The Method for Determining Optimal Analysis Length of Vibration Data Based on Improved Multiscale Permutation Entropy

Jianwei Zhang, Ziyu Li, Peng Yan, Yang Li, and Jinlin Huang

1 School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450046, Henan, China
2 Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education, Wuhan University, Wuhan 430072, Hubei, China
3 Guangdong Research Institute of Water Resources and Hydropower, Guangzhou 510635, Guangdong, China

Correspondence should be addressed to Jianwei Zhang; zjwcivil@126.com

Research on damage diagnosis or safety monitoring based on structural vibration response is one of the hot issues in the engineering field. The characteristic information of the structure is obtained by analyzing the structure response data. In the process of data analysis, the choice of data length is very important, which is related to the validity of the structure monitoring results. At present, the selection of data length is usually subjective, which reduces the rigor of the structure monitoring process. Therefore, a method based on improved multiscale permutation entropy (IMPE) is proposed to determine the optimal data analytical length (ODAL) of vibration data. This method creatively applies multiscale permutation entropy (MPE) to the field of data length analysis when processing nonlinear and nonstationary signals and optimizes MPE with the help of the improved coarse-grained method to obtain IMPE. IMPE is sensitive to different data lengths, and the entropy changes with the increase of the data length and tends to be stable. Here, the stable value is defined as a standard entropy. The entropy satisfying 97% of the standard entropy is used as the effective entropy, and the corresponding data length value of the effective entropy is selected as the ODAL of the vibration data. This method is suitable for many fields, provides a reliable data analytical length for data analysis, and has good engineering practicability.

1. Introduction

The structural vibration data of large-scale water conservancy projects is the main carrier that reflects the structural vibration characteristics [1–3]. With the continuous upgrading and transformation of the level of monitoring automation, in the face of massive monitoring data, selecting effective vibration data analytical length and extracting data feature information with the help of efficient and reliable data processing methods is an important basis for realizing real-time safety monitoring of hydraulic structures. The analysis and application of long time series data involve many fields [4–6]. In the field of data-driven feature information extraction from monitoring data, Dong et al. used the ensemble empirical mode decomposition (EEMD) method to analyze the relationship between different vibration sources and the structural operation safety of offshore wind turbines and realized the safety monitoring of offshore wind turbine [7]. Lian et al. used the FM-CEEMD decomposition method to process the time series of modal vibration information of high double arch dams, which can effectively avoid the loss of modal information and improve the accuracy of modal parameter identification [8]. Deng et al. used a variety of evolutionary algorithms to identify the characteristics of the complex multiobjective optimization problem of airport gate allocation and obtain the optimal solution [9, 10]. Chen et al. applied a wavelet method to model vibration response data to get the amplitude-dependent nonlinear damping and stiffness of the HAPB system, which provides the basis for vibration response...
analysis and prediction [11]. These scholars extracted the required characteristic information from the monitoring data with the help of various methods and obtained many achievements.

However, the data length, as another important factor influencing the analysis results, has been seldom concerned. Usually, the length of vibration analysis data in the process of damage diagnosis and safety monitoring of structures is selected by the subjective intention of individuals with great arbitrariness. It is impossible to exclude the influence of different data lengths on the subsequent results, which is certainly misleading. The data length determines the richness of the signal it contains. If the data length is too long, there will be disadvantages such as time-consuming and cumbersome calculations, which will increase the burden of safety monitoring of the hydraulic structure. The selection of shorter data will cause the loss or incompleteness of feature information, making misjudgment in the safety monitoring process. Therefore, choosing the optimal data analytical length (ODAL) that can retain complete feature information and has high calculation efficiency is the key step to ensure the accuracy and real-time performance of the hydraulic structure safety monitoring. In view of this, finding a method that can effectively evaluate the richness of data features is a prerequisite for determining ODAL.

Multiscale permutation entropy (MPE) method, proposed by Aziz and Arif on the basis of permutation entropy, is a new method for detecting dynamic mutation and time series permutation [12]. It is characterized by simple calculation, high sensitivity, and strong antinoise ability and can effectively reflect the small changes in the time series of nonlinear and nonstationary signals [13]. Because of its sensitivity in detecting dynamic mutations, this method has become popular in data analysis and has been widely used in biomedical, mechanical damage diagnosis, stock analysis, and other fields [14–19]. With the development and extension of MPE, the improved multiscale permutation entropy (IMPE) has been proposed and applied in many fields. Humeau-Heurtier et al. put forward improved coarse-graining method to improve the MPE, which reduces the dependence of the analysis results on the data length and greatly improves the accuracy of data analysis [20]. Azami and Escudero applied the IMPE in the field of biomedical signal analysis, and it is found that IMPE improves the reliability of entropy estimation compared with the MPE [21]. Zhang et al. found that the IMPE can more accurately dig out the depth information of the time series, and the entropy value obtained is more consistent and stable [22]. Minhas and Singh analyzed the potential features of bearing nonlinear vibration signals and implemented bearing fault diagnosis with the help of IMPE and found that IMPE was more accurate than other state-of-the-art permutation-based feature extraction methods [23]. In view of this, we believe that IMPE can effectively represent different permutation-based feature extraction methods [23]. In view of potential features of bearing nonlinear vibration signals and information of the timeseries, and the entropy value obtained is entropy estimation compared with the MPE [21]. Zhang et al. and Escudero applied the IMPE in the field of biomedical signal and greatly improves the accuracy of data analysis [20]. Azami reduce the dependence of the analysis result on the data length improved coarse-graining method to improve the MPE, which applied in many fields. Humeau-Heurtier et al. put forward scale permutation entropy (IMPE) has been proposed and the development and extension of MPE, the improved multi-namic mutations, this method has become popular in data tionary signals [13]. Because of its sensitivity in detecting dy-sensitivity, and strong antinoise ability and can effectively reflect mutation [12]. It is characterized by simple calculation, high method for detecting dynamic mutation and time series per-selecting the ODAL are presented.

In this section, the principles of IMPE and the method on selecting the ODAL are presented.

2. Principle of IMPE and Method of Selecting ODAL

In this section, the principles of IMPE and the method on selecting the ODAL are presented.

2.1. Multiscale Permutation Entropy. Based on the permutation entropy algorithm, the multiscale permutation entropy algorithm multiscales one-dimensional time series, calculates the permutation entropy of coarse-grained time series at different scales, and finally forms the multiscale permutation entropy.

First, consider a one-dimensional time series \( \{X(i); i = 1, 2, \ldots, n\} \); the following sequence was obtained by coarse-graining processing:

The coarse granulation process is shown in

\[
y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{j} X(i), \quad j = 1, 2, \ldots, \left\lceil \frac{n}{s} \right\rceil, (1)
\]

where \( y_j^{(s)} \) is coarse-grained sequence and scale factor \( s \) determines the degree of coarse-grained sequence. The length of coarse-graining time series decreases with the increase of scale factor \( s \). When \( s = 1 \), the coarse-grained sequence is the original sequence; \( \lceil n/s \rceil \) denotes rounding up \((n/s)\).

Reconstruct the coarse-grained sequence \( y_j^{(s)} \) to obtain the reconstructed coarse sequence \( y_j^{(s)} \) and the reconstructed components \( Y_j^{(s)} \):

\[
Y_j^{(s)} = \left\{ y_j^{(s)}, y_{j+1}^{(s)}, \ldots, y_{j+(m-1)s}^{(s)} \right\}, \quad l = 1, 2, \ldots, n - (m - 1)s, (2)
\]

where \( m \) and \( \tau \) denote the embedding dimension and time delay, respectively. \( Y_j^{(s)} \) represents the \( l \)-th reconstructed component.
Permutation entropy is defined as follows:

\[ z_{i_t(l_t-1)} \leq z_{i_t(l_t-2)} \leq \cdots \leq z_{i_t(l_m-1)}; \]  

(3)

where \( l_1, l_2, \ldots, l_m \) denote the index of the column of each element in the reconstructed component \( Y_{t}^{(s)} \).

For any coarse-grained sequence \( y_{i_t}^{(s)} \), a set of symbol sequences \( s(r) = (l_1, l_2, \ldots, l_m) \) can be obtained for \( r = 1, 2, \ldots, R \) and \( R \leq m! \). Set the probability of the occurrence of each symbol sequence as \( P_r \), \( r = 1, 2, \ldots, R \). Permutation entropy is defined as follows:

\[ H_{P(m)} = -\sum_{r=1}^{R} P_r \ln P_r. \]  

(4)

Normalize the above permutation entropy:

\[ H_p = \frac{H_{P(m)}}{\ln (m!)} , \quad 0 \leq H_p \leq 1. \]  

(5)

The \( H_p \) value represents the random degree of time series \( y_{i_t}^{(s)} \). The smaller the value is, the smaller the time series complexity is. On the contrary, the greater the value is, the more random the time series and the greater the complexity are.

In summary, after the coarse-graining of original time series \( \{X(i); i = 1, 2, \ldots, n\} \), the reconstructed coarse sequence \( y_{i_t}^{(s)} \) are obtained, and the permutation entropy of each coarse-grained sequence is calculated to obtain MPE, that is, \( H_{mp}(X) = \{H_p(1), H_p(2), \ldots, H_p(s)\} \).

2.2. Improved Coarse-Graining Method. Coarse-graining is an indispensable part in the calculation of MPE. The essence of the traditional coarsening method is as follows: for scale factor \( s \), the original time series is divided into \( s \) nonoverlapping windows, the mean value of data points in each window is calculated, and a new set of time series is formed from the obtained mean value. The specific process diagram is shown in Figure 1(a). The disadvantage is that when the scale factor \( s \) is too large, the time series is too short and contains too few data points, leading to the inaccurate estimation of MPE. In order to solve this problem, Humenau-Heurtier et al. [20] proposed the coarse-graining process based on the moving average and applied it to the calculation of sample entropy. In this way, the optimized coarse-graining method was applied to the calculation of MPE to improve the accuracy of results. The specific process diagram is shown in Figure 1(b).

On the given scale factor \( s \), the corresponding coarse-grained sequence is obtained by moving average method. The formula is as follows:

\[ z_{i_t}^{(s)} = \frac{1}{s} \sum_{i=s}^{j+s-1} X(i), \quad j = 1, 2, \ldots, (n-s+1), \]  

(6)

where \( z_{i_t}^{(s)} \) is coarse-grained sequence.

The sequence length after coursing is \( L \):

\[ L = (n-s+1). \]  

(7)

If the lengths of the original time series \( n = 600 \) and \( s = 15 \) are taken, the length of the shortest coarse-grained sequence obtained by the improved coarse-graining method is 586, and the length of the shortest coarse-grained sequence obtained by the traditional coarse-graining method is 40. Therefore, the improved coarse-graining method greatly improves the validity of the result.

2.3. Parameter Selection of the Phase Space Reconstruction. In order to ensure the validity of entropy values at each scale, the embedding dimension \( m \) and the delay time \( \tau \) need to be calculated separately before calculating the entropy value of each coarse-grained sequence. As an important step before entropy value calculation, the selection of phase space reconstruction parameters is closely related to the accuracy of signal analysis results, which can be roughly divided into independent determination and joint determination methods. Both methods are feasible, but for the detection of abnormal conditions, the independent determination method is more accurate [24]. Here, the embedding dimension \( m \) and delay time \( \tau \) are obtained by false nearest neighbor (FNN) and mutual information (MI) methods. The criterion for selecting two parameters is that the appropriate dimension \( m \) is the corresponding dimension when the percentage of pseudo-proximity points in the phase space tends to zero; the optimal delay time \( \tau \) is the delay time corresponding to the minimum value for the first time. In the measured data, \( m \geq 2, \tau \geq 1 \).

Inaccurate selection of \( m \) value will lead to the failure of dynamic mutation in time series, and inaccurate selection of \( \tau \) value will lead to too large or too small degree of correlation between data. Determining the phase space reconstruction parameters \( m \) and \( \tau \) by using the above two methods can effectively reduce the error of the permutation entropy calculation results and improve the accuracy of the analysis results.

2.4. Optimal Length Selection of Data Analysis. The determination of ODAL based on IMPE uses the sensitivity of the entropy value to the sudden change of dynamical system to find the suitable length of data, so as to solve the problem of data length selectivity in data analysis. Generally, it can be divided into the following four parts: coarse-graining, parameter selection and phase space reconstruction, calculation of entropy value, and selection of sequence length. The specific process diagram is shown in Figure 2, and the detailed steps are as follows:

1. Lay out sensor devices at the key positions of the measured structures to obtain the vibration data of the structures \( \{X(i); i = 1, 2, \ldots, n\} \).

2. Extract data information of different lengths \( N \), and appropriate scale factor \( s \) is selected and vibration data are coarse-graining processed. On the given scale factor \( s \), the corresponding coarse-grained sequence \( z_{i_t}^{(s)} \) is obtained by moving average.
(3) Determine the phase space reconstruction parameters \( m \) and \( \tau \) of coarse-grained data by false nearest neighbor (FNN) and mutual information (MI) methods, respectively, and the phase space reconstruction is carried out.

(4) Calculate the entropy value \( \text{PE}_1, \text{PE}_2, \ldots, \text{PE}_s \) of each time series after coarse-graining, and get multiple scales permutation entropy \( \text{MPE}_s = \{\text{PE}_1, \text{PE}_2, \ldots, \text{PE}_s\} \). Mean of multiscale permutation entropy is used as the basis to measure the complexity of vibration data, in which \( \text{MPE} = (\text{PE}_1 + \text{PE}_2 + \cdots + \text{PE}_s) / s \).

(5) Calculate MPE \( (N_1) \), MPE \( (N_2) \), \ldots, MPE \( (N_i) \), \ldots, MPE \( (N_n) \) of different length \( N \) vibration data under the same vibration state. Now, the larger the data length \( N \) is, the larger the entropy value is. MPE tends to a certain fixed value until the data reaches a certain length. When \( \text{MPE}(N_n) - \text{MPE}(N_{n-1}) = 0 \), MPE \( (N_n) \) is taken as the standard entropy value, and the data length \( N_n \) corresponding to MPE \( (N_n) \) is taken as the standard data length \( N \).

(6) Compare MPE \( (N_1) \), MPE \( (N_2) \), \ldots, MPE \( (N_i) \), \ldots, MPE \( (N_n) \) with MPE \( (N_n) \), respectively, and select MPE \( (N_i) \) which is greater than 97% of the MPE \( (N_n) \), and then the shortest data length corresponding to MPE \( (N_i) \) is defined as the ODAL of the vibration data.

3. Simulation Test

3.1. Verify the Improved Coarse-Graining Method. In order to test the superiority of the improved coarse-graining method in MPE, taking the white noise in the condition of short data length as an example, the entropy values of time series at different scales before and after the optimization of coarse-granulation method are calculated, respectively. The specific experimental data and results are shown in Figure 3.

White noise is a time series of purely random processes with a theoretical entropy value of 1. As a time series of pure random process, the theoretical entropy of white noise is 1. Since the accuracy of the entropy value in the actual test has a certain dependence on the data length, the actual permutation entropy of white noise ranges from 0.90 to 0.97 under two shorter data lengths \( N = 200 \) and \( N = 500 \). It can be seen from the graph that the measured entropy value of white noise signal decreases with the increase of scale factor after the original and improved coarse-graining methods; that is, the shorter the data length is, the smaller the entropy value is, the greater the difference

Figure 1: Improvement of coarse-graining. (a) Illustration of coarse-grained time series. (b) An improved illustration of coarse-graining time series.
between the actual value and the theoretical value is. The difference is that, after the improvement of coarse-graining method, the measured entropy value is less affected by the scale factor. With the increase of scale factor, the decrease speed of entropy value slows down and the accuracy increases. Compared with the original method, it shows that the entropy value of the improved coarse-grained method is closer to the theoretical value under the condition of the same data length and scale factor.

In addition, according to the experimental results, whether before or after the improved coarse-graining method, the entropy value of white noise signal with a data length of 500 is more accurate than the entropy value of white noise signal with a length of 200. It also shows that, besides the coarse-graining method, the sequence length is also an important factor affecting the accuracy of the entropy value.
3.2. Test the Influence of Data Length on IMPE. It can be known from the above simulation experiments that the sequence length will affect the accuracy of the entropy value. If the sequence length is too short, the credibility of the entropy value is low. If the sequence length is too long, there are various disadvantages, such as cumbersome calculation, being time-consuming, and blurring of mutation. Therefore, finding a sequence length with accurate calculation results, moderate length and favorable calculation are a key part to ensure the validity of the analysis results. At the same time, it also provides a reference standard for the selection of data length in subsequent signal analysis. Here, the white noise, which has a standard entropy value, is taken as an example for analysis. The specific test data and results are shown in Figure 4.

As shown in Figure 4, the entropy curves of white noise with length \( N = 200, 500, 1000, 2000, 3000, 4000, \) and 5000 on scale \( s \in [1, 15] \), respectively. With the increase of sequence length, the measured entropy value gradually approaches the true value 1; with the increase of scale factor, the longer the sequence is affected by scale factor, the more accurate the entropy value is. The results of IMPE are very sensitive to data length, and the scale are associated with the data length. In addition, it can be seen that when the data length of white noise reaches 4000–5000, the entropy changes very slightly with the increase of data length and finally stabilizes at 0.998. It is shown that the accuracy of actual entropy value does not increase with the infinite increase of data length, and its accuracy has a certain limit. Therefore, in signal analysis, only within a certain length range, the more the data length is, the more accurate the analysis is.

The purposes of entropy value calculation are signal analysis and evaluation, which have higher requirements for the accuracy of the results. However, the higher the accuracy requirement is, the richer the data information is; that is, the more the data length is, the more difficult the calculation is. Therefore, when the data length \( N \) reaches a certain level, the change of the entropy value of IMPE with the increase of \( N \) can be ignored, and the \( N \) and IMPE at this time are set to the standard data length and the standard entropy value. In the actual calculation, the standard data length is used as the analysis length and the accuracy is the highest, but usually the data length is large, the entropy value calculation and analysis take a long time, and the efficiency is low. Therefore, under the condition that the analysis effect is not affected, the entropy value satisfying the standard value of 97% accuracy is taken as the effective entropy value, and the corresponding data length is taken as the ODAL. For example, in the above experiment, when the white noise data length reaches 5000 or more, the entropy value no longer increases with the increase of the data length, and the entropy value reaches 0.998, which is a little different from the theoretical value. Therefore, the entropy value 0.998 corresponding to the data length \( N = 5000 \) was taken as the standard entropy value, and the entropy value of 0.968 corresponding to the data length satisfying 97% accuracy of the standard value was selected as the reasonable analysis length. The result shows that the data length of white noise \( N = 3000 \) has an entropy value of 0.971, which satisfies the accuracy requirement, so the ODAL of white noise is 3000.

3.3. Construct Simulation Signal. In order to verify the reliability of the method for determining the ODAL, the simulation pure signal \( f_1(t) \) and white noise \( f_2(t) \) are constructed. Their expressions are as follows:

\[
f_1(t) = 20e^{-(\pi/2)t} \sin(15t) + 15e^{-(\pi/3)t} \sin(20t),
\]

\[
f_2(t) = \text{rand}(m),
\]

where \( t \) is time, \( m \) is the number of samples, and \( \text{rand}(m) \) is white noise, which follows the standard normal distribution, and \( f = 100 \) Hz is sampling frequency.

Add white noise \( f_2(t) \) with SNR of 5%, 10%, 20%, 30%, 40%, and 50% to pure signal \( f_1(t) \). According to the above method, select noise signal with data length of \( N = 500, 800, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, \) and 5000, and the entropy curve is calculated, as shown in Figure 5.

According to Figure 5, the entropy value of the noise signal increases with the increase of the data length. When the data length reaches 4000–5000, the entropy value is basically stable at 0.998, which has the same change law as the white noise signal. Therefore, the data length \( N = 2000 \) corresponding to the entropy value that satisfies the standard entropy value of 97% accuracy can be selected as the optimal analysis length of the vibration measurement data of the noisy signal. When the sequence length reaches the optimal data length, the entropy of the noisy signals with different SNR tends to the same stable entropy value. Therefore, it can be proved that the multiscale permutation entropy has a strong antinoise ability and good robustness, which can effectively judge the state of the signal.
4. Engineering Example

4.1. Example 1

4.1.1. Engineering Situation. The Jingtaichuan project in Gansu Province is an electric water-lifting irrigation project with high head, large flow, and multiechelon. During the pipeline operation for many years, the effects of material aging, the fluctuating load of water flow, traffic, and so on [25], which lead to structural damage and aging of the pipeline in some extent, directly affect the safety of pipeline operation. Taking the seventh pump pipeline of Jingtaichuan Project Phase II as the research object, the site layout chart and the pipeline sensor placement are shown in Figure 6.

4.1.2. Collection of Vibration Information. The vibration information of pipe is collected by sensors mounted on the surface of the pipe. In order to ensure the collection effect and get the system characteristic information, 22 sensors are placed in the pipeline of the pumping station [26]. At the instant of the switching unit, the water flow excitation and self-excitation force at the connection of the branch and main pipeline are particularly complex. Therefore, three speed sensors are placed at the connection, namely, horizontal X direction, horizontal Y direction, and vertical Z direction, to accurately capture the dynamic characteristics of the system, while different numbers of sensors are arranged at other positions. The excitation of pipeline vibration test is obtained by means of switching pumps, and two general types of vibration states are the moment of the switch machine and the stable operation of the unit, which are divided into (A ~ K) 11 kinds of states. The 11 test states are shown in Table 1. In order to fully reflect the vibration characteristics of pipelines under various working conditions, the vibration signals are selected as the research objects, that is, the sensors No. 15 (corresponding to the X direction), at the intersections where the switching moments of (1–3) turbines are more obviously affected. The specific experimental conditions are as follows.

4.1.3. Determination of m and τ. Phase space reconstruction is an indispensable part of drawing recursive graphs and calculating quantitative indicators. Therefore, the key step of phase space reconstruction is to select reasonable m and τ. The phase space reconstruction parameters m and τ are selected by FNN and MI methods, respectively. The appropriate dimension m is the corresponding dimension when the percentage of pseudo-proximity points in the phase space tends to zero (at least less than 5%); the optimal delay time τ is the delay time corresponding to the minimum value for the first time.

The phase space reconstruction parameters of the pipeline in the seventh pumping station of Jingtaichuan Project II have little difference among the above (A ~ K) 11 states, which are stable in m = 4 and τ = 4 states, respectively. Therefore, the calculation parameters of IMPE selected in this paper are m = 4, τ = 4, and s = 15. Here, the parameter selection diagram of condition B is shown in Figure 7.

4.1.4. Analysis of Examples. Pipeline in the pump off and on moment will be subject to a larger excitation force from the system itself; that is, the pump vibration will have a significant impact on the operation of the pressure pipeline. Then with the smooth operation of the machine, the impact of water excitation is reduced, and the amplitude of the pipeline is reduced. To obtain the ODAL of vibration data of pipeline under different operating conditions, the vibration information obtained from no. 15 sensor is taken to research. Here, one working condition is selected in the four states of startup, shutdown, stable operation, and total shutdown to show the change of signal entropy value with the sequence length.

The entropy value curve of the pipeline in Jingtaichuan Project under working conditions A, B, C, and J is shown in Figure 8. For each working condition, 12 different sequence lengths were selected, such as $N = 200, 500, 1000, 2000, 3000$,
4000, 5000, 6000, 7000, 8000, 9000, and 10000. From Figure 8, the entropy values of signals in the four states are different, but the measured entropy values at the switching machine moment (working conditions B and J) are much lower than those at stable operation (working conditions A and C). The entropy value tends to stabilize when the data length increases to a certain level, and this trend is consistent with the pattern presented by the white noise signal. Results
show that it is feasible to select the length of signal analysis by using IMPE. In addition, IMPE is also effective in expressing the relationship between structural state and vibration. The results are as follows: the entropy values reach a steady state when the data length $N = 6000$ under four working conditions, and the entropy values are 0.535, 0.620, 0.884, and 0.850, respectively, for stable operation, full stop, shutdown moment, and startup instant. The ODAL selected according to the accuracy requirement is $N = 2000$. Therefore, the ODAL of vibration data of pipeline is $N = 2000$.

To test the influence of data length on IMPE, the lengths $N = 1000$ and $N = 2000$ are selected to analyze the same vibration signal, which shows that the data length is different, and the effectiveness of signal monitoring is also different. Figure 9(a) is the vibration signal characteristics from opening to stable operation of the pump. Figure 9(b) is the variation curve of the entropy value under two sequence lengths.

It can be seen from Figure 9, when the sequence length is 2000 and 1000, the maximum-minimum entropy difference is 0.09 and 0.055, respectively. When the sequence length is 2000, the vibration signal has obvious mutation. Through comparative analysis, it can be concluded that the sensitivity of IMPE to sudden changes in the signal is reduced, which is not conducive to state detection. The reason is that the data length determines the signal richness. Results show that the shorter the data length is, the greater the impact on the accuracy of signal analysis will be. To avoid the misjudgment in the state detection, choosing a reasonable data length becomes an important part to ensure the correctness of analysis consequence.

4.2. Example 2. In order to verify the practicability of the method in actual projects, the vibration of the spillway section 5 of the Three Gorges dam is studied. The arrangement of measuring points is shown in Figure 10. The vibration signals of 1~6 channels of horizontal and vertical dynamic displacement sensors of 1#, 2#, 3#, and 4#
measuring points are selected as the research object, which fully reflect the vibration characteristics of the dam body.

The entropy value change curve of the vibration signals of the six channels in the no. 5 overflow dam section of the Three Gorges Project is shown in Figure 11. It can be seen from the figure that when the sequence length \( N = 200, 500, 1000, 2000, 3000, \) and 4000, the entropy value of the vibration signal in each channel is different. The entropy values of horizontal dynamic displacement vibration (channels 1–4) are smaller than those measured in vertical dynamic displacement entropy (channels 5 and 6). The entropy value gradually increases with the increase of data length and finally reaches a stable state. It shows that it is feasible to select the signal analysis length by IMPE method. When the data length \( N = 2000 \), the entropy value of each channel reaches a stable state and meets the 97% accuracy requirement, and the optimal data analytical length under each working condition is calculated as \( N = 1000 \). This shows

Figure 9: Vibration signal and permutation entropy value change at the pump opening moment. (a) Time course graph of pipeline vibration. (b) Diagram of entropy change under different sequence lengths. Note: the black line is the change of the entropy value of the subsequence \( N = 1000 \); the red line is the change of the entropy value of the subsequence \( N = 2000 \).

Figure 10: Measuring point layout of no. 5 overflow section of Three Gorges dam.
that the method of IMPE can be effectively applied in practical engineering and has great value in selecting the appropriate length of vibration data.

5. Conclusion

Different length of signal will contain different richness of characteristic information. Combined with the superiority of multiscale permutation in nonlinear signal mutation detection, a method to determine the ODAL of vibration data based on improved multiscale permutation entropy (IMPE) is proposed. The method is applied to the optimal data analytical length selection of white noise simulation signals and actual engineering structural vibration monitoring data. The main conclusions are as follows:

(1) By analyzing the simulated signal with different SNR, it is proved that the improved multiscale permutation entropy has strong antinoise ability and good robustness, which can effectively avoid the influence of mixed noise on the accuracy of calculation results.

(2) The simulation and the analysis results of specific projects show that the length of the data is closely related to the accuracy and stability of the improved multiscale permutation entropy. The entropy value corresponding to the optimal data analytical length meets the standard entropy value of 97%, which can meet the requirements of engineering accuracy.

(3) The proposed method provides a reliable data length for signal analysis, eliminates the randomness and subjectivity caused by artificial selection of data length, and improves the accuracy and efficiency of hydraulic structural vibration safety monitoring.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Additional Points

*Highlights*. A method for determining optimal analysis length of vibration data based on improved multiscale permutation entropy is proposed. A white noise signal is used to investigate the superiority of IMPE compared to MPE. The analysis on the simulated signal with different SNR proves the IMPE has strong antinoise ability and good robustness. Simulation analysis and real engineering cases are used to determine the effectiveness of obtaining the optimal data analytical length based on IMPE. The entropy value corresponding to the optimal data analytical length that meets the standard entropy value of 97% can satisfy the requirements of engineering accuracy.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51679091), Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education (Grant no. RMHSE1902), the Program for Science and Water Conservancy Science and Technology Innovation Project in Guangdong Province (Grant no. 2020-18), and Science and Technology Program of Guangzhou (Grant no. 2020-ky34).

References

[1] J. Dong, S. Yi, W. Deyu, and Y. Jiaxiang, "Data analysis and processing for Maglev vibration test systems," in *Proceedings of the 2015 IEEE International Conference on Electronic Measurement & Instruments*, vol. 3, pp. 1162–1166, Qingdao, China, June 2015.

[2] Z.-C. Wang, F. Wu, and W.-X. Ren, "Stationarity test of vibration signals with surrogate data and time-frequency analysis," *Advances in Structural Engineering*, vol. 20, no. 8, pp. 1143–1154, 2017.

[3] Z. Chen, X. Fu, Y. Xu, C. Y. Li, B. Kim, and K. T. Tse, "A perspective on the aerodynamics and aeroelasticity of tapering: partial reattachment," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 212, pp. 1–19, 2021.

[4] J. Hu and W. Zheng, "Multistage attention network for multivariate time series prediction," *Neurocomputing*, vol. 383, pp. 122–137, 2020.

[5] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, and J. Xu, "An improved differential evolution algorithm and its application in optimization problem," *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, 2021.

[6] J. Zhang, G. Hou, K. Cao, and B. Ma, "Operation conditions monitoring of flood discharge structure based on variance
[7] X. Dong, J. Lian, and H. Wang, “Vibration source identification of offshore wind turbine structure based on optimized spectral kurtosis and ensemble empirical mode decomposition,” Ocean Engineering, vol. 172, no. 15, pp. 199–212, 2019.

[8] J. Lian, Q. Rong, X. Dong, and H. Wang, “Structural model parameter identification method based on an improved HHT for suppressing mode mixing,” Journal of Vibration and Shock, vol. 37, no. 18, pp. 1–8, 2018.

[9] W. Deng, J. Xu, Y. Song, and H. Zhao, “Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem,” Applied Soft Computing, vol. 100, pp. 1–16, 2020.

[10] W. Deng, J. Xu, H. Zhao, and Y. Song, “A novel gate resource allocation method using improved PSO-based QEA,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–9, 2020.

[11] Z. Chen, K. T. Tse, K. Kwok, A. Kareem, and B. Kim, “Measurement of unsteady aerodynamic force on a galloping prism in a turbulent flow: a hybrid aeroelastic-pressure balance,” Journal of Fluids and Structures, vol. 102, pp. 1–16, 2021.

[12] G. Q. Rao, F. Z. Feng, A. W. Si, and J. L. Xie, “Method for optimal determination of parameter in permutation entropy algorithm,” Journal of Vibration and Shock, vol. 33, no. 1, pp. 73–79, 2014.

[13] J. Zhang, Q. Jiang, T. Wang et al., “Analysis of vibration characteristics of pipeline of pumping station based on prototype observation,” Transactions of the Chinese Society of Agricultural Engineering, vol. 33, no. 1, pp. 77–83, 2017.

[14] J. Zhang, Z. Bao, X. Liu, and T. Wang, “Sensor placement optimization method for pressure pipeline suitable for grading pumping station,” Transactions of the Chinese Society of Agricultural Engineering, vol. 32, no. 4, pp. 113–118, 2016.