Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 3739551, 11 pages
https://doi.org/10.1155/2021/3739551

Research Article

Outlier Detection Based on Multivariable Panel Data and K-Means Clustering for Dam Deformation Monitoring Data

Jintao Song[1,2], **Shengfei Zhang**[1,2], **Fei Tong**[1,2], **Jie Yang**[1,2], **Zhiquan Zeng**[3] and **Shuai Yuan**[1,2]

[1]*School of Water Resources and Hydro-Electric Engineering, Xi'an University of Technology, Xi'an 710048, China*
[2]*State Key Laboratory of Eco-Hydraulics in Northwest Arid Region, Xi'an 710048, China*
[3]*Powerchina Huadong Engineering Corporation, Hangzhou 310014, China*

Correspondence should be addressed to Jintao Song; zibet998@126.com

Received 4 November 2021; Revised 29 November 2021; Accepted 6 December 2021; Published 21 December 2021

Academic Editor: Cemal Ozer Yigit

A dam is a super-structure widely used in water conservancy engineering fields, and its long-term safety is a focus of social concern. Deformation is a crucial evaluation index and comprehensive reflection of the structural state of dams, and thus there are many research papers on dam deformation data analysis. However, the accuracy of deformation data is the premise of dam safety monitoring analysis, and original deformation data may have some outliers caused by manual errors or instruments aging after long-time running. These abnormal data have a negative impact on the evaluation of dam structural safety. In this study, an analytical method for detecting outliers of dam deformation data was established based on multivariable panel data and K-means clustering theory. First, we arranged the original spatiotemporal monitoring data into the multivariable panel data format. Second, the correlation coefficients between the deformation signals of different measuring points were studied based on K-means clustering theory. Third, the outlier detection rules were established through the changes of the correlation coefficients. Finally, the proposed model was applied to the Jinping-I Arch Dam in China which is the highest dam in the world, and results indicate that the detection method has high accuracy detection ability, which is valuable in dam safety monitoring applications.

1. Introduction

Since the 19th century, there have been many dam-break events in the world, such as Malpasset Dam (France, 1959), Vajont Dam (Italy, 1963), and Banqiao Dam (China, 1975), which brought heavy disasters and huge economic losses to the relevant countries [1, 2]. Dam safety monitoring is an effective means to monitor dam structure safety [3]. Thus, governments and dam engineering researchers began to attach great importance to dam safety monitoring, including dam deformation, seepage, stress and strain, etc. Among the various monitoring subjects, deformation is an comprehensive reflection of dam safety behaviors which can be effectively assessed through the analysis of dam deformation data [4].

The research on dam deformation analysis has experienced a stage from qualitative analysis to quantitative analysis and focuses on the causes and statistical model of dam deformation [5]. After the 20th century, with the gradual development of artificial intelligence, artificial intelligence algorithm is used to simulate the input-output relationship of dam deformation, and many high-precision analysis models are established [6–8].

A large number of dam safety analysis studies have been carried out based on the original deformation monitoring data. However, the accuracy of original deformation monitoring data is the foundation of dam safety analysis. At present, dam deformation data are mainly obtained through automatic system acquisition or manual reading, which may have some outliers due to the monitoring instrument aging, artificial error, structural state change, etc [9]. Deformation outlier usually deviates from the normal value, which affects the correctness of dam safety evaluation. Therefore, detecting the outliers of deformation data should be conducted before

the dam safety analysis using original deformation data. The outlier values are often considered an important reflection of the changes of structural safety behavior.

To date, the outlier detecting methods of dam safety monitoring data mainly include the process line method, statistical test method, and mathematical model method [10]. However, there are some problems in these methods. The process line method, as a simple and effective method, relies on human subjective judgment and is easy to be affected by subjective experience. The statistical test method is constructed by means of mathematical statistics. Different probability distribution models often lead to different results. The mathematical model method is aimed at establishing the regression equation between the dependent variable and the influence factors; due to the complexity of the actual influencing factors, the accuracy and applicability of the model are considered the key of the method.

Thus, most outlier detection methods focus on the change process of the monitoring data from the time level, ignoring its spatial relationship. A large number of monitoring data from different dams show that the deformation series of different measuring points have a significant correlation [11, 12]. Therefore, it is needed to study how to comprehensively consider the space-time relationship of dam deformation in outlier detection model and effectively obtain accurate dam deformation data.

The multivariable panel data integrate the temporal and spatial characteristics of deformation series and can describe the dynamic characteristics of dam deformation [13]. K-means clustering is a widely used clustering algorithm which is suitable for big data clustering and can be used in dam deformation detection [14, 15]. Thus, in this article, combining multivariable panel data with K-means clustering theory, the inherent distribution regularity of deformation data was obtained, and outliers of dam deformation were detected quantitatively based on the changing laws of the deformation relevance.

2. Methodology

2.1. Multivariable Panel Data. Traditional dam deformation clustering analysis usually uses monitoring data at a fixed time and contains one monitoring subject [12, 16, 17]. However, fixed time-based clustering analysis neglects dynamic change tendency of deformations, and single clustering subject cannot effectively reflect the overall situation of deformation. For example, the deformation value of a point includes both horizontal deformation and vertical deformation in the operation time. Multivariable panel data are transformed into a three-dimensional data format, including the dimension of monitoring subject, measuring point information, and time. Hence, multivariable panel data can effectively reflect the space-time information of dam deformation and provide a suitable database for clustering analysis of dam deformation.

Suppose that there are N deformation measuring points in the dam, and each measuring point has T measuring values and M monitoring subjects; then, the multivariable panel data are shown as follows.

Multivariable panel data are transformed into a three-dimensional data format, and from the dimension of measuring point number, Table 1 shows that the multivariable panel data could be represented by the matrix as follows:

$$y_i = \begin{bmatrix} x_{i1}(1) & x_{i1}(2) & \cdots & x_{11}(T) \\ x_{i2}(1) & x_{i2}(2) & \cdots & x_{i2}(T) \\ \vdots & \vdots & & \vdots \\ x_{iM}(1) & x_{iM}(2) & \cdots & x_{iM}(T) \end{bmatrix}, \quad i = 1, 2, \ldots, N. \tag{1}$$

From the dimension of measuring subject and time, the representation of matrix is similar to equation (1), which is not expressed again. When the dam deformation data are converted to multiindex panel data format, the following study is how to use clustering theory to analyze the correlation between each dam deformation measurement series.

2.2. K-Means Clustering Algorithm. After K-means clustering algorithm was proposed, it has been widely studied and applied in different disciplines. K-means clustering has been extensively applied in the field of dam safety evaluation, including clustering of displacement, seepage, and stress [18–21]. This algorithm has the characteristics of being simple and efficient, and thus it is still one of the most widely used clustering algorithms at present.

For a given dataset x containing P elements with the dimension of d, $x = \{x_1, x_2 \ldots x_P\} x_i \in R^d$.

The target of K-means clustering algorithm is to divide the elements of dataset into K categories. $x = \{c_k, k = 1, 2, \ldots K\}$. Each partition represents a category expressed as c_k, and each c_k has a center point μ_k. In order to quantify the similarity between dataset elements, K-means algorithm generally selects Euclidean distance as the measurement standard. Thus, the distance between the element and center point in the category c_k is expressed as follows:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2. \tag{2}$$

The key of K-means clustering is to minimize $J(c)$ which is the sum of distance, and the expression of $J(c)$ is

$$\begin{aligned} J(c) &= \sum_{k=1}^{K} J(c_k) = \sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} d_{ki} \|x_i - \mu_k\|^2, \end{aligned} \tag{3}$$

where $d_{ki} = 0$ if $x_i \notin c_i$ or $d_{ki} = 1$ if $x_i \in c_i$. According to the least square method and Lagrange principle, the cluster center point μ_k should be taken as the average value of element of the category c_k.

Combined with multivariable panel data theory and K-means clustering algorithm, we can quantitatively analyze the correlation of dam deformation and divide the dam deformation into several zones, and the specific process of dam deformation clustering is as follows:

Table 1: Multivariable panel data.

Time	1	⋯	t	⋯	T
Point no.	Subject				
	$1 \cdots j \cdots M$	⋯	$1 \cdots j \cdots M$	⋯	$1 \cdots j \cdots M$
1	$x_{11}(1) \ldots x_{1M}(1)$	⋯	$x_{11}(t) \cdots x_{1M}(t)$	⋯	$x_{11}(T) \cdots x_{1M}(T)$
⋯	⋯	⋯	⋯	⋯	⋯
i	$x_{i1}(1) \cdots x_{iM}(1)$	⋯	$x_{i1}(t) \cdots x_{iM}(t)$	⋯	$x_{i1}(T) \cdots x_{iM}(T)$
⋯	⋯	⋯	⋯	⋯	⋯
N	$x_{N1}(1) \cdots x_{NM}(1)$	⋯	$x_{N1}(t) \cdots x_{NM}(t)$	⋯	$x_{N1}(T) \cdots x_{NM}(T)$

i is the number of measuring points, j is the number of measuring subjects, t is the monitoring time, and $x_{ij}(t)$ represents the deformation value of the measuring point and monitoring subject numbered i and j, respectively, at time t ($i = 1, 2, \ldots N$, $j = 1, 2, \ldots M$, $t = 1, 2, \ldots T$).

(1) Transform the deformation data into multivariable panel data format.

(2) Randomly take K elements from panel data as the center point of category c_k.

(3) Calculate the distance between the remaining elements and the cluster centers, respectively, and then classify the remaining elements according to the distance from the center point.

(4) According to the clustering results by step 2, the cluster centers are recalculated by taking the mean of elements in each cluster.

(5) Repeat step 2 and step 3 to calculate the new cluster center continuously.

(6) The process is stopped until the clustering result does not change.

(7) Output the dam deformation clustering result.

3. Outlier Detection Method

After the K-means clustering method is used to derive the deformation clustering features, the dam is divided into several deformation zones. The deformation zone numbered i has n_i measuring points. The outlier detection matrix of this deformation zone can be expressed as follows:

$$K_i = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n_i} \\ d_{21} & d_{22} & \cdots & d_{2n_i} \\ \vdots & \vdots & & \vdots \\ d_{n_i1} & d_{n_i2} & \cdots & d_{n_in_i} \end{bmatrix}, \tag{4}$$

where d_{ab} is the correlation coefficient of the deformation data between the measuring point a and b. The coefficients can be calculated for each measuring point in this deformation zone as follows:

$$d_{ab}(t) = \sqrt{\sum_{j=1}^{M} \left(\Delta x_a(j,t) - \Delta x_b(j,t)\right)^2}, \tag{5}$$

where M is the number of monitoring subjects. In order to better identify outliers through deformation change analysis, we focus on the variation of deformation values which are $\Delta x_a(j,t) = x_a(j,t) - x_a(j,t-1)$. $x_a(j,t)$ is the deformation value of the monitoring subject of j at time t for the monitoring point a.

When the change laws of deformation signals of a deformation zone are normal, the coefficient between different measuring points in the same deformation zone is close. When the change laws of deformation data at one measuring point become abnormal, remarkable changes will occur in the coefficients of this measuring point with other measuring points. Thus, the detection of deformation outliers can be transformed into the detection of the outliers of correlation coefficient. The core problem of outlier recognition is the formulation of outlier standard. The detection criteria of abnormal value of dam deformation are proposed below.

In order to quantify the variation difference of deformation data of each measuring point, the correlation coefficient vector of measuring point m can be estimated as

$$\overrightarrow{O_m}(t) = \left(d_{m1}(t), d_{m2}(t), \ldots, d_{mn}(t)\right), \tag{6}$$

where m is the number of the measuring point and n is the total number of measuring points in the zone. $\overrightarrow{O_m}(t)$ corresponds to the m row vector of the outliers detection matrix K at time t. The vector module is defined to quantify the change amplitude of the correlation coefficient vectors $\overrightarrow{O}_m$, that is,

$$\left|\overrightarrow{O_m}(t)\right| = \sqrt{d_{m1}^2(t) + d_{m2}^2(t) + \cdots d_{mn}^2(t)}. \tag{7}$$

If the vector module of a measuring point is significantly bigger than the other measuring points in the zone, the deformation of this measuring point may have outliers. In order to evaluate the degree of significant difference quantitatively, the research introduces the 3σ-rule which is the classical outlier analysis method.

For the dataset $X = \left\{|\overrightarrow{O_1(t)}|, |\overrightarrow{O_2(t)}|, \cdots |\overrightarrow{O_m(t)}|\right\}$, $t = 1, 2, \ldots, T$,

If X satisfies Gaussian distribution, then the parameter of Gaussian distribution is

$$\begin{cases} \widehat{\mu} = \overline{X} = \dfrac{\sum_{t=1}^{T}\sum_{m=1}^{n}\left|\overrightarrow{O}_m(t)\right|}{nT}, \\ \widehat{\sigma}^2 = \dfrac{\sum_{t=1}^{T}\sum_{m=1}^{n}\left(\left|\overrightarrow{O}_m(t)\right| - \widehat{\mu}\right)^2}{nT}. \end{cases} \tag{8}$$

In the 3σ-rule, if data are bigger than $\widehat{\mu} + 3\widehat{\sigma}$ or smaller than $\widehat{\mu} - 3\widehat{\sigma}$, they are considered as outliers. However, when $|\overrightarrow{O}_m|$ is smaller than $\widehat{\mu} - 3\widehat{\sigma}$, it shows that the deformation law of this measuring point is more similar to that of other measuring points in the same zone. Thus, the outlier detection rule is

$$\left|\overrightarrow{O}_m(t)\right| > \widehat{\mu} + 3\widehat{\sigma}. \tag{9}$$

If the deformation data of a measuring point meet the above formula, it is considered that the data of the measuring point at time t are abnormal. The outlier identification process of dam deformation is designed as shown in the flowchart (Figure 1) below.

4. Case Study

4.1. Project Overview. Jinping-I concrete arch dam is located at the Yalong River, which is the key river in Sichuan Province in China. The height of the dam is 305 m, which is the highest in the world. The normal water level of the reservoir is 1880 m, and the total storage is 7.76 billion cubic meters [22]. The power station is mainly used for power generation and flood storage.

According to the requirements of dam monitoring specifications, a large number of monitoring instruments are embedded in the dam to comprehensively monitor the deformation, seepage, stress, and strain of the structure. In order to analyze the outliers of dam deformation, the following focuses on the conditions of dam deformation monitoring. Figure 2 shows the layout of deformation measuring points along the height direction.

The dam began to store water after it was completed in December 2012. On December 1, 2012, it began to store water for the first time, and the reservoir water level began to rise from 1648.37 m. On June 15, 2013, the diversion bottom outlet was closed, and the second stage of water storage began, and the reservoir water level began to rise from 1712.48 m. On August 26, 2013, the third stage of water storage began, and the water level reached 1838.66 m on November 20. On December 11, the water level reached 1880 m, and then the water level remained stable. Since the dam deformation during the impoundment period is relatively important for dam safety and may have outliers, the dam deformation data from December 2012 to December 2014 are selected as the object for outlier detection analysis.

4.2. Clustering Analysis of Dam Deformation Data. The first step of outlier analysis is to study the zoning characteristics of the dam using multivariable panel data and K-means panel clustering analysis method. Due to the large number of measuring points, this section takes PL9-3, PL11-5, and PL13-5 as an example to show the multivariable panel data of dam deformation. Table 2 shows the multivariable panel data of Jinping-I Dam during the impoundment stage (2012.12.1–2014.12.11). δ_x and δ_y are radial displacement and tangential displacement, respectively. The deformation data of typical dam segments are shown in Figures 3 and 4.

According to the K-means clustering algorithm, $k = 8$ is selected in this paper in order to divide the deformation of the dam into 8 zones which could fully reflect the zoning characteristics of the dam. Each monitoring point in the same deformation zone has highly similar deformation law. Figure 5 shows the results of deformation zones (zones I, II, III, IV, V, VI, VII, and VIII) in the dam. When deformation zones of the dam are determined, the proposed outlier detection analysis method of dam deformation data could be used to evaluate whether the deformation data of a monitoring point have outliers or not.

4.3. Outlier Detection Analysis of Dam Deformation Data. Due to the large number of measuring points in the dam, when detecting the outliers of deformation data, deformation zone I in Figure 5 is taken as an example. For convenience of numbering, PL11-3, PL11-4, PL13-3, and PL13-4 in deformation zone I are set as measuring points 1, 2, 3, and 4, respectively, and d_{ab} is the correlation coefficient between measuring points a and b in the deformation series. Figure 6 represents the calculated results of correlation coefficients in this deformation zone during the impoundment period.

The module values of correlation coefficient vectors of all measuring points in zone I could be calculated based on equation (7). The changing laws of module values of different monitoring points in zone I are expressed in Figures 7–10. In order to evaluate whether the deformation data of each measuring point in deformation zone I contain outliers, according to the proposed outlier detection method, the threshold value of 3σ-rule should be calculated. According to the calculated results of deformation data in zone I, $\widehat{\mu} = 0.38$, $\widehat{\sigma} = 0.61$. Therefore, the detection criterion for outliers in equation (9) could be expressed as $|\overrightarrow{O}_m(t)| > \widehat{\mu} + 3\widehat{\sigma} = 0.38 + 3 \times 0.61 = 2.21$. On the basis of the diagnostic criteria of outliers, outliers can be identified for the deformation measured values of each measuring point in deformation zone I, and the identification results are marked in Figures 7–10.

4.4. Comparative Analysis of Outlier Detection Results. In order to verify the effectiveness of the proposed outlier detection method, this section selects the detection results of two methods as a comparison. One is the pauta criterion (3σ rule), and the other is the manual inspection. Pauta criterion is a widely used statistical test method for outliers, that is, to test whether the measured value is more than 3 times the sample standard deviation. The manual inspection method verifies whether there are manual recording errors in the

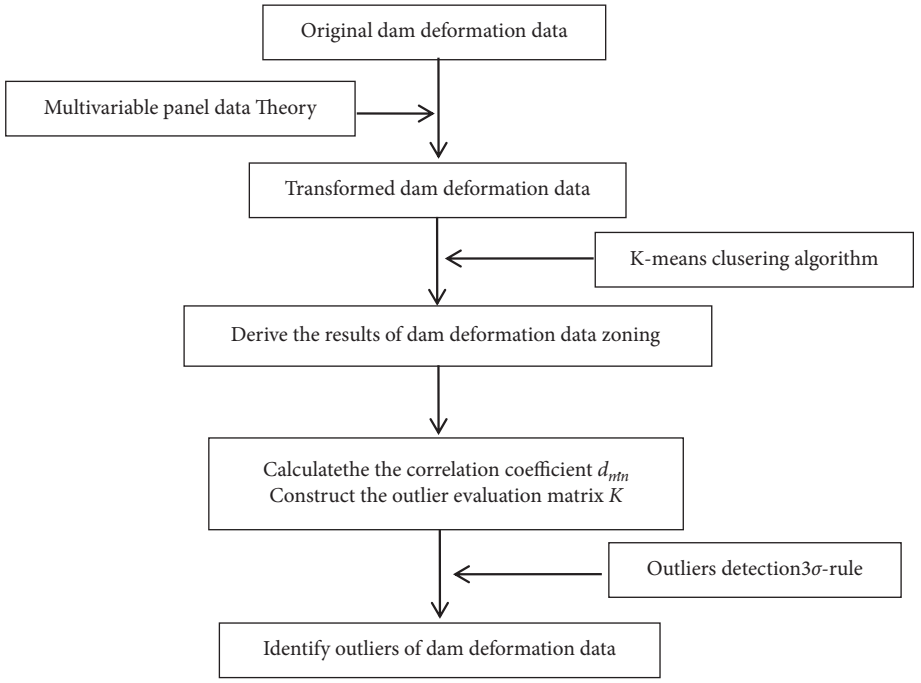

Figure 1: Flowchart of dam deformation outlier detection.

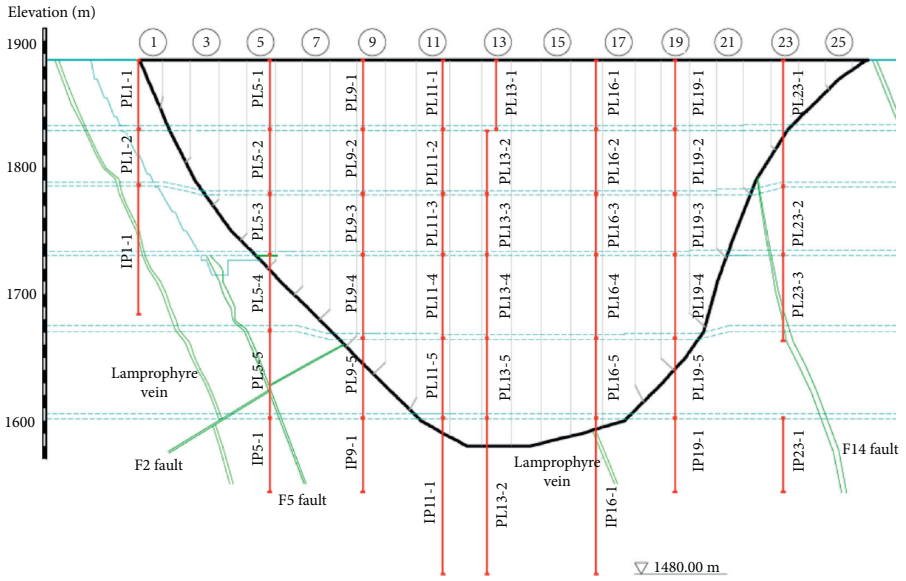

Figure 2: Distribution diagram of deformation measuring points on Jinping-I Dam.

Table 2: Multivariable panel data of Jinping-I Dam during impoundment stage.

Time	2012.12.1	⋯	2013.10.1	⋯	2014.12.11
Point no.	Subject				
	δ_x, δ_y (mm)	⋯	δ_x, δ_y (mm)	⋯	δ_x, δ_y (mm)
PL9-3	−0.21, −0.12	⋯	7.00, 0.90	⋯	28.80, 5.43
⋯	⋯	⋯	⋯	⋯	⋯
PL11-5	−0.05, −0.01	⋯	14.86, 2.69	⋯	27.18, 5.04
⋯	⋯	⋯	⋯	⋯	⋯
PL13-5	0.03, −0.08	⋯	18.20, 1.68	⋯	32.09, 3.24

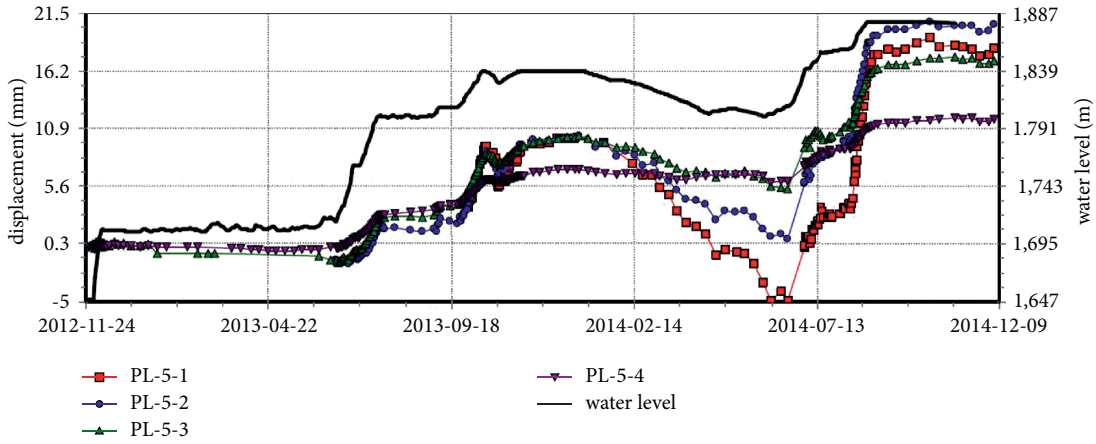

Figure 3: Measured radial displacement process lines of dam segment #5.

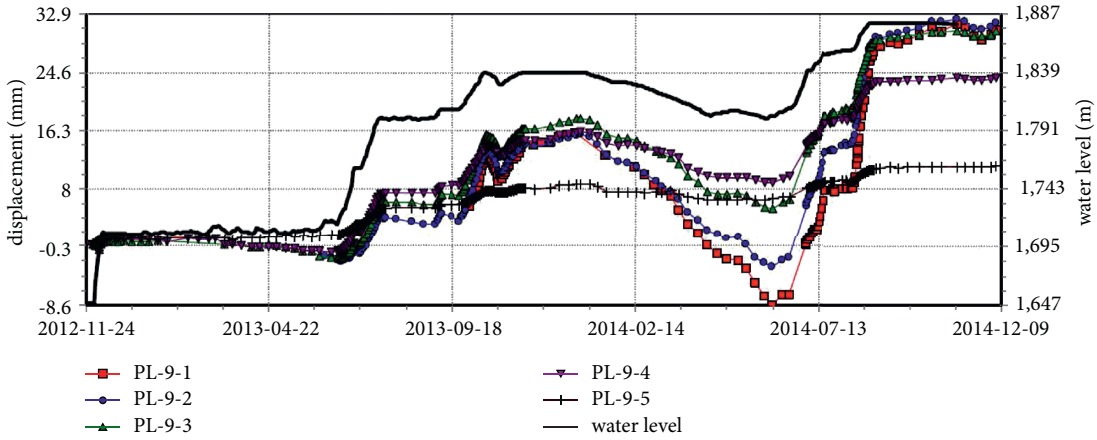

Figure 4: Measured radial displacement process lines of dam segment #9.

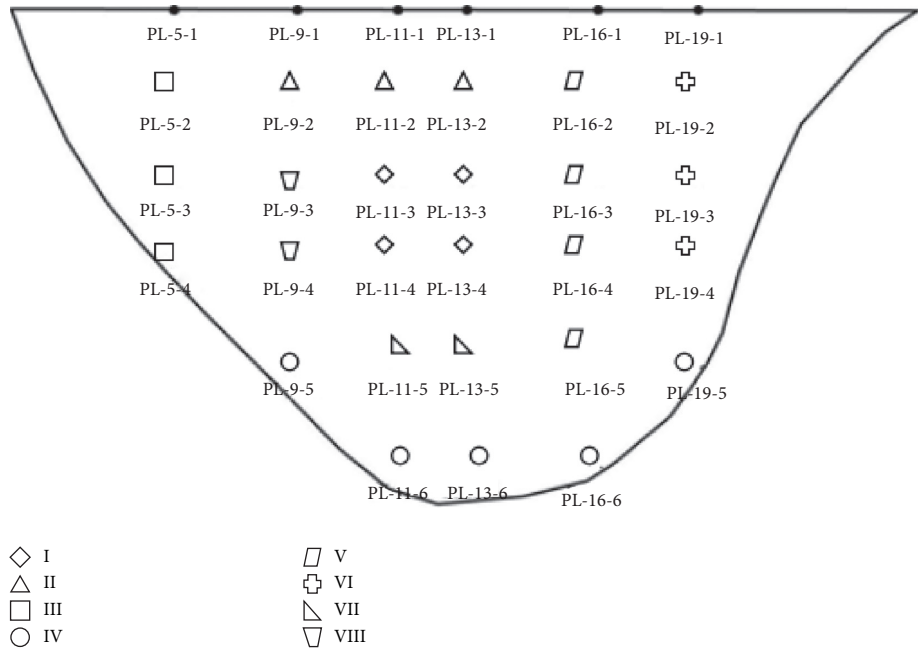

Figure 5: Schematic of deformation zoning of the Jinping-I Dam.

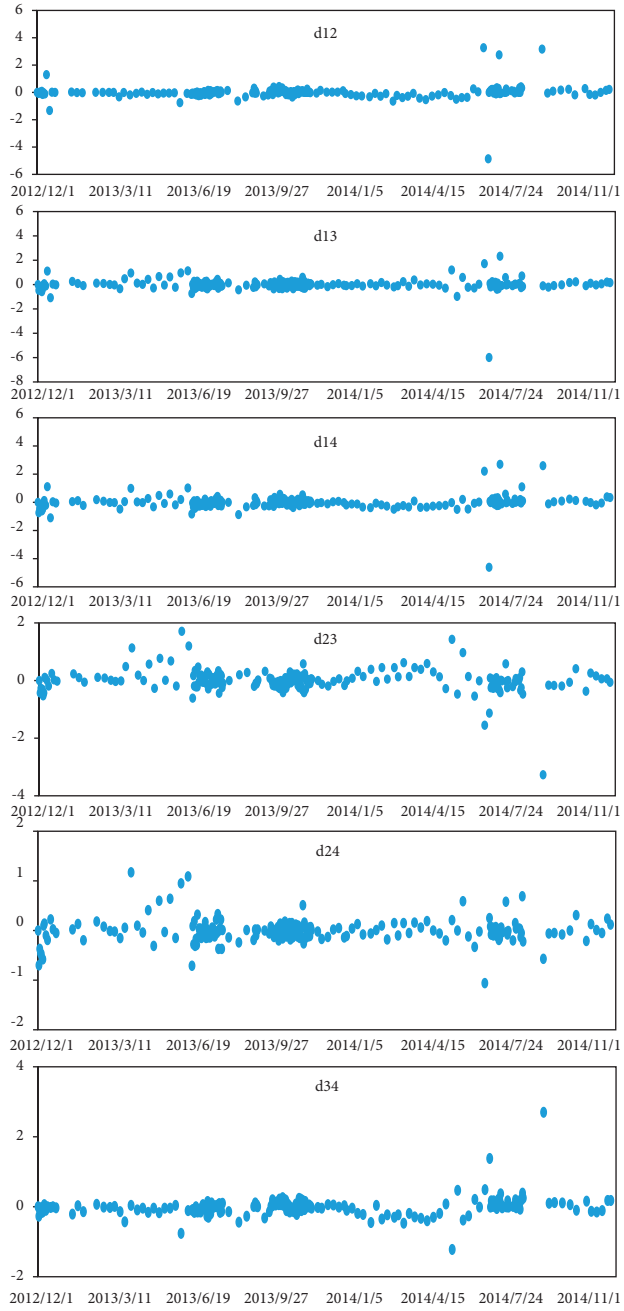

Figure 6: Results of correlation coefficients in deformation zone I during impoundment period.

monitoring data through the verification of the original monitoring data by the monitoring recorder.

4.4.1. Outlier Detection Results of the Proposed Method. In most of the time period, the change rule of deformation vector modulus of each measuring point is relatively stable which means $|\overrightarrow{O}_m(t)| < 2.21$, and the change rule of deformation data is normal. However, it can be seen from Figures 7–10 that the four observation points have the same time of abnormal points, which are June 19, 2014, July 15, 2014, and August 14, 2014. The original monitoring data of deformation variation at four measuring points are shown in Figures 11–14. It can be seen from the figures that the abnormal time of large deformation variation is completely consistent with the above outlier analysis. The outliers of each measuring point at three time points are marked on the graph.

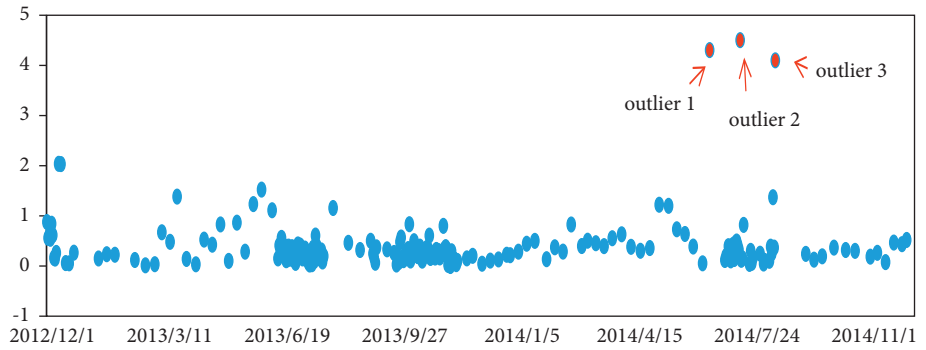

Figure 7: Distribution graph of module values of PL11-3 during impoundment period.

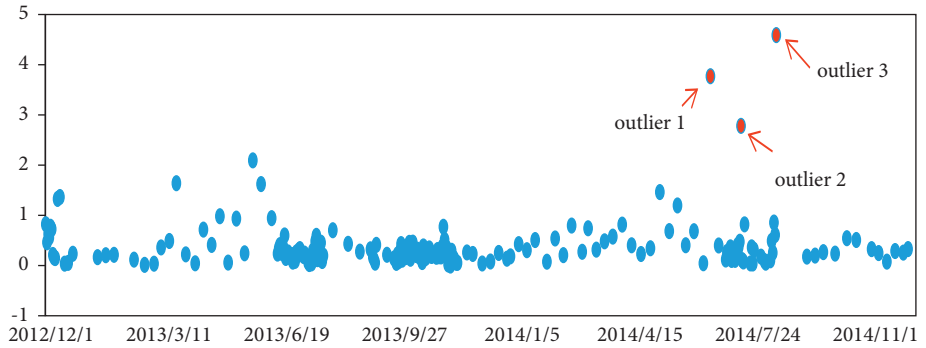

Figure 8: Distribution graph of module values of PL11-4 during impoundment period.

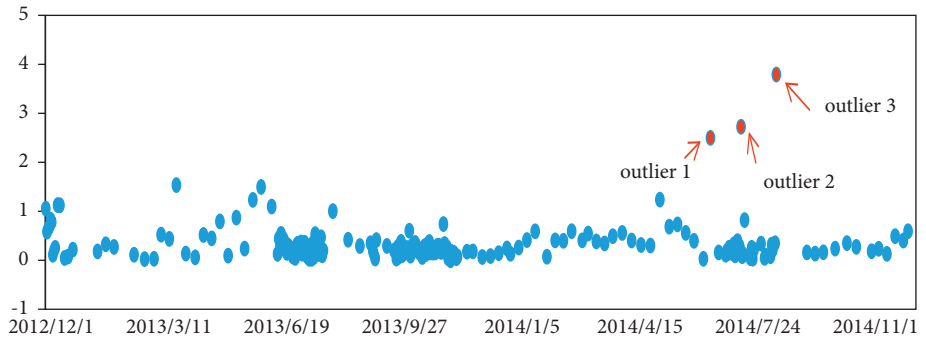

Figure 9: Distribution graph of module values of PL13-3 during impoundment period.

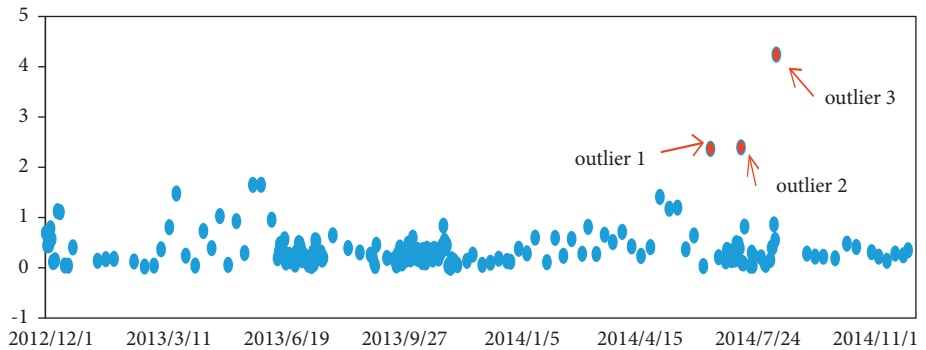

Figure 10: Distribution graph of module values of PL13-4 during impoundment period.

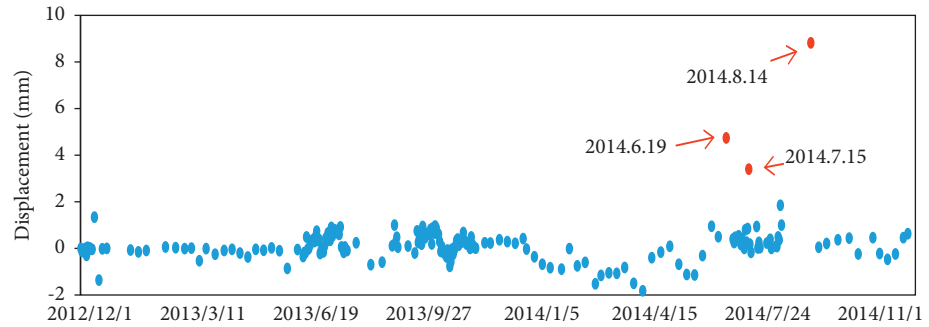

Figure 11: Distribution graph of deformation variation of PL11-3 during impoundment period.

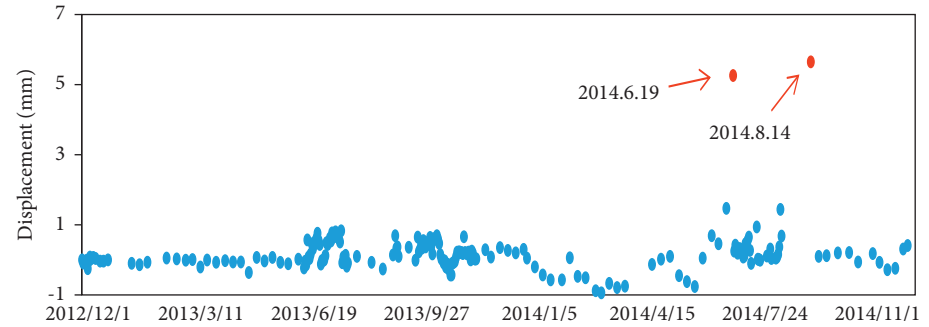

Figure 12: Distribution graph of deformation variation of PL11-4 during impoundment period.

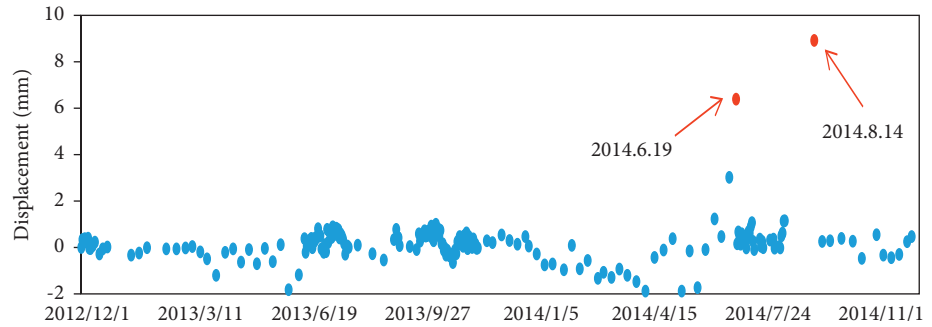

Figure 13: Distribution graph of deformation variation of PL13-3 during impoundment period.

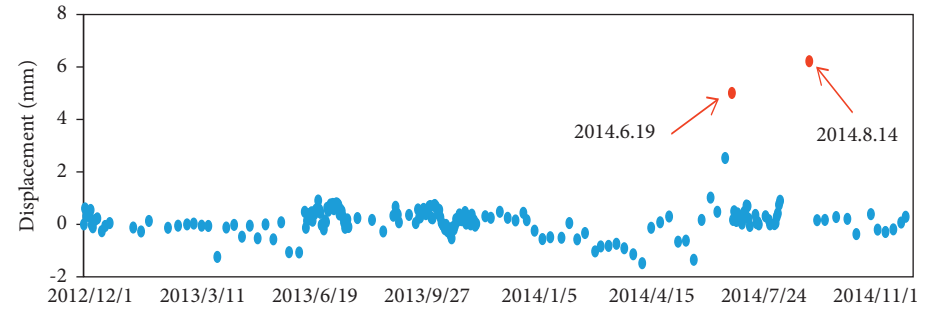

Figure 14: Distribution graph of deformation variation of PL13-4 during impoundment period.

Table 3: Statistical parameters of deformation data.

Statistical parameters	PL11-3	PL11-4	PL13-3	PL13-4
Mean value	0.19	0.19	0.19	0.20
Standard deviation	0.88	0.64	0.94	0.71

Table 4: Deformation change of measuring points on abnormal date.

Date	PL11-3	PL11-4	PL13-3	PL13-4
2014.6.19	1.91(Y)	3.14(Y)	3.367(Y)	2.70(Y)
2014.7.15	0.57(Y)	−1.47(N)	−1.95(N)	−1.61(N)
2014.8.14	5.99(Y)	3.53(Y)	5.90(Y)	3.91(Y)

4.4.2. Outlier Detection Results of the 3σ Method. According to the monitoring data of four measuring points, the mean value and standard deviation of each sample are calculated, respectively. The calculation results are shown in Table 3.

Relying on the 3σ rule, the calculation index of each sequence is calculated by the following formula:

$$d = |y - \mu| - 3\sigma. \tag{10}$$

If $d < 0$, it means that the difference between the measured value and the sample mean does not exceed 3 times the standard deviation, and the measured value is normal; if $d \geq 0$, it means that the difference between the measured value and the sample mean exceeds 3 times the standard deviation, and the measured value is abnormal.

In order to compare and verify the effectiveness of the proposed method, the d value corresponding to the above three abnormal dates is analyzed. The calculation results of the three abnormal dates are shown in Table 4. Y in the bracket indicates that the measured value is an abnormal data through detection; N indicates that the measured value is a normal data through detection.

By comparing the results of Table 4 and Figures 11–14, it can be found that the outliers detected by the two methods are completely consistent.

4.4.3. Outlier Detection Results of the Manual Inspection Method. The monitoring data are checked by the monitoring recorder. On June 19, 2014, and August 14, 2014, due to the staff's routine inspection of the monitoring instruments, the interval time between the measured values is not one day, but seven days, during which there are no monitoring data, so the difference between the adjacent monitoring data is large, and thus the deformation data in these two time periods are judged as abnormal values. These abnormal values are not caused by structural changes or monitoring errors, but due to the lack of measured values. On July 15, 2014, the deformation variation of PL11-3 is 3.40 mm, and the deformation variation of the other three measuring points is less than 1.00 mm; By checking with the monitoring data management department of the project, it is found that the original data should be 0.20 mm, and the abnormal value is caused by manual recording error. Hence, the proposed method effectively identifies the outliers of dam deformation data.

5. Conclusion

In this study, multivariable panel data theory and K-means clustering algorithm are combined to construct an outlier detection model for dam deformation monitoring data. The conventional outlier recognition mainly aims at the single measurement point sequence and uses the probability and statistics method to define and diagnose outliers. The proposed method detects abnormal data of dam deformation through clustering analysis and effectively considers the relevance between different measuring points, which avoids the influence of short data sequence or difficult expression of probability function in conventional outlier recognition. The research on the characteristics of deformation zoning plays a positive role in studying the overall deformation behavior and safety evaluation of the dam. Through the analysis of a typical dam project, it is found that the proposed method can effectively identify the outliers in the dam deformation data and provide reliable information foundation for dam researchers and management personnel.

Data Availability

This paper involves the deformation monitoring data of Jinping Dam. Because the dam is the highest dam in China, the data are confidential, so they cannot be disclosed.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant nos. 52109166 and 52039008).

References

[1] J. Yang, J. Qu, Q. Mi, and Q. Li, "A CNN-lstm model for tailings dam risk prediction," *IEEE Access*, vol. 8, Article ID 206491, 2020.

[2] H. M. Amin, "Risk, reliability, resilience and beyond in dam engineering: a state-of-the-art review," *International Journal of Disaster Risk Reduction*, vol. 31, pp. 806–831, 2018.

[3] A. De Sortis and P. Paoliani, "Statistical analysis and structural identification in concrete dam monitoring," *Engineering Structures*, vol. 29, no. 1, pp. 110–120, 2007.

[4] X. He, C. Gu, Z. Wu, and H. Su, "Dam risk assistant analysis system design," *Science in China–Series E: Technological Sciences*, vol. 51, no. S2, pp. 101–109, 2008.

[5] B. Wei, B. Liu, D. Yuan, Y. Mao, and S. Yao, "Spatiotemporal hybrid model for concrete arch dam deformation monitoring considering chaotic effect of residual series," *Engineering Structures*, vol. 228, pp. 101–112, 2021.

[6] C. Y. Kao and C. H. Loh, "Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches," *Structural Control and Health Monitoring*, vol. 20, no. 3, pp. 282–303, 2013.

[7] W. Chen, X. Wang, Z. Cai, and C. Liu, "DP-GMM clustering-based ensemble learning prediction methodology for dam deformation considering spatiotemporal differentiation," *Knowledge-Based Systems*, vol. 222, pp. 23–29, 2021.

[8] Y. Li, K. Min, Y. Zhang, and L. Wen, "Prediction of the failure point settlement in rockfill dams based on spatial-temporal data and multiple-monitoring-point models," *Engineering Structures*, vol. 243, pp. 112–119, 2021.

[9] H. Su, Z. Wen, Z. Chen, and S. Tian, "Dam safety prediction model considering chaotic characteristics in prototype monitoring data series," *Structural Health Monitoring*, vol. 15, pp. 629–639, 2016.

[10] S. Park, N. S. Park, S.-s. Kim, G. Jo, and S. Yoon, "Outlier detection of water quality data using ensemble empirical mode decomposition," *Journal of Korean Society of Environmental Engineers*, vol. 43, no. 3, pp. 160–170, 2021.

[11] Y. Hu, C. Shao, C. Gu, and Z. Meng, "Concrete dam displacement prediction based on an ISODATA-GMM clustering and random coefficient model," *Water*, vol. 11, no. 4, pp. 714–720, 2019.

[12] J. Hu and F. Ma, "Zoned deformation prediction model for super high arch dams using hierarchical clustering and panel data," *Engineering Computations*, vol. 20, pp. 15–22, 2020.

[13] L. Cheng, T. Zhang, L. Chen et al., "Investigating the impacts of urbanization on PM2.5 pollution in the yangtze river delta of China: a spatial panel data approach," *Atmosphere*, vol. 11, no. 10, pp. 1058–1067, 2020.

[14] Z. L. Jiang, N. Guo, Y. Jin et al., "Efficient two-party privacy-preserving collaborative k-means clustering protocol supporting both storage and computation outsourcing," *Information Sciences*, vol. 518, pp. 168–180, 2020.

[15] K. Behzadian, A. H. Eghbali, K. Behzadian, F. Hooshyaripor, R. Farmani, and A. P. Duncan, "Improving prediction of dam failure peak outflow using neuroevolution combined with K-means clustering," *Journal of Hydrologic Engineering*, vol. 22, no. 6, 2017.

[16] C. Bo, T. Hu, Z. Huang, and C. Fang, "A spatio-temporal clustering and diagnosis method for concrete arch dams using deformation monitoring data," *Structural Health Monitoring*, vol. 18, pp. 24–32, 2018.

[17] J. Hu and F. Ma, "Comparison of hierarchical clustering based deformation prediction models for high arch dams during the initial operation period," *Journal of Civil Structural Health Monitoring*, vol. 11, pp. 1–18, 2021.

[18] H. Q. Guo, B. Wen, and X. F. Bai, "Study of seepage properties of fractured rock mass based on improved K-means clustering algorithm," *Applied Mechanics and Materials*, vol. 405-408, pp. 310–315, 2013.

[19] M. F. M. Yunoh, S. Abdullah, M. H. M. Saad, Z. M. Nopiah, and M. Z. Nuawi, "K-means clustering analysis and artificial neural network classification of fatigue strain signals," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 39, no. 3, pp. 757–764, 2017.

[20] J. Z. C. Lai, T. J. Huang, and Y. C. Liaw, "A fast K-means clustering algorithm using cluster center displacement," *Pattern Recognition*, Elsevier Science Inc, vol. 42, 2009.

[21] S. Wang, C. Xu, Y. Liu, and B. Wu, "Mixed-coefficient panel model for evaluating the overall deformation behavior of high arch dams using the spatial clustering," *Structural Control and Health Monitoring*, vol. 28, no. 10, 2021.

[22] S. Y. Wu, W. Cao, and J. Zheng, "Analysis of working behavior of Jinping-I Arch Dam during initial impoundment," *Water Science and Engineering*, vol. 9, no. 3, p. 9, 2016.