A Coupled Model for Dam Foundation Seepage Behavior Monitoring and Forecasting Based on Variational Mode Decomposition and Improved Temporal Convolutional Network

Yantao Zhu,1,2 Zhiduan Zhang,3 Chongshi Gu,1,2 Yangtao Li,1,2 Kang Zhang,1,2 and Mingxia Xie4

1State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210024, China
2College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210024, China
3Center for Rivers & Lakes Protection and Construction Safety, Changjiang Water Resources Commission, Wuhan, China
4Changjiang Survey, Planning, Design and Research Co., Ltd., Wuhan 430010, China

Correspondence should be addressed to Yantao Zhu; zyt50@163.com

Received 1 October 2022; Revised 8 November 2022; Accepted 21 November 2022; Published 17 April 2023

1. Introduction

The concrete dam is one of the main types of high dam constructions in the world. Among high dams above 200 m, the concrete dam accounts for more than 60%. Therefore, it is of great significance to ensure the safety of concrete dams. The behavior of dam foundation seepage is one of the key factors affecting the safety and stability of concrete dams [1–4]. On the one hand, the uplift pressure generated by seepage directly acting on the dam foundation is one of the important unfavorable loads affecting the structural stability of the gravity dam; on the other hand, under the long-term effect of seepage, the joint fissure of the rock mass around the dam foundation may transform into a weak interlayer [4–6]. The seepage pressure is an important manifestation of dam foundation seepage, so it is of great significance to establish a high-performance concrete dam foundation seepage pressure prediction model [7, 8]. However, due to the complex environment of dam foundation seepage in reality, there exist many influencing factors that cause the dam foundation seepage pressure to have obvious nonstationary and nonlinear characteristics [9–12]. Furthermore, the seepage pressure data may be contaminated by noise [13]. All of the above bring challenges to an accurate prediction [14, 15].
Thus, it is necessary to develop new methods to get better performance of the dam foundation seepage pressure prediction model.

Under the action of environmental factors, external loads, and other factors, the original monitoring data inevitably suffer from certain noise interference [16–19]. For the dam monitoring data contaminated by noise, traditional denoising methods include wavelet analysis, empirical mode decomposition (EMD), seasonal-trend decomposition based on Loess (STL), and so on. For example, Li et al. [20] used the STL method to decompose the horizontal displacement monitoring sequence of the concrete dam and used the method of calculating the maximum signal-to-noise ratio to denoise the signal. The noise components of the measured signal are generally located in the high-frequency portion of the signal. However, the wavelet analysis only further decomposes the low-frequency part of the signal but does not continue to decompose the high-frequency part. For non-stationary signals, when the useful signal is drowned by the noise, the wavelet threshold denoising method is not ideal [21]. The EMD decomposition is prone to end effects and modal confusion problems. By applying these methods, the denoising of the signals will also lose the detailed characteristics of the high-frequency portion. Therefore, it is necessary to find a more refined denoising method.

The wavelet packet analysis is an extension of the wavelet analysis, which also decomposes the high-frequency part of the signal, in a more refined way. The extraction effect of high-frequency information is better in wavelet packet analysis than in wavelet analysis under the condition of low frequency [22]. Variational mode decomposition (VMD) is an adaptive and completely nonrecursive signal decomposition estimation method proposed by Dragomiretskiy and Zosso in 2014. This method has a solid foundation of mathematical theory, overcomes the end effect and mode confusing defects of the EMD method, and has better noise robustness.

Consequently, this paper combines the advantages of VMD decomposition and wavelet packet analysis. Firstly, the contaminated signals are decomposed into several components by VMD, and the components with more noise are denoised by the wavelet packet threshold. Finally, the signal pieces are reconstructed to obtain the denoised signal.

In terms of dam monitoring quantities’ prediction, statistical model methods such as stepwise regression were first used. However, their index selection is subjective, and the results are not good for high-dimensional nonlinear data [3, 23–25]. After that, the emergence and development of machine learning techniques provide a new way for dam monitoring quantities prediction. For example, Malekloo et al. [22] used the ELM algorithm to build an accurate and easy-to-train gravity dam displacement monitoring model. Su et al. [26] fitted the rough set theory and support vector machine theory to obtain the relationship between dam safety operation and influencing factors. The classical machine learning method has a good effect on small sample sets; however, for large sample sets, the classical methods frequently converge slowly and easily fall into local optimal values.

As a branch of machine learning, deep learning has a good capability of feature extraction and data fitting [17, 18, 27, 28]. It has been widely used in image, speech, and natural language processing [29, 30], and it mainly includes several major types of model structures such as convolutional neural network (CNN) and recurrent neural network (RNN). Dam monitoring quantities’ prediction is a typical time series prediction problem [31, 32]. In this respect, the currently commonly used deep learning algorithms are RNN and its variants are LSTM, etc. For example, Wei et al. [33] established RNN and trained into dynamic predictors of landslide displacement using a training algorithm named reservoir computing. The biggest weakness of RNN methods (RNN and its improved version) is that the limited number of network layers (usually 2–3 layers) can easily cause overfitting. Consequently, the extraction of information is not concise enough, which leads to long processing time for large-scale data. Therefore, it is necessary to find and study new prediction methods.

Time convolutional network (TCN) is a new deep learning algorithm to solve time series prediction problems [34]. It combines the best practices extracted from CNN, which merge the advantages of the traditional CNN model and RNN model: parallel data processing as the CNN model to extract key information, and a similar processing mechanism as the RNN model for time series data, which has certain memory. Its performance in a variety of tasks and data sets is comparable to or even exceeded RNN models. At present, TCN has been well applied in many fields, such as speech recognition and machine translation, but it has not been used in the dam engineering field.

This paper proposes an improved temporal convolutional network (ITCN) model suitable for dam foundation seepage pressure data prediction. A hysteresis experiment is carried out to obtain the optimal model by correlating the receptive field size of the ITCN model with the hysteresis of the dam foundation seepage pressure. Finally, the optimal ITCN dam foundation seepage pressure prediction model of each measurement point is obtained after the training process. To evaluate and compare the effectiveness of the proposed model, we adopted two deep learning models, which are the RNN model and the RNN model’s variant LSTM model, which is more common in the field of time series prediction, and the stepwise regression statistical model, which is more commonly used in the field of traditional dam engineering, as the benchmark methods. The MSE, RMSE, MAE, and MAPE were used as prediction accuracy evaluation indicators. The verification results confirm the flexibility and high prediction accuracy of the proposed prediction model.

The rest of the study is organized as follows. Section 2.1 introduces the basic principle and specific method of the VMD-wavelet packet denoising. Section 2.3 introduces the relevant influencing factors of concrete dam foundation seepage pressure and the selection of multidimensional input factors of each prediction model. Section 2.4 introduces the structure, function, and improvement process of each part of the proposed ITCN model, as well as the overall
2. Methodology

2.1. Principle of VMD

2.1.1. Construction of the Variational Problem. This method assumes that the signal can be decomposed into different intrinsic mode function (IMF) components, and each IMF component is regarded as an FM-AM signal. When the original signal is decomposed in K-level, the FM and AM signal \( u_k(t) \) can be expressed as

\[
    u_k(t) = A_k(t) \cos[\varphi_k(t)],
\]

where \( A_k(t) \) is the instantaneous amplitude of \( u_k(t) \), and \( \varphi_k(t) \) is the instantaneous phase of \( u_k(t) \), whose derivative \( \dot{\varphi}_k(t) \) is the instantaneous frequency \( \omega_k(t) \) of \( u_k(t) \), and \( \omega_k(t) \geq 0 \).

The bandwidth of each IMF component is estimated by performing Hilbert transform on all IMF components, then transforming the analytic signal to its corresponding baseband, and estimating it by the Gaussian smoothing method to calculate the L2 norm of its gradient. So far, we can construct a constrained variational problem, that is, under the condition that the sum of the modal components is equal to the input signal \( f \), and the sum of the bandwidth of each modal component is minimized.

2.1.2. Variational Problem Solving. By introducing both quadratic penalty terms and Lagrangian multipliers, the problem can be turned into unconstrained. Then the alternate direction method of multipliers (ADMM) is used to turn the problem into a series of suboptimization problems. The approximate solutions of IMF component \( u_k \), center frequency \( \omega_k \), and Lagrangian multipliers \( \lambda(t) \) are obtained as follows:

\[
    \tilde{u}_{k}^{n+1}(\omega) = \frac{\tilde{f}(\omega) - \sum_{k} \tilde{u}_{k}(\omega) + \tilde{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_{0})^2},
\]

\[
    \omega_{k}^{n+1} = \frac{\int_{0}^{\infty} \omega |\tilde{u}_{k}(\omega)|^2 d\omega}{\int_{0}^{\infty} |\tilde{u}_{k}(\omega)|^2 d\omega},
\]

\[
    \tilde{\lambda}^{n+1}(\omega) = \tilde{\lambda}^{n}(\omega) + \tau \left( \tilde{f}(\omega) - \sum_{k} \tilde{u}_{k}^{n+1}(\omega) \right),
\]

where \( \tilde{f}(\omega) \), \( \tilde{u}_{k}(\omega) \), \( \tilde{\lambda}^{n+1}(\omega) \), and \( \tilde{u}_{k}^{n+1}(\omega) \) are the Fourier transform of \( f(t) \), \( u_k(t) \), \( \tilde{\lambda}^{n+1}(\omega) \), and \( u_k^{n+1}(\omega) \), respectively, and \( n \) is the current iteration times.

Let IMF component \( u_k \), center frequency \( \omega_k \), and Lagrangian multipliers \( \lambda(t) \) alternate towards optimization iterations. When the accuracy convergence criterion \( \sum ||\tilde{u}_{k}^{n+1} - \tilde{u}_{k}^{n-1}||_2^2/||\tilde{u}_{k}^{n-1}||_2^2 < \varepsilon \) is satisfied, output the final set of \( \tilde{u}(\omega) \). Then \( k \) IMF components decomposed can be obtained, and the remaining undecomposed part of the signal is the residual component.

To sum up, the algorithm flowchart of VMD decomposition is shown in Figure 1.

2.2. Principle of Wavelet Packet Threshold Denoising

2.2.1. Wavelet Packet Analysis. In mathematics, a wavelet packet is composed of a set of linear combined wavelet functions; therefore, the choice of wavelet basis function directly affects the effect of wavelet packet denoising. Currently, there are hundreds of wavelet basis functions, and according to [35], the best basis function in wavelet denoising is the db4 wavelet. Therefore, this study chooses the db4 wavelet as the basic function of wavelet packet decomposition.

The wavelet packet decomposition tree can intuitively show the decomposition process of the wavelet packet. When the signal \( S \) is decomposed into three levels, its wavelet packet decomposition tree is shown in Figure 2.

In Figure 2, \( A \) represents the low-frequency component of the signal, \( D \) represents the high-frequency component of the signal, and their subscript represents the decomposition level (scale). The components of different decomposition scales can be combined with each other. Therefore, many decomposition structures can constitute the wavelet packet basis library. Each wavelet packet basis can completely save all the energy of the signal. However, the reflected signal characteristics are different; therefore, it is necessary to determine a set of optimal wavelet packet basis.

In the wavelet packet basis library, the wavelet packet basis which minimizes the cost function is the optimal wavelet packet basis. Commonly used cost functions include gate threshold coefficient, relative energy, and entropy criterion [20]. In this study, Shannon entropy is used as the cost function.

2.2.2. Denoising Threshold Selection

(1) Threshold Estimation Method. In the process of wavelet packet denoising, how to choose threshold \( T \) is a key problem. Common threshold estimation methods include Stein’s rigrsure threshold, Sqtwolog threshold, heursure threshold, and minimax threshold. Through experiments, for the data in this paper, the rigsure threshold can retain more signal characteristics; therefore, the rigsure threshold is selected as the threshold \( T \) in this paper, and its basic principle is as follows.

Let \( S \) be the signal to be denoised, \( Q = \{Q_1, Q_2, \cdots, Q_N\} \), and \( Q_1 \leq Q_2 \leq \cdots \leq Q_N \). Among them, \( N \) is the number of elements in signal \( S \), each element in \( Q \) is the square of each element in signal \( S \), and they are arranged in an ascending order. Here, define the risk vectors \( R = \{R_1, R_2, \cdots, R_N\} \), whose elements are as follows:
Ri = (N−2i + (N−i)QN + ∑_{k=1}^{i} Q_1) / N, i = 1, 2, ..., N. (2)

We take the minimum value R_a of each element in R as the risk value according to its subscript a, and the corresponding threshold is determined as follows:

\[ T = \sigma \sqrt{Q_a}, \]  

where \( \sigma \) is the noise signal deviation [22].

(2) Threshold Function. After the threshold, \( T \) is determined in the process of threshold denoising for each node coefficient, and two threshold functions proposed by Li et al. [23], hard threshold and soft threshold, are widely used. Compared with the hard threshold method, the soft threshold denoising method can achieve the optimal estimation to ensure that the denoised signal is as smooth as the original signal; therefore, the soft threshold function is used in this study.

2.3. Seepage Pressure Monitoring Model. For each prediction model, the multidimensional input factors are the same, and the differences are that the statistical model additionally includes previous items. Here, the first step is to introduce the statistical model factor selection.

According to the analysis of actual engineering data, the changes in upstream and downstream water levels, rainfall, and dam foundation temperature all have a certain impact on the concrete dam foundation seepage pressure; in addition, considering the change in the overall internal environment, the aging factor should also be selected. Therefore, the statistical model of dam foundation seepage pressure is as follows:

\[ Y = Y_{Hu} + Y_{Hd} + Y_{p} + Y_{T} + Y_{\theta}. \]  

The selection of each component factor in the statistical model is introduced as follows.

2.3.1. Upstream Water Level Component \( Y_{Hu} \). Since the dam foundation seepage pressure has a certain hysteresis relative to the change of upstream water level, the influence of upstream water level in the previous month needs to be considered, which is

\[ Y_{Hu} = \sum_{i=1}^{5} a_{ui} H_{ui}, \]  

where \( \bar{H}_{ui} \) is the average upstream water level on the monitoring day and \( 1 \sim 4 \) days before, \( 5 \sim 10 \) days before, \( 11 \sim 20 \) days before, and \( 21 \sim 30 \) days before the monitoring day (\( i = 1 \sim 5 \)) and \( a_{ui} \) is the regression coefficients.

2.3.2. Downstream Water Level Component \( Y_{Hd} \). The downstream water level component is similar to the upstream water level component, that is,
However, the deep learning methods of ITCN, LSTM, and RNN in this paper, all have a certain degree of memory. Therefore, there is no need to add previous term in their multidimensional input factors. In addition, they are consistent with the statistical model.

2.4. Improved Temporal Convolutional Network (ITCN). Dam foundation seepage pressure prediction is a typical time series prediction problem, and the TCN model is suitable for processing prediction tasks with time series structures [36]. However, in reality, the environment of dam foundation seepage is complex: there are many influencing factors, and there is a limited hysteresis relative to environmental changes. So, the original TCN model structure is not suitable for the prediction of dam foundation seepage pressure. Consequently, based on the TCN model, this paper proposed an improved temporal convolutional network (ITCN) model suitable for dam foundation seepage pressure data prediction. The model structure retains the basic structure of the TCN model as a whole, that is, dilated causal convolution structure, residual block, and fully convolutional network (FCN); in addition, according to the characteristics of dam foundation seepage pressure data, the following improvements have been made:

(1) To better extract the features of the input multidimensional factor data and to adapt the subsequent convolutional layer, a fully connected layer is added in front of the model to increase the dimension of the data.
(2) For the characteristics of dam foundation seepage pressure with limited hysteresis and environmental complexity, this study refers to the idea of bottleneck residual block to improve the residual block in the original TCN model.

The structure of each part of this improved temporal convolutional network (ITCN) is introduced in detail as follows.

2.4.1. Fully Connected Layer. Dam foundation seepage pressure prediction is a typical multidimensional input time series prediction problem. To extract advanced features of the input multidimensional factor data \([x_i]\) and to adapt the subsequent convolutional layer, a fully connected layer is added in front of the ITCN model. The number of neurons in the fully connected layer \(k\) is greater than the input data dimension \(n\). Then, use the \(k\)-dimensional data that increased by the fully connected layer as the input data of the subsequent convolutional layer, as shown in Figure 3:

\[
Y_{t|d} = \sum_{i=1}^{5} a_{di} P_{di},
\]

2.3.3. Rainfall Component \(Y_p\). The dam foundation seepage pressure also has a certain hysteresis relative to rainfall, that is,

\[
Y_p = \sum_{i=1}^{6} d_i P_i,
\]

where \(P_i\) is the average rainfall on the monitoring day and 1 day before, 2 day before, 3 ~ 4 days before, 5 ~ 15 days before, and 16 ~ 30 days before the monitoring day \((i = 1 ~ 6); d_i\) is the regression coefficient.

2.3.4. Temperature Component \(Y_T\). To fully reflect the influence of temperature change on dam foundation seepage pressure, we choose to use the measurement temperature at the measuring point and the sine wave periodic function as the temperature component:

\[
Y_T = \sum_{i=1}^{2} \left[ b_1 \left( \sin \frac{2 \pi t_i}{365} - \sin \frac{2 \pi t_{i+1}}{365} \right) + b_2 \left( \cos \frac{2 \pi t_i}{365} - \cos \frac{2 \pi t_{i+1}}{365} \right) \right],
\]

where \(t\) is the cumulative number of days from the initial measurement date to the monitoring date; \(t_i\) is the cumulative number of days from the initial measurement date to the first measurement date of the data series taken for modeling; \(i = 1, 2\) is the annual cycle and semiannual cycle; \(b_0, b_1, b_2\) are regression coefficients.

2.3.5. Aging Component \(Y_\theta\). Due to the deposition in front of the dam and the change of impervious body’s impervious effect, the composition of aging component is complex, and generally, the following forms are adopted:

\[
Y_\theta = d_1 (\theta - \theta_0) + d_2 (\ln \theta - \ln \theta_0) + d_3 (e^{-\theta} - e^{-\theta_0}),
\]

where \(\theta\) is the previous \(t\) divided by 100; \(\theta_0\) is the previous \(t_0\) divided by 100; and \(d_1, d_2, d_3\) are regression coefficients.

In summary, the statistical model for concrete dam foundation seepage pressure prediction is

\[
Y = \sum_{i=1}^{5} a_{wi} \bar{P}_{wi} + \sum_{i=1}^{5} a_{di} \bar{P}_{di} + \sum_{i=1}^{6} d_i P_i + b_0 Y_0
\]

\[
+ \sum_{i=1}^{2} \left[ b_1 \left( \sin \frac{2 \pi t_i}{365} - \sin \frac{2 \pi t_{i+1}}{365} \right) + b_2 \left( \cos \frac{2 \pi t_i}{365} - \cos \frac{2 \pi t_{i+1}}{365} \right) \right],
\]

\[
+ d_1 (\theta - \theta_0) + d_2 (\ln \theta - \ln \theta_0) + d_3 (e^{-\theta} - e^{-\theta_0}) + a_0,
\]

where \(a_0\) is a constant term.

It should be noted that, as can be seen from the formula, to reflect the hysteresis of the impact factor, the statistical model adopts the method of adding the previous term. However, the deep learning methods of ITCN, LSTM, and RNN in this paper, all have a certain degree of memory. Therefore, there is no need to add previous term in their multidimensional input factors. In addition, they are consistent with the statistical model.

2.4. Improved Temporal Convolutional Network (ITCN). Dam foundation seepage pressure prediction is a typical time series prediction problem, and the TCN model is suitable for processing prediction tasks with time series structures [36]. However, in reality, the environment of dam foundation seepage is complex: there are many influencing factors, and there is a limited hysteresis relative to environmental changes. So, the original TCN model structure is not suitable for the prediction of dam foundation seepage pressure. Consequently, based on the TCN model, this paper proposed an improved temporal convolutional network (ITCN) model suitable for dam foundation seepage pressure data prediction. The model structure retains the basic structure of the TCN model as a whole, that is, dilated causal convolution structure, residual block, and fully convolutional network (FCN); in addition, according to the characteristics of dam foundation seepage pressure data, the following improvements have been made:

(1) To better extract the features of the input multidimensional factor data and to adapt the subsequent convolutional layer, a fully connected layer is added in front of the model to increase the dimension of the data.
(2) For the characteristics of dam foundation seepage pressure with limited hysteresis and environmental complexity, this study refers to the idea of bottleneck residual block to improve the residual block in the original TCN model.

The structure of each part of this improved temporal convolutional network (ITCN) is introduced in detail as follows.

2.4.1. Fully Connected Layer. Dam foundation seepage pressure prediction is a typical multidimensional input time series prediction problem. To extract advanced features of the input multidimensional factor data \([x_i]\) and to adapt the subsequent convolutional layer, a fully connected layer is added in front of the ITCN model. The number of neurons in the fully connected layer \(k\) is greater than the input data dimension \(n\). Then, use the \(k\)-dimensional data that increased by the fully connected layer as the input data of the subsequent convolutional layer, as shown in Figure 3:

2.4.2. Improved Residual Block. As a typical time series prediction problem, the dam foundation seepage pressure prediction can be expressed as follows: we use the input multidimensional factor data sequence \(X = \{ [x_1], [x_2], \cdots, [x_i], \cdots, [x_T] \}\) to predict the seepage pressure sequence \(Y = \{ y_1, y_2, \cdots, y_i, \cdots, y_T \}\). Obviously, the output prediction sequence of seepage pressure values needs to meet the limitation of causality condition, which is
the predicted value of seepage pressure at time $t$ that is only related to the multidimensional factor data sequence $\{[x_1], [x_2], \ldots, [x_k]\}$ at time $t$ and before and has nothing to do with the “future” data sequence $\{[x_{t+1}], [x_{t+2}], \ldots, [x_n]\}$. For this reason, here, the concept of fully convolutional network and causal convolution is first introduced.

2.4.3. Fully Convolutional Network. Compared with the traditional convolutional network, the fully convolutional network (FCN) [26] uses convolution layers instead of fully connected layers in the last few layers, building a complete fully convolutional network to achieve intensive prediction. In other words, the element-level prediction of the sequence can be achieved under one-dimensional convolution, which is the significance of introducing the fully convolutional network into time series prediction problems.

In addition, compared with the low-level convolution network, the convolution network at the high level has a larger receptive field, which can sense the historical information in a longer time frame, with good sensitivity to the changes in characteristics, and this is very helpful in building long-term memories.

The limitation of causal conditions is uniquely specific to the problem of causal time series prediction. Based on the fully convolutional network, the causal convolution structure [27, 28] is developed, as shown in Figure 4. Causal convolution can be regarded as cutting the fully convolutional network in half, only performing convolution operations on the input at the current time $t$ and the previous time.

To some extent, time convolutional network can be simply expressed as $\text{TCN} = \text{1DFCN} + \text{causal convolutions}$; thus, the CNN model is transformed into a model suitable for processing causal time series data.

The receptive field of ordinary causal convolutional networks is linearly dependent on the network depth. To extract information from historical data over a long period, one requires a fairly deep network structure or a large convolution kernel, which will greatly increase the computational burden of the model. Therefore, TCN introduces the dilated causal convolution structure [28].

The dilated causal convolution increases the receptive field of the model by adding holes to the standard causal convolution. The size of the hole is the number of intervals between the points of the convolution kernel. The larger the hole, the larger the receptive field of the convolution kernel, and the longer historical information the convolution output is related to.

Generally, to avoid the gridding effect, the hole size $D$ of each layer is set as an exponential form of the hyperparameter dilation rate $d (1, d^1, d^2, \ldots, d^k)$, and the dilation rate $d$ should not be smaller than the size of the convolution kernel. A three-layer dilated causal convolutional network with a convolution kernel size of 2 and $d = 2$ is shown in Figure 4. As can be seen, by adding holes, the receptive field of the model expands exponentially with the increase of the network depth.

For a single convolution kernel, the size of the receptive field of dilated causal convolution is

$$ R_i = (C - 1)D + 1, $$

where $i$ is the number of layers where the convolution kernel is located; $C$ is the size of the convolution kernel; and $D$ is the size of the hole.

In this paper, we deduce that for the dilated causal convolution network with depth $n$, the final receptive field size is

$$ R_n = (C - 1) \sum_{i=1}^{n} d^{n-i} + 1. $$

In addition, it was found during the experiment that in order to quickly calculate the approximate range of the receptive field, the following approximate formula could be used:

$$ R_n = C d^{n-1}. $$

The gradual increasing and deepening of the network may cause network degradation. The residual block [29], by adding a shortcut connection to the redundant layer of the network, can realize identity mapping and make the deep network equivalent to the shallower optimal network structure.

Several layers of networks containing a shortcut connection are called residual blocks. In a convolutional network, the residual block can be expressed as

$$ O(x) = \text{Activation}(x + F(x, W_H)), $$

where $\text{Activation()}$ represents the activation function and $W_H$ represents the convolution kernel weight to be learned.
This paper combined with the characteristics of the dam foundation seepage pressure prediction improves the residual block in the original TCN model. The practice has proved that the residual block needs at least two layers of networks to achieve a good improvement. In the residual block of the original TCN, the two layers of networks are both dilated causal convolutional layers. Although this kind of residual block can greatly increase the receptive field of the model, it also limits the development of network depth. In practical engineering projects, dam foundation seepage pressure has hysteresis relative to environmental changes, but such hysteresis has a certain limit. Meanwhile, because the environment of the dam foundation seepage is complex, there are many influencing factors, and the dam foundation seepage pressure problem is very complicated. In other words, the model does not need to have a huge receptive field; instead, the depth of the model is required.

Therefore, this paper refers to the idea of the bottleneck residual block [29]. After many trials, we designed an improved residual block, whose structure is shown in Figure 5:

As shown in Figure 5, for the residual mapping part \( F(x) \) on the left, the improvement is that we replace the first dilated causal convolutional layer in the original residual block with a standard convolutional layer that has no causal relationship; instead, the second layer maintains the dilated causal convolution unchanged. In this way, the depth of the network can be doubled while ensuring a certain model receptive field to improve the feature extraction ability of the model. At the same time, the number of output channels in the standard convolutional layer is set to be consistent with that in the dilated causal convolutional layer, which means the first layer of convolution performs the dimension increase or decrease processing on the input data to ensure the number of input channels is equal to the number of output channels for the second layer.

In addition, the first standard convolution layer and the second dilated causal convolution layer both use the rectified linear unit (ReLU) [30] as the activation function. The dropout regularization layer [31] is retained after each layer, which is the same process as in the original TCN. In addition, the second layer is processed by weight normalization.

For the identity mapping part \( x \) on the right, that is, the shortcut connection, if the input data and output data have the same dimensions, add input \( x \) and output \( F(x) \) directly; if the input data and output data have different dimensions, add \( 1 \times 1 \) convolution, adjust the number of filters, and make sure that the tensors are added in the same scale.

2.4.4. The Whole Architecture of the ITCN Model. In the design of the ITCN model, several residual blocks are stacked to form a deep residual network. The whole architecture of the ITCN model can be expressed as input of the multidimensional factor sequence \( [x_i] \) into the fully connected layer for dimensional increase processing, and then input the result into the deep residual network; after the dimensionality decreases the processing of several residual blocks, finally, output the one-dimensional dam foundation seepage pressure prediction sequence \( Y \). The whole architecture of the ITCN model is shown in Figure 6:

The whole architecture of the ITCN model proposed in this paper for dam foundation seepage pressure data prediction has the following characteristics:

1. This model uses a flexible convolution architecture: According to the number of output interfaces, the input sequence of any dimension can be mapped to the output sequence of fixed dimensions freely, with the ability of multidimensional input and multidimensional output. This proposed model can therefore solve the problem of the diversity of influencing factors on dam foundation seepage pressure.

2. By using the dilated causal convolution structure, this model has the memory ability and is capable of solving the hysteretic problem in the prediction of dam foundation seepage pressure.

3. Through the stack dilated convolution layer and parameter sharing mechanism, the proposed model greatly reduces the computational burden of the model.

4. By improving the residual blocks, while ensuring a certain model receptive field, this model can extract higher dimensional features, thus solving the limited hysteresis and large complexity due to environmental changes of the dam foundation seepage pressure.

2.5. The Basic Workflow of the Proposed Model. For the dam foundation seepage pressure prediction model based on VMD-wavelet packet denoising and ITCN proposed in this study, the basic workflow is shown below:

Step 1: For the measured data of dam foundation seepage pressure at each measuring point, judge whether it has been contaminated by noise.
Step 2: we perform the VMD decomposition experiment on the output data of Step 1. According to the decomposition situation of different decomposition levels, determine the final decomposition level and the components that need to be denoised.

Step 3: for the components selected by Step 2, according to the percentage of energy recovery per $l2$ value of the data after denoising and the retained signal characteristics, determine the level of wavelet packet decomposition and the denoising threshold. Then perform wavelet packet threshold denoising.

Step 4: we reconstruct the signal with the components denoised by the wavelet packet threshold and the remaining components that do not need denoising. Then, the dam foundation seepage pressure data after VMD-wavelet packet denoising is obtained.

Step 5: after preparing the denoised dam foundation seepage pressure data and multidimensional factor data composed of relevant environmental factors, normalize these data, and divide them into the training set, validation set, and test set.

Step 6: we set the corresponding parameters of the model and initialize them.

Step 7: we perform hysteresis experiments on different measuring points and find the optimal receptive field that is closest to the real lag time of dam foundation seepage pressure at the measuring point to obtain the optimal ITCN dam foundation seepage pressure prediction model for each measuring point and input the multidimensional factor data at the future moment. The future changes of dam foundation seepage pressure can consequently be predicted.

The basic flowchart can be expressed, as shown in Figure 7:

3. Case Study

3.1. Project Overview. Uplift pressure is one of the important unfavorable loads affecting the structural stability of a gravity dam. Compared with the arch dam, the gravity dam has a wider foundation, longer seepage path, and more obvious hysteresis. Therefore, this paper selects a high gravity dam with dam foundation uplift pressure measuring points in the first and second rows behind the impervious curtain in the middle dam block, as the research object of dam foundation seepage pressure.

This high gravity dam is an RCC gravity dam whose maximum dam height is 168 m and is divided into 24 dam blocks. The two dam foundation uplift pressure measuring points selected in the middle dam block (13#) are numbered UP13 and UP27 respectively, as shown in Figure 8. In terms of measured data, we selected the period (2015-2-10 ~ 2018-12-31) as the research period. During this period, the monitoring frequency was guaranteed to be once a day, and a total of 1421 groups of dam foundation seepage pressure data were obtained.

In terms of the environmental variables, four environmental variables including upstream water level, downstream water level, rainfall, and dam foundation temperature were selected; among which, the dam foundation temperature was selected to be the measurement temperature of
corresponding measuring points. The relevant environmental variables and the actual measurement process line of the dam foundation seepage pressure at each measuring point are shown in Figure 9.

3.2 VMD-Wavelet Packet Denoising. It can be seen from the measured process line that the measured signal of the UP27 measuring point has been severely contaminated by the noise. Therefore, before training the ITCN model, it is necessary to perform the VMD-wavelet packet denoising processing.

3.2.1 VMD Decomposition. According to the basic flow, firstly, the decomposition level $K$ of VMD needs to be determined. In general, when $K$ is less than 3, the decomposition level of the signal will be insufficient, and there will be mode mixing; when $K$ is too large, false signal components will appear. So this paper starts with $K = 3$ and gradually increases the decomposition levels to find the optimal $K$ value.

After comprehensive comparison, for the measured signal of the dam foundation seepage pressure at the UP27 measuring point, carry out VMD 5-level decomposition. The noise components can be better concentrated in IMF1~3 components and residual components, as shown in Figure 10. Therefore, let the number of decomposition levels $K = 5$, and select IMF1~3 components and residual components for wavelet packet threshold denoising.

3.2.2 Wavelet Packet Threshold Denoising. As mentioned earlier, in this paper, when performing wavelet packet threshold denoising, the basis function chooses to be the db4 wavelet; the cost function uses Shannon entropy; the threshold estimation method uses string threshold; and the threshold function uses soft threshold. Therefore, it is only necessary to determine the level of wavelet packet decomposition for each component.

The key basis to determine the level of wavelet packet decomposition is the percentage of energy recovery $\text{perf}_2$ of the data after denoising, whose expression is as follows:

$$\text{perf}_2 = 100 \times \frac{\|x_d\|^2}{\|x\|^2},$$

(15)

where $x$ is the signal before denoising and $x_d$ is the signal after denoising.

The smaller the $\text{perf}_2$ is, the less information is retained after denoising, and the better the denoising effect is.
For each component, the search range of the best decomposition level is set to 1 ~ 7 and record the perf2 value after each denoising in the following figure.

As shown in Figure 11, for IMF1 components, when the decomposition level is 2, the minimum perf2 value is reached; the other components reach the minimum perf2 value when the decomposition level is 3. The signal image after denoising also retains many signal features and achieves a good denoising effect. Therefore, the optimal decomposition level of the IMF1 component is determined to be 2, and the optimal decomposition level of the remaining components is all 3.

Reconstruct the signal with the denoised components and IMF4~5 components to get the dam foundation seepage pressure data of UP27 measuring point after VMD-wavelet packet denoising, as shown in Figure 12:

3.3. Model Training and Prediction. After preparing the non-noise dam foundation seepage pressure data and multidimensional factor data composed of relevant environmental factors, etc, input data into the initial ITCN model. By performing the hysteresis experiment, optimize the model structure. The optimal ITCN dam foundation seepage pressure prediction model is obtained by training. All the prediction models in this paper are based on the Tensor-Flow2.0.0a platform using Python3.7 language.

3.3.1. Hysteresis Experiment. The measured data and seepage theory analysis show that the dam foundation seepage pressure has a certain hysteresis relative to environmental changes, which means the dam foundation seepage pressure data at some time are closely related to the factor data in the previous time period. When the range of the model's receptive field is closer to the length of this lag time, the model can fully capture the historical information within this time range to achieve the best learning effect.

The dam foundation seepage pressure of different dam types and positions has different hysteresis, so for different measuring points, we need to change the size of the receptive field of the model to find the optimal receptive field that is closest to the real lag time of the dam foundation seepage pressure at this measuring point to obtain the optimal ITCN dam foundation seepage pressure prediction model structure for each measuring point.

According to formulas (12) and (13), in the ITCN network, the parameters that determine the receptive field are convolution kernel size C, dilation rate d, and the number of residual network layers n. As mentioned earlier, due to the particularity of the dam foundation seepage pressure, the model does not need to have a huge receptive field, but the depth of the model is required. Therefore, the convolution kernel size C and dilation rate d are set to the minimum value of 2. The size of the model’s receptive field is changed by adjusting the number of residual network layers. Simultaneously, the filters in each layer of the residual network should be successively reduced to 1 to ensure a smooth dimensional decreasing process. The size of the model’s receptive field and the filters of each layer for different residual network layers are shown in Table 1.

To compare the learning effects of different receptive field models, we record the mean square error (MSE) mean value of multiple training results of the validation set. The MSE expression is as follows:
where $y_i$ is the measured value, $\hat{y}_i$ is the predicted value of the model, and $m$ is the length of the validation set data.

Record the hysteresis experiment results of each measuring point as shown in Figure 13. It should be noted that since it is in a normalized state at this time, the MSE magnitude is small:

As can be seen from Table 1 and Figure 13, with the increase in the number of residual network layers, the model receptive field gradually increases, and the learning effect for dam foundation seepage pressure at each measuring point is gradually enhanced. When a certain optimal receptive field is reached, the best learning effect is achieved subsequently. However, with a further increase of the receptive field of the model, the learning effect is no longer better overall.

The optimal receptive field of the model reflects the real lag time of dam foundation seepage pressure at the measuring point. The UP13 measuring point in the first row behind the impervious curtain showed a lag of about 8 days, and the UP27 measuring point in the second row behind the impervious curtain showed a lag of 8 ~ 16 days. This is roughly consistent with the actual cognition and reflects the basic seepage law that the longer the seepage path is, the more obvious the hysteresis is.

3.3.2. Model Prediction. Through the hysteresis experiment, we obtained the optimal model structure for each measuring point. Based on this training model, when the MSE value of the validation set is the minimum, we stopped the training to obtain the optimal ITNC dam foundation seepage pressure prediction model for each measuring point.
To evaluate and compare the effectiveness of the ITCN dam foundation seepage pressure prediction model, we adopted two deep learning methods, the RNN model and its variant LSTM model, which are more common in the field of time series prediction, and also the stepwise regression statistical model, which is more commonly used in the field of traditional dam engineering, as the benchmark methods. Use the denoised data for training to predict the change of dam foundation seepage pressure in the last half year.
To comprehensively evaluate the prediction effect, in terms of prediction accuracy evaluation indicators, in addition to mean squared error (MSE), this paper also considers root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The formula of each evaluation indicator is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}, \quad (17)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i|, \quad (18)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%. \quad (19)$$

Finally, the prediction results of the dam foundation seepage pressure models at each measuring point are shown in Figure 14.

In addition, to verify the improvement effect of the VMD-wavelet packet denoising on model prediction, the original data of UP27 measuring point without denoising were used to train the ITCN dam foundation seepage pressure prediction model for comparison. The comparison results are shown in the following charts.

First of all, it can be seen intuitively from the process lines of the prediction results in Figure 14, among all the models, the ITCN model proposed in this study achieves the best fit between the predicted values and the true seepage pressure values. Two other deep learning models such as LSTM and RNN models have slightly weaker fitting performance. The stepwise regression model has the worst fitting performance although it also reflects the right overall...
trend, and it provides large mismatches in some periods between the predicted values and the real values. The evaluation indicator values of the predicted results in Table 2 further confirm this statement: each indicator value of the ITCN model is smaller than that of the other models. The indicator values of the ITCN model are only 50%–90% of those of LSTM and RNN models, and 15%–40% of those of the stepwise regression model, and the values are all small. The above results prove that the proposed prediction model has a strong fitting ability, small overall error, and high accuracy.

Secondly, it can be seen in Figure 15 and Table 3. After the VMD-wavelet packet denoising processing was performed on the dam foundation seepage pressure data contaminated by the noise, the predicted value is closer to the true value. Each evaluation indicator has been reduced by 25–50%. This indicates that the influence of the noise has been reduced to some extent, and the fitting effect of the model has been improved obviously. But it should be noted that after the denoising of the UP27 measuring point data, the accuracy of the model still has a certain gap with the UP13 measuring point model. This is because in addition to the severe noise contamination, the UP27 measurement point data also have obvious abnormal fluctuations. These fluctuations, not related to environmental factors, are caused by the measuring point itself. Therefore, it is necessary to check the UP27 measuring point in time to eliminate relevant abnormalities to further improve the prediction accuracy.

Overall, the verification results for this engineering example show that the VMD-wavelet packet denoising method can effectively eliminate the influence of noise contamination on the prediction model; the prediction accuracy of ITCN model is better than that of the other models, and the prediction accuracy of each deep learning model is better than that of the stepwise regression statistical model. The prediction accuracy of each model is ITCN > LSTM > RNN > stepwise regression.
Figure 14: The process line of the prediction results of different dam foundation seepage pressure models at each measuring point.

Table 2: Evaluation indicator value for the prediction result of the dam foundation seepage pressure model at each measuring point.

<table>
<thead>
<tr>
<th>UP13</th>
<th>ITCN</th>
<th>LSTM</th>
<th>RNN</th>
<th>Stepwise regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.004</td>
<td>0.008</td>
<td>0.008</td>
<td>0.035</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.067</td>
<td>0.091</td>
<td>0.087</td>
<td>0.188</td>
</tr>
<tr>
<td>MAE</td>
<td>0.054</td>
<td>0.072</td>
<td>0.065</td>
<td>0.165</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.24%</td>
<td>0.32%</td>
<td>0.28%</td>
<td>0.72%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UP27</th>
<th>ITCN</th>
<th>LSTM</th>
<th>RNN</th>
<th>Stepwise regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0013</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0063</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.036</td>
<td>0.039</td>
<td>0.039</td>
<td>0.080</td>
</tr>
<tr>
<td>MAE</td>
<td>0.0298</td>
<td>0.0302</td>
<td>0.0331</td>
<td>0.0748</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.30%</td>
<td>0.31%</td>
<td>0.34%</td>
<td>0.76%</td>
</tr>
</tbody>
</table>

Figure 15: The process line of ITCN model prediction results before and after denoising at the UP27 measurement point.
4. Conclusion and Discussion

To establish a higher performance dam foundation seepage pressure prediction model, this work firstly performs VMD-wavelet packet denoising for the data contaminated by noise. Then, we propose and deeply research the ITCN dam foundation seepage pressure prediction model. The conclusions are as follows:

(1) Considering the dam foundation seepage pressure data contaminated by noise, VMD decomposition is utilized to perform wavelet packet threshold denoising on the components with more noise. Experimental results prove that the signal is reconstructed to obtain the dam foundation seepage pressure data after denoising to eliminate the influence of noise and improve the prediction accuracy of the model.

(2) An improved TCN model is used to build the prediction model according to some characteristics of dam foundation seepage pressure data. This ITCN model also retains the advantages of flexible architecture, free adjustment, and a small amount of calculation.

(3) In addition, as the ITCN model has a certain memory, it has a good learning effect for the dam foundation seepage pressure with hysteresis. This work relates the receptive field size of the ITCN model to the hysteresis of the dam foundation seepage pressure by changing the receptive field of the model to research the real lag time of dam foundation seepage pressure.

(4) Through the hysteresis experiment, we obtained the optimal model structure, based on which, the optimal ITCN dam foundation seepage pressure prediction model was obtained by training. The prediction results show that the prediction accuracy of the ITCN model is better than that of LSTM, RNN, and stepwise regression models.

However, some limitations must be addressed. For newly built dams, more consideration should be given to applying machine learning methods to study the construction of monitoring models under the condition of a small number of samples. Furthermore, data augmentation methods should also be introduced to increase the richness of the data.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Yantao Zhu developed the idea, curated the data, wrote the original draft, and carried out funding acquisition. Zhiduan Zhang collected the resources and supervised the study. Chongshi Gu carried out funding acquisition and supervised the study. Kang Zhang developed the software, visualized the study, and developed the methodology. Yangtao Li and Mingxia Xie validated the study.

Acknowledgments

This work was supported by the National Key R&D Program of China (2022YFC3005401), the National Natural Science Foundation of China (U2040223 and U2243223), China Postdoctoral Science Foundation (2022M720998), the Natural Science Foundation of Jiangsu Province (BK20220978), Jiangsu Water Science and Technology Project (20220224), and Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB176).

References

Structural Control and Health Monitoring 17


