


## Article

# A Novel Intelligent Learning Method for Identifying Gross Errors in Dam Deformation Monitoring Series

Chunhui Fang <sup>1</sup>, Xue Wang <sup>2</sup>, Jianchao Li <sup>3</sup>, Luobin Wu <sup>1</sup>, Jiayi Wang <sup>4,\*</sup>  and Hao Gu <sup>4,\*</sup>

<sup>1</sup> Research Institute of Water Sciences, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China; fangchh@zjweu.edu.cn (C.F.); 19511243587@163.com (L.W.)

<sup>2</sup> College of Water Conservancy Engineering, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311231, China; wxczsxzg@sina.com

<sup>3</sup> Agriculture and Rural Affairs Bureau (Water Conservancy Bureau) of Keqiao District, Shaoxing City, Shaoxing 312030, China; topview666@126.com

<sup>4</sup> College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China

\* Correspondence: wjy0423@hhu.edu.cn (J.W.); ghao@hhu.edu.cn (H.G.)

**Abstract:** In view of the problem that traditional dam outlier identification methods mostly rely on single-monitoring-point models and do not fully consider the spatio-temporal correlation characteristics of deformation between monitoring points, which can easily lead to the misdiagnosis of outliers, this paper proposes a novel Ward-VMD-BiLSTM-Iforest method for identifying gross errors in dam deformation monitoring. By integrating spatio-temporal clustering, variational mode decomposition (VMD), and BiLSTM neural networks, the method effectively identifies outliers while avoiding the misclassification of data mutations caused by environmental changes. Compared to traditional models (GRU, LSTM, and BiLSTM), the HHO-BiLSTM model demonstrates superior performance, achieving an  $R^2$  of 0.97775 at TCN08, with a reduced MAE and better accuracy. In comparison with the Raida and Romanovsky criteria, the proposed method achieves 100% precision and 100% recall, significantly improving detection accuracy and reducing misjudgment. This method provides an effective and reliable solution for dam deformation outlier detection.

**Keywords:** dam deformation; gross error; spatio-temporal clustering; deep learning; monitoring model; isolation forest algorithm



Academic Editor: Chin H Wu

Received: 17 December 2024

Revised: 3 January 2025

Accepted: 6 January 2025

Published: 8 January 2025

**Citation:** Fang, C.; Wang, X.; Li, J.; Wu, L.; Wang, J.; Gu, H. A Novel Intelligent Learning Method for Identifying Gross Errors in Dam Deformation Monitoring Series. *Water* **2025**, *17*, 148. <https://doi.org/10.3390/w17020148>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In order to understand the deformation state of a dam, a large number of deformation monitoring points are often set up on and around the dam body, and deformation data are regularly obtained to construct the deformation field of the dam. In fact, the deformation data of the dam comes from deformation monitoring points with different spatial coordinates. These deformation data are distributed in a certain spatial area, with both temporal and spatial properties, and are typical spatio-temporal data. In addition, due to the different degrees of integrity of the dam structure, the deformation of the dam monitoring points affects and correlates with each other. With the advancement of spatio-temporal data mining technology, the detection and discovery of implicit information in spatio-temporal data has been developed and has played a certain role in practical applications. At present, there are many analyses for the temporal characteristics of dam deformation data, but there are few studies specifically targeting its spatio-temporal characteristics.

Since traditional dam deformation analysis focuses on the change law of a single monitoring point, its data representation method is a one-dimensional time series; currently, dam deformation analysis considers multiple monitoring points, and emerging data

representation methods such as panel data show great potential. In addition, as typical spatio-temporal data, dam deformation data have multiple characteristics such as temporality, spatiality, complexity, massiveness, multidimensionality, and uncertainty, which brings many challenges to deformation spatio-temporal data mining. In order to avoid sudden damage to the dam during its service and cause serious engineering accidents, resulting in major loss of life and property, it is necessary to monitor the dam's operating status in real time and analyze the monitoring data regularly to ensure the long-term stable operation of the dam [1–4]. In actual engineering, the dam deformation data obtained by various means have data pollution problems such as incomplete measurement values, incorrect precision, repeated redundancy, and error anomalies, which greatly reduce the accuracy and efficiency of spatio-temporal data mining. This is due to the influence of uncertain factors such as instrument failure and human interference, which often leads to gross errors in safety monitoring data [5,6]. In order to improve the quality of the spatio-temporal data to be mined, it is necessary to perform corresponding preprocessing on the gross errors.

The research on data gross error processing has gone through several stages [7,8]. The initial gross error processing methods were mainly based on simple statistical tests. In this stage, for example, methods such as the Grubbs test [9], the Dixon test [10], and the Chauvenet [11] criterion were limited by the computing technology at the time. Such methods were usually inefficient when processing large-scale data and high-dimensional data such as dam monitoring. Traditional methods for identifying and processing gross errors in dam safety monitoring data mainly include three categories: the process line discrimination method, statistical test method, and regression model discrimination method [12,13]. The process line discrimination method relies on expert experience to determine whether the safety monitoring data deviate from the data process line for gross error discrimination. However, this method is time-consuming and labor-intensive and has low processing efficiency [14]. In recent years, many international scholars have conducted in-depth research on dam abnormal data processing methods (i.e., gross error processing methods). Machine learning and data mining technologies have begun to be applied to gross error processing [15,16], which has improved the quality of dam monitoring data gross error processing. For example, Rong et al. [17] developed a multi-point anomaly recognition model by combining an improved local outlier factor (LOF) and mutual verification considering spatio-temporal correlation. Wang et al. [18] developed a new framework based on self-supervised learning for the abnormal data detection and classification of rockfill dam deformation data, which includes an abnormal data detection method based on transformers and synthetic abnormal data. Song et al. [19] proposed a new dynamic detection method for dam abnormal data based on SSA-NAR for the outlier detection of dam monitoring data. This combined method does not rely on the relationship between the effect quantity and the environmental quantity in traditional dam safety theory, but only uses the time series of the effect quantity for change mining, which can avoid the impact of missing or abnormal influence quantities. Mao et al. [20] studied a monitoring data preprocessing and classification method for the gross error identification of hydraulic structures. This method uses linear regression and wavelet analysis techniques to effectively distinguish various types of data in the data set. Zhou et al. [21] proposed a new model based on a generative adversarial network and variational autoencoder for anomaly detection in dam monitoring data. In order to detect abnormal data in real time and quickly provide high-quality data for subsequent data analysis, Li et al. [22] proposed a multiple local anomaly coefficient method. Meng et al. [23] proposed a method for processing dam abnormal displacement monitoring data based on a matrix operation and the cuckoo search algorithm. Xiao et al. [24] established a structural safety monitoring model that can adaptively identify various types of anomalies in dam deformation monitoring data in view of the shortcomings of

conventional monitoring models such as difficulty in selecting influencing factors and poor ability to resist outlier interference. Hu et al. [25] proposed a method combining dynamic time warping and local anomaly factors to identify time series anomalies on different time scales. Chen et al. [26] proposed a gross error identification method based on the fuzzy C-means clustering algorithm (FCM) segmentation and density clustering algorithm (Ordering Points To Identity the Clustering Structure, OPTICS) combined with the local anomaly factor (LOF) algorithm. In addition, signal decomposition theory has also been applied to the processing of gross errors in dam safety monitoring information [27], mainly including a wavelet transform [28–30], empirical mode decomposition [31,32], variational mode decomposition [33,34], etc. However, the current gross error identification methods still have practical problems such as insufficient identification accuracy and excessive human judgment factors.

In response to the limitations of traditional methods for dam deformation outlier detection, this paper investigates a comprehensive characterization method for dam deformation data, focusing on techniques suitable for spatio-temporal data mining. Existing models often rely on single monitoring points and overlook the complex spatio-temporal correlations between different monitoring points, which can lead to the misdiagnosis of outliers. To address these challenges, this study proposes a novel framework that uses the Ward method for clustering dam deformation data, followed by variational mode decomposition (VMD) to decompose the time series of deformation data. In building on the decomposed components, an HHO-optimized BiLSTM model is applied for time series analysis, ensuring more accurate and robust predictions. The residual sequences extracted from the model are further analyzed using the isolation forest algorithm for effective abnormal data detection.

This research introduces several novel contributions: first, the combination of spatio-temporal clustering with the VMD-based decomposition and HHO-optimized BiLSTM model significantly enhances the ability to capture complex deformation patterns over time and between monitoring points. This integrated approach outperforms traditional models, which often fail to account for such correlations. Second, the use of isolation forest for anomaly detection in the residual sequences provides a more reliable and precise method for identifying gross errors, reducing both false positives and missed detections. Through engineering examples, the proposed method's effectiveness in both data characterization and preprocessing is validated, demonstrating a clear advantage over existing techniques.

## 2. Ward-VMD-BiLSTM-IForest Method

### 2.1. Ward Space Point Clustering Method

#### 2.1.1. Basic Indicators

According to the deformation data of each monitoring point in each period, the spatial monitoring points are partitioned to determine the deformation spatial distribution characteristics of the dam in each period. This section uses the absolute deformation  $x_{nt} = \delta_{nt}$ , deformation increase  $y_{nt} = \delta_{nt} - \delta_{n,t-1}$ , and relative deformation increase  $z_{nt} = (\delta_{nt} - \delta_{n,t-1}) / (\delta_{n,\max} - \delta_{n,\min})$  to construct the spatial monitoring point deformation similarity distance partition index. The implementation process is as follows:

For the information of all time sections in the same divided period, this paper uses two partitioning indicators, the absolute deformation similarity distance (ADD) and the deformation increase similarity distance (DGD) of the spatial monitoring points, to reflect the difference in absolute deformation and deformation increase between spatial monitoring points  $p$  and  $q$  ( $p$  or  $q = 1, 2, \dots, N$ ), where the expression of the full-time average absolute

deformation distance partition index  $d_{S,pq}(ADD)$  between spatial monitoring points  $p$  and  $q$  ( $p$  or  $q = 1, 2, \dots, N$ ) is

$$d_{S,pq}(ADD) = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{pt} - x_{qt})^2}, x_{pt} = \delta_{qt}, x_{qt} = \delta_{qt} \quad (1)$$

where  $T$  represents the number of time sections.

The expression of the full-time average deformation that increases distance partition index  $d_{S,pq}(DGD)$  between spatial monitoring points  $p$  and  $q$  ( $p$  or  $q = 1, 2, \dots, N$ ) is

$$d_{S,pq}(DGD) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_{pt} - y_{qt})^2}, y_{pt} = x_{pt} - x_{p,t-1}, y_{qt} = x_{qt} - x_{q,t-1} \quad (2)$$

Similarly, in considering the different total amplitudes of the full time series of measurements at different monitoring points, it is necessary to add the full-time average relative deformation increase distance partition index  $d_{S,pq}(RDGD)$  between the spatial monitoring points  $p$  and  $q$  ( $p$  or  $q = 1, 2, \dots, N$ ), which is expressed as

$$d_{S,pq}(RDGD) = \sqrt{\frac{1}{T} \sum_{t=1}^T (z_{pt} - z_{qt})^2}, z_{pt} = \frac{y_{pt}}{\delta_{p,\max} - \delta_{p,\min}}, z_{qt} = \frac{y_{qt}}{\delta_{q,\max} - \delta_{q,\min}} \quad (3)$$

where  $d_{S,pq}(ADD)$ ,  $d_{S,pq}(DGD)$ , and  $d_{S,pq}(RDGD)$ , respectively, represent the similarity distances of absolute deformation, deformation increase, and relative deformation increase between spatial monitoring points  $p$  and  $q$  in a certain division period. If the three deformation partitioning indicators of the  $p$ -th monitoring point and the  $q$ -th monitoring point in a certain division period are closer, the two monitoring points are classified into the same spatial deformation partition. Similarly, if the  $p$ -th monitoring point  $x_{pt}$  (or  $y_{pt}$  or  $z_{pt}$ ) is closer to the  $q$ -th monitoring point  $x_{qt}$  (or  $y_{qt}$  or  $z_{qt}$ ), the two monitoring points are classified into the same deformation partition.

In addition, in order to characterize the influence degree and comprehensive influence of the above three zoning indicators on the deformation characteristics of each spatial monitoring point in a certain period of time of the dam, this paper constructs a comprehensive distance zoning indicator of the deformation of the spatial monitoring point  $d_{S,pq}(CD)$ , and its expression is

$$d_{S,pq}(CD) = \sqrt{\beta_1 d_{S,pq}^2(ADD) + \beta_2 d_{S,pq}^2(DGD) + \beta_3 d_{S,pq}^2(RDGD)}, \beta_1 + \beta_2 + \beta_3 = 1 \quad (4)$$

where  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  represent the weights of the above three deformation similarity distance partition indicators.

Similarly, in the spatial monitoring point partition of the same division period, Equations (1)–(3) need to be standardized, thereby obtaining the standardized comprehensive distance partition index in Equation (4). The standardized calculation formula for the full-space monitoring point data of each time section is

$$Z_S(X_{nt}) = \frac{X_{nt} - \mu_t}{\sigma_t} (t = 1, 2, 3, \dots, T) \quad (5)$$

where  $X_{nt} = x_{nt}$  or  $y_{nt}$  or  $z_{nt}$ ;  $\mu_t = \frac{\sum_{n=1}^N X_{nt}}{N}$ ;  $\sigma_t = \sqrt{\frac{\sum_{n=1}^N (X_{nt} - \mu_t)^2}{N-1}}$ .

### 2.1.2. Clustering Method

Spatial point clustering is based on the Ward cluster linkage method, and its basic principles are as follows:

When assuming that the  $N$  deformation monitoring points of the dam body have been divided into  $k$  categories, denoted as  $G_1, G_2, \dots, G_k$ ,  $N_l$  represents the number of monitoring points in category  $G_l$ ,  $X_{il}$  represents the index value of monitoring point  $i$  ( $i = 1, 2, \dots, N_l$ ) in  $G_l$ ,  $\bar{X}_l$  represents the index center of  $G_l$ , and then the sum of squares of deviations  $W_l$  of monitoring points in  $G_l$  and the total sum of squares of deviations of category  $k$  are [35]

$$W_l = \sum_{i=1}^{N_l} (X_{il} - \bar{X}_l)^2 \quad (6)$$

$$W = \sum_{l=1}^k W_l \quad (7)$$

When  $k$  is determined, the classification that minimizes  $W$  should be selected.

Based on the classification idea of the Ward method, for the dam deformation panel data containing  $T$  periods and  $N$  monitoring points, the sum of squares of deviations of  $N_l$  monitoring points within the panel data classification  $G_l$  can be obtained as

$$W_{tl} = \sum_{i=1}^{N_l} [\beta_1 \cdot (x_{itl} - \bar{x}_{tl})^2 + \beta_2 \cdot (y_{itl} - \bar{y}_{tl})^2 + \beta_3 (z_{itl} - \bar{z}_{tl})^2] \quad (8)$$

where  $W_{tl}$  represents the sum of squares of deviations of the monitoring points at time  $t$  in  $G_l$ ;  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are the entropy weights;  $x_{itl}$ ,  $y_{itl}$ , and  $z_{itl}$  correspond to the three parameters in Formulas (1), (2) and (3), respectively; and  $\bar{x}_{tl}$ ,  $\bar{y}_{tl}$ , and  $\bar{z}_{tl}$  represent the average values of the parameters of the  $N_l$  monitoring points at time  $t$  in  $G_l$ , respectively.

Depending on the  $k$  value, different sums of squared deviations can be obtained, and the corresponding hierarchical clustering dendrogram can be drawn.

### 2.2. Processing of Measured Effect Sizes Based on Variational Mode Decomposition (VMD)

The VMD algorithm mainly decomposes the measured signal  $f(t)$  into  $K$  intrinsic mode components (IMFs) by solving the variational constraint equation that minimizes the sum of the modal bandwidths. Figure 1 shows the VMD algorithm processing process  $u_k(t)$ . In the VMD algorithm, IMF is the amplitude modulation frequency modulation function [36], i.e.,

$$u_k(t) = A_k(t) \cos[\varphi_k(t)] \quad (9)$$

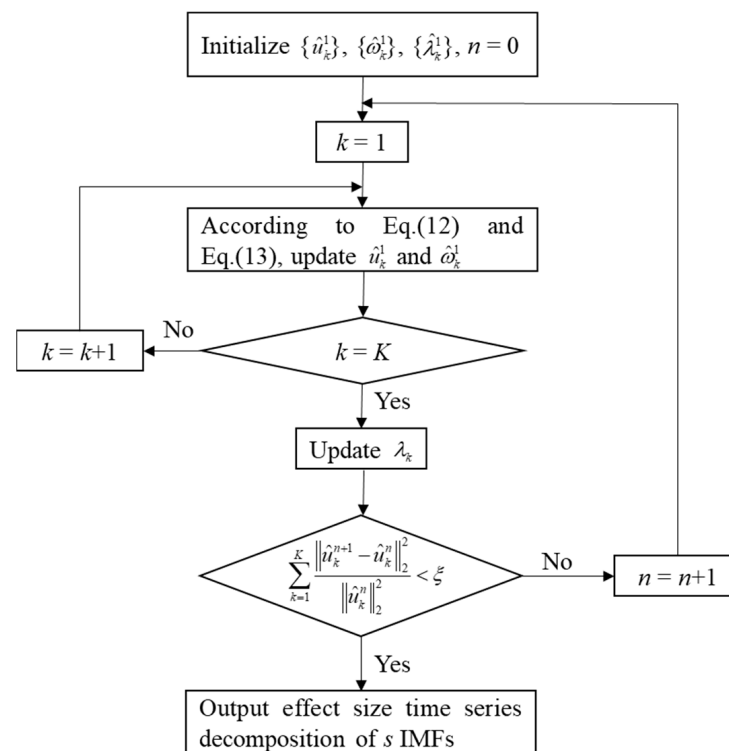
where  $A_k(t) \geq 0$  is the amplitude; and  $\varphi_k(t)$  is the phase.

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \partial_t \left\| \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^K u_k(t) = f(t) \end{cases} \quad (10)$$

To solve the constrained variational problem in Equation (10), the update iteration of the modal component can be equivalent to solving the following quadratic optimization problem:

$$u_k^{n+1} = \operatorname{argmin} \left\{ \alpha \left\| \partial_t \left[ \left( \sigma(t) + \frac{j}{\pi t} \right) \times u_k^n(t) \right] \exp(-j\omega_k t) \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k^n(t) + \frac{\lambda^n(t)}{2} \right\|_2^2 \right\} \quad (11)$$

where  $n$  is the current iteration number.



**Figure 1.** Processing flow of measured effect quantities based on variational mode decomposition.

By transforming Equation (11) into the frequency domain through a Parseval/Plancherel Fourier isometry transformation, we can obtain the frequency domain solutions of the modal function and the center frequency, which are

$$u_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (12)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^n(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^n(\omega)|^2 d\omega} \quad (13)$$

In the iterative update process, the update strategy of the multiplication operator is

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau[\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega)] \quad (14)$$

For a given discrimination accuracy  $\xi > 0$ , when Equation (15) is satisfied, the iterative update terminates. The frequency domain solution of each modal component obtained with Equation (12) is converted to the time domain using an inverse Fourier transform, and the time domain signals of  $K$  intrinsic mode functions can be obtained.

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \xi \quad (15)$$

### 2.3. HHO-BiLSTM Monitoring Model

#### 2.3.1. HHO Optimization Algorithm

The Harris Hawk optimization algorithm (HHO) simulates the three main stages of the Harris Hawk hunting process, i.e., the process of determining the most dangerous failure path of the cascade dam group, which mainly includes the search stage, the search and development stage, and the development stage. The specific implementation principle of this method is as follows:

In the search phase of the HHO, the position of the Harris Hawk is updated mainly using Equation (16) [37], i.e.,

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)|, & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(L_B + r_4(U_B - L_B)), & q < 0.5 \end{cases} \quad (16)$$

where  $X(t)$  is the position vector of the eagle at algorithm iteration number  $t$ ;  $X_{rand}(t)$  is the individual position vector of the randomly selected eagle;  $X_{rabbit}(t)$  is the prey position vector;  $r_1, r_2, r_3, r_4$ , and  $q$  are random variables in the range of  $(0,1)$ ;  $U_B$  and  $L_B$  are the upper and lower bounds of the search space, respectively; and  $X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t)$  is the average position vector of the eagle group.

During the search and exploitation phases, Harris's Hawks exhibit different behaviors depending on the potential for prey to escape, which is quantified by the so-called "escape energy", whose energy factor  $E$  is expressed as

$$E = 2E_0(1 - \frac{t}{T}) \quad (17)$$

where  $E_0$  is the initial value of the escape energy; and  $T$  is the maximum number of iterations of the algorithm.

Using Equation (17), Harris's Hawk searches for prey.

When the energy factor is  $|E| < 1$ , it indicates that the prey is not physically fit. At this time, the eagle flock switches to hunting mode, which can be divided into the following four situations:

When  $|E| \geq 0.5$  and  $r \geq 0.5$ , the Harris Hawk conducts soft encirclement and trapping, i.e.,

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)| \quad (18)$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \quad (19)$$

where  $J$  represents the jumping energy of the prey when it escapes; and  $\Delta X(t)$  represents the difference between the positions of the hawk and the prey at the  $t$ th iteration.

When  $|E| < 0.5$  and  $r \geq 0.5$ , the Harris Hawk conducts hard encirclement and capture, i.e.,

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \quad (20)$$

When  $|E| \geq 0.5$  and  $r < 0.5$ , the Harris Hawk performs a dive-style soft encirclement and capture, i.e.,

$$X(t+1) = \begin{cases} Y, & f(Y) < f(X(t)) \\ Z, & f(Z) < f(X(t)) \end{cases} \quad (21)$$

$$Y = X_{rabbit}(t) - E|JX_{rabbit} - X(t)| \quad (22)$$

$$Z = Y + SL_F(D) \quad (23)$$

where  $S$  represents a  $D$ -dimensional random vector;  $f(\cdot)$  represents the fitness function; and  $L_F(\cdot)$  represents the Harris Hawk flight function.

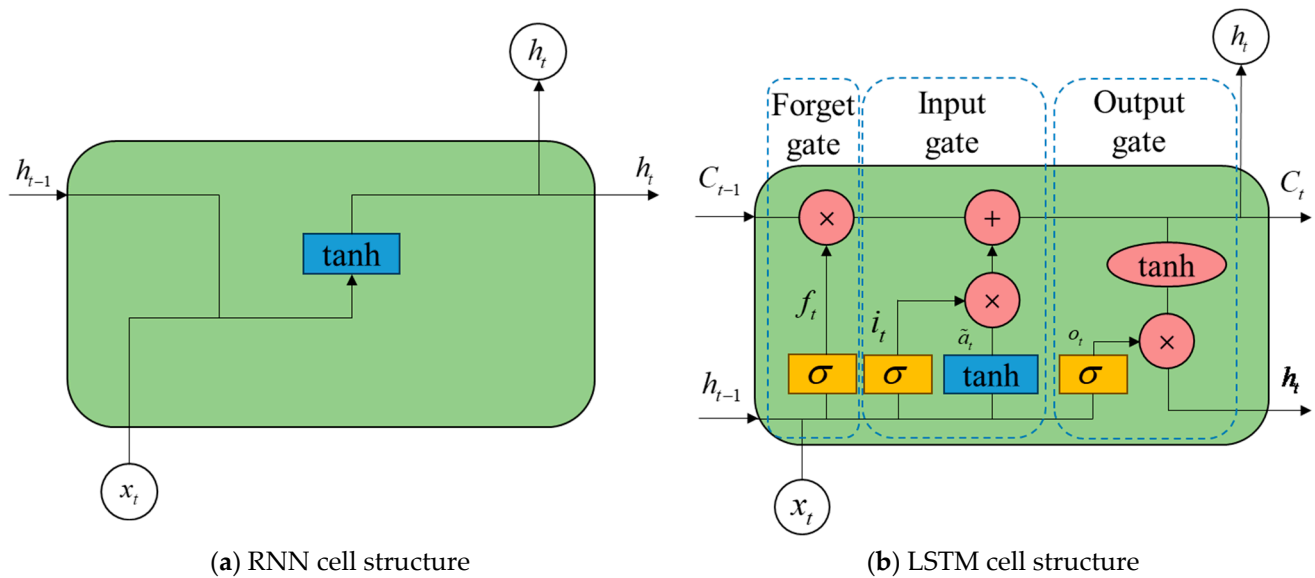
When  $|E| < 0.5$  and  $r < 0.5$ , the Harris Hawk conducts a dive-style hard encirclement and capture, i.e.,

$$Y = X_{rabbit}(t) - E|JX_{rabbit} - X_m(t)| \quad (24)$$

$$X(t+1) = \begin{cases} Y, & f(Y) < f(X(t)) \\ Z, & f(Z) < f(X(t)) \end{cases} \quad (25)$$

### 2.3.2. BiLSTM Monitoring Model

The bidirectional long short-term neural network is composed of a forward Long Short-Term Memory (LSTM) and a backward LSTM. The bidirectional LSTM networks are connected to each other, and the data are input, passed, and spliced from both the forward and reverse directions to obtain the final result. An LSTM is a special Recurrent Neural Network (RNN). Compared with in an RNN, the unit structure of an LSTM is more complex, as shown in Figure 2. Its unit structure is mainly composed of a cell state, forget gate, input gate, and output gate.



**Figure 2.** RNN and LSTM cell structures.

The forget gate determines the amount of information passing through the dam prototype monitoring data and retains the valid information [36].

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (26)$$

where  $W_f$  is the weight matrix of the forget gate;  $b_f$  is the bias term;  $\sigma$  is the sigmoid activation function,  $\sigma(x) = 1/(1 + e^{-x})$ ;  $t$  represents the current moment; and  $h_{t-1}$  represents the output state of the previous moment.

The function of the input gate is to input the new monitoring data information into the cell at the current moment. The sigmoid layer determines the proportion (degree) of the new information selected  $i_t$ , and the tanh layer is used to generate a new candidate vector  $\tilde{a}_t$ , which represents the storage of the new information learned.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (27)$$

$$\tilde{a}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (28)$$



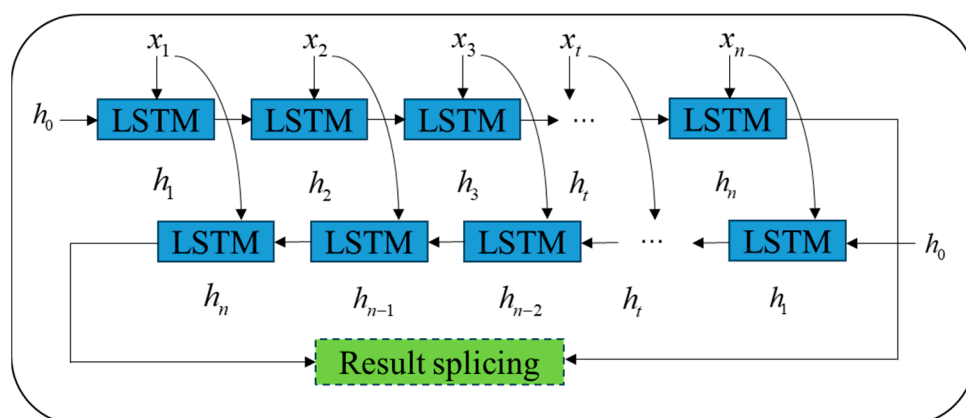
Based on the selection of the cell state at the previous moment and the input information at the current moment by the forget gate and the input gate, the cell state can be updated as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (29)$$

$$h_t = o_t \times \tanh(C_t) \quad (30)$$

In order to deeply explore the potential correlation and time-varying regularity of dam deformation, we use statistical analysis to extrapolate historical time series to predict its development trend. The training and verification data set of the BiLSTM model mainly depends on the order of time occurrence. The BiLSTM inputs  $x_t$  at each time step  $t$  and provides it to the forward and reverse LSTMs, as shown in Figure 3. It also outputs the hidden layers  $h_{tq}$  and  $h_{tf}$  through forward and reverse calculations based on two separate hidden layers, and finally obtains the final output  $h_t$  through output combination splicing.

$$h_t = f(h_{tq}, h_{tf}) \quad (31)$$



**Figure 3.** BiLSTM model structure.

## 2.4. Isolation Forest Detection of Abnormal Data

### 2.4.1. The Concept of Comprehensive Residuals

The clustered deformation's similar monitoring points are input into the HHO-BiLSTM model, the deformation prediction value of each monitoring point is output, the model prediction residual of each monitoring point is calculated, and the correlation of the model prediction residuals between the clustered similar monitoring points is used as the judgment basis to construct the concrete dam deformation abnormal value discrimination index based on the comprehensive residuals of similar monitoring points. For a monitoring point  $i$  in a partition, assuming that the other deformation sequence's similar monitoring points in the same partition are  $1, 2, \dots, n$ , by calculating the correlation coefficient  $r_i$  ( $i = 1, 2, \dots, n$ ) between the residuals of the other deformation monitoring points and the residual of the monitoring point, the comprehensive residual index of the studied monitoring point is constructed according to Equation (32).

$$s_{comprehensive} = w_1 s_1 + w_2 s_2 + \dots + w_n s_n \quad (32)$$

where

$$w_i = r_i / (r_1 + r_2 + \dots + r_n) \quad (33)$$

where  $s_{comprehensive}$  is the comprehensive residual;  $s_i$  is the residual of the  $i$ th monitoring point; and  $w_i$  is the proportion of the residual of the  $i$ th monitoring point in the comprehensive residual.

### 2.4.2. Isolation Forest Algorithm

After obtaining the high-frequency modal components, the isolation forest algorithm needs to be used to identify the outliers. Isolation forest (iForest) is an integrated parameter-free, unsupervised, fast anomaly detection method with the advantages of linear time complexity and high accuracy. It can identify outliers from high-frequency modal components. Its basic principle is to realize the anomaly detection of comprehensive residual components through a set of randomly generated isolation trees (iTrees). The main steps of isolation tree construction are as follows:

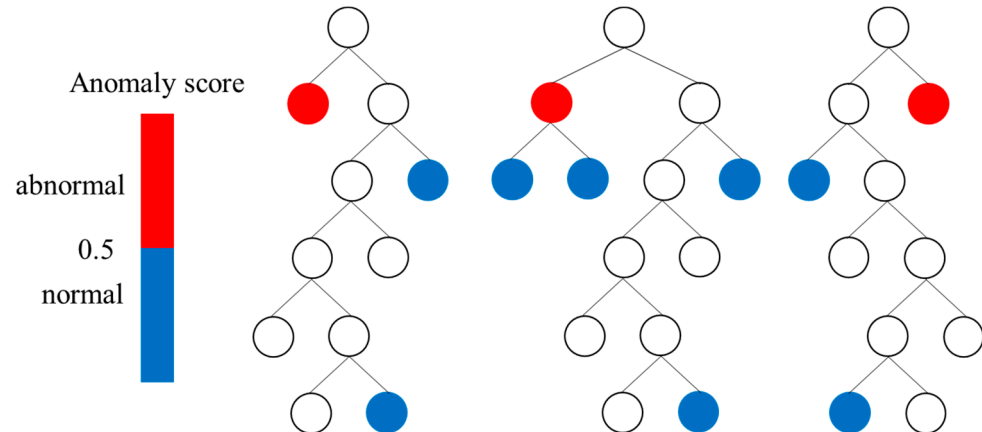
Step 1: Assume that in high-frequency component  $\mathbf{I}$ , the high-frequency modal component sequence of the safety monitoring information of monitoring point  $c$  is  $\mathbf{I}_c(t) = (y_1, y_2, \dots, y_n)$ ; randomly select  $\tau$  sample points from it to form a sample subset  $S = \{s_1, s_2, \dots, s_\tau\}$ ; and put it into the root node of the tree. The characteristic dimension of the sample subset is  $\beta$ .

Step 2: In the current sample subset  $S$ , randomly specify a feature  $\xi$  and a split value  $p$ , which is between the maximum and minimum values of the specified dimension.

Step 3: Generate a hyperplane from this cutting value, divide the sample subset into 2 subspaces, put the data of  $s(\xi) < p$  in the left child node of the current node, and put the data of  $s(\xi) \geq p$  in the right child node of the current node.

Step 4: Recursively repeat Step 1 to Step 3 in the child node and continuously construct new child nodes until one of the following conditions is met: ① there is only one data in the child node, which means that it cannot be cut any further; ② the tree has reached the set height. Then, stop node division and form an isolated tree.

Step 5: Loop Step 1 to Step 4 until the number of isolated trees reaches the specified number and then build an isolation forest from these isolated trees, as shown in Figure 4.



**Figure 4.** Isolation forest algorithm logic diagram.

In the process of tree node division, outliers in the high-frequency component will be isolated earlier, i.e., outliers are closer to the root node of the isolated tree, while normal values are farther away from the root node. Therefore, the outlier function  $s(y, n)$  is used to obtain the outlier score of high-frequency component sample  $y$  in isolated forest  $\mathcal{F}$  [38]:

$$s(y, n) = 2^{-\frac{E(h_f(y))}{c(n)}} \quad (34)$$

where

$$c(n) = 2H(n-1) - 2(n-1)/n \quad (35)$$

$$H(\tau-1) = \ln(\tau-1) + \lambda \quad (36)$$

$$E(h_t(y)) = \frac{\sum_{t \in \mathcal{F}} h_t(y) + \sum_{t \in \mathcal{F}} c(|l_t(y)|)}{|\mathcal{F}|} \quad (37)$$

where  $h_t(y)$  is the high-frequency component sample, with  $y$  as the path length from the external node to the root node of the isolated tree, i.e., the number of edges;  $E(h_t(y))$  is the average value of  $h_t(y)$  in the isolated tree set, indicating the average number of times sample  $y$  is divided;  $c(n)$  represents the average length of the search path;  $H(\cdot)$  is the harmonic number;  $\lambda$  is the Euler constant;  $|\mathcal{F}|$  is the total number of isolated trees; and  $|l_t(y)|$  represents the number of samples in the isolated tree  $t$  that are in the same leaf node  $l_t(y)$  as the sample  $y$ .

According to the abnormal score, the high-frequency component sample  $y$  is tested for abnormal values: if  $s(y, \psi) < 0.5$ , the sample is considered to be a normal value; if  $s(y, \psi) \approx 0.5$ , the safety monitoring information sample is considered to have no obvious abnormality; if  $s(y, \psi) \approx 1$ , the high-frequency component sample  $y$  is considered to be an abnormal value, and the corresponding original data point  $x$  is suspected to be a gross error. Since the isolation forest algorithm detects abnormal values from a mathematical perspective and cannot explain the physical causes of abnormal values, it may misjudge the sudden change caused by the sudden change in environmental quantities (water level or temperature) as a gross error, so it is necessary to further distinguish the suspected gross errors.

In view of the fact that gross errors are unrelated to the structural properties of the dam and environmental changes, the synchronization between the outliers and the sudden changes in environmental quantities can be used to determine whether the outliers are caused by sudden changes in environmental quantities. The core idea is to decompose the environmental quantity monitoring sequence using the variational mode decomposition (VMD) method, extract the high-frequency modal components, and use the isolated forest algorithm to detect anomalies. If both the environmental quantity and the original data are detected as outliers at a certain moment, it is considered that the abnormal change in the original data is caused by the sudden change in the environmental quantity and should be retained; conversely, if only the original data are detected as outliers, and the environmental variables do not show abnormalities, the original data are considered to be gross errors and should be eliminated.

### 2.5. Coupled Ward-VMD-BiLSTM-Iforest Algorithm

The coupled Ward-VMD-BiLSTM-Iforest algorithm mainly includes the following stages: the clustering and partitioning of monitoring points, extraction of residual components by variational mode decomposition, calculation of fitting values and predicted values by HHO-BiLSTM modeling, calculation of anomaly scores, and anomaly determination. The specific process is shown in Figure 5.

The basic steps of this algorithm are as follows:

Step 1: Monitoring point clustering and partitioning: Use the Ward clustering algorithm to cluster and partition the monitoring points, so that the monitoring points in each area have similar deformation process lines.

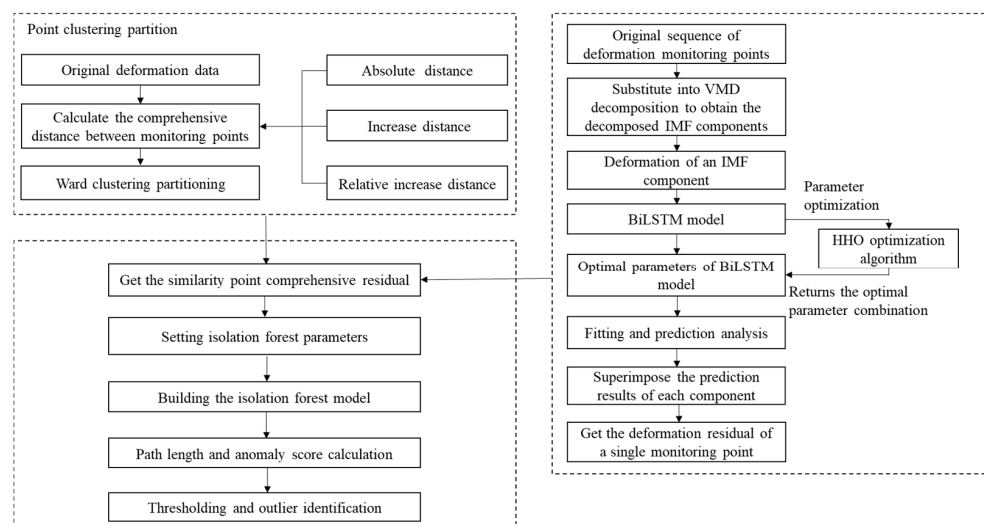
Step 2: Perform center frequency analysis on the original effect size sequence  $S(t)$  of the dam, obtain the optimal number  $s$  of IMF components through the center frequency method, and use  $s$  frequencies with significant energy as the initial value of the center frequency of VMD, i.e., determine the number of decomposed IMF components  $K = s$ .

Step 3: Input the original effect size time series of the dam into VMD and decompose it into  $s$  IMF components.

Step 4: Use the HHO-BiLSTM algorithm to fit and predict each IMF component and combine the measured values to extract the residual sequence.

Step 5: Anomaly score calculation: According to the isolation forest algorithm, multiple random numbers are constructed and their path lengths are calculated, and finally, the anomaly score is calculated.

Step 6: Abnormality determination: Determine a threshold for anomaly scores based on the contamination rate and identify points with anomaly scores exceeding the threshold as outliers.



**Figure 5.** Ward-VMD-BiLSTM-Iforest algorithm flow chart.

### 3. Case Study

#### 3.1. Project Overview

The Ertan Arch Dam Project (Figure 6) is located in the lower reaches of the Yalong River in the southwest of Sichuan Province, China, at the junction of Yanbian County and Miya County in Panzhihua City. The maximum dam height of a concrete hyperbolic arch dam in Sichuan Province is 240 m, the dam's top elevation is 1205 m, and the lowest elevation of the foundation surface is 965 m. The dam crest arc length is 774.69 m, the arch crown top thickness is 11 m, the arch crown bottom thickness is 55.7 m, the arch end maximum thickness is 58.51 m, the arch ring maximum center angle is  $91.49^\circ$ , and the upstream face maximum overhang is 0.18. The total storage capacity of the Ertan Reservoir is 5.8 billion  $\text{m}^3$ , the regulating storage capacity is 3.37 billion  $\text{m}^3$ , and the normal water storage level is 1200 m, while the catchment area is 116,400  $\text{km}^2$ .

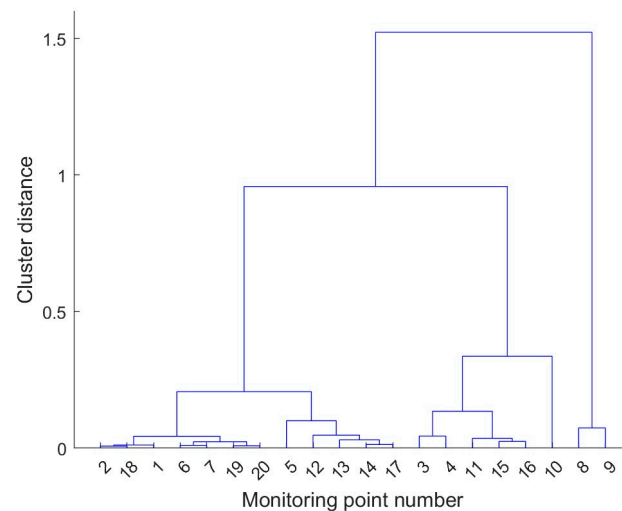


**Figure 6.** Airscape of the Ertan Dam Project.

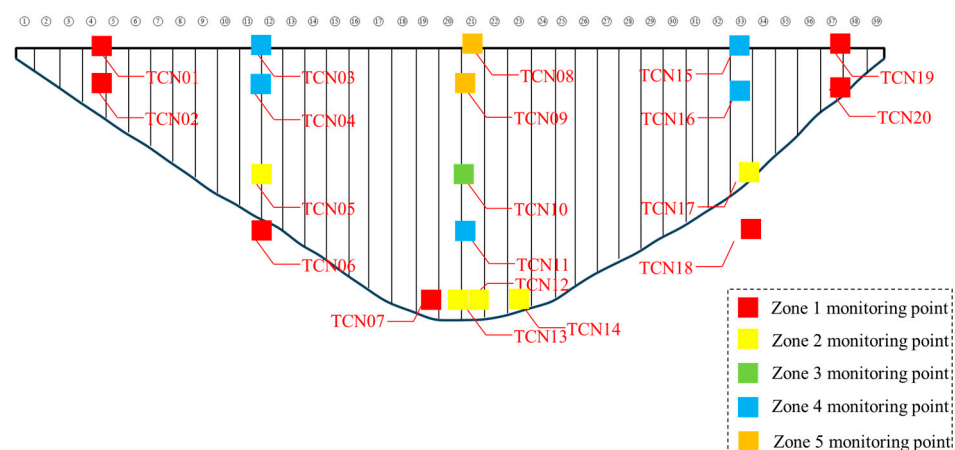
In order to monitor the horizontal displacement of the crown, spandrel and 1/4 arch ring of the C dam, 8 inverted vertical lines and 10 vertical lines were arranged, totaling 20 monitoring points. To verify the effectiveness of the proposed method, the displacement monitoring data of the vertical line monitoring points of the dam from January 2000 to December 2019 were selected for analysis.

### 3.2. Cluster Partitioning

The deformation of 29 monitoring points was divided into similarities using the Ward algorithm. The resulting tree diagram of the dam deformation monitoring points is shown in Figure 7, and the spatial measuring point clustering results are shown in Figure 8.



**Figure 7.** Clustering dendrogram of dam deformation monitoring points.



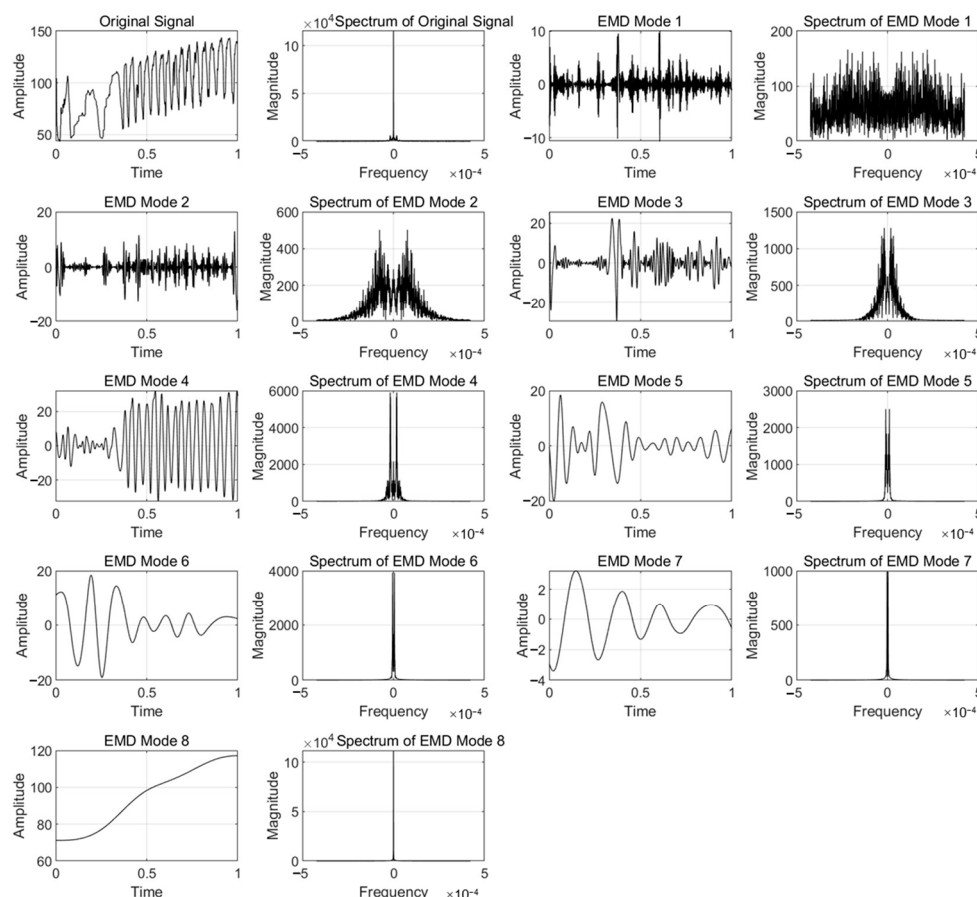
**Figure 8.** Arch dam deformation clustering results using Ward algorithm.

### 3.3. Variational Mode Decomposition

Take the monitoring point in zone 5 as an example; the values of  $\alpha$  and  $K$  are 2000 and 6, respectively. The decomposition result of the monitoring point TCN08 is shown in Figure 9.

The modal components of the monitoring points are sorted from low frequency to high frequency. The  $IMF_1$  component of the monitoring point is a trend item indicating that the overall deformation trend of the dam is in an upward state; the  $IMF_2 \sim IMF_5$  components of the monitoring point are periodic items with strong periodicity and high stability; the

IMF<sub>6</sub> component of the monitoring point is a random item with weak periodicity and low stability.



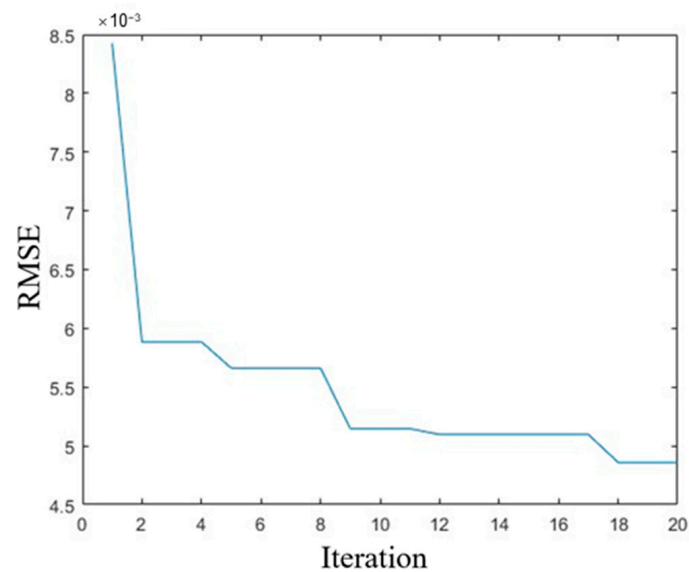
**Figure 9.** VMD algorithm decomposition results.

### 3.4. HHO-BiLSTM Modeling Results

The HHO is used to optimize the hyperparameters of the model, the population size is set to 5, and the maximum number of iterations is 20. Figure 10 shows the convergence curve of the optimization algorithm. Using the HHO to optimize the BiLSTM model can effectively improve the ability to obtain the optimal parameters of the model.

In order to compare the improvement of the prediction accuracy between the HHO-BiLSTM model and the traditional prediction model, the prediction results of the HHO-BiLSTM model are compared with the prediction results and evaluation indicators of the GRU, LSTM, BiLSTM models, and the monitoring data of multiple monitoring points in the partition are used to verify the applicability of the model. The number of hidden nodes, initial learning rate, and maximum number of iterations of the LSTM and BiLSTM were set to 100, 0.001, and 50, respectively; the number of hidden layer units, maximum number of training cycles, and initial learning rate of the GRU model were set to 100, 100, and 0.001, respectively. The evaluation indicators of the prediction results of different models are shown in Table 1. It can be seen from Table 2 that the MAE, MAPE, RMSE, and  $R^2$  of the HHO-BiLSTM model at each measurement point are better than those of the BiLSTM. In taking TCN08 as an example, the  $R^2$  of the prediction result of the HHO-BiLSTM model is 0.97775, which is 0.2882%, 0.827%, and 0.4438% higher than those of the GRU model, LSTM model, and BiLSTM model, respectively; and the MAE is reduced by 0.1633, 0.5153, and 0.2258 mm compared with the GRU, LSTM, and BiLSTM models.





**Figure 10.** Convergence curve of HHO.

**Table 1.** Comparison of performance rating indicators of prediction results of each model for two monitoring points.

Monitoring Point	Model	MAE/mm	MAPE/%	RMSE/mm	R <sup>2</sup>
TCN08	GRU	2.5499	2.7902	4.1076	0.97494
	LSTM	2.9019	3.1654	4.5149	0.96973
	BiLSTM	2.6124	2.7937	4.2296	0.97343
	HHO-BiLSTM	2.3866	2.5987	3.8705	0.97775
TCN09	GRU	2.3118	2.5411	3.6669	0.97257
	LSTM	2.6750	3.0831	4.1213	0.96535
	BiLSTM	2.5221	2.8816	3.9101	0.96881
	HHO-BiLSTM	2.0200	2.3189	3.2907	0.97791

**Table 2.** The residuals' statistical details for all monitoring points.

Monitoring Point Number	Max	Min	Average	Standard Deviation
TCN01	1.4789	−3.8927	0.0453	0.2918
TCN02	0.8203	−3.3909	−0.0886	0.2007
TCN03	19.3066	−11.1765	0.1164	2.1726
TCN04	15.9820	−8.5553	0.0434	1.6451
TCN05	4.2837	−1.7233	0.2766	0.4337
TCN06	0.6098	−0.3481	0.0018	0.0690
TCN07	2.8155	−2.0961	0.0462	0.2780
TCN08	36.5070	−14.0944	−0.1023	3.8266
TCN09	33.0171	−11.2725	1.4294	3.4833
TCN10	18.7243	−7.3364	0.2044	1.8325
TCN11	9.9552	−3.3071	0.5236	0.9141
TCN12	1.5967	−0.8582	−0.1068	0.1935
TCN13	2.9326	−1.0125	0.0106	0.2974
TCN14	1.1475	−1.9070	0.0250	0.2276
TCN15	11.6986	−5.2325	−0.4554	1.4173
TCN16	10.1748	−4.0106	−0.4286	1.1257
TCN17	2.2240	−0.8804	0.1093	0.2397

Table 2. Cont.

Monitoring Point Number	Max	Min	Average	Standard Deviation
TCN18	0.3007	−0.2903	−0.0205	0.0584
TCN19	1.6671	−0.8759	0.1108	0.2652
TCN20	2.2952	−3.0117	0.0025	0.1860

In order to build the HHO-BiLSTM model and test the prediction accuracy of the model, 70% of the monitoring data were used as the training set and 30% as the test set. The HHO was used to automatically optimize the number of nodes in the first and second hidden layers of the model, the number of training times, and the learning rate, and the optimization ranges were set to [1, 100], [1, 100], [1, 50], and [0.001, 0.01]. The residual series equals the true value series minus the fitted/predicted series. The residuals' statistical details for all monitoring points are shown in Table 2.

In taking the monitoring points in zone 5 as an example, six gross errors were artificially constructed in the test set. The modeling prediction results, original sequences, and residual sequences of the HHO-BiLSTM of the TCN08 and TCN09 monitoring points are shown in Figures 11 and 12.

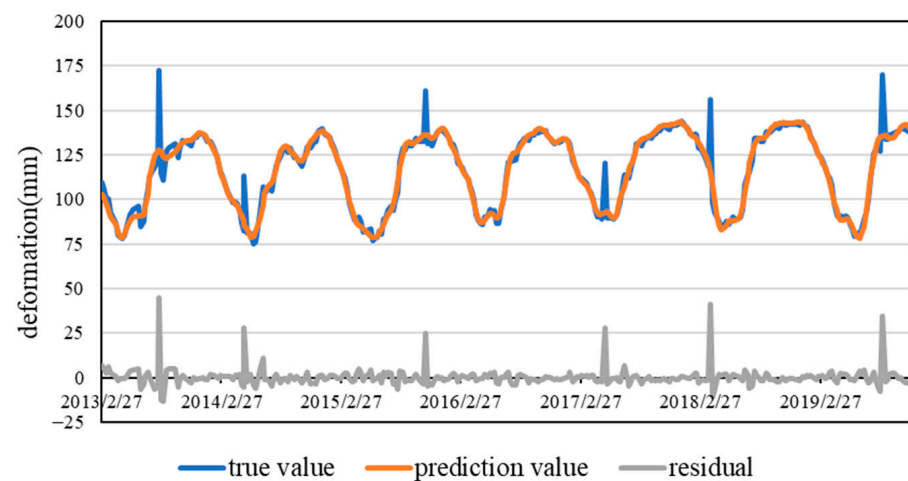


Figure 11. HHO-BiLSTM modeling results of TCN08 monitoring point.

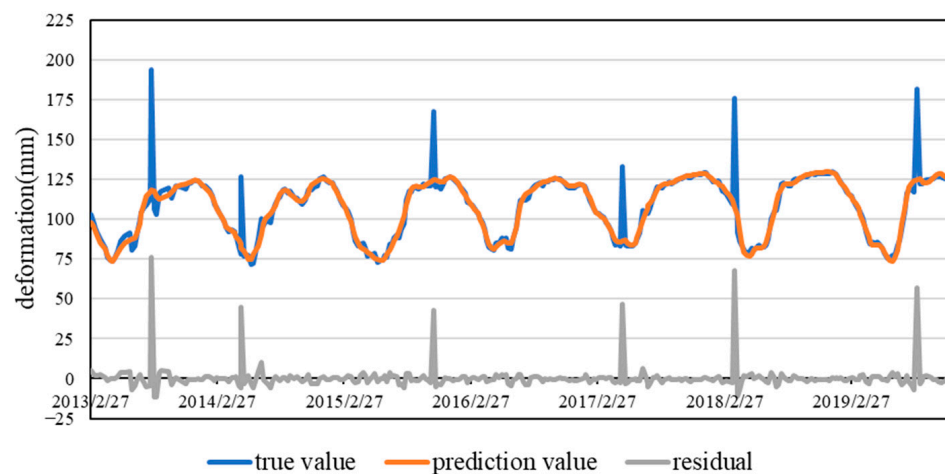
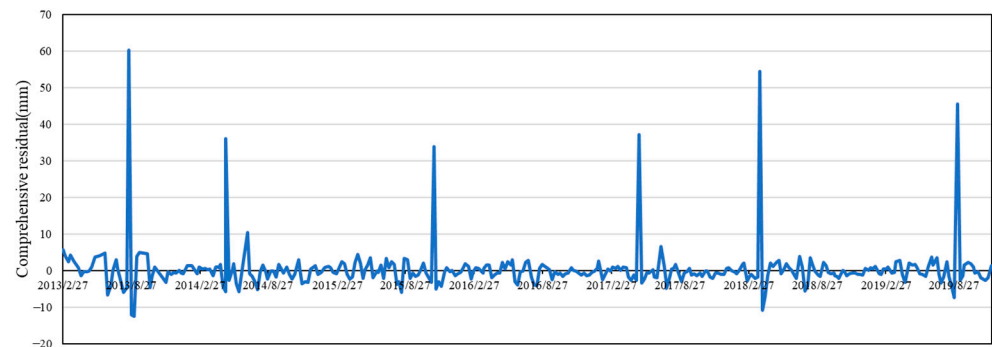


Figure 12. HHO-BiLSTM modeling results of TCN09 monitoring point.



### 3.5. Calculation of Comprehensive Residuals

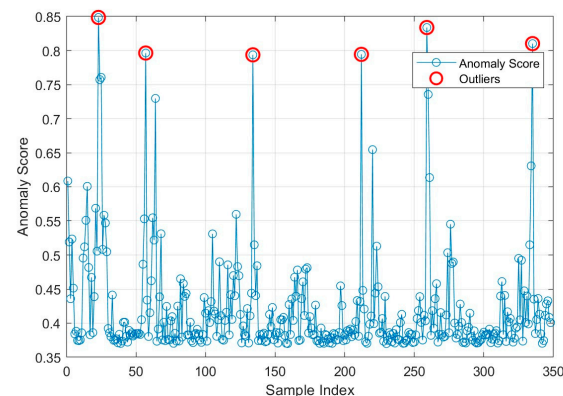
According to the comprehensive residual identification criteria, the residual correlation  $r$  between the TCN09 and TCN08 monitoring points was calculated to be 0.9822. The proportions of the TCN08 and TCN09 monitoring points in the comprehensive residuals of TCN08,  $w_1$  and  $w_2$ , were 0.5045 and 0.4955, respectively. The comprehensive residual sequence of TCN08 is shown in Figure 13.



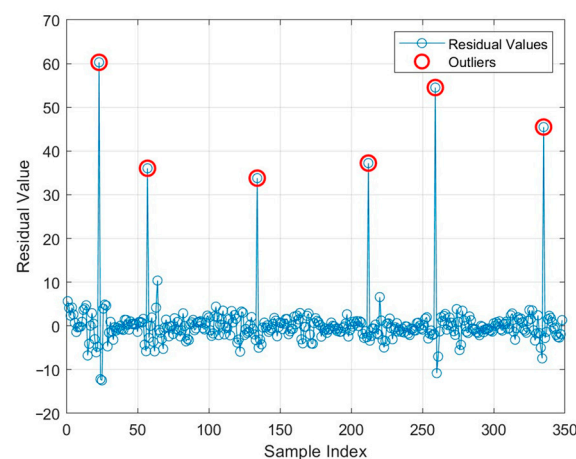
**Figure 13.** TCN08 comprehensive residual sequence.

### 3.6. Final Recognition Result of Isolation Forest Algorithm

With the isolation forest algorithm, a group of isolated trees of the residual components are randomly generated to form an isolation forest, and the anomaly score of the residual sequence is calculated, as shown in Figures 14 and 15.



**Figure 14.** Isolation forest anomaly scores for TCN08.



**Figure 15.** Isolation forest algorithm gross error identification results.

According to the judgment criteria of the isolation forest algorithm for anomaly scores, all measurements with anomaly scores greater than 0.77 are included in the suspected gross error set. At the same time, with the same steps as above, the environmental quantity (air temperature and reservoir water level) monitoring sequence is multi-scale decomposed using the variational mode decomposition method, and the reservoir water level and air temperature high-frequency modal component sequence is tested for anomalies using the isolation forest algorithm. It was found that there are no outliers in the temperature monitoring sequence. Therefore, it is only necessary to compare the dates of the outlier set in the reservoir's water level monitoring sequence with those of the suspected gross error set for synchronization, and then identify the gross errors in the original radial displacement monitoring sequence. It can be seen that the six gross error values added to the displacement monitoring sequence of each measuring point in this paper have been detected, and there is no missed detection or false detection, which verifies the effectiveness of the proposed method for identifying gross errors in safety monitoring information.

In order to verify the superiority of the proposed method, the detection results are compared and analyzed with those of the Raida criterion and the Romanovsky criterion. The precision rate *P*, recall rate *R*, weighted evaluation index *F*, and other evaluation indicators were calculated to compare and analyze the results. Table 3 shows the evaluation indicators of the outlier recognition results of different methods at the TCN08 monitoring point. As can be seen from Table 3, the traditional statistical test method of the Raida criterion has a recall rate of 100%, but its precision rate is only 50%. The Romanovsky criterion has a precision rate of 100%, but its recall rate is less than 50%. The weighted evaluation index is low, and the detection effect is poor. The proposed method comprehensively considers the spatio-temporal clustering characteristics of dam deformation and the high-precision prediction model. Its precision rate and recall rate are both 100%, and the recognition effect is better than that of the other comparison methods. In summary, the proposed method has a good performance improvement in the accuracy, missed judgment, and misjudgment of dam deformation outlier recognition.

**Table 3.** Evaluation indexes of outlier recognition results of different methods for TCN08 monitoring point.

Identification Method	P/%	R/%	F/%
The method proposed in this paper	100.00	100.00	100.00
The Raida criterion [3]	50.00	100.00	66.67
The Romanovsky criterion [3]	100.00	33.33	50.00

#### 4. Conclusions

This paper proposes a method for identifying outliers in concrete dam deformation data based on spatio-temporal clustering and deep learning, combining the Ward clustering method, VMD algorithm, HHO-BiLSTM algorithm, and isolation forest (Iforest) algorithm. A set of outlier identification criteria were constructed based on the comprehensive residuals of associated monitoring points. The prediction accuracy and outlier identification accuracy of this method were verified through case study. The deformation patterns of each monitoring point in the deformation partitions obtained using the defined comprehensive distance and the Ward clustering method exhibit higher similarity than those obtained through traditional clustering methods and are more consistent with the regional distribution characteristics of arch dam deformation. The main conclusions are as follows:

(1) The Ward-VMD-BiLSTM-Iforest outlier identification method proposed in this paper can effectively identify all outliers in the data set and will not misidentify data

mutations caused by environmental variable changes as outliers, providing a reliable and effective new approach for outlier identification in dam deformation monitoring data.

(2) The proposed Ward-VMD-BiLSTM-Iforest outlier identification method demonstrates high recognition accuracy, with experimental results showing that this method outperforms traditional methods, which achieves 100% precision and 100% recall. Moreover, by comparing data from different monitoring points, the efficiency of outlier detection is significantly enhanced.

(3) Although the proposed outlier identification method achieves accurate identification, the selection of the Iforest value threshold requires manual judgment and determination based on specific circumstances. How to adaptively select the appropriate threshold remains an important direction for future research.

**Author Contributions:** Conceptualization, J.W. and H.G.; methodology, C.F. and J.W.; software, X.W.; validation, J.L.; formal analysis, C.F.; investigation, L.W.; resources, X.W.; data curation, J.L.; writing—original draft preparation, C.F. and J.W.; writing—review and editing, C.F. and H.G.; visualization, L.W.; supervision, H.G.; project administration, H.G.; funding acquisition, H.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 52379122); National Natural Science Foundation of China (No. 52209159); Hohai University Central Universities Basic Research Operating Expenses Project (No. B230201011); and Plan Project of Water Conservancy Science and Technology of Zhejiang Province (Project nos. RC2183 and RC2457).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Milillo, P.; Perissin, D.; Salzer, J.T.; Lundgren, P.; Lacava, G.; Milillo, G.; Serio, C. Monitoring Dam Structural Health from Space: Insights from Novel InSAR Techniques and Multi-Parametric Modeling Applied to the Pertusillo Dam, Basilicata, Italy. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 221–229. [\[CrossRef\]](#)
2. Gu, C.; Su, H.; Wang, S. Advances in Calculation Models and Monitoring Methods for Long-Term Deformation Behavior of Concrete Dams. *J. Hydroelectr. Eng.* **2016**, *35*, 1–14. (In Chinese)
3. Wang, Y.; Gu, C.; Shi, L.; Gu, H.; Zhang, J.; Lu, X.; Wu, Y.; Zhu, M. Gross Error Identification Method for Deformation Monitoring Data of Concrete Dams Based on Improved IGG III-ELM Method. *Adv. Sci. Technol. Water Resour.* **2023**, *43*, 89–95. (In Chinese)
4. Zhao, E.; Li, Z.; Yuan, D. Deep Learning Model for Deformation Prediction of Dam Based on Dual-Stage Attention Mechanism. *J. Hohai Univ. Nat. Sci.* **2023**, *51*, 44–52. (In Chinese)
5. Lei, S.; Wei, L.; Tao, P.; Zhang, M. Application of Multi-Source Heterogeneous Data in Water Conservancy Monitoring Informatization. *Geospatial Inf.* **2023**, *21*, 34–38. (In Chinese)
6. Wang, L.; Wang, X.; Zhang, J.; Yu, J.; Wang, J. Comprehensive Evaluation Method of Dam Safety Considering Fusion Weight Optimization and Conflicting Information Sources. *J. Tsinghua Univ. Sci. Technol.* **2023**, *63*, 1566–1575. (In Chinese)
7. Smiti, A. A Critical Overview of Outlier Detection Methods. *Comput. Sci. Rev.* **2020**, *38*, 100306. [\[CrossRef\]](#)
8. Hawkins, D.M. *Identification of Outliers*; Chapman and Hall: London, UK, 1980.
9. Grubbs, F.E. *Sample Criteria for Testing Outlying Observations*; University of Michigan: Ann Arbor, MI, USA, 1949.
10. Dixon, W.J. Analysis of Extreme Values. *Ann. Math. Stat.* **1950**, *21*, 488–506. [\[CrossRef\]](#)
11. Chauvenet, W. *A Manual of Spherical and Practical Astronomy: 2: Theory and Use of Astronomical Instruments*; Lippincott: Philadelphia, PA, USA, 1863.
12. Fei, Y. *Error Theory and Data Processing*; Machinery Industry Press: Beijing, China, 2004. (In Chinese)
13. Zhou, Y.; Gan, X.; Li, D. Research on Gross Error Identification Technology of Dam Safety Monitoring Data. *J. Yangtze River Sci. Res. Inst.* **2011**, *28*, 16–20. (In Chinese)
14. Wang, J.; Wu, Y.; Zheng, D. Analysis of Dam Monitoring Data Based on Multi-Sensor Information Fusion. *J. Wuhan Univ. Eng. Ed.* **2004**, *1*, 32–35. (In Chinese)
15. Wang, H.; Bah, M.J.; Hammad, M. Progress in Outlier Detection Techniques: A Survey. *IEEE Access* **2019**, *7*, 107964–108000. [\[CrossRef\]](#)

16. Sikder, M.N.K.; Batareseh, F.A. Outlier Detection Using AI: A Survey. *AI Assur.* **2023**, *231*, 291.
17. Rong, Z.; Pang, R.; Xu, B.; Zhou, Y. Dam Safety Monitoring Data Anomaly Recognition Using Multiple-Point Model with Local Outlier Factor. *Autom. Constr.* **2024**, *159*, 105290. [\[CrossRef\]](#)
18. Wang, L.; Wang, X.; Zhang, J.; Wang, J.; Yu, H. A Self-Supervised Learning-Based Approach for Detection and Classification of Dam Deformation Monitoring Abnormal Data with Imaging Time Series. *Structures* **2024**, *68*, 107148. [\[CrossRef\]](#)
19. Song, J.; Chen, Y.; Yang, J. A Novel Outlier Detection Method of Long-Term Dam Monitoring Data Based on SSA-NAR. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 6569367. [\[CrossRef\]](#)
20. Mao, Y.; Li, J.; Qi, Z.; Yuan, J.; Xu, X.; Jin, X.; Du, X. Research on Outlier Detection Methods for Dam Monitoring Data Based on Post-Data Classification. *Buildings* **2024**, *14*, 2758. [\[CrossRef\]](#)
21. Zhou, Y.; Shu, X.; Bao, T.; Li, Y.; Zhang, K. Dam Safety Assessment through Data-Level Anomaly Detection and Information Fusion. *Struct. Health Monit.* **2023**, *22*, 2002–2021. [\[CrossRef\]](#)
22. Li, B.; Bai, X.; Li, J.; Wang, L. Outlier Detection of Gravity Dam Deformation Monitoring Data Based on the Multiple Local Outlier Coefficient Method. *Math. Probl. Eng.* **2022**, *2022*, 7157844. [\[CrossRef\]](#)
23. Meng, Z.; Wang, Y.; Zheng, S.; Wang, X.; Liu, D.; Zhang, J.; Shao, Y. Abnormal Monitoring Data Detection Based on Matrix Manipulation and the Cuckoo Search Algorithm. *Mathematics* **2024**, *12*, 1345. [\[CrossRef\]](#)
24. Xiao, S.; Cheng, L.; Ma, C.; Yang, J.; Xu, X.; Chen, J. An Adaptive Identification Method for Outliers in Dam Deformation Monitoring Data Based on Bayesian Model Selection and Least Trimmed Squares Estimation. *J. Civ. Struct. Health Monit.* **2024**, *14*, 763–779. [\[CrossRef\]](#)
25. Hu, J.; Ma, F.; Wu, S. Anomaly Identification of Foundation Uplift Pressures of Gravity Dams Based on DTW and LOF. *Struct. Control Health Monit.* **2018**, *25*, e2153. [\[CrossRef\]](#)
26. Chen, L.; Gu, C.; Zheng, S.; Wang, Y. A Method for Identifying Gross Errors in Dam Monitoring Data. *Water* **2024**, *16*, 978. [\[CrossRef\]](#)
27. Li, B.; Yang, J.; Hu, D. Dam Monitoring Data Analysis Methods: A Literature Review. *Struct. Control Health Monit.* **2020**, *27*, e2501. [\[CrossRef\]](#)
28. Chan, J.C.; Ma, H.; Saha, T.K.; Ekanayake, C. Self-Adaptive Partial Discharge Signal De-Noising Based on Ensemble Empirical Mode Decomposition and Automatic Morphological Thresholding. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 294–303. [\[CrossRef\]](#)
29. Lau, L. Wavelet Packets Based Denoising Method for Measurement Domain Repeat-Time Multipath Filtering in GPS Static High-Precision Positioning. *GPS Solut.* **2017**, *21*, 461–474. [\[CrossRef\]](#)
30. Wang, D.; Xiong, Y.; Xu, S. High-Precision Dynamic Single-Epoch Positioning Algorithm Based on Window Wavelet Denoising. *J. Wuhan Univ. Inf. Sci.* **2015**, *40*, 779–784. (In Chinese)
31. Bian, K.; Wu, Z. Data-Based Model with EMD and a New Model Selection Criterion for Dam Health Monitoring. *Eng. Struct.* **2022**, *260*, 114171. [\[CrossRef\]](#)
32. Zhang, C.; Fu, S.; Ou, B.; Liu, Z.; Hu, M. Prediction of Dam Deformation Using SSA-LSTM Model Based on Empirical Mode Decomposition Method and Wavelet Threshold Noise Reduction. *Water* **2022**, *14*, 3380. [\[CrossRef\]](#)
33. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [\[CrossRef\]](#)
34. Luo, Y.; Huang, C.; Zhang, J. Research on Engineering Structure State Characteristics Analysis Based on Improved IVMD. *J. Basic Sci. Eng.* **2021**, *29*, 873–888. (In Chinese)
35. Hu, T. Research on the Spatiotemporal Data Mining Model of Concrete Dam Deformation. Master's Thesis, Hohai University, Nanjing, China, 2017. (In Chinese)
36. Wei, B.; Xu, W.; Yuan, D.; Xu, F. A Displacement Time-Series Prediction Method for Concrete Dams Based on the Reliability Identification and Mining of Observation Data. *J. Appl. Basic Eng. Sci.* **2024**, *32*, 1049–1066. (In Chinese)
37. Huang, Z.; Gu, C.; Peng, J.; Wu, Y.; Gu, H.; Shao, C.; Zheng, S.; Zhu, M. A Statistical Prediction Model for Sluice Seepage Based on MHHO-BiLSTM. *Water* **2024**, *16*, 191. [\[CrossRef\]](#)
38. He, Q. Research on Risk Situation Analysis and Response Decision-Making Method for Watershed Dam Group Based on Measured Data. Ph.D. Thesis, Hohai University, Nanjing, China, 2024. (In Chinese)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.