Outlier Detection of Slope Deformation Monitoring Data based on WMA-3σ

Bin Li[1,2,3], Xingping Bai[1,2,*], Huanhuan Gao[1], Ting Liu[1]

[1]Northwest Engineering Corporation Limited, Power China, Xi'an, China
[2]Xi'an Key Laboratory of Clean energy digital technology, Xi'an, China
[3]Xi'an University of Technology, Xi'an, China

*Corresponding author's e-mail:278335788@qq.com

Abstract. The outliers in slope deformation monitoring data often contain important information. The influence of external environment, the failure of slope structure and the failure of monitoring instrument are the important reasons for the outliers. Rapid and accurate detection of outliers is not only the basic work of data analysis and calculation, but also an important measure to find out whether the slope is safe in time. Slope deformation monitoring data is time series. The short-term changes is smooth and stable with strong autocorrelation. In this paper, an adaptive weight calculation method was proposed for Weighted Moving Average (WMA) algorithm. The algorithm can estimate the measured data with high precision without being affected by outliers. Then, the difference sequence between the estimated data and the measured data was calculated, and the mirror processing was proposed for the difference sequence. In order to eliminate the asymmetric distribution of the difference sequence caused by the trend of the measured data. Finally, the outliers of the difference sequences after mirror processing were detected using the 3σ criterion. Thus the outlier detection of the measured data is realized. Through example analysis, WMA-3σ method can accurately detect outliers in the measured data. It has important reference significance for real-time and efficient outlier detection and intelligent data analysis.

Keywords: Slope deformation, Monitoring data, Outlier detection, Weighted moving average, 3σ criterion.

1 Introduction

Slope safety problems are often encountered when constructing projects in high mountains and valleys. Monitoring instruments are usually embedded in the slope to collect data on slope deformation, groundwater, cracks, microearthquakes, etc. The acquisition of monitoring data in the project is only the foundation. How to obtain useful information from the data to ensure slope safety is the key. Through the monitoring data process line analysis, comparative analysis, regression analysis and so on, it can better understand and master the changing process and safety state of the slope. Then, appropriate measures are taken to ensure slope safety. However, in the process of monitoring

P. Xiang et al. (eds.), *Proceedings of the 2023 5th International Conference on Hydraulic, Civil and Construction Engineering (HCCE 2023)*, Atlantis Highlights in Engineering 26,
https://doi.org/10.2991/978-94-6463-398-6_19

data collection, there will inevitably be a few outliers. These outliers may be caused by the interference of the surrounding environment or the instrument failure, or by the change of the safety state of the slope. The reasons are more complicated. The existence of outliers directly affects the result of data analysis and the accuracy of calculation. Therefore, outlier detection is not only the basic work of data analysis and calculation, but also an important measure to find out whether the slope is safe in time. The high quality data after eliminating outliers is an important content of slope deformation prediction and early warning analysis, which is of great value for discovering potential safety hazards of slope in time and ensuring slope safety [1-3].

Outlier detection methods are usually based on clustering or classification [4-6], based on wavelet analysis [7-8], based on statistics [9-10] or based on a variety of other algorithms [11-13]. However, no matter what kind of algorithm, it is necessary to determine a suitable threshold to judge whether the data is an outlier [14-15]. In hydropower engineering, statistical method is often used to detect the outlier of slope monitoring data, among which the 3σ criterion is the most widely used [16-17]. The 3σ method needs to assume that the data meets the normal distribution, and then calculate the standard deviation σ of the data. If a certain data exceeds the range of 3 times standard deviation (3σ), the occurrence of the data is considered as a small probability event and it is regarded as an outlier. It is usually necessary to establish a regression model by using statistical regression, neural network, time series or other methods to calculate the estimated value of the measured data, and then obtain a stable difference sequence according to the estimated value and the measured value, and finally establish the 3σ criterion for the difference sequence. When new data is obtained, it is only necessary to calculate whether the difference value between the new value and the estimated value satisfies the 3σ criterion to determine whether the new value is an outlier [18-20]. However,in the process of establishing the regression model, it is necessary to ensure that the training data are all normal data. Otherwise, the outliers in it will have an adverse impact on the modeling. Therefore, for the measured data obtained for the first time, outlier detection should be carried out before modeling or other data analysis. If the amount of data is small, manual detection can be used, but when faced with a large number of data and intelligent detection needs, manual detection is difficult to meet the work requirements. Therefore, outlier detection is the basic work of data analysis and is very important.

In this paper, for the features of slope deformation monitoring data, an outlier detection algorithm, WMA-3σ, is proposed for slope deformation monitoring data. For the WMA method, an adaptive weight calculation method is proposed, which can avoid the influence of outliers in the measured data on the estimated value. Thus, a stable and reliable difference sequence can be obtained from the estimated value and the measured value. In addition, the trend in the measured data will lead to the asymmetric distribution of the difference sequence. So, this paper put forward the mirror processing to solve this problem, and finally adopted 3σ criterion to detect the outliers for the difference sequence after mirror processing. Through engineering example analysis, the estimated value calculated by WMA-3σ method is not affected by the outliers in the measured data, and the accuracy is high, the outlier detection results were good. This method can find the hidden danger of slope safety in time, and provides a reference for the intelligent analysis of engineering data.

2 WMA and 3σ criterion

2.1 Weighted Moving Average

Moving averages are the basis for weighted moving averages. Moving average is a method of forecasting time series. In a set of data, there are n consecutive values, all of which have the same weight, and then the value of the next moment can be calculated according to the n values and its weight. The formula is as follows.

$$F_i = \frac{1}{n}(x_{i-1} + x_{i-2} + \cdots + x_{i-n}) \quad (1)$$

Where: $_i$ is the predicted value at the i time; x_{j-n} is the measured value at time i-n; n is the range of values.

Weighted moving average assigns different weights to the n consecutive pieces of data to predict the data in the next moment. The formula is as follows.

$$F_i = \frac{1}{n}(\omega_1 x_{i-1} + \omega_2 x_{i-2} + \cdots + \omega_n x_{i-n}) \quad (2)$$

Where: is the weight of the measured value at time i-n, $\sum_{j=1}^{n} \omega_j = 1$.

The weight calculation of weighted moving average method is very important, and whether the weight is reasonable directly affects the accuracy of the prediction result. In general, the latest data is the most important for predicting the future, so the weight should be higher. This is also in line with the slope deformation process.

2.2 The 3σ criterion

The applicable premise of the 3σ criterion is to assume that the data series meets the normal distribution. This method needs to calculate the mean value μ and standard deviation σ of the data, and then select a suitable interval according to a certain probability. If a data exceeds this interval, it is not a random noise but an outlier. On the contrary, it is a normal data. The 3σ criterion is defined as follows: The probability that the data is distributed in (μ-σ, μ+σ) is 0.6827; in (μ-2σ, μ+2σ) is 0.9545; in (μ-3σ, μ+3σ) is 0.9973. The 3σ criterion means that almost all data is in the (μ-3σ, μ+3σ) range, and the probability of exceeding this range is less than 0.3%.

Let a data series be $\{x_1, x_2, \ldots, x_j, x_{j+1}, \ldots, x_t\}$, whose mean value is calculated as follows.

$$\mu = \frac{1}{t}\sum_{i=1}^{t} x_i \quad (3)$$

The standard deviation is:

$$\sigma = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(x_i - \mu)^2} \tag{4}$$

3 WMA-3σ method

3.1 Basic Principle

Aiming at the problem of the first outlier detection of measured data, the WMA-3σ outlier detection method is proposed in this paper. Considering the characteristics of slope deformation monitoring data with strong autocorrelation and stable change in the short term, the WMA-3σ algorithm is first used to estimate the measured data at a certain time and the previous c time. Then the difference between the estimated data and the measured data is detected by outlier based on the 3σ criterion.Finally, the outlier detection of the measured value can be realized.

3.2 Calculation Step

Before detecting outliers using WMA-3σ algorithm, it is necessary to determine the parameter c, which indicates that a value in measured data is significantly correlated with the first c values, so that the weighted moving average method can be used to estimate the value based on the first c values. Usually, the c value can be determined by correlation analysis, expert experience analysis, or a combination of methods. After determining the parameter c, the outlier test can be performed using WMA-3σ method. The specific steps are as follows.

(1)Calculate the estimated value

Let a set of measured data series be expressed as X={x_1, x_2,..., x_n}, the estimated value based on weighted moving average method is calculated as follows.

$$x_i' = \sum_{j=1}^{c} \omega_j x_{i-j} \tag{5}$$

Where: x_i' is the estimated value of the measured value x_i at the i time; x_{j-k} is the measured data at time i-j; ω_j is the weight of x_{j-k}, $\sum_{j=1}^{c} \omega_j = 1$.

As can be seen from equation (5), the estimated value is the sum of c measured values multiplied by the corresponding weights.The c measured values are the first c values at time i. The weight calculation is the most critical, and its detailed steps are as follows.

1. Calculate differences $\Delta_j x_i$ between measured value x_i and the first c measured values, the absolute value are expressed as$|\Delta_j x_i|$.

$$\Delta_j x_i = x_i - x_{i-j} \quad (j=1, 2, \cdots, c) \tag{6}$$

2. Calculate the sum s of these absolute values $|\Delta_j x_i|$.

$$s = \sum_{j=1}^{c} |\Delta_j x_i| \tag{7}$$

3. The ratio of s to $\left|\Delta_j x_i\right|$ is calculated separately, and c variables w_j are obtained.

$$w_j = \frac{s}{|\Delta_j x_i|} \tag{8}$$

4. Calculate the sum S of these variables w_j.

$$S = \sum_{j=1}^{c} w_j \tag{9}$$

5. The ratio of w_j to S is calculated separately, and c weights ω_j are obtained.

$$\omega_j = \frac{w_j}{S} \tag{10}$$

As can be seen from the above formula, when calculating the estimated value at different times, the weight is adjusted adaptively according to the distance between the measured value at this time and the measured value at the previous time c. The smaller the distance, the larger the weight, the larger the distance, the smaller the weight. This is to reduce the weight of outliers on the estimate value, thereby improving the accuracy of the estimate value. When $x_i = x_{i-j}$, the formula (8) can not be calculated. According to the principle of the algorithm, the weight of x_{i-j} is the largest. At this point,let the estimate value $x_i' = x_i$.

In a set of measured values, the initial value of the estimated value corresponds to the measured value at the time c+1. In order to ensure that the sequence length of the estimated value and the measured value is the same.It also need to calculate the initial estimated value for the first c time. The calculation method is still according to formula (5) to formula (10). It is only necessary to calculate the measured value in reverse order, that is, to calculate the estimated value with the last c measured values of x_i and their corresponding weights.

(1)Calculate the difference value

Calculate the difference value between the estimated value and the measured value to form the difference sequence $\Delta X = \{\Delta x_1, \Delta x_2, \cdots, \Delta x_n\}$.

$$\Delta x_i = x_i - x_i' \tag{11}$$

(2)Mirror processing

Take the opposite of Δx_i and form the sequence $\Delta X' = \{-\Delta x_1, -\Delta x_2, \cdots, -\Delta x_n\}$, and then merge the two sequences ΔX and $\Delta X'$ into one sequence $\Delta X''$, which has length $2n$.

(3)Calculate the threshold

Calculate the standard deviation σ for the sample $\Delta X''$according to equations (3) and (4), and then the threshold T is obtained.

$$T = \pm k\sigma \tag{12}$$

Where: k is the standard deviation coefficient, usually is 3. If the outlier test results are not satisfactory, the k value can be adjusted by combining expert experience.

(4)Outlier detection

Judge the difference value between the estimated value and the measured value, when $|\Delta x_i|$>T, x_i is a outlier, otherwise, it is a normal value.

The WMA-3σ method is aimed at the outlier detection of the measured data. It focuses on the absolute value of the difference value between the data, so as to determine whether there is an outlier. Therefore,the above calculation formula does not consider the unit of data.

3.3 Other Notes

WMA-3σ method has the characteristics of high computational efficiency, real-time detection and strong robustness. The calculation process of WMA-3σ method calls less sample data, and the process of threshold calculation and outlier detection is simple, so its detection efficiency is high. When there is a large number of historical data, the threshold calculated by WMA-3σ is representative to a certain extent. It can not only detect the outlier of the historical data, but also monitor the outlier of the measured value in the future, which has a certain real-time performance. When there are outliers in measured value, the algorithm can greatly reduce the weight of the outlier's contribution to the estimated value, so as to reduce the influence of the outlier on the accuracy of the estimated value. Therefore, the method has certain robustness.

4 Example

A high slope was equipped with surface and deep deformation monitoring instruments to monitor the deformation process of the slope. In this paper, outliers were detected and analyzed by using the monitoring data of slope deep deformation. The data were obtained by the rod displacement meter arranged in the adit. The data range is from January 2020 to October 2022. The monitoring data showed a slow monotonically increasing change trend, and there are small fluctuations and step-like changes in local areas. This is due to the inevitable noise error in the measurement process of the instrument, as well as the reason of insufficient accuracy. In this paper, WMA-3σ method was used to detect the outliers of the data. Since the outliers of the measured data are not obvious, five outliers are added as the data to be detected in this paper, and the data change process is shown in Figure 1.

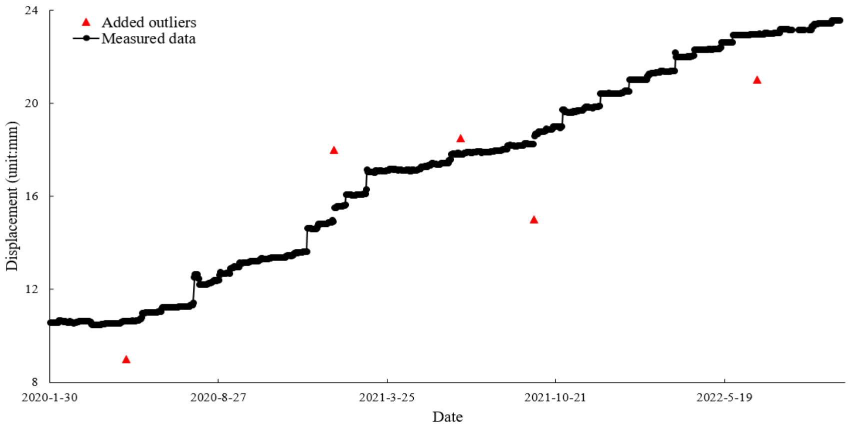

Fig. 1. Process of the measured data and added outliers

According to relevant analysis and expert experience analysis, parameter *c* and *k* in WMA-3σ method were selected as 3. The data sequences of the estimated value and the value to be detected in the calculation process is shown in Figure 2. The difference sequence between the data to be detected and the estimated value is shown in Figure 3. The difference sequence after mirror processing is shown in Figure 4, and the outlier test result of the data to be detected is shown in Figure 5.

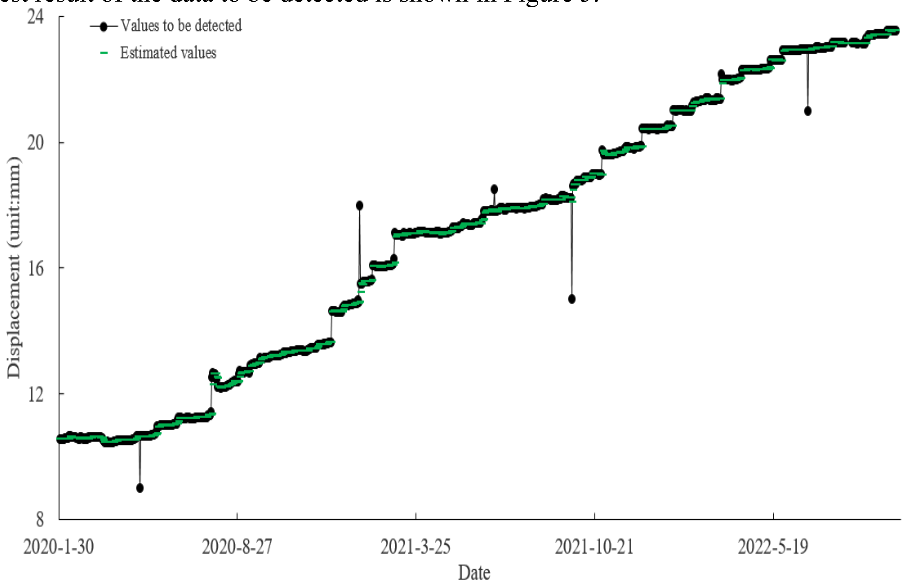

Fig. 2. The data sequences of the estimated value and the value to be detected

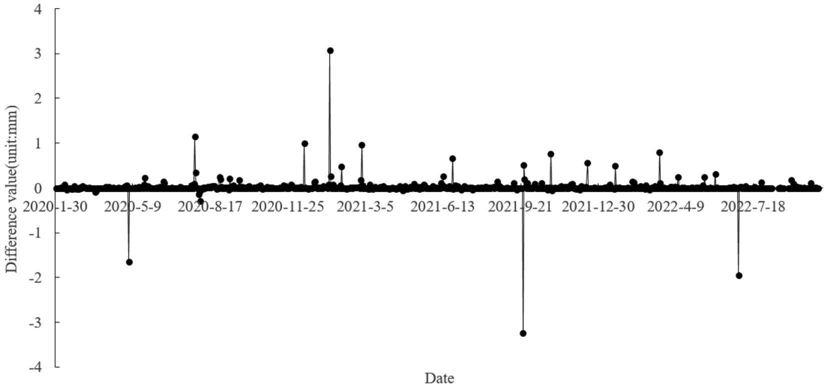

Fig. 3. The difference sequence between the data to be detected and the estimated value

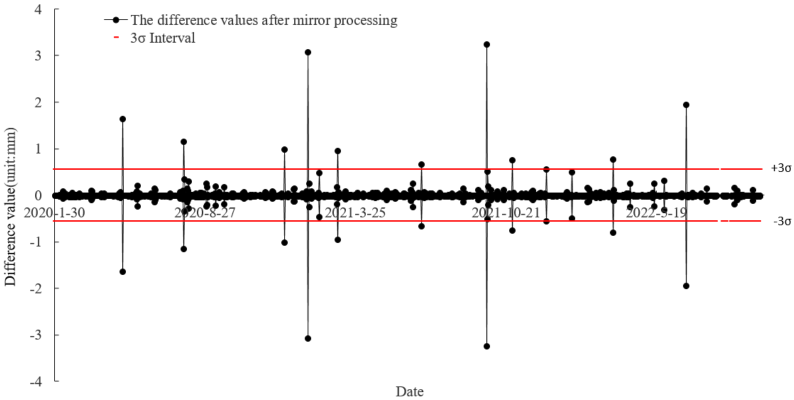

Fig. 4. The difference sequence after mirror processing

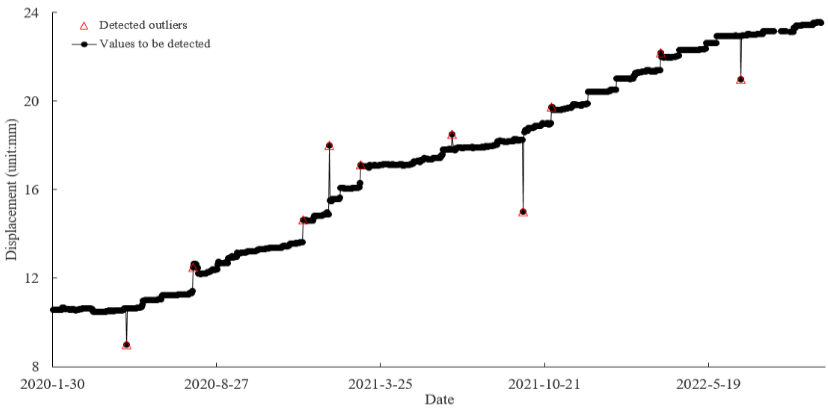

Fig. 5. Outlier detection result of the data to be detected

As can be seen from Figure 2, the change process and rule of the data to be detected and the estimated value are very close, and the two sets of data are almost overlap. However, the outliers in the data to be detected do not appear in the estimated value. This shows that WMA-3σ method is not affected by outliers in the data to be detected, and can estimate the data to be detected very accurately. As shown in Figure 3, in the difference sequence between the data to be detected and the estimated value, in data with large absolute values, the number of positive numbers is significantly greater than that of negative numbers. This is caused by the monotonically increasing trend in the data to be detected. This situation is not conducive to the calculation of 3σ criterion. This is because the 3σ criterion assumes that the sample data is normally distributed and should have a certain symmetry. This does not satisfy the conditions for the application of the 3σ criterion, that is, the data needs to be normally distributed and symmetrical. Therefore, mirror processing is applied to the difference sequences, and the difference sequence processed has symmetry, as shown in Figure 4. The 3σ threshold was calculated for the difference sequences after mirror processing, and 10 data were outside the threshold interval, that is, they are identified as outliers. As can be seen from Figure 5, the 10 outliers in the difference sequence correspond to the outliers in the data to be detected, including 5 added outliers and 5 step outliers. Compared with the previous data, the step outliers have a large variation and is also an outlier, so it can be detected by WMA-3σ method. It can be seen from the above test results that the method has a good test effect and can detect the outliers of slope deformation monitoring data.

5 Conclusions

According to the regularity and characteristics of the deformation monitoring data of high slope, this paper put forward the outlier detection method of WMA-3σ. Through the systematic research and example calculation of this method, the following conclusions are obtained.

(1) The slope deformation data changes gently and stably in the short term, with strong autocorrelation. This paper put forward an adaptive weight calculation method for weighted moving average algorithm. The method can well calculate the estimated value of the measured value. The variation rule, trend and local variation characteristics of the estimated value are highly similar to the measured data, and are not affected by the outliers in the measured data.

(2) The monitoring data of slope deformation has the tendency of monotonic change in the long run. This makes the total amount of positive change greater than the total amount of negative change, which leads to the asymmetry of the difference sequence between the estimated value and the measured value. In order to better meet the application conditions of 3σ criterion, this paper put forward the mirror processing for the difference sequence, which can solve the asymmetry problem well, so as to better apply the 3σ criterion to calculate the threshold.

(3) WMA-3σ method can accurately detect outliers in measured data, including single outliers and step outliers. In general, according to the monitoring data of different measuring points or different instruments, parameter k can be properly adjusted by trial

calculation or expert experience to obtain satisfactory outlier detection results. It has important reference value for outlier detection and intelligent data analysis of all kinds of engineering monitoring data.

(4) Outlier detection has always been the focus and difficulty in all kinds of data analysis, and the threshold selection is the key and core issue. The occurrence of outliers is also complex and changeable. So it is difficult to establish a single method for detection of all kinds of outliers. This paper combined WMA method and 3σ criterion, and improved them.The WMA-3σ method proposed can be used to detect outliers of slope deformation monitoring data in general conditions.

Acknowledgments

This work was financially supported by Natural Science Basic Research Program of Shaanxi (Program No. 2021JM-331); Shaanxi Provincial Education Commission(Program NO. 23JY058).

References

1. Qi Yanli,Tian Gang,Bai Mingzhou,et al. Research on construction deformation prediction and disaster warning of karst slope based on mutation theory[J].Scientific Reports, 2022,Volume 12, Issue 1:15182-15182.
2. Xu Ying,Huang Huibao,Li Yanling,et al. A three-stage online anomaly identification model for monitoring data in dam[J].Structural Health Monitoring, 2022, Volume 21, Issue 3:1183-1206.
3. Wu Shuangshaung. The Deformation Characteristics, the Methods for Prediction and Threshold Warning of the Gapa Landslide under Large Variations in Reservoir Level[D]. China University of Geosciences, Wuhan,2023.
4. Jiang Qijia, Jiang Zhongming, Tang Dong, et al.Gross Error Detection of Slope Safety Monitoring Data Based on SSA-DBSCAN[J].Journal of Yangtze River Scientific Research Institute,2022,39(04):85-90+98.
5. Li Yi, Zhao Erfeng, He jing. Abnormal data identification and reconstruction model of dam deformation monitoring[J]. Advances in Science and Technology of Water Resources,2023,43(02):109-114.
6. Demir Selçuk;Sahin Emrehan Kutlug.Application of state-of-the-art machine learning algorithms for slope stability prediction by handling outliers of the dataset[J] Earth Science Informatics, 2023,Volume 16,Issue 3:2497-2509.
7. Xiong Yan. Research on improvement of LSTM model and application in prediction of landslide vertical displacement[D]. Chongqing Jiaotong University,Chongqing,2023.
8. Liang Yongping, Li Sheng, Lai Guoquan. Application of EEMD-Wavelet in Extracting Deformation Information of High Slope[J/OL]. Journal of Safety and Environment: [2023-11-24]:1-9.
9. Rémi Domingues, Maurizio Filippone, Pietro Michiardi, et al. A comparative evaluation of outlier detection algorithms: Experiments and analyses[J]. Pattern Recognition,2018,74:406-421.
10. Xu BeiBei, Cui Chenfeng. Research of outlier detection on automatic monitoring data of dam[J]. Journal Geomatics,2015,40(2):59-61.

11. Chen Ron, Han Baoan, HanBaohua, et al. Prediction method of soil slope stability based on DE-MELM[J].Safety and Environmental Engineering,2020,27(04):87-93.
12. Wang Jianmin. Intelligent analysis and prediction of mine slope deformation monitoring data based on gaussian process[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(09):1206.
13. Lu Nan, Dong Yuanfeng, Guo Shitai, et al. Improvement IQR error detection method of slope deformation monitoring of open-pit mine[J]. METAL MINE,2018(08):114-117.
14. Karlo Martinović,Kenneth Gavin,Cormac Reale,et al. Rainfall thresholds as a landslide indicator for engineered slopes on the Irish Rail network[J]. Geomorphology, Volume 306, Issue 2018, 40-50.
15. Yang Zongji, Wang Liyong, Shi Lili, et al. Research of monitoring and early warning methods for rainfall-induced landslides based on multivariate thresholds [J]. Chinese Journal of Rock Mechanics and Engineering,2020,39(02):272-285.
16. Liu Gang, Ye Lixin, Chen Qiyu, et al. Abnormal event detection of city slope monitoring data based on multi-sensor information fusion [J]. Bulletin of Geological Science and Technology,2022,41(02):13-25.
17. Gu Yanchang, Wu Yunxing, Huang Haibing, et al. Prediction model of dam safety behavior based on genetic algorithm optimized support vector machine [J]. Journal of Hohai University(Natural Sciences),2020,48(05):419-425.
18. Li Haifeng, Sun Jietao. Study on local early warning monitoring model and monitoring index of slope operation safety in pivotal area [J]. Water Resources and Hydropower Engineering,2023,54(10):18-33.
19. Tang Xianqi, Yang Haiyun, Wu Fan, et al. Determination of monitoring index for concrete dam based on improved POT model[J]. HYDRO-SCIENCE AND ENGINEERING,2021(02):117-123.
20. Song Jintao;Chen Yongchao;Yang Jie. A Novel Outlier Detection Method of Long-Term Dam Monitoring Data Based on SSA-NAR[J].Wireless Communications and Mobile Computing,Volume 2022, Issue,2022.