected by analyzing the trend of different factors [2]. Si et al. [3] used support vector machine to train original monitoring data. This method improved the precision and detection accuracy

for seepage behavior and provided reasonable trends from different factors. Xiang et al. [4] introduced a particle swarm optimization algorithm during the modeling process, which analyzed the lag time of water level in osmometers and optimized the expressions of water and rainfall factors. Thus, this algorithm was optimized, and the accuracy of the model was improved. Gamse and Oberguggenberger [5] used a “coordinate time series” method to analyze the seepage monitoring data of earth dams, thereby avoiding the over-fitting phenomenon in the model. The expressions for water level, temperature, and aging factor were reasonable, and the monitoring model detected abnormal seepage behavior.

The second type of detection method combines finite element calculation and monitoring data. The seepage parameters of dam body and foundation are calculated by back analysis, and the seepage field of the dam is simulated [6, 7]. Abnormal seepage behavior is detected by comparing the calculated and measured values of the finite element model. Zhang et al. [8] established a finite element model for earth dam. The seepage parameters for the different parts of the dam were calculated by using seepage monitoring data, and the seepage field of the dam was simulated under various operating conditions. A comparison of the calculated and measured values obtained from the osmometers revealed that core wall failure causes the abnormal seepage behavior of the dam. Ren et al. [9] conducted an inversion of dam seepage parameters based on the finite element model of an earth dam and the monitoring data from the osmometers. The seepage field of a dam was simulated under normal water levels. The result provided a basis for seepage behavior detection under normal operating conditions. Chi et al. [10] introduced a neural network algorithm for the inversion of seepage parameters of earth dam, thereby improving the accuracy of inversion results and providing reliable seepage behavior detection.

The third type of detection method uses knowledge of experts to evaluate seepage behavior, in which the seepage monitoring data at a single point are scored. A fusion algorithm is used to integrate the score of different points, and the overall seepage behavior of an earth dam is detected based on fusion results. Yang [11] evaluated the monitoring data of seepage pressure by using expert experience. The membership matrix of each monitoring point was established, and an analytic hierarchy process was used to integrate the monitoring projects. Then, the overall seepage of the dam was detected. Shao and Xin [12] introduced a projection pursuit method into the fusion process, in which the weights of different monitoring points were assigned during the fusion process, thereby improving the detection accuracy.

The existing methods provide the means for seepage behavior detection of earth dams from different perspectives. However, the safety monitoring of modern dams is mainly based on automatic monitoring, and the amount of monitoring data constantly increases. The information contained in the data has become increasingly abundant. Therefore, the existing methods exhibit the following shortcomings. (1) The methods mainly focus on data with only one monitoring point, which is called the single-point detection method. This method can only detect local seepage behavior. Data with

different points should be fused if the overall seepage behavior must be detected; this technique is called the multiple-point detection method. (2) The overall seepage behavior is qualitatively detected based on experts' experiences and detection results of local seepage behavior. The subjectivity of this method is strong, and the experts' experiences affect the detection results. Therefore, the quantitative detection method for the overall seepage behavior of dams should be investigated.

In summary, new research methods should be developed to explore the potential information in massive monitoring data and establish an efficient and accurate model for seepage behavior detection given the increase in monitoring data. Therefore, this study utilizes principal component analysis (PCA) and mutual information (MI) in data mining technology to extract massive seepage monitoring data with multiple points. The PCA is used for information extraction, and MI is used to describe the correlation between the principal component (PC) and upstream water level. The detection model for seepage behavior is established based on MI distribution, thus providing a new means of accurately detecting the overall seepage behavior of earth dams.

2. Method for Establishing the Model

2.1. Modeling Process. Data mining [13, 14] refers to the process of discovering hidden information from massive amounts of data. The correlation study on the data and the extraction of key information from massive monitoring data is the primary content for establishing a seepage detection model given the increasing volume of monitoring data.

The PCA is an important data mining algorithm [15] that extracts one or a few PCs from a plurality of variables to replace the original variables through the correlation between the data based on the principle of minimizing data information loss. Currently, PCA is extensively applied in data analysis [16, 17]. Several seepage monitoring cross sections are arranged in the dam, and osmometers are arranged in the sections to monitor the seepage behavior of dams. The monitoring data are called the water level of osmometers. The fusion of the data from multiple osmometers should be conducted to detect the overall seepage behavior quantitatively. The locations of several sections and working conditions are similar. Therefore, the data from these osmometers are similar and are correlated. In this study, the PCA is used to reconstruct one or a few integrated variables (PC) that reflect the basic characteristics of primitive variables when each osmometer is considered a primitive variable. PCs contain key information from primitive variables and provide the basis for quantitative detection of the overall seepage behavior.

The upstream water level is an important factor that affects the seepage of earth dams [18]. For earth dams, the earth that is used to fill the dam does not prevent seepage, and the upstream water enters the earth. To prevent seepage damage in the dam, a core wall made from impervious materials is constructed in the dam to block the seepage and ensure its safety. Therefore, the water level of osmometers arranged in the front core wall is close to the upstream water

FIGURE 1: Modeling process for seepage detection of earth dams.

level considering the poor antipermeability of the earth. The PCs of these osmometers are also close to the upstream water level. The correlation between the PC and upstream water level is strong. The water level of osmometers that were arranged behind the core wall is significantly reduced given the antipermeability of the core wall. The PCs of these osmometers are also significantly reduced. Therefore, the correlation between the PC and upstream water level is low. MI [19] is used to quantitatively describe the correlation between the upstream water level and PC after extracting the PC of osmometers. A considerable amount of MI results in a strong correlation between upstream water level and PC. Compared with the traditional correlation coefficient, MI simultaneously describes the linear and nonlinear relationships between the variables. In addition, MI is extensively used to describe the correlation of variables [20, 21].

MI between the upstream water level and PC should fluctuate in a rational region when the core wall is intact, thereby indicating that the seepage behavior of the dam is normal. The seepage quantity in the dam increases, and the PC becomes abnormal when the core wall is damaged given the effect of upstream water level, thus leading to an abnormal MI. This condition indicates that the seepage behavior of the dam is abnormal. The MI fluctuation range, that is, the detection model, can be obtained by analyzing the MI distribution of historical data. The seepage behavior of the dam is normal if the MI falls within this range. The seepage behavior of the dam is abnormal if the MI falls outside this range.

The advantage of MI in detecting seepage behavior is that MI can be used to eliminate the interference of osmometer failure. In general, several abnormal values are found in the data of osmometers. The PC fuses the data from different osmometers. Therefore, the PC also contains abnormal data. These abnormal values may reflect the abnormal seepage behavior. However, these abnormal values may be caused by osmometer failure. The abnormal values caused by osmometer failure may interfere in detecting seepage behavior and lead to misdiagnosis. The MI represents the correlation between PC and upstream water level. If the abnormal data are caused by osmometer failure, then the MI will not be abnormal because the abnormal data are not caused by the upstream water level. Therefore, MI eliminates the interference from osmometer failure and improves the detection accuracy.

In summary, this study uses PCA to extract PCs from massive seepage measurement and MI to describe the

correlation between the PC and the upstream water level. The detection model for seepage behavior is constructed based on the MI distribution. A flowchart that illustrates the modeling process is depicted in Figure 1.

2.2. Extracting PCs of Effect Variables. The number of seepage monitoring points is assumed as n ; that is, the number of primitive variables is n , and each point contains m times of observed value. Therefore, these observed values can form the following $n \times m$ matrix:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \quad (1)$$

where $\mathbf{X}_i$ ($i = 1, 2, \dots, n$) is the row vector that denotes the monitoring data sequence of the i th monitoring point and x_{ij} ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$) denotes the j th monitoring data of the i th monitoring point.

In matrix $\mathbf{X}$, the working environments of these primitive variables (seepage monitoring points) are similar. Therefore, the measured data of these points $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T$ exhibit a strong correlation. The PCA is used to reconstruct n irrelevant integrated variables (PC) when the number of primitive variables is n . Score matrix $\mathbf{F}$ can be expressed as

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_n \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \mathbf{L}\mathbf{X}, \quad (2)$$

where $\mathbf{F}_i$ is the i th PC and $\mathbf{L}$ is the score coefficient matrix. l_{ij} is the coefficient of the j th primitive variables in the i th PC and reflects the relevance between the j th primitive variables $\mathbf{X}_j$ and the i th PC, $\mathbf{F}_i$. A large absolute value of l_{ij} leads to a high correlation between $\mathbf{F}_i$ and $\mathbf{X}_j$. Hence, considerable information on $\mathbf{X}_j$ can be explained by $\mathbf{F}_i$. If l_{ij} is positive, then the correlation between $\mathbf{F}_i$ and $\mathbf{X}_j$ is positive. If l_{ij} is negative, then the correlation between $\mathbf{F}_i$ and $\mathbf{X}_j$ is negative. $\mathbf{F}_i$ demonstrates the following properties:

- ① $\mathbf{F}_i$ is uncorrelated with $\mathbf{F}_j$ ($i \neq j$; $i, j = 1, 2, \dots, n$).
- ② $\mathbf{F}_1$ has the largest variance in all linear combinations of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$. $\mathbf{F}_2$ has the biggest variance in all linear

combinations of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, which are uncorrelated with $\mathbf{F}_1, \dots, \mathbf{F}_n$ has the largest variance in all linear combinations of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, which are uncorrelated with $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{n-1}$.

Equation (2) denotes that calculating $\mathbf{L}$ is an important step. PC can be obtained by calculating $\mathbf{L}$. Assume that the covariance matrix of primitive variables is expressed as

$$\mathbf{C} = \begin{bmatrix} \text{cov}(\mathbf{X}_1, \mathbf{X}_1) & \text{cov}(\mathbf{X}_1, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_1, \mathbf{X}_n) \\ \text{cov}(\mathbf{X}_2, \mathbf{X}_1) & \text{cov}(\mathbf{X}_2, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_2, \mathbf{X}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\mathbf{X}_n, \mathbf{X}_1) & \text{cov}(\mathbf{X}_n, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_n, \mathbf{X}_n) \end{bmatrix}. \quad (3)$$

The eigenvalue decomposition of $\mathbf{C}$ can be expressed as

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (4)$$

where $\mathbf{\Lambda}$ is the diagonal matrix; that is, $\mathbf{\Lambda} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_n]$. λ_i is the eigenvalue of $\mathbf{C}$, which is the variance of the i th PC. $\mathbf{U}$ is the eigenvector matrix, which can be written as $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_n)$. $\mathbf{u}_i$ can be written as $(u_{1i}, u_{2i}, \dots, u_{ni})^T$ ($i = 1, 2, \dots, n$). Thus, $\mathbf{L} = \mathbf{U}^T$ can be confirmed.

Assume that $\mathbf{a}_i = (a_{1i}, a_{2i}, \dots, a_{ni})^T$ is an orthogonal vector, which makes the i th PC $\mathbf{F}_i = a_{1i}\mathbf{X}_1 + a_{2i}\mathbf{X}_2 + \cdots + a_{ni}\mathbf{X}_n = \mathbf{a}_i^T \mathbf{X}$. The variance of $\mathbf{F}_i$ is calculated as follows:

$$\begin{aligned} \text{var}(\mathbf{F}_i) &= \mathbf{a}_i^T \mathbf{C} \mathbf{a}_i = \mathbf{a}_i^T \cdot \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \cdot \mathbf{a}_i = \sum_{i=1}^n \lambda_i \mathbf{a}_i^T \mathbf{u}_i \mathbf{u}_i^T \mathbf{a}_i \\ &\leq \lambda_i \sum_{i=1}^n \mathbf{a}_i^T \mathbf{u}_i \mathbf{u}_i^T \mathbf{a}_i = \lambda_i \mathbf{a}_i^T \mathbf{a}_i = \lambda_i. \end{aligned} \quad (5)$$

The second property of $\mathbf{F}_i$ indicates that $\mathbf{F}_i$ has the largest variance in all linear combinations of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, which are uncorrelated with $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{i-1}$. Therefore, $\text{var}(\mathbf{F}_i)$ can reach the largest variance when $\mathbf{a}_i = \mathbf{u}_i$; that is, $\mathbf{F}_i = u_{1i}\mathbf{X}_1 + u_{2i}\mathbf{X}_2 + \cdots + u_{ni}\mathbf{X}_n = \mathbf{u}_i^T \mathbf{X}$. Thus, $\mathbf{L} = \mathbf{U}^T$.

If the number of primitive variables is n , then less than n PCs can be reconstructed. The ability of these n PCs to explain primitive variables is different. Therefore, z (where $z < n$) PCs should be extracted from n PCs that best describe the properties of primitive variables. The values of λ_i are sorted from large to small, and the value of z is typically determined based on the cumulative variance contribution rate η , which is calculated as follows:

$$\eta = \frac{\sum_{i=1}^z \lambda_i}{\sum_{i=1}^n \lambda_i} \times 100\%. \quad (6)$$

In general, if η is greater than 95%, then over 95% of the original information can be explained by former z PCs. Therefore, $\eta \geq 95\%$ is set as the discriminant index for extracting z PCs from n PCs. In engineering applications, the number of PCs can be properly adjusted according to the specific circumstances.

2.3. MI Calculation. Assume that the i th PC of seepage is $\mathbf{F}_i$ and the upstream water level is $\mathbf{Y}$. Then, the MI I_i between $\mathbf{F}_i$ and $\mathbf{Y}$ is calculated as

$$I_i = \sum_{f_i \in \mathbf{F}_i} \sum_{y \in \mathbf{Y}} g(f_i, y) \log \frac{g(f_i, y)}{g(f_i)g(y)}, \quad (7)$$

where $g(f_i)$ and $g(y)$ are the probability density functions of $\mathbf{F}_i$ and $\mathbf{Y}$, respectively, and $g(f_i, y)$ is the joint probability density function of $\mathbf{F}_i$ and $\mathbf{Y}$. If the correlation between $\mathbf{F}_i$ and $\mathbf{Y}$ is high, then I_i will be considerable. Moreover, if $\mathbf{F}_i$ and $\mathbf{Y}$ are not related, then I_i will be zero. $\mathbf{F}_i$ and $\mathbf{Y}$ may not follow the fixed-form distribution type. Hence, kernel density estimation (KDE) method [22] is used to estimate the probability density functions of $\mathbf{F}_i$ and $\mathbf{Y}$. In this method, $g(f_i)$ and $g(y)$ are expressed as

$$\begin{aligned} g(f_i) &= \frac{1}{mh} \sum_{d=1}^m K\left(\frac{f_i - F_{id}}{h}\right); \\ g(y) &= \frac{1}{mh} \sum_{d=1}^m K\left(\frac{y - Y_d}{h}\right), \end{aligned} \quad (8)$$

where m are the measuring times, F_{id} is the d th measured value of the i th PC, Y_d is the d th measured value of the upstream water level, and K is the kernel function. Gaussian kernel function [22] is generally used and expressed as

$$\begin{aligned} K\left(\frac{f_i - F_{id}}{h}\right) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(f_i - F_{id})^2}{2h^2}\right); \\ K\left(\frac{y - Y_d}{h}\right) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y - Y_d)^2}{2h^2}\right). \end{aligned} \quad (9)$$

The joint probability density function $g(f_i, y)$ is expressed as

$$g(f_i, y) = \frac{1}{m} \sum_{d=1}^m \frac{1}{h^2} \cdot K\left(\frac{f_i - F_{id}}{h}, \frac{y - Y_d}{h}\right). \quad (10)$$

In (8)–(10), h is the bandwidth used to control the smoothness and fitting accuracy of the probability density curve. If the value of h is high, then the probability density curve is smooth with a low fitting precision. If the value of h is small, then the smoothness of probability density curve decreases but the fitting precision increases. In general, the value of h is determined through the comprehensive analysis of smoothness and fitting accuracy.

2.4. Detection Model of Seepage. Assume that the length of the dam safety monitoring data is r years. If the unit is years, then the detection model is constructed based on the annual variation of MI. In (7), the MI between the PCs and the upstream water level can be obtained. These MI values form the following MI matrix:

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \vdots \\ \mathbf{I}_z \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1r} \\ I_{21} & I_{22} & \cdots & I_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ I_{z1} & I_{z2} & \cdots & I_{zr} \end{bmatrix}, \quad (11)$$

where $\mathbf{I}_i$ ($i = 1, 2, \dots, z$) is the row vector, which indicates the MI of i th seepage PC and upstream water level in different years. I_{ij} ($i = 1, 2, \dots, z; j = 1, 2, \dots, r$) denotes the MI between the i th seepage PC and the upstream water level in the j th year. Assume that the mean of $\mathbf{I}_i$ ($i = 1, 2, \dots, z$) is $\bar{I}_i$ ($i = 1, 2, \dots, z$). Then, mean vector $\bar{\mathbf{I}} = [\bar{I}_1, \bar{I}_2, \dots, \bar{I}_z]^T$ can be obtained.

Assume that MI is independent and obeys the normal distribution. The seepage detection model can be established by analyzing the distribution of $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ ($j = 1, 2, \dots, r$). The statistics T^2 can be constructed as follows:

$$T^2 = \frac{r}{r+1} \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right)^T \cdot \mathbf{S}^{-1} \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right), \quad (12)$$

where $\mathbf{S}$ is the following covariance matrix:

$$\mathbf{S} = \begin{bmatrix} \text{cov}(\mathbf{I}_1, \mathbf{I}_1) & \text{cov}(\mathbf{I}_1, \mathbf{I}_2) & \cdots & \text{cov}(\mathbf{I}_1, \mathbf{I}_z) \\ \text{cov}(\mathbf{I}_2, \mathbf{I}_1) & \text{cov}(\mathbf{I}_2, \mathbf{I}_2) & \cdots & \text{cov}(\mathbf{I}_2, \mathbf{I}_z) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\mathbf{I}_z, \mathbf{I}_1) & \text{cov}(\mathbf{I}_z, \mathbf{I}_2) & \cdots & \text{cov}(\mathbf{I}_z, \mathbf{I}_z) \end{bmatrix}. \quad (13)$$

Based on statistical theory [15], T^2 follows the $((r-1)z/(r-z))F(z, r-z)$ distribution in which

$$T^2 \sim \frac{(r-1)z}{r-z} F(z, r-z). \quad (14)$$

Therefore, possibility P where the MI value $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ in the j th year falls into the confidence region $100(1-\alpha)\%$ satisfies the following equation:

$$P \left(T^2 \leq \frac{(r-1)z}{r-z} F_\alpha(z, r-z) \right) = 1 - \alpha, \quad (15)$$

and the confidence region satisfies the following inequality:

$$T^2 \leq \frac{(r-1)z}{r-z} F_\alpha(z, r-z). \quad (16)$$

The region is a confidence interval when $z = 1$; the region is a confidence ellipse when $z = 2$; the confidence region is a confidence ellipsoid when $z = 3$; the region is a hyperellipsoid when $z > 3$.

The range of the confidence region is determined by the eigenvalue of covariance matrix $\mathbf{S}$ and significance level α . $\mathbf{S}$ is symmetric and positively definite and has z real eigenvalues that are greater than zero. The eigenvalues of $\mathbf{S}$ are expressed as

$$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_z \geq 0. \quad (17)$$

The confidence interval $100(1-\alpha)\%$ of the MI value $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ in the j th year is centered on mean vector $\bar{\mathbf{I}}$. The lengths of each half axis are expressed as

$$\begin{aligned} & \sqrt{\frac{\delta_1 z (r^2 - 1)}{r(r-z)}} F_\alpha(z, r-z), \\ & \sqrt{\frac{\delta_2 z (r^2 - 1)}{r(r-z)}} F_\alpha(z, r-z), \dots, \\ & \sqrt{\frac{\delta_z z (r^2 - 1)}{r(r-z)}} F_\alpha(z, r-z). \end{aligned} \quad (18)$$

From the statistical theory [15], significance level α is typically set as 0.05 and 0.01. Therefore, the distribution of $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ satisfies the following equations:

$$\begin{aligned} & \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right)^T \\ & \cdot \mathbf{S}^{-1} \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right) \leq \frac{z(r^2 - 1)}{r(r-z)} \end{aligned} \quad (19)$$

$$\cdot F_{0.05}(z, r-z),$$

$$\begin{aligned} & \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right)^T \\ & \cdot \mathbf{S}^{-1} \left([I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T - \bar{\mathbf{I}} \right) \leq \frac{z(r^2 - 1)}{r(r-z)} \end{aligned} \quad (20)$$

$$\cdot F_{0.01}(z, r-z).$$

Equations (19) and (20) are considered the detection models for seepage behavior of earth dams. For the MI value $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ in the j th year, the probability of falling in the range of (19) is 0.95, and the probability of falling outside the range of (20) is 0.01. Based on the small probability principle, the event is considered a small probability event when its probability is less than 0.01. If a small probability event occurs, then appropriate attention must be provided. The seepage behavior of earth dams is divided into three states, namely, normal, early warning, and abnormal, when the preceding mentioned theories are combined with engineering experience in seepage monitoring; these states are described as follows:

(1) $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ falls within the range of (19) ($P = 0.95$); that is, the seepage behavior of earth dam is normal.

(2) $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ falls outside the range of (20) ($P = 0.01$), thereby indicating a small probability event; that is, the seepage behavior is abnormal, and corresponding engineering measures should be immediately taken.

(3) The region between (19) and (20) is a transition region between the normal and abnormal statuses. If $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ falls within this range ($P = 0.04$), then it warrants an early warning status. The trend of $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ should be observed in the future. If $[I_{1j}, I_{2j}, \dots, I_{ij}, \dots, I_{zj}]^T$ tends toward the abnormal range, then appropriate engineering measures should be taken.

FIGURE 2: Shenzhen Reservoir Project.

FIGURE 3: Layout of the osmometers in the Shenzhen Reservoir Project.

3. Case Study

3.1. Description of the Project. The Shenzhen Reservoir (Figure 2) is located downstream of Shawan River in Shenzhen City, Guangdong Province, China. This reservoir is a water conservation project with functions of flood control, water supply, and power generation. The main building includes the main dam, the left auxiliary dam, and the right auxiliary dam. This main dam is an earth dam that has a core wall with a shell material that is gravelly, silty, and clayey, and the core wall for antiseepage is made of concrete. Four seepage monitoring cross sections (MXF, MXG, MXS, and MXL) are arranged to monitor the dam seepage behavior and antiseepage effect of the core wall. Then, 20 osmometers are placed on the cross sections, where five osmometers are placed in each cross section. The osmometers in front of the core wall are called prewall osmometers, which are numbered as MXF1, MXG1, MXS1, and MXL1 to facilitate early recognition. The osmometers behind the core wall are called back-wall osmometers and are numbered as MXF2–MXF5, MXG2–MXG5, MXS2–MXS5, and MXL2–MXL5. The locations of the osmometers are exhibited in Figure 3.

In general, the current study uses prewall osmometers (MXF1, MXG1, MXS1, and MXL1) and the first osmometers of the back-wall (MXF2, MXG2, MXS2, and MXL2) as representative monitoring points. Seepage behavior is detected through a data mining method that uses the monitoring data, which are obtained from the osmometers from January 1,

FIGURE 4: Process lines of the prewall osmometers.

1995, to December 31, 2014. The process lines for prewall osmometers MXF1, MXG1, MXS1, and MXL1 are demonstrated in Figure 4. Meanwhile, the process lines for back-wall osmometers MXF2, MXG2, MXS2, and MXL2 are displayed in Figure 5.

The qualitative analysis in Figure 4 denotes that the measured values of the prewall osmometers and the variations are similar. The qualitative analysis presented in Figure 5 indicates that the measured values of MXF2, MXG2, and MXS2 in the first osmometers of the back-wall are the same. However, the fluctuation of the MXL2 value from 2005 to 2010 significantly increases, thereby demonstrating an

TABLE 1: Covariance matrix of the prewall osmometers.

	MXF1	MXG1	MXS1	MXL1
MXF1	1.00	0.94	0.91	0.92
MXG1	0.94	1.00	0.98	0.97
MXS1	0.91	0.98	1.00	0.98
MXL1	0.92	0.97	0.98	1.00

TABLE 2: Covariance matrix of the first back-wall osmometers.

	MXF2	MXG2	MXS2	MXL2
MXF2	1.00	0.76	0.70	0.21
MXG2	0.76	1.00	0.85	-0.01
MXS2	0.70	0.85	1.00	0.05
MXL2	0.21	-0.01	0.05	1.00

FIGURE 5: Process lines of the back-wall osmometers.

abnormal phenomenon where the measured values of MXL2, MXF2, MXG2, and MXS2 are inconsistent.

The possible causes of the abnormal measured values of MXL2 include the following: (1) the core wall in the MXL monitoring section being damaged, thus resulting in an abnormal seepage behavior; (2) osmometer failures, such as external water infiltration in the MXL2 osmometer, and abnormal operation of the MXL2 osmometer. This study uses the data mining method to establish the detection model for seepage behavior. The seepage behavior is detected. Then, the causes of abnormal MXL2 data are speculated.

3.2. Mining Seepage Monitoring Data

3.2.1. Extraction of PCs in the Prewall Osmometers. The covariance matrix C_p for prewall osmometers MXF1, MXG1, MXS1, and MXL1 is calculated by using (3), as displayed in Table 1. The covariance matrix C_b for back-wall osmometers MXF2, MXG2, MXS2, and MXL2 is also calculated, as presented in Table 2.

From Tables 1 and 2, the covariance of the prewall osmometers is determined between 0.91 and 0.98; this covariance indicates a high correlation among the values. The covariance of MXF2, MXG2, and MXS2 is between 0.70 and 0.85, and the correlation is also high. However, the covariance of MXL2, MXF2, MXG2, and MXS2 is between -0.01 and

0.21, thereby indicating that MXL2 is weakly correlated with the first back-wall osmometers on the other monitored sections.

The eigenvalues and their variance contribution rate and the cumulative variance contribution rates of C_p and C_b can be calculated by using (4) and (6) as summarized in Table 3.

Table 3 indicates that the eigenvalue of F_{p1} in the prewall osmometers is considerably larger than the eigenvalues of the other PCs. In addition, the variance contribution rate of F_{p1} reaches 96.59%, which is higher than the threshold of 85.00%, thereby denoting that the main information from the original information of MXF1, MXG1, MXS1, and MXL1 can be explained using F_{p1} . Hence, F_{p1} can be used to describe the seepage characteristics of MXF1, MXG1, MXS1, and MXL1. The values X_{p1} , X_{p2} , X_{p3} , and X_{p4} represent MXF1, MXG1, MXS1, and MXL1, respectively. In (2), the expression of F_{p1} can be expressed as

$$F_{p1} = 0.253X_{p1} + 0.251X_{p2} + 0.252X_{p3} + 0.244X_{p4}. \quad (21)$$

In (21), the coefficients of X_{p1} , X_{p2} , X_{p3} , and X_{p4} are extremely close; these coefficients indicate that the measured data of MXF1, MXG1, MXS1, and MXL1 are similar. Therefore, F_{p1} can express MXF1, MXG1, MXS1, and MXL1.

Table 3 also indicates that 63.85% of the original measured information can be explained by using the first PC F_{b1} in the first back-wall osmometers, whereas 25.55% of the original measured information can be explained by using the second PC F_{b2} . The cumulative variance contribution rate of F_{b1} and F_{b2} reaches 89.64%. Although the cumulative variance contribution rates of F_{b1} and F_{b2} are below the threshold, the information of F_{b3} and F_{b4} is significantly reduced. Therefore, the original measured information of MXF2, MXG2, MXS2, and MXL2 can be represented by F_{b1} and F_{b2} . Let X_{b1} , X_{b2} , X_{b3} , and X_{b4} be the measured data of MXF2, MXG2, MXS2, and MXL2, respectively, to calculate the expressions of F_{b1} and F_{b2} by using (2):

$$F_{b1} = 0.311X_{b1} + 0.326X_{b2} + 0.312X_{b3} + 0.051X_{b4}, \quad (22)$$

$$F_{b2} = 0.152X_{b1} - 0.201X_{b2} - 0.132X_{b3} + 1.181X_{b4}. \quad (23)$$

TABLE 3: Eigenvalues and the variance contribution rates of C_p and C_b .

		PC	F_{p1}	F_{p2}	F_{p3}	F_{p4}
Prewall osmometers	Eigenvalues		3.86	0.10	0.02	0.01
	Variance contribution rate (%)		96.59	2.58	0.51	0.32
	Cumulative variance		96.59	99.17	99.68	100
	Contribution rate (%)					
		PC	F_{b1}	F_{b2}	F_{b3}	F_{b4}
Back-wall osmometers	Eigenvalues		2.56	1.02	0.29	0.13
	Variance contribution rate (%)		63.85	25.55	7.35	3.26
	Cumulative variance		63.85	89.40	96.75	100
	Contribution rate (%)

FIGURE 6: Process lines for F_{p1} , F_{b1} , and F_{b2} and upstream water level.

In (22), the coefficients of MXF2, MXG2, and MXS2 are close and significantly higher than the coefficient of MXL2, thereby indicating that F_{b1} mainly explains the original measured information of MXF2, MXG2, and MXS2. In (23), the coefficient of MXL2 is higher than the absolute value of the coefficients of MXF2, MXG2, and MXS2, thus denoting that F_{b2} mainly explains the original measured information of MXL2.

The values of F_{p1} during the monitoring period from January 1, 1995, to December 31, 2014, are calculated using (21), whereas those of F_{b1} and F_{b2} are calculated using (22) and (23), respectively. The process lines for F_{p1} , F_{b1} , and F_{b2} and upstream water level are illustrated in Figure 6.

In this figure, the process lines for F_{p1} and upstream water level coincide, thereby showing a strong correlation between F_{p1} and the upstream water level. The correlations between F_{b1} and the upstream water level and between F_{b2} and the upstream water level are not evident. In addition, F_{b1} mainly describes the seepage characteristics of MXF2, MXG2, and MXS2, thus exhibiting the strong regularity of process line. By contrast, F_{b2} mainly describes the seepage characteristics of MXL2. Therefore, a significant fluctuation in its process line is observed during the period of 2005–2010.

The PCA in data mining combines information that highly correlates and separates anomalous data. The key information for the original osmometers is expressed by

F_{p1} , F_{b1} , and F_{b2} , thereby reducing the number of original variables and providing the basis for quantitative detection.

3.2.2. MI between PCs and Upstream Water Level. Additional analysis is conducted by calculating the MI between PC and upstream water level to establish the detection model for seepage behavior and determine the cause of abnormal measurements of MXL2. Let I_{p1} , I_{b1} , and I_{b2} denote the MI values between F_{p1} and upstream water level, between F_{b1} and upstream water level, and between F_{b2} and upstream water level, correspondingly, during the period of 1995–2014. In (8), the probability density function of each PC and upstream water level can be obtained from the KDE when the bandwidth is set to 1.0, 0.5, and 0.1. The image is displayed in Figure 7.

In this figure, the probability density function can accurately simulate the distribution of PC and the upstream water level when the bandwidth is set to 0.1. The MI values I_{p1} , I_{b1} , and I_{b2} under this bandwidth during the period of 1995–2014 (i.e., 20 years) are calculated using (7). The matrix of I is expressed as

$$I = \begin{bmatrix} I_{p1} \\ I_{b1} \\ I_{b2} \end{bmatrix} = \begin{bmatrix} I_{p1,1} & I_{p1,2} & \cdots & I_{p1,j} & \cdots & I_{p1,20} \\ I_{b1,1} & I_{b1,2} & \cdots & I_{b1,j} & \cdots & I_{b1,20} \\ I_{b2,1} & I_{b2,2} & \cdots & I_{b2,j} & \cdots & I_{b2,20} \end{bmatrix}; \quad (24)$$

($j = 1, 2, \dots, 20$).

The process lines for the MI values are depicted in Figure 8.

MI reflects the correlation among variables, and MXF1, MXG1, MXS1, and MXL1 are placed in front of the core wall. Therefore, a high correlation theoretically exists between F_{p1} and the upstream water level, thus indicating that I_{p1} is large. However, the core wall plays the main role for antiseepage. If the seepage behavior of the earth dam is normal, then the correlations between F_{b1} and the upstream water level and between F_{b2} and the upstream water level should be significantly reduced; these conditions indicate that I_{b1} and I_{b2} are small. If the seepage behavior is abnormal, then the correlations between F_{b1} and the upstream water level and between F_{b2} and the upstream water level will increase, thereby indicating that I_{b1} and I_{b2} will exhibit a significant increase.

FIGURE 7: Probability density functions of the PC and upstream water level by KDE.

In Figure 8, $\mathbf{I}_{p1}$ varies within the range of $[1.17, 2.44]$, and $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$ vary within the range of $[1.30 \times 10^{-1}, 6.32 \times 10^{-1}]$ during the period of 1995–2014. $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$ are significantly lower than $\mathbf{I}_{p1}$. Therefore, we can qualitatively consider that the seepage behavior is reasonable.

3.3. Detection Model of Seepage Behavior. The result of Kolmogorov-Smirnov [23] analysis shows that $\mathbf{I}_{p1}$ follows a normal distribution $N(1.86, 0.31^2)$, $\mathbf{I}_{b1}$ follows a normal distribution $N(0.31, 0.11^2)$, and $\mathbf{I}_{b2}$ follows a normal distribution $N(0.27, 0.10^2)$. The detection model is established using the distribution of MI values to quantitatively analyze the seepage behavior. The measured value of the MXL2 osmometer in 2005–2010 is evidently abnormal; that is, $\mathbf{I}_{b2}$ may not reflect the real MI between MXL2 ($\mathbf{F}_{b2}$) and the upstream water

level. Therefore, the detection model is established based on the distribution of $\mathbf{I}_{p1}$ and $\mathbf{I}_{b1}$, which reflects the real MI of the measured value and the upstream water level.

In (15) and (16), the confidence region is an ellipse when the number of PCs = 2. The means for $\mathbf{I}_{p1}$ and $\mathbf{I}_{b1}$ are 1.86 and 0.31. Significance level α is set to 0.05 and 0.01. Then, the two confidence ellipses can be obtained using (19) and (20). The equations are expressed as follows:

$$\left(\begin{bmatrix} \mathbf{I}_{p1,j} \\ \mathbf{I}_{b1,j} \end{bmatrix} - \begin{bmatrix} 1.86 \\ 0.31 \end{bmatrix} \right)^T \begin{bmatrix} 0.0911 & 0.0041 \\ 0.0041 & 0.0116 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} \mathbf{I}_{p1,j} \\ \mathbf{I}_{b1,j} \end{bmatrix} - \begin{bmatrix} 1.86 \\ 0.31 \end{bmatrix} \right) = 7.87, \quad (25)$$

FIGURE 8: Process lines of MI.

$$\left(\begin{bmatrix} I_{p1,j} \\ I_{b1,j} \end{bmatrix} - \begin{bmatrix} 1.86 \\ 0.31 \end{bmatrix} \right)^T \begin{bmatrix} 0.0911 & 0.0041 \\ 0.0041 & 0.0116 \end{bmatrix}^{-1} \cdot \left(\begin{bmatrix} I_{p1,j} \\ I_{b1,j} \end{bmatrix} - \begin{bmatrix} 1.86 \\ 0.31 \end{bmatrix} \right) = 13.32. \quad (26)$$

Equations (25) and (26) are considered the detection model for seepage behavior, and their images are exhibited in Figure 9. In this model, the seepage behavior can be determined based on the positions of $I_{p1,j}$, $I_{b1,j}$, and $I_{b2,j}$ ($j = 1, 2, \dots, 20$) in the ellipses. (1) If the MI falls within the range of (25), then the seepage behavior is normal. (2) If the MI falls within the range of (25) and (26), then the seepage behavior signals an early warning. (3) If the MI falls outside the range of (26), then the seepage behavior is abnormal.

The MI values ($I_{p1,j}$, $I_{b1,j}$) and ($I_{p1,j}$, $I_{b2,j}$) from 1995 to 2014 are plotted in Figure 9. In this figure, the values in other years ($I_{p1,j}$, $I_{b1,j}$) and ($I_{p1,j}$, $I_{b2,j}$) are in a normal state, except for the value of ($I_{p1,j}$, $I_{b1,j}$) in 2004, which is in the early warning state. This result indicates that the seepage behavior is normal. Therefore, the significant fluctuation of MXL2 in 2005–2010 may be caused by equipment failure.

3.4. Verifying the Speculation. The MXL2 osmometer was tested and analyzed through an engineering method to verify the speculation.

(1) The technical performance of the MXL2 osmometer was tested, and the results showed that the current service status of the MXL2 osmometer is qualified.

(2) The piezometer sensitivity in the MXL2 osmometer was also tested, and the results showed that the piezometer sensitivity in the MXL2 osmometer is unqualified. A certain degree of clogging occurred in the piezometer.

(3) The working records of the MXL2 piezometer were investigated and analyzed. The results showed that the dam surface was transformed in 2004. However, the piezometer in the MXL2 osmometer was poorly maintained, thereby causing rainfall infiltration. The piezometer was punched and cleaned at the beginning of 2011, and piezometer maintenance was conducted. Thus, the measured results of MXL2 after 2011 were normalized.

FIGURE 9: Detection model of seepage behavior.

In summary, the evident fluctuation of MXL2 in 2005–2010 was mainly attributed to the unqualified piezometer sensitivity and rainfall infiltration in the piezometer of the MXL2 osmometer. The results of engineering test and detection model are consistent with each other, thereby confirming the speculation.

4. Conclusion

Seepage behavior is an important factor that affects the safety of earth dams. In this study, the PCA and MI methods are organically combined to detect the overall seepage behavior of earth dams. The monitoring data from different monitoring sections are effectively synthesized and mined. The detection model can eliminate the interference of osmometer failure and improve the accuracy of the detection, thereby providing a new method for detecting the overall seepage behavior of earth dams.

The main contributions of this paper are as follows: (1) The PCA method is applied to fuse the data of correlated osmometers, thus promoting the development of seepage detection from a single point to multiple points. (2) The detection model is established by MI distribution, which supports the improvement of seepage detection from being a qualitative method to being a quantitative method. In particular, the method can be extended to detect the behavior of concealed engineering such as core wall, foundation, and steel structure.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 51379162) and Water Conservancy

Science and Technology Innovation Project of Guangdong Province (2016-06).

References

- [1] ICOLD, *Dam failures statistical analysis*, International Commission on Large Dams, Paris, France, 1995.
- [2] G. Q. Liang, M. S. Zheng, and B. Y. Sun, "Analysis model and method of seepage observation data for earth rock-fill dams," *Journal of Hydraulic Engineering*, vol. 2, pp. 83–87, 2003.
- [3] C. D. Si, J. J. Lian, and Y. Zheng, "Genetic support vector machine model for seepage safety monitoring of earth-rock dams," *Journal of Hydraulic Engineering*, vol. 38, no. 11, pp. 1340–1346, 2007.
- [4] 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.
- [5] S. Gamse and M. Oberguggenberger, "Assessment of long-term coordinate time series using hydrostatic-season-time model for rock-fill embankment dam," *Structural Control and Health Monitoring*, vol. 24, no. 1, Article ID e1859, 2017.
- [6] L. Wang and Q. Xu, "Analysis of three dimensional random seepage field based on Monte Carlo stochastic finite element method," *Rock and Soil Mechanics*, vol. 35, no. 1, pp. 287–292, 2014.
- [7] C.-B. Zhou, W. Liu, Y.-F. Chen, R. Hu, and K. Wei, "Inverse modeling of leakage through a rockfill dam foundation during its construction stage using transient flow model, neural network and genetic algorithm," *Engineering Geology*, vol. 187, pp. 183–195, 2015.
- [8] J. Zhang, J. Wang, and H. Cui, "Causes of the abnormal seepage field in a dam with asphaltic concrete core," *Journal of Earth Science*, vol. 27, no. 1, pp. 74–82, 2016.
- [9] J. Ren, Z.-Z. Shen, J. Yang, and C.-Z. Yu, "Back analysis of the 3D seepage problem and its engineering applications," *Environmental Earth Sciences*, vol. 75, no. 2, article no. 113, pp. 1–8, 2016.
- [10] S. Chi, S. Ni, and Z. Liu, "Back Analysis of the Permeability Coefficient of a High Core Rockfill Dam Based on a RBF Neural Network Optimized Using the PSO Algorithm," *Mathematical Problems in Engineering*, vol. 2015, Article ID 124042, 2015.
- [11] H. P. Yang, "Fuzzy Comprehensive Evaluation of Dam Safety Based on AHP-Entropy Combination Weight Method," *Yellow River*, vol. 35, no. 6, pp. 116–118, 2013.
- [12] L. F. Shao and Y. Y. Xin, "Safety Evaluation of Earth-Rock Dam Based on Projection Pursuit Analysis and Normal Cloud Model," *Water Resources and Power*, vol. 33, no. 12, pp. 81–84, 2015.
- [13] J. Peral, A. Mate, and M. Marco, "Application of Data Mining techniques to identify relevant Key Performance Indicators," *Computer Standards Interfaces*, vol. 54, no. SI, pp. 76–85, 2017.
- [14] J. Choi, B. Kim, H. Hahn et al., "Data mining-based variable assessment methodology for evaluating the contribution of knowledge services of a public research institute to business performance of firms," *Expert Systems with Applications*, vol. 84, pp. 37–48, 2017.
- [15] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, Prentice Hall, Upper Saddle River, NJ, USA, 6th edition, 2007.
- [16] H. Chen, B. Jiang, N. Lu, and Z. Mao, "Multi-mode kernel principal component analysis-based incipient fault detection for pulse width modulated inverter of China Railway High-speed 5," *Advances in Mechanical Engineering*, vol. 9, no. 10, 2017.
- [17] L. Qiang, S. J. Qin, and C. Tianyou, "Decentralized fault diagnosis of continuous annealing processes based on multilevel PCA," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 687–698, 2013.
- [18] J. P. He, *Safety Monitoring Theory and Its Application of dam*, China Water Power Press, Beijing, 2010.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.
- [20] P. A. Legg, P. L. Rosin, D. Marshall, and J. E. Morgan, "Feature Neighbourhood Mutual Information for multi-modal image registration: An application to eye fundus imaging," *Pattern Recognition*, vol. 48, no. 6, pp. 1937–1946, 2015.
- [21] G. M. Bidelman and S. P. Bhagat, "Objective detection of auditory steady-state evoked potentials based on mutual information," *International Journal of Audiology*, vol. 55, no. 5, pp. 313–319, 2016.
- [22] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, UK, 1986.
- [23] W. Daniel Wayne, *Applied Nonparametric Statistics*, PWS-Kent, Boston, Mass, USA, 2nd edition, 1989.