Investigation on sensitive warning index of earth dam by use of dimensionality reduction methods

FANG Weihua, ZHANG Na, JIN Youjie
Nanjing Automation Institute of Water Resource & Hydrology, Jiangsu Nanjing, 210012, China

Abstract. The main damage cause of homogeneous earth dams is overtopping failure. In order to uncover the overall change rule of measured data when comes to overtopping failure and to establish foundation for earth dam warning indicators, this paper presents a new exact method of homogeneous earth dam overtopping rushed burst data analysis based on KPCA. Sliding window was used to construct the time series of measured data and two indicators were obtained which were the number of principal components and matrix norm of PCA or KPCA based on the 85% contribution rate. In-situ test showed that, Corresponding dam state deteriorated and final burst, these two indicators increased, decreased, and finally increased. Different size of sliding window and norm calculation shows that the two indicators are not sensitive to the window size and type of norm. Therefore, these two indicators can be taken as the first and second level sensitive warning indicators for the earth dam safety condition.

1 Introduction
According to relevant statistics, overtopping failure is the main cause of middle and small size dam break. Numerical simulation methods not only fail to consider the complex stress path and the constitutive model, but also depend on calculation parameters, node distribution, and grid generated to some degree. Therefore, studying on the evolution of the dam safety behavior with measured data in the process of dam break, and extracting the corresponding early warning index should have great significance on reducing safety risks of earth dam.

There are some methods to extract the warning index based on measured data, such as confidence interval method, typical low probability method of monitoring effect quantities, limit state method and extreme value theory, etc. However, these methods are put forward for single measure point, which consider neither information redundancy or complementary between multiple-spot, nor strain localization of earth-rock body in unstable failure. In this work, with the key point of dam state expression dimension, we consider the strain localization and fractal dimension during dam failure, analyze dam-break measured data by dimensional reduction, and make calculation analysis of various schemes with PCA(Principal Component Analysis) and KPCA (Kernel Principal Component Analysis). Finally, we put forward the novel method of dam failure warning index based on the number of principal components and corresponding matrix norm. The analysis of measured data shows that the proposed method is feasible.

2 Data Analysis Methods
The way to reduce dimension mainly includes Non-negative Matrix Decomposition, Self-organizing Map, Isometric Mapping, Locally Linear Embedding, PCA and KPCA, etc. Because of its clear concept and simple calculation, the PCA method has been applied to analyze
dam safety monitoring data widely \cite{2-4}. By using the multidimensional confidence ellipsoid to monitor multiple effect quantity, literatures \cite{5-6} have changed the traditional model of single effect analysis and improved the efficiency of data analysis and structure safety monitoring. Based on the certain factors (influence) among dam safety monitoring physical variables, we propose the hypothesis that when the dam state is stable, these factors are stable; when the dam state changes, these factors are unstable. And then the change of the dam state is represented by factors’ number or features.

2.1 PCA

(1) The structure of window data is $n \times p$ ($n$ is measurement time, $p$ is point’s number) which is composed of $p$-dimensional stochastic row vector, and then standardize it:

$$x^*_i = \frac{x_i - \overline{X}_i}{\sqrt{\text{Var}(\overline{X}_i)}} \quad (i = 1, 2, ..., n; j = 1, 2, ..., p)$$

where $\overline{X}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \quad (j = 1, 2, ..., p)$.

(2) Calculate $p \times p$-dimensional covariance matrix (also called correlation matrix) $\Sigma$ of data matrix $X^*$,

$$\Sigma = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

where $r_{ij} = \frac{\sum_{i=1}^{n} (x_{ij} - \overline{X}_i)(x_{kj} - \overline{X}_j)}{(n-1)\sum_{i=1}^{n} (x_{ij} - \overline{X}_i)^2 \sum_{j=1}^{n} (x_{kj} - \overline{X}_j)^2}$.

(3) Obtain eigenvalues and corresponding eigenvectors $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ from right upper triangular matrix $\Sigma^{up}$ of $\Sigma$, the corresponding orthogonal eigenvectors are denoted by $\omega_1, \omega_2, \cdots, \omega_p$, and selecting the number of eigenvalue $m$ whose cumulative contribution rate of variance $\sum_{i=1}^{m} \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$ accounted for approximately 85%.

2.2 KPCA

(1) Reference to type (1), we standardize the window data $n \times p$ matrix composed of monitoring data, which are measurement points $p$ and measurement times $n$.

(2) Obtain the $p \times p$-dimensional kernel matrix $K = (k_{ij})_{p \times p}$ according to kernel function and kernel parameter after optimal selection.

(3) Get the new matrix from centralizing the kernel matrix $K$, obtain the top $m$ eigenvalue $\lambda$ and corresponding eigenvectors $\beta$ whose cumulative contribution rate of variance accounted for approximately 85% from the new matrix, and then construct principal component matrix $\Sigma^{up}$.

2.3 Determination of early warning index

(1) Make the number of principal components and kernel principal components $m$ as the early warning index, and analyze its evolution law associated with dam state.

(2) Make the principal components and kernel principal components matrix norm as the early warning index, and analyze its evolution law associated with dam state.
3 Example Verification

3.1 Engineering situation
A homogeneous cohesive soil earth dam with largest water-filled volume of the upstream reservoir is $10 \times 10^4 \text{m}^3$, 9.7m largest height, 3m top width, 120m long and 27.25m wide. The gradients of the dam in upper and lower reaches of the dam are both 1:2.5. The dam is composed of silty loam with 11.5% of clay, and its compactness is 97%. The cohesion value $C$ ranges from 7.5 to 39.5 kPa, and the incipient frictional velocity of the undisturbed soil is 0.06 cm/s. The test process, which lasts 46 days, includes dam filling (instrument installation), first water impoundment, stabilized level, secondary impoundment and dam failure, etc. More detail can be found in [13] and [14].

3.2 Measurement Point Choice and Pretreatment of Measuring Data
To analyze data, we select sixty-three measuring locations and values: one upstream water level data, one downstream water level data, thirteen seepage pressure data, one seepage discharge data, two water content data, twelve surface and internal deformation data, thirty-three dam temperature data.

With the consideration of the time of monitoring point input, we choose the monitoring period from October 16, 2012 to November 17, 2012 to analyze the whole process of before/at/after dam failure.

In order to use the monitoring data sufficiently, avoid interpolating too much and avoid deleting unnecessarily, we selected 30min/frequency. We selected the whole Lagrange Interpolation method for the monitoring data sequence which has lower frequency. We selected the monitoring data of corresponding measurement times for the monitoring data sequence which have higher frequency.

3.3 PCA
The window of data ($96 \times 63$) is composed of values measured every two days (48 times a day) from 63 measurement points. There are 63 measurement points and 96 measurement times corresponding to the numbers of column vector and row vector, and there are 32 windows from October 16, 2012 to November 17, 2012.

![Figure 1 Time-series of PCA analysis results in 96 x 63/window](image)

Figure 1 (horizontal coordinates denotes time window, and longitudinal coordinates denotes cumulative contribution of variance of principal components) shows the results of the PCA: The number of principal components which can reach the certain cumulative contribution increased from October 16th to October 31st, gradually decreased from October 31st to November 14th, and then increased suddenly on November 17th. Meanwhile, the cumulative contribution rate of first principal component has presented a process of decreasing slowly firstly, then increasing slowly and finally decreasing suddenly. So the change in number of principal components reflects the dam state: it initially increased, then decreased with the deterioration of earth dam state, and suddenly increased at final burst. According to fractal theory, the decrease of principal components’ number before earth dam failure corresponds to the whole dimension reduction, and the sudden increase in November 17th reflects the discretization and dimension expanding of earth dam structure.
3.4 KPCA

3.4.1 Selection of Kernel Function. We view radial basis function, polynomial kernel function and sigmoid kernel function, etc as common kernel functions. The method of selecting kernel function includes Cross-Validation method, multiple experimental method, and hybrid kernel function.

In order to select the appropriate kernel function and kernel parameter, we choose two days (October 16th and October 17th) as test samples to extract the principal components. The discrimination of the rationality of parameters depends on the cumulative contribution rate of first principal component, and then the optimal kernel function is selected by comparing three kinds of kernel functions.

(1) Polynomial kernel function

\[ K(x, y) = (s(x_1, x_2) + b)^d \quad s, d \in R \quad \text{and} \quad c \geq 0 \]  

Using MATLAB 7.0, the contribution rate of first kernel principal component was defined as 75.71%. When \( s = 0.05, c = 0, d = 3 \). If \( s \) and \( c \) remain unchanged, when \( d = 2, d = 3, d = 4 \), the contribution rate of first kernel principal component is 76.80%, 78.68%, 79.60% respectively.

When \( s = 0.06, c = 0, d = 3 \), the largest contribution rate of first kernel principal component of polynomial kernel function is 85.71%.

(2) Neural network kernel function

\[ K(x, y) = \tanh(a(x \cdot y) + b) \quad (a, b \in R) \]  

When \( a = 0.01, b = 1 \), the largest contribution rate of first kernel principal component is 84.40%.

(3) Radial basis function

\[ K(x_1, x_2) = \exp \left( -\frac{||x_1 - x_2||^2}{2\sigma^2} \right) (\sigma \geq 0) \]  

According to the contribution rate of first kernel principal component, which corresponds to different values of \( \sigma \) in RBF (shows in Figure 2 and Table 1), when \( \sigma \geq 75 \), the contribution rate of first kernel principal component turns to a stable value of 87.5%, so we selected \( \sigma = 75 \).
By comparing the three kinds of kernel function above, we select radial basis function for KPCA, because it has the largest contribution rate of first kernel principal component, only one parameter and better stability during calculation.

\[ K(x_1, x_2) = \exp \left( -\frac{|x_1 - x_2|^2}{2 \times 75^2} \right) \]  

(6)

### 3.4.2 Kernel Principle Component Extraction

![Figure 3: Time-series of KPCA analysis results in 96×63/window](image)

Under a certain cumulative contribution rate, Figure 3 shows time-series of the number of kernel principal components in 96*63 per window. Figure 3 is similar to Figure 1, the number of kernel principal components increased from October 16th to October 30th, gradually decreased from November 3rd to November 14th, and increased suddenly on November 17th. Meanwhile, the cumulative contribution of first kernel principal component has presented a change process that has firstly decreased slowly, then increased slowly, and finally decrease suddenly. As depicted in the table, the kernel principal components’ number firstly increased, then decreased with the deterioration of dam state, and suddenly increased at the final burst. The consistent results of Figure 3 and Figure 1 show that the PCA or KPCA method can be applied to analyze the measured data of dam break.

### 4 Sensitivity Analysis of Early Warning Index

#### 4.1 Sensitivity of Windows

**4.1.1 Window Selection**

(1) Window(96×28)
There are 32 windows, from values measured every two days (48 times a day) of 28 representative measuring points, whose size is 96×28. Figure 4 shows time-series of KPCA based on 85% cumulative contribution rate.

Figure 4 Time-series of KPCA analysis results in 96×28/window

Figure 4 shows that the number of kernel principal components changed since October 18th, and came to the maximum number on October 27th and on October 30th. It decreased since November 12th, and then increased suddenly on November 17th. We can draw a conclusion that the 96×28 per window and 96×63 per window are basically consistent.

(2) Window(144×63)
Comparing the principal component window(96×63) and window(96×28), we find that the number of principal components of two class has the same trend in reflection of dam failure, but has difference in the discrimination of deterioration of earth dam state.

In order to analyze the sensitivity of two indexes to window size, we chose every three days as one window. Thus, the 144×63 per window has 63 measurement points and 144 measurement times. There are 16 windows during the analysis time.

We obtained the number and matrix of kernel principal component with cumulative contribution rate ≥85%, by using KPCA on 16 consecutive windows (144×63 per window) based on MATLAB. Figure 5 shows this time-series.

Figure 5 Time-series of KPCA analysis results in 144×63/window

As shown in Figure 5, the number of kernel principal components changes since October 18th, and came to the maximum number on October 30th. It decreased since November 11th, and then increased suddenly on November 17th. We can draw a conclusion that Figure 5 is basically consistent with Figure 1 and Figure 3.

(3) Window(144×28)
During the analysis time, there are 16 windows composed of measured data every three days, which size is 144×28, from 28 representative measurement points and 144 measurement times.
We obtained the number and matrix of kernel principal component which cumulative contribution rate ≥85% by using KPCA on 16 consecutive windows (144×28/window) based on MATLAB. The cumulative contribution processes of each kernel principal component are shown in Figure 6.

Figure 6 shows that the number of kernel principal components changed since October 18th, and came to the maximum number on October 30th. It decreased since November 11th, and then increased suddenly on November 17th. We can conclude that Figure 6 is basically consistent with former windows.

4.1.2 Sensitivity Evaluation. Table 2 shows a comprehensive comparison among different windows. We can find that the number of kernel principal components and the curve of contribution rate of first kernel principal component can reflect the evolution of the dam state. They are not sensitive to window selection. Therefore, the number of kernel principal components and the curve of contribution rate of first kernel principal component can be used as the warning indexes of dam-break.

| Window class | Window size | Method | Begin to change | Drastic change | Stable change | Sudden change |
|--------------|-------------|--------|-----------------|----------------|---------------|---------------|
| 1            | 96×63       | I      | 17              | 30             | 11            | 17            |
|              |             | II     | 17              | 30             | 11            | 17            |
| 2            | 96×28       | I      | 18              | 27             | 10            | 17            |
|              |             | II     | 17              | 30             | 12            | 17            |
| 3            | 144×63      | I      | 18              | 30             | 11            | 17            |
|              |             | II     | 18              | 30             | 11            | 17            |
| 4            | 144×28      | I      | 22              | 30             | 11            | 17            |
|              |             | II     | 24              | 30             | 11            | 17            |

Table 2 Date comparison of four windows results
(I: Number of principal components; II: Curve of principal component curve)

4.2 Norm Dependence

4.2.1 Norm Calculation. We calculated kernel principal component matrix norm of row norm, column norm and F-norm, using the formulas 7, 8 and 9.

\[
\|X\|_\infty = \max_i \sum_j^n |x_{ij}| \tag{7}
\]

\[
\|X\|_1 = \max_j \sum_i^n |x_{ij}| \tag{8}
\]

\[
\|X\|_F = \left(\sum_{i,j} x_{ij}^2\right)^{1/2} \tag{9}
\]

Calculate each kernel principal component matrix norm of four windows with three matrix norm.
calculation formulas above. Figure 7 shows the norm curve.

Figure 7 Comparison of three matrix norm in different windows
4.2.2 Sensitivity Evaluation. As shown in Figure 7, the number of kernel principal components have increased from October 17th to October 31st, decreased from October 31st to November 11th gradually, and increased suddenly on November 17th. Meanwhile, the kernel principal component matrix norm has presented a change process of decreasing slowly, then increasing slowly and finally decreasing suddenly. This is consistent with the process of dam failure. We can also find that the calculation results are consistent in spite of using different norm, which means the change of norm is not sensitive to the computational method of specific norm. Thus the principal component norm can be used as the early warning index of dam failure.

5 Conclusions
The homogeneous earth dam is regarded as a complex continuous-discrete system. We need more than ten-dimensional safety monitoring data to reflect the evolution of dam safety state accurately. However, the high-dimensional data has poor visibility and is hard to understand intuitively. In this work, a new one-dimension monitoring index is proposed by making full use of PCA or KPCA for dimension-reduction, which includes the number of the principal components/kernel principal components and the principal component/kernel principal component matrix norm whose variance cumulative contribution rate is accounted for about 85%. The measured data analysis of practical engineering shows that the index has the advantages of simple treating process, accurate reflection of dam safety and various robust parameters. Specifically, with the changes in the form of earth dam, the number of the principal components or kernel principal component and the principal component or kernel principal component matrix norm increase firstly, then decrease, and suddenly increase when a dam collapse. The changes of the two indices above are consistent with the process of dam-break and not sensitive to the specific window or the type of matrix norm. The next step is to study the dependence of the indices to the cumulative contribution rate. Furthermore, in order to strengthen the theoretical basis of this article, the theoretical basis from engineering mechanics mechanism and mathematical physics sense of PCA or KPCA should be further examined.

Acknowledgements
This work was supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK20151070) and the National Natural Science Foundation of China (Grant No.51479054).

References
[1] Fang W H. Dam Safety Monitoring: Problems, Views & Methodologies. Nanjing: Hohai University Press, 2012, 22-30. (in Chinese)
[2] Zheng F G. Application of principal component regression analysis in determining the component proportion of dam effect variable. Water Power, 2010, 36(12): 84-87. (in Chinese)
[3] Huang W D, Xu C, Duan Y G, Deng C F. Application of principal component regression to dam safety monitoring. Hydropower Automation and Dam Monitoring, 2009, 33(1): 57-59. (in Chinese)
[4] Mei Y T, Xu H L, Wang F. Time-varying prediction model of dam monitoring data based on principal component analysis. Water Power, 2011, 37(10): 100-103. (in Chinese)
[5] Yu H, Wu Z R, Bao T F, et al. Multivariate analysis in dam monitoring data with PCA. Sci China Tech Sci, 2010, 53: 1088–1097.
[6] Wang Hui. An application research on face recognition based on KPCA feature extract and SVM. Dissertation of Masteral Degree. Hefei: Hefei University of Technology, 2006. (in Chinese)
[7] Du Z M, Tu H, Geng G H. KPCA method research and application process. Computer Engineering and Applications, 2010, 46(7): 8-10. (in Chinese)
[8] Scholkopf B, Smola A Muller K R. Non-linear analysis as a kernel eigenvalue Problem[J]. Natural on computation, 1998, 10(5): 1299-1319.
[9] She P B, Mike S, Burge J Cetal. In Put Space Versus Feature Space in Kernel-based Methods[J]. IEEE Trans on Neural Networks, 1999, 10(5): 1000-1017.
[10] Gaudat Gand Anouar F. Generalized discriminate analysis using a kernel approach[J]. *Neural Network*, 2000, 12(10): 2385-2404.

[11] Honeine P, Richard C. Solving the pre-image problem in kernel machines: A direct method[J]. *Machine Learning for Signal Processing*, 2009, 53(4): 1-6.

[12] Liu G G, Liu H, Huang D. Design method and application with kernel function for pattern analysis. *Journal of East China University of Science and Technology (Natural Science Edition)*, 2007, 33(3): 405-409. (in Chinese)

[13] Chen S S. Simulation of the failure mechanism and break process of earth rock dam. Beijing: *China Water&Power Press*. 2012, PP: 30-38. (in Chinese)

[14] Zhang J Y, Li Y, Xuan G X, et al. Overtopping breaching of cohesive homogeneous earth dam with different cohesive strength. Sci China Ser E-Tech Sci, 2009, 52 (10): 3024—3029.

[15] Wang X F, Qiu J, Liu G J. Kernel Function Optimization in Kernel Principle Component Analysis and Its Application to Feature Extraction of Gear Faults. *Journal of Vibration, Measurement & Diagnosis*, 2007, 27(1): 62-64. (in Chinese)