



Article Vibration Prediction and Evaluation System of the Pumping Station Based on ARIMA–ANFIS–WOA Hybrid Model and D-S Evidence Theory

Shuo Wang ¹, Liaojun Zhang ^{2,*} and Guojiang Yin ³

- ¹ College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China; qqwang@hhu.edu.cn
- ² College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
- ³ Safety Monitoring Center of Hydraulic Metal Structure of the Ministry of Water Resources, PRC,
- Hohai University, Nanjing 210098, China; gjyin@hhu.edu.cn Correspondence: ljzhang@hhu.edu.cn

Abstract: Research on the vibration response prediction and safety early warning is of great significance to the operation and management of pumping station engineering. In the current research, a hybrid prediction method was proposed to predict vibration responses of the pumping station based on a single model of autoregressive integrated moving average (ARIMA), a combined model of the adaptive network-based fuzzy inference system (ANFIS) and whale optimization algorithm (WOA). The performance of the developed models was studied based on the effective stress vibration data of the blades in a shaft tubular pumping station. Then, the D-S evidence theory was adopted to perform safety early warning of the operation state by integrating the displacement, velocity, acceleration and stress indicators of the vibration responses of the pumping station. The research results show that the proposed prediction model ARIMA–ANFIS–WOA exhibited better accuracy in obtaining both linear and nonlinear characteristics of vibration data than the single prediction model and hybrid model with different optimization algorithms. The D-S evidence fusion results quantitatively demonstrate the safe operation state of the pumping station. This research could provide a scientific basis for the real-time analysis and processing of data in pumping station operation and maintenance systems.

Keywords: hybrid prediction model; vibration; pumping station; ARIMA-ANFIS-WOA

1. Introduction

As turbulent water flows into the flow channel of a pumping station, the unit and the concrete house will generate long-term continuous vibrations. These vibrations may cause severe damage to the unit instruments and pump house structure, as well as adversely affecting the health of the staff. Moreover, the relatively strong vibrations can pose a threat to the overall safety and functionality of the pumping station project. Therefore, research on the evaluation and prediction of these vibrations is crucial in aiding managers' decision-making process, minimizing potential damage and ensuring the safe operation of the pumping station. For this purpose, it is vital to select suitable methods for establishing a vibration prediction model and a safety early warning model for the pumping station.

There are numerous models available for predicting structural vibrations. Forsat [1] proposed a higher-order shear deformation beam theory to predict the vibrations of hyperelastic beams. Mirjavadi [2] employed the Timoshenko beam theory to predict the thermal vibrational behavior of 2D functionally graded porous microbeams. Due to their complex construction, pumping stations may not be suitable for the application of the mathematical model. Methods based on an artificial neural network (ANN) and its intelligent optimization are in full swing in predicting the structural behavior of pumping houses and hydropower houses. In 2007, Lian et al. [3] utilized the back-propagation (BP) neural



Citation: Wang, S.; Zhang, L.; Yin, G. Vibration Prediction and Evaluation System of the Pumping Station Based on ARIMA–ANFIS–WOA Hybrid Model and D-S Evidence Theory. *Water* 2023, *15*, 2656. https:// doi.org/10.3390/w15142656

Academic Editor: Helena M. Ramos

Received: 6 June 2023 Revised: 16 July 2023 Accepted: 19 July 2023 Published: 22 July 2023



Copyright: © 2023 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/). network method to predict the vibration displacement amplitude of the Three Gorges Hydropower Station, and the predictions were compared with the measured data. Miao et al. [4] used a radial basis function (RBF) neural network to predict the vibration acceleration amplitude of a pier. Since 2014, scholars have conducted a series of studies on predicting the vibration responses of powerhouses. Xu et al. [5] combined the generalized regression neural network (GRNN) and the fruit fly optimization algorithm (FOA) to predict the radial displacement amplitude of the flood discharge surface hole cover plate under various working conditions. The results confirmed the superior prediction ability and learning speed of the FOA-GRNN method compared to BP and ELMAN neural networks. Xu et al. [6] used the survival-of-the-fittest and step-by-step selection particle swarm optimization algorithm (SSPSO) to optimize the smoothing parameter P of GRNN, and they carried out prediction research on vibration problems of hydropower stations under various load conditions. Their results demonstrate that SSPSO-GRNN outperforms PSO–GRNN, self-competitive PSO–GRNN and GA–PSO–GRNN in terms of prediction accuracy, convergence performance and generalization ability. Wang et al. [7] established a model based on relevant vector machine (RVM) regression to predict the vertical displacement standard deviation of a large underground hydropower station under different loads. The results indicate that the RVM model has higher prediction accuracy than the support vector machine (SVM) model. Based on the vibration response data of the same powerhouse, Liu and Du [8] confirmed that the vibration prediction model based on an RBF neural network and improved bat algorithm (IBA) is superior to the RVM model in terms of prediction accuracy and generalization ability. Song et al. [9] utilized the improved firefly algorithm (IFA) and BP neural network to predict the amplitude of hydropower house vibrations. The results demonstrate that the prediction accuracy and convergence speed of the IFA–BP model are significantly improved compared to the BP and FA–BP models. All of the proposed prediction methods and results are highly significant for the advancement of vibration research in pumping station engineering. However, current research on predicting vibrations in pumping houses and hydropower houses mainly focuses on the vibration amplitude under specific working conditions, with fewer studies addressing the prediction of time trend responses of structural vibrations.

Vibration trend prediction can accurately depict the operational behavior of a unit through a time series of vibration responses, surpassing the limited scope of vibration amplitude analysis for specific working conditions. In the field of vibration trend prediction research, the autoregressive integrated moving average (ARIMA) has gained widespread adoption as a statistical method for accurate time-series prediction. ARIMA effectively handles non-stationary time series by employing lag value regression of the dependent variable and the present value of the random error term, thereby harnessing trend characteristics, dynamic information and series persistence to forecast future trends [10,11]. The adaptive network-based fuzzy inference system (ANFIS) represents an intelligent prediction model that combines the principles of fuzzy inference and artificial neural networks (ANNs). ANFIS incorporates fuzzy rules into its inference system to handle uncertainty related to influencing factors and utilizes ANN for simulation and prediction tasks. Compared with the single prediction model, ANFIS offers distinct advantages, such as simplicity in expressing fuzzy logic and the capability for self-learning in a neural network. These merits have contributed to the successful utilization of ANFIS across various domains. Milan et al. [12] employed ANFIS and optimized ANFIS methods to predict the optimal exploitation of groundwater resources. Tran et al. [13] developed an ANFIS-based prediction model for assessing the processing performance of the thrust and surface roughness in biological composites. Sharifi et al. [14] used the ANFIS method to evaluate the intelligent performance of the agricultural surface water distribution system, yielding superior prediction results when compared to ANN and FIS methods.

Single prediction algorithms are usually simple in principle and easy to implement, and they can provide the time-varying characteristics of the prediction object from different angles, but they have limitations, such as incomplete information reflection and limited

scope of application. To overcome these limitations, hybrid models combine the strengths of multiple models to compensate for the shortcomings of a single model. The research of Armstrong [15] confirmed that the hybrid prediction model presents greater advantages in solving short-term prediction problems. In the specific context of dam monitoring [16–18], hydrological forecasting [19–22] and other hydraulic engineering fields, several hybrid prediction models have been proposed and have achieved higher precision results. For example, in a study by Luo et al. [23], a constrained PSO-SVR model was developed for centrifugal pumps. The research results demonstrated well-predicted performance under multiple operating conditions, compared to experimental results. Similarly, Huang et al. [24] proposed a hybrid neural network model that incorporates multiple geometrical parameters and operation conditions to predict the energy performance of centrifugal pumps. These hybrid prediction models show promising results in the field of pump energy performance prediction and can contribute to better decision making and optimization for pump operations in various industries and applications.

The pumping station is a complex system with different types of components and sources of vibration. Therefore, it is essential to assess the safety level using predicted vibration data. Safety early warning is a comprehensive research subject that involves multiple projects and levels and provides a higher level of safety prediction.

Compared with other engineering production fields, research on safety early warning in the field of hydraulic engineering had a later start. This delay can be attributed to the challenges in using analytical relations or mathematical models to describe the strong nonlinearity, fuzziness and complexity of the hydraulic system. In the early 1980s, the United States made a preliminary attempt to introduce risk analysis technology into the safe operation and maintenance of dams, which was achieved through the introduction of risk early warning theory. In 1984, the international dam conference further promoted the application of risk theory in dam management. Concurrently, the United States and Western Europe, together with other countries, improved the risk early warning technology and developed diverse safety early warning theories [25]. The research on safety early warning of hydraulic projects has been conducted using the risk early warning theory. Sang et al. [26] proposed an extended cloud model (ECM) combined with the extended analytic hierarchy process (EAHP) to assess the overall safety trend of dams and select a safety trend warning indicator. He et al. [27] proposed an integrated variable fuzzy evaluation model to evaluate the social and environmental impact of dam breaks. Yang et al. [28] presented a systematic approach for analyzing the law and early warning of vertical displacements in sluice clusters located in coastal soft soil.

Intelligent safety management technology is increasingly being applied to practical projects, and the concept of risk early warning has gained widespread recognition and attention. Previous safety early warning studies focused on reservoirs, dams and sluices in hydraulic projects, primarily from a risk analysis perspective, and they yielded favorable outcomes. Unfortunately, there has been a lack of attention paid to safety early warning systems for pumping stations and hydropower houses, resulting in a dearth of research in this area. Therefore, addressing these gaps in research is essential to guarantee the safe operation and improve the productivity of pumping stations and hydropower houses. The D-S evidence theory is a commonly used information fusion technology. D-S evidence theory introduces the probability distribution function, confidence function and likelihood function, lessening the reliance on traditional probability theory's prerequisites of prior probabilities, conditional probability and unified identification framework. The D-S evidence theory has wide application in fault detection [29] as well as in safety evaluation for the ocean environment [30] and the cloud platform [31]. In the field of hydraulic engineering, Chen et al. [32] enhanced the evidence distance measure method in D-S evidence theory by utilizing the belief Wasserstein-1 distance (BWD) and applied it to dam health diagnosis. Their findings indicate that the proposed method achieves significantly higher accuracy when compared to existing approaches. Xu et al. [33] proposed a D-S evidence theory based on neural networks for turbine fault diagnosis. They utilized the BP and

RBF networks to form the initial diagnosis layer and observed that the proposed method yields superior diagnostic results compared to single diagnosis methods. Recently, the D-S evidence method has been sparsely employed in the field of pumping stations.

In light of the demand for real-time data processing and analysis in the management system of pumping stations, this study aims to evaluate the level of vibration in the pumping station by developing a system for predicting vibrations and providing early warnings for safety purposes. To achieve this goal, a hybrid model is proposed, which combines the ARIMA single model, the ANFIS model and the whale optimization algorithm (WOA) to predict vibration responses. Furthermore, the D-S evidence theory is used to conduct safety early warning research. The prediction of the vibration trend of the effective stress on the blades is compared with that of a single prediction model and other models that employ different optimization algorithms. The vibration data collected will also serve as a source of data for the safety early warning system. In the early warning model, the fusion indicators for analyzing the vibration. The probability of each safety level is then quantitatively calculated and evaluated. This research could provide guidance for vibration control and attention of the pumping station project.

2. Calculation Method and Procedure

2.1. Autoregressive Integrated Moving Average Algorithm

The ARIMA model, proposed by Box and Jenkins [34], is a time-series analysis method commonly known as the Box–Jenkins model. The ARIMA model utilizes the historical information of a time series and employs a linear combination of predictions from multiple white noise processes. It can be applied to both stationary and non-stationary time series after differencing to achieve stability. The mathematical expression of the ARIMA model is as follows.

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$
(1)

where μ is a constant, ε_t is the white noise sequence, θ_i is the moving average coefficient, p is the order of autoregression and q is the order of moving average.

The ARIMA model is established for predicting the vibration response of pumping stations. The procedure is performed as follows.

- (1) The time-series data of vibration response are obtained and preprocessed. This may involve removing any outliers or missing data points, as well as normalizing the data.
- (2) The stationarity test is conducted after the data preprocessing step. If the test fails, differential processing shall be carried out until it passes the stationarity test.

The ADF-KPSS joint test is a statistical test commonly used to assess both stationarity and long memory in a time series. The ADF test controls for high-order sequence correlation by including lagged difference terms of the dependent variable in the regression equation. This test is used to determine if the time-series data exhibit a unit root, which implies non-stationarity. Assuming that the vibration response y_t follows an AR(p) process, the model used for the unit root test can be represented by Equations (2)–(4).

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^p \theta_i \Delta y_{t-1} + e_t \tag{2}$$

$$\Delta y_{t} = \mu + \rho y_{t-1} + \sum_{i=1}^{p} \theta_{i} \Delta y_{t-1} + e_{t}$$
(3)

$$\Delta y_{t} = \mu + \beta t + \rho y_{t-1} + \sum_{i=1}^{p} \theta_{i} \Delta y_{t-1} + e_{t}$$
(4)

where μ is the drift term and *t* is the time trend term. In the testing process, the test model is selected based on the characteristics of the sequence. If $\rho = 0$ in Equations (2)–(4), then the null hypothesis of a unit root exists; if ρ is significantly less than 0, the null hypothesis of a unit root in the test sequence is rejected.

The KPSS test, on the other hand, examines whether the data have a trend or exhibit long memory behavior. The null hypothesis of the KPSS test is that the sequence $\{y_t\}$ is stationary, and the alternative hypothesis is that the sequence $\{y_t\}$ is non-stationary. The principle is to remove the intercept term and trend term from the residual estimate sequence $\{\hat{e}_r\}$ and construct the *LM* statistic. The basis for the existence of a unit root in the original sequence is whether there is a unit root in the test $\{\hat{e}_r\}$.

$$S(t) = \sum_{r=1}^{t} \hat{e}_r \tag{5}$$

The construction of the KPSS statistic *LM* is as follows:

$$LM = \sum S(t)^{2} / (T^{2} f_{0})$$
(6)

where f_0 is the residual spectral density when f = 0, and $S(t)^2$ is a consistent estimate of the residual variance. The stationarity of the sequence can be determined by comparing with the critical value.

If the sequence rejects the ADF test but accepts the KPSS null hypothesis, the sequence is stationary. If the sequence simultaneously rejects the ADF and KPSS null hypotheses, the sequence may exhibit long memory and further testing is required.

(3) The autocorrelation and partial correlation coefficients of the series are calculated to determine the order of the ARIMA model. Information criteria, including Akaike's Information Criterion (*AIC*) and Bayesian Information Criterion (*BIC*), are adopted to select the optimal ARIMA model.

$$AIC = -2\log(L) + 2K \tag{7}$$

$$BIC = -2\log(L) + K\log(T) \tag{8}$$

where *L* is the likelihood of time series, *K* is the number of estimated parameters and *T* is the size of the time series. The model with the least *AIC* is the best model. *BIC* criterion penalizes the number of parameters more than *AIC*. The best model is selected similar to the *AIC* criterion by choosing the model with the lowest *BIC* value.

(4) A residual sequence independence test is conducted. The Durin–Watson test, also known as DW test, is used to test for first-order autocorrelation of residuals in regression analysis, especially in the case of time series. Assuming the residual is e_t the equation for the autocorrelation of each residual is $e_t = \rho e_{t-1} + V_t$. The null hypothesis for the test is $\rho = 0$, and the alternative hypothesis is $\rho \neq 0$. The test statistic *d* is shown in Equation (9).

$$d = \frac{\sum_{t=2}^{l} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$
(9)

Since *d* is approximately equal to 2(1 - p), the closer the value of this statistic is to 2, the better. If it is less than 1, it indicates the presence of autocorrelation in the residuals.

2.2. Adaptive Network-Based Fuzzy Inference System

ANFIS is a combined prediction model, which was proposed by Jang [35]. Combining fuzzy logic and the neural network organically, ANFIS has the decision-making judgment ability of a fuzzy system and the self-learning ability of a neural network. It automatically

generates if-then rules, learns from sample data and adapts the parameters of the neural network model. The front parameters are adjusted using the BP algorithm in the reverse transmission of ANFIS. The rear parameters are adjusted using the Least Square Method (LSM) in the forward transmission. The combined algorithm based on the BP algorithm and LSM improves the calculation efficiency and prediction accuracy of the ANFIS model.

Based on the Takagi–Sugeno model [36], the structure of two input–single output ANFIS is shown in Figure 1. Corresponding if–then rules are expressed as follows.

- (1) If x_1 is A_1 and x_2 is B_1 , then $y = p_1x_1 + q_1x_2 + r_1$;
- (2) If x_1 is A_2 and x_2 is B_2 , then $y = p_2x_1 + q_2x_2 + r_2$.



Figure 1. Two input-single output ANFIS structure.

The ANFIS network consists of the following layers: fuzzification layer, rule inference layer, normalization layer, defuzzification layer and output layer. The fuzzification layer transforms each precise input into several fuzzy subsets, each represented by a membership function indicating the degree of belonging. The nodes in the fuzzification layer are adaptive nodes, and their calculation formula is described in Equation (10).

$$O_{i}^{(1)} = \begin{cases} \mu_{A_{i}}(x_{1}) = \left[1 + \left(\frac{x_{1} - c_{1i}}{\delta_{1i}}\right)^{2b_{1i}}\right]^{-1} = e^{\frac{-0.5(x_{1} - c_{1i})^{2}}{\delta_{1i}^{2}}} \\ \mu_{B_{i}}(x_{2}) = \left[1 + \left(\frac{x_{2} - c_{2i}}{\delta_{2i}}\right)^{2b_{2i}}\right]^{-1} = e^{\frac{-0.5(x_{2} - c_{2i})^{2}}{\delta_{2i}^{2}}} \quad i = 1, 2 \end{cases}$$
(10)

where $O_i^{(1)}$ is the node output of the first fuzzification layer. x_i (i = 1, 2) is the precise input of node *i*. A_i (or B_i) is the fuzzy subset corresponding to x_i . μ_{Ai} and μ_{Bi} are the membership functions of A_i and B_i , respectively. { δ_i , b_i , c_i } are antecedent parameters, whose values are related to the shape of the membership function.

The rule inference layer multiplies the output of the membership function signals by the fuzzification layer to obtain the excitation intensity value of the if–then fuzzy rules. The node output $O_i^{(2)}$ is expressed in Equation (11).

$$O_i^{(2)} = \omega_i = \prod_{i=1}^2 O_i^{(1)} = \mu_{A_i}(x_1)\mu_{B_i}(x_2) \quad i = 1,2$$
(11)

The normalization layer is responsible for normalizing the excitation intensity value output by the rule-reasoning layer. Specifically, the percentage of the excitation intensity of the corresponding node is calculated, as well as the sum of all the excitation intensities. The output $O_i^{(3)}$ of the third normalization layer is described in Equation (12).

$$O_i^{(3)} = \overline{\omega}_1 = \omega_i / \sum_{i=1}^2 \omega_i i = 1,2$$
 (12)

The nodes of the defuzzification layer and the fuzzification layer are adaptive nodes. The role of the defuzzification layer is to convert fuzzy variables into precise variables. Based on if–then fuzzy rules, the normalized parameters are weighted and summed to obtain accurate output $O_i^{(4)}$.

$$O_i^{(4)} = \overline{\omega}_l f_i = \overline{\omega}_l (p_i x_1 + q_i x_2 + r_i)$$
(13)

where $\{p_i, q_i, r_i\}$ are the subsequent parameters, which will be constantly adjusted during the training process.

The output layer is used to sum all input signals, represented by fixed nodes marked with ' Σ '. The output $O_i^{(5)}$ is shown in Equation (14).

$$O_i^{(5)} = \sum_{2}^{i=1} \overline{\omega}_i f_i \tag{14}$$

In ANFIS, the generation of fuzzy variables usually adopts the unsupervised learning clustering analysis method. The samples are divided into different types of subspaces according to the similarity. There are generally three methods of generating fuzzy variables in ANFIS, namely, grid partition (GP), subtractive clustering (SC) and fuzzy C-means clustering (FCM).

(1) Grid partition algorithm

The GP algorithm is a clustering method that transforms data samples into grid cells. The data sample is divided into grid cells using parallel lines along the membership function axis. The correlation of each grid cell is then calculated and compared to the threshold of the data cluster to determine whether to merge it with the surrounding grid and form a data cluster, thus achieving the purpose of classification.

The GP clustering algorithm overcomes the limitations of other clustering algorithms that are sensitive to the shape and size of the cluster. It reduces model training time by connecting the subspaces divided based on the data dimension in a grid-based manner. However, GP clustering has poor scalability, and the accuracy of the GP algorithm is easily influenced by noisy sample data, resulting in relatively rough results.

(2) Subtractive clustering algorithm

The SC algorithm is a density-based clustering algorithm that was proposed by S. Chiu [37] in 1994. The SC algorithm assumes that any data point may be the cluster center. The probability of the data point as the cluster center is evaluated based on the data point density near each point. The data point with the highest density is selected as the cluster center, while data points with lower density are excluded. After the first cluster center is selected, the next cluster center is selected from the remaining data points using the same method. This process continues until the density near the data points is lower than the defined threshold.

Generally, it is assumed that all data points are located in a hypercube with a unit of 1, meaning that each one-dimensional coordinate of the data point is between 0 and 1. The density D_i of the data point x_i is defined as [38]:

$$D_i = \sum_{j=1}^n \exp\left(\frac{x_i - x_j}{\left(\frac{r_a}{2}\right)^2}\right)$$
(15)

where r_a represents the influence radius of the data point density range. Obviously, the more data points within the influence radius, the greater the density D_i , and the greater the probability that the data point will become the cluster center.

After calculating and comparing the density of all data points, the highest density of data points is selected as the first cluster center X_i , and its density is defined as D_{Xi} . The density of the remaining data points is then adjusted based on D_{Xi} as follows.

$$D_{i} = D_{i} - D_{X_{i}} \sum_{j=1}^{n} \exp\left(-\frac{x_{i} - x_{j}}{\left(\frac{r_{b}}{2}\right)^{2}}\right)$$
(16)

where r_b is constant, which is defined as $r_b = \eta r_a$. The inhibition factor η should be greater than 1 to prevent the distance between different cluster centers from being too close.

It is more suitable to use the SC algorithm to divide the input space when the number of input variables is greater than three. This approach results in fewer fuzzy rules compared to the adaptive grid method. It also provides a more reasonable division of the input space with reduced training time. Moreover, the fuzzy rules could be increased one by one, which prevents gaining over-fitting results, improves the generalization and accuracy ability of the model.

(3) Fuzzy C-means algorithm

The FCM clustering algorithm, originally proposed by J. C. Dunn [39] in 1974 and improved by J. Bezdek [40] in 1981, has established itself as a highly accurate and widely applicable method in many clustering algorithms [41,42]. The core of the FCM algorithm is to perform iterative calculations and update the cluster center point based on the minimum cost function.

FCM decomposes the sample data $\{x_1, x_2, ..., x_n\}$ into *k* fuzzy groups and determines the cluster center $\{c_1, c_2, ..., c_k\}$ for each fuzzy group based on the minimum cost function. The membership value of the *j*th data point x_j to the *i*th cluster center c_i is denoted as u_{ij} , and it ranges from 0 to 1. The sum of the entire membership matrix is 1 after data normalization, namely:

$$\sum_{j=1}^{k} u_{ij} = 1 \quad j = 1, 2, \cdots, n$$
(17)

The cost function of FCM is typically represented by Equation (18).

$$J(\mathbf{U}, c_1, \cdots, c_k) = \sum_{i=1}^k J_i = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m d_{ij}^2$$
(18)

where U is the membership matrix, c_i is the *i*th cluster center point, d_{ij} is the Euclidean distance from the *j*th data point x_j to the *i*th cluster center c_i and *m* is the weighted index and ranges in $[1, +\infty)$.

Lagrange multiplier λ_j (j = 1, 2, ..., n) is brought into Equation (18) to solve the necessary conditions for J to reach the minimum value.

$$J(\mathbf{U}, c_1, \dots, c_k, \lambda_1, \dots, \lambda_n) = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j \left(1 - \sum_{i=1}^k u_{ij} \right)$$
(19)

The expression for the cluster center c_i (Equation (20)) and membership degree u_{ij} (Equation (21)) is obtained by taking the derivative of cost function *J* with respect to c_i and u_{ij} , respectively. The FCM clustering algorithm iteratively solves the problem. The cost function *J* stops when it becomes less than the threshold or reaches the maximum number of iterations, resulting in the determination of the final clustering center *c* and membership matrix *U*.

$$c_{i} = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_{j}}{\sum_{j=1}^{n} u_{ij}^{m}}$$
(20)

$$=\frac{1}{\sum_{z=1}^{k} \left(\frac{d_{ij}}{d_{zj}}\right)^{\frac{2}{m-1}}}$$
(21)

2.3. Optimization Algorithms

WOA, proposed by S. Mirjalili and A. Lewis [43], is a metaheuristic optimization algorithm based on humpback whale hunting behavior. Humpback whales like to prey on fish and shrimps near the water surface in the form of circular contraction and spiral rise, as shown in Figure 2. In contrast to other traditional single algorithms, WOA offers several advantages, including a simple structure, easy implementation and high convergence accuracy. Despite being a relatively new optimization algorithm introduced in recent years, WOA has been widely applied in fault diagnosis [44,45] and other fields, with successful prediction outcomes.

 u_{ij}



Figure 2. Schematic diagram of hunting behavior of humpback whales.

A mathematical model is established to describe the predatory behavior of whales. Assuming that the current optimal candidate solution is target prey, the search agent will update the current position towards the target prey. The whale's prey encirclement behavior is expressed in Equations (22) and (23).

$$D = |\boldsymbol{C} \cdot \boldsymbol{X}^*(t) - \boldsymbol{X}(t)|$$
(22)

$$\mathbf{X}(t+1) = \mathbf{X}^*(t) - \mathbf{A} \cdot D \tag{23}$$

where *D* represents the distance between the search agent and the target prey. X and X^* are the current position and optimal position vector of the whale, respectively. t is the number of iterations. *A* and *C* are vector factors and are expressed as follows.

$$\mathbf{A} = 2a \cdot \mathbf{r}_a - a \tag{24}$$

$$C = 2 \cdot r_c \tag{25}$$

where *a* decreases linearly from 2 to 0 during iteration, and r_a and r_c are random vectors between 0 and 1.

Humpback whales surround their prey along the spiral path and emit bubbles. The whale's bubble net hunting strategy is shown as follows.

$$\mathbf{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \mathbf{X}^*(t)$$
(26)

where $D' = |\mathbf{X}^*(t) - \mathbf{X}(t)|$ is the line length from whale to prey, *b* is the shape parameter of the logarithmic spiral and *l* is a random number between -1 and 1.

Assuming that the probability of humpback whales using the ring contraction and spiral rise mechanisms to update the position is 50%, respectively, then:

$$\mathbf{X}(t+1) = \begin{cases} \mathbf{X}^{*}(t) - \mathbf{A} \cdot D & \text{if } p < 0.5\\ D' \cdot e^{bl} \cdot \cos(2\pi l) + \mathbf{X}^{*}(t) & \text{if } p \ge 0.5 \end{cases}$$
(27)

where p is the random number of selection probability in [0, 1].

Random prey search is required to update the position of the Humpback whales, as shown in Equations (28) and (29).

$$D = |\boldsymbol{C} \cdot \boldsymbol{X}_{\text{rand}}(t) - \boldsymbol{X}(t)|$$
(28)

$$\mathbf{X}(t+1) = \mathbf{X}_{\text{rand}} \ (t) - \mathbf{A} \cdot \mathbf{D}$$
⁽²⁹⁾

where $X_{rand}(t)$ is a randomly selected search agent position vector.

Two additional algorithms are adopted to optimize the weight coefficients in the hybrid models, which are Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). GA is an intelligent optimization algorithm proposed based on evolutionary theory and genetic principles. It simulates the principles of survival of the fittest and natural selection by designing a specific population to survive in a particular environment, using three operations, selection, crossover and mutation, to obtain the optimal individual and, ultimately, the optimal solution to a problem. Selection, crossover and mutation are the core operations of genetic algorithms.

The PSO algorithm originated from the study of bird flock predation behavior. It is a global optimization algorithm that utilizes cooperation and information sharing among individuals in a population to find the optimal solution with good global search ability. The updating formulas for the velocity and position of the particles in the population are given by:

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 \left(pbest_{id}^k - X_{id}^k \right) + c_2 r_2 \left(gbest_{id}^k - X_{id}^k \right)$$
(30)

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1}$$
(31)

where V_{id}^k and X_{id}^k represent the current velocity and position of particle *i*, *pbest*_{id}^k and *gbest*_{id}^k represent the individual best and global best, V_{id}^{k+1} and X_{id}^{k+1} represent the newly updated velocity and position of the particle, ω is the inertia weight, c_1 and c_2 are non-negative constant learning factors and r_1 and r_2 are random numbers between 0 and 1.

2.4. Calculation Procedure of ARIMA-ANFIS-WOA Hybrid Model and Evaluation Criterion

There are certain limitations in the traditional single-time series prediction model in terms of prediction accuracy. Based on the ARIMA single prediction model, ANFIS combined prediction model and WOA, the ARIMA–ANFIS–WOA hybrid prediction model is proposed and applied to predict the vibration response of the pumping station.

The hybrid prediction theory allocates appropriate weight coefficients to different prediction methods for the same prediction problem. The final prediction results of the hybrid model are achieved by superimposing the results of the single prediction model. For the ARIMA–ANFIS–WOA hybrid prediction model, $Y_t = \{y_t\}$ is assumed as the actual time-series data. \hat{Y}_t is the final prediction result of the hybrid prediction model. \hat{Y}_t is expressed as follows.

$$\hat{Y}_t = w_1 f_{1t} + w_2 f_{2t}$$
 $t = 1, 2, \cdots, m$ (32)

$$w_1 + w_2 = 1 \tag{33}$$

where f_{1t} and f_{2t} represent the prediction results of the first and second prediction model at time *t*, respectively. *m* is the maximum prediction time. w_1 and w_2 are the weight coefficients of the first and second prediction results, respectively.

The calculation process of the ARIMA–ANFIS–WOA hybrid prediction model is shown in Figure 3. The specific implementation steps are described as follows.



Figure 3. Calculation procedure of ANFIS-ARIMA-WOA hybrid prediction model.

Step 1: Select and normalize the time series of the vibration response of the pumping station. Since the time history response of the effective stress of the blades exhibits both linear and nonlinear characteristics, we choose the effective stress curve calculated from numerical simulation for the prediction study. The dataset is divided into training set and test set.

Step 2: Train and predict the time series at each time point using the ANFIS model and the ARIMA model, respectively. The prediction results at this step serve as intermediate results during the prediction process of the hybrid model. In the prediction study based on the ANFIS model, we employ three different algorithms, namely GP, SC and FCM, to generate the fuzzy structure, and we compare the prediction results to determine the optimal ANFIS results.

Step 3: WOA is used to obtain the weight coefficients of intermediate results. Firstly, WOA is initialized to calculate the individual fitness of whales. The optimal position is recorded, and parameters *a*, *A* and *C* are updated. Then, the probability *p* value discrimination is conducted. If $p \ge 0.5$, spiral motion will be performed according to Equation (26). Otherwise, if |A| > 1, random search will be performed according to Equations (28) and (29). If $|A| \le 1$, prey surrounding will be conducted according to Equations (22) and (23). Next, the fitness *f* is calculated and compared with the optimal fitness f_{best} . If $f < f_{best}$, the position is updated, and the next iteration is proceeded until the optimal solution is achieved. Otherwise, the next iteration proceeds without updating the position.

Step 4: Combine the intermediate prediction results of the ANFIS model and ARIMA model with the optimal weight coefficient obtained from the WOA algorithm to establish the ANFIS-ARIMA-WOA hybrid prediction model for predicting the vibration response of the pumping station.

Step 5: Verify the prediction accuracy of the hybrid model. If the prediction results of the ARIMA-ANFIS-WOA model satisfy the accuracy requirements, the final prediction results are generated, and a corresponding performance evaluation is conducted. Otherwise, return to step 3 to optimize the weight coefficient until the desired prediction accuracy is attained.

The model's prediction accuracy is evaluated using root mean square error (*RMSE*), mean absolute error (*MAE*), standard deviation (*SD*) and correlation coefficient (*R*). The calculation formula is provided in Equations (34)–(37) [46–48]. *RMSE* and *SD* assess the accuracy and stability of the model, respectively. Consequently, smaller values of *MAE*, *RMSE* and *SD* indicate better prediction results, while a larger *R* value signifies greater predictive ability.

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{N}}$$
(34)

$$MAE = \frac{1}{N} \sum_{k=1}^{N} |y_k - \hat{y}_k|$$
(35)

$$SD = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y_k - \mu)^2}$$
 (36)

$$R = \frac{\operatorname{cov}(y_t, y_t)}{\sqrt{V(y_t)}\sqrt{V(y_t)}}$$
(37)

where *N* is the number of sample data, y_k is the predicted value of the model, \hat{y}_k is the mean value of the predicted value and μ is the mean value of the sample data.

2.5. D-S Evidence Theory

The D-S evidence theory, proposed by A.P. Dempster and G. Shafer, is a decisionmaking method used to address uncertain problems. This theory has found widespread application in various fields, such as data fusion, risk assessment and modern decision making, where a rigorous reasoning process and robust fusion of multiple sources of information are required. In essence, the D-S evidence theory involves establishing an identification framework that represents all possible outcomes of a decision problem. Subsequently, subsets within this framework are evaluated, generating a trust function. The trust function assigns a truth value to propositions recognized by the framework. In cases where multiple subsets exist, the trust functions obtained from each subset can be combined using Dempster's composition rule, thereby yielding the evidence synthesis result for each subset.

The accuracy of fusion results relies heavily on the basic probability distribution function of the evidence. Since evidence exhibits characteristics of discrete probability distribution, the basic probability value for small probability events is set at 5%. The solution of the basic probability value $m_i(A_j)$ of the evidence m_i for the proposition A_j could be referred to Equation (38).

$$m_i(A_j) = \begin{cases} 0.05, x_i < a_{ij1}, x_i > a_{ij2} \\ [x_i - (a_{ij1} - u_{ij})] / (2u_{ij}), a_{ij1} \le x_i \le u_{ij} \\ (a_{ij2} + u_{ij} - x_i) / (2u_{ij}), u_{ij} < x_i \le a_{ij2} \end{cases}$$
(38)

where x_i represents the monitoring value of the evidence m_i . a_{ij1} and a_{ij2} denote the lower and upper limits of the range for proposition A_j , respectively. These limits are determined based on threshold parameter results. The average value of the range of proposition A_j is represented by u_{ij} , as shown in Equation (39).

$$u_{ij} = \left(a_{ij1} + a_{ij2}\right)/2 \tag{39}$$

Normalization is performed for $m_i(A_i)$ obtained from Equation (38).

$$m_i'(A_j) = \frac{1 - 0.05z}{\sum\limits_{j=1}^n m_i(A_j)} m_i(A_j), n = 1, 2, \dots, s, u$$
(40)

where z is the number of small probability propositions in the evidence. $m_i'(A_j)$ is the normalized probability distribution value.

Figure 4 illustrates the safety warning process for the vibration response of the pumping station based on the D-S evidence theory. The process begins by defining evaluation indexes and criteria, establishing an evaluation system for the vibration response of the pumping station. Specifically, displacement, velocity, acceleration and stress are employed as evaluation indexes. Subsequently, a D-S evidence set is constructed using vibration response prediction data for the pumping station. The threshold value for the vibration response is determined based on the vibration control standard of the pumping station. Basic probability distribution is then conducted on the evidence set to create an early safety warning identification framework for vibration. Finally, the identification results of each piece of evidence are integrated and reasoned using the D-S evidence theory. This involves assessing basic probability, trust measure, likelihood measure and other indicators. The evidence fusion results of the operation state of the pumping station units are evaluated.



Figure 4. Flow chart of pumping station vibration response safety early warning based on D-S evidence theory.

3. Prediction Results of the Vibration Responses and Evaluation

3.1. Data Source and Collection

The data for the prediction study are derived from the numerical simulation results of the pumping station. As shown in Figure 5, a large-shaft tubular pumping station model is established, including the pump unit, the concrete structure and the fluid inside the flow channel. A two-way iterative FSI method is employed to explore the vibration features and provide a data source for prediction study. The parameters of the pump are as follows: impeller diameter $D_2 = 3.25m$, single unit design flow $Q = 30m^3/s$, lift H = 0.96m, rated speed n = 105rpm and total installed capacity 6250 kW.

Compared with the time series of other variables, including displacement and acceleration, the time series of the effective stress express both linear and nonlinear features. Therefore, the effective stress results are chosen as the data source. The span of the data is from 6.0 to 10.0 s with a total sample of 800. Thus, 640 sets of data ranging from 6.0 to 9.2 s are taken as training samples, which accounts for 80% of the total sample. The remaining datasets ranging from 9.2 to 10.0 s are taken as test samples, accounting for 20% of the total sample.



Figure 5. Numerical model of the pumping station. (a) Structure grid; (b) fluid grid.

3.2. Prediction Results of ARIMA Model

The ARIMA prediction model is established for vibration response prediction of the pumping station. The ADF and KPSS tests are carried out, and the results show that the significance test level p value is 0.0614, greater than the confidence level value 0.05, which indicates that the sequence is unstable and fails to pass the significance test. The first-order difference operation is carried out and p = 0.001 is obtained, which passes the significance test.

The ARMA (p, q) model can be identified by observing the autocorrelation function (ACF) and partial autocorrelation function (PACF) diagram of the sequence, though it is not the final basis for order determination. The ACF and PACF after the first-order difference are shown in Figure 6. The red dots are the values of ACF and PACF, and the area between blue lines corresponds to 95% confidence interval. It shows that ACF and PACF gradually tend to zero after the lag of zero order, presenting tailing feature.



Figure 6. (a) ACF and (b) PACF diagram of ARIMA model.

In order to determine the order of the ARIMA model more accurately, *AIC* and *BIC* criteria are adopted to select the order violently. In this research, the minimum sum of *AIC* and *BIC* information is used as the criterion to determine the model order. The effective stress prediction model is finally determined as ARMA (7, 4). The residual sequence is basically a standard normal distribution. The DW coefficient is 1.9959, very close to 2,

which indicates no first-order correlation in the residual sequence. Therefore, the residual sequence is considered as a white noise sequence and passes the residual test. The effective stress prediction curve based on the ARIMA model is shown in Figure 7, where *E* is the absolute error between the original value and the prediction. The effective stress curve suddenly changes at t = 9.590s and t = 9.675s, and the prediction error becomes relatively large. Since the ARIMA principle is based on moving averages and autoregression, the predicted results are generally close to the historical average. Therefore, the prediction value fits well at other time points with good prediction effects.



Figure 7. Prediction curve of effective stress based on ARIMA model.

3.3. Prediction Results of ANFIS Model

The configuration of parameters in the prediction model can significantly impact the prediction results. For the GP–ANFIS model, the key parameters are the time lag, the number and type of membership functions. There is no standardized time lag for input variables. In this research, the time lag value τ is set within a range of $1\Delta t \sim 6\Delta t$; specifically, the vibration response sequences of $t - 1\Delta t$, $t - 2\Delta t$, $t - 3\Delta t$, $t - 4\Delta t$, $t - 5\Delta t$ and $t - 6\Delta t$ periods are taken as input. The vibration response of the prediction period $1\Delta t$ is taken as the output. The number of membership functions is set to 2, 3, 4, 5 and 6. Membership function types include Triangular, Bell, Trapezoid and Gaussian. In the SC–ANFIS model, the influence radius (*IR*) is an important parameter, and its value is set within a range of 0.20~0.90 in the current research. In the FCM–ANFIS model, the weighted exponent *m* ranging from 1 to 9 is significant for the prediction results.

GA is employed to optimize the key parameters for predicting the effective stress vibration trend. Preset amounts of parameters play a crucial role in determining the behavior and performance of the GA, which typically include parameters, such as the iteration number, population size, crossover rate and mutation rate. In the current research, each parameter is tested with three different values to find the optimal value. The grid search method is employed to determine the optimal value. The parameters in GA are set as follows: 100 iterations, an initial population of 200, crossover probability of 0.90 and mutation probability of 0.01. The parameter optimization process curve of the ANFIS model based on GA is shown in Figure 8. As the number of iterations increases, RMSE gradually decreases and stabilizes at 55 iterations. The convergence curve of the GP– ANFIS model has a larger slope compared to the SC-ANFIS model. The FCM-ANFIS model exhibits fast convergence speed, high accuracy and good prediction performance. FCM is an unsupervised fuzzy clustering method that integrates the essence of fuzzy theory. Compared to the poor clustering scalability of GP and inflexibility of SC, the FCM algorithm provides more flexible clustering results without human intervention in the implementation process. Therefore, for nonlinear vibration data with irregular effective stresses, FCM-ANFIS achieves high prediction accuracy and efficiency.



Figure 8. Parameter optimization process curve of ANFIS model based on different fuzzy structures.

The optimal parameter settings for the prediction models, namely GP-ANFIS, SC-ANFIS and FCM-ANFIS, are provided in Table 1. For the blade effective stress prediction, the optimal input variable parameter settings for the GP-ANFIS model consist of a $4\Delta t$ time lag and two Gaussian membership functions (refer to Figure 9). In the case of the SC-ANFIS model, the optimal influence radius (*IR*) value is determined to be 0.23. As for the FCM-ANFIS model, the optimal weighted exponent is set to *m* = 3.83. The correlation coefficient (*R*) fitting result is depicted in Figure 10, demonstrating a good overall fitting effect. Notably, the FCM-ANFIS model achieves a maximum *R* value of 0.9896. Likewise, Figure 11 presents the prediction results of the effective stress for different fuzzy structures of the ANFIS model. By considering the obtained *RMSE* values during the parameter optimization process (shown in Figure 8), it is evident that the FCM-ANFIS model outperforms the GP-ANFIS and SC-ANFIS prediction models in terms of prediction accuracy.

Prediction Model	Parameter Optin	Optimal Parameter Setting		
	Time lag $ au$	$\{1\Delta t, 2\Delta t, 3\Delta t, 4\Delta t, 5\Delta t, 6\Delta t\}$	$4\Delta t$	
GP-ANFIS model	Membership function number	{2, 3, 4, 5, 6}	2	
	Membership function type	{Triangular, Bell, Trapezoid, Gaussian}	Gaussian	
SC-ANFIS model	Influence radius IR	[0.20, 0.90]	0.2272	
FCM-ANFIS model	Weighted exponent <i>m</i>	[1, 9]	3.8268	





Figure 9. Gaussian membership function of GP-ANFIS model with optimal parameter setting.



Figure 10. Fitting results of the correlation coefficient *R* based on (**a**) GP–ANFIS model; (**b**) SC–ANFIS model; (**c**) FCM–ANFIS model.



Figure 11. Prediction results of the effective stress based on different fuzzy structures of ANFIS model.

3.4. Prediction Results of ARIMA-FCM-ANFIS-WOA Hybrid Model

The FCM clustering method is adopted to generate the fuzzy structure in the ANFIS model. Table 2 shows the parameter settings in the FCM–ANFIS model and ARIMA–FCM–ANFIS–WOA model. The input membership function of the FCM–ANFIS model is Gaussian, the output membership function is linear, the fuzzy structure is Takagi–Sugeno, the number of fuzzy rules is 10, the maximum number of iterations is 1000, the initial time step is 0.01 and the time decline rate and growth rate are 0.9 and 1.1, respectively. In the ARIMA–FCM–ANFIS–WOA hybrid model, WOA is set to have 100 iterations and 100 whales.

Prediction Model	Parameter	Value	
FCM-ANFIS	Input membership function type	Gaussian	
	Output membership function type	Linear	
	Fuzzy structure	Takagi-Sugeno	
	Fuzzy rule number	10	
	Maximum number of epochs	1000	
	Initial time step	0.01	
	Time step reduction rate/growth rate	0.9/1.1	
ARIMA-FCM-ANFIS-WOA	Number of iterations	100	
	Number of whales	100	

Table 2. Parameter settings for prediction models.

Figure 12 shows the effective stress prediction curve based on the ARIMA–FCM– ANFIS–WOA model. The overall fitting degree of the hybrid model is relatively high. The hybrid model improved the prediction performance near the curve mutation, reducing the maximum absolute error to less than 0.5 MPa. The hybrid model reduces the risk of misjudgment from a single prediction model and yields more accurate and reliable evaluation results by combining the predictions from multiple time-series models using weight coefficients. The ARIMA-ANFIS hybrid model can handle different types of data as it leverages the linear data processing capability of ARIMA and the nonlinear data processing capability of ANFIS. The weight coefficient of the ARIMA-ANFIS method is optimized using the metaheuristic optimization algorithm, WOA, during the parameter optimization process. The proposed ARIMA-ANFIS hybrid method can capture both linear and nonlinear features of time series, overcoming the limitations of a single method that cannot capture all information of time series.



Figure 12. Prediction curve of effective stress based on ARIMA-FCM-ANFIS-WOA model.

3.5. Prediction Evaluation of Different Models

Two commonly used optimization algorithms, namely GA and PSO, are employed to optimize the weight coefficients in the hybrid model. Table 3 shows the prediction results obtained from different prediction models, including ARIMA, GP–ANFIS, SC–ANFIS, FCM–ANFIS, ARIMA–FCM–ANFIS–GA, ARIMA–FCM–ANFIS–PSO and ARIMA–FCM–ANFIS–WOA. The prediction results of the test set show that the ARIMA–FCM–ANFIS–WOA model exhibits relatively smaller values for *RMSE*, *MAE* and *SD* compared with the combined ANFIS model. Furthermore, the correlation coefficient *R* reaches an impressive value of 0.9915. The precision achieved by the hybrid model surpasses that of both the single ARIMA model and the combined ANFIS model, which shows that the hybrid model performs more effectively in reducing the probability of miscalculation and bias risk of a single model. Additionally, the calculation results derived from GA, PSO and WOA in the hybrid model are similar. Overall, the ARIMA–ANFIS–WOA model demonstrates

Training Set Test Set Model RMSE (MPa) MAE (MPa) SD (MPa) R RMSE (MPa) MAE (MPa) SD (MPa) R 0.1465 0.9263 0.0850 0.0852 0.9831 ARIMA 0.1465 0.0215 0.0072 0.0900 0.9816 0.0714 0.9841 0.0081 0.0901 0.0709 0.0049 GP-ANFIS SC-ANFIS 0.1015 0.0103 0.1015 0.9748 0.0729 0.0055 0.0727 0.9870 FCM-ANFIS 0.0880 0.9811 0.0707 0.9896 0.0077 0.0881 0.0047 0.0708 ARIMA-FCM-ANFIS-GA 0.1017 0.0103 0.1017 0.9766 0.0666 0.0044 0.0668 0.9914 ARIMA-FCM-ANFIS-PSO 0.1018 0.0104 0.1019 0.9761 0.0698 0.0049 0.0699 0 9898 0.9765 0.9915 ARIMA-FCM-ANFIS-WOA 0.1021 0.0104 0.1022 0.0665 0.0044 0.0667

a higher prediction accuracy in comparison to the hybrid model that integrates GA and PSO algorithms.

Table 3. Prediction results of the effective stress of the blades usin	ng different	prediction models.
--	--------------	--------------------

3.6. Vibration Prediction Results Based on ARIMA-FCM-ANFIS-WOA Hybrid Model

In order to comprehensively understand the operational status of the pumping station and provide a data source for the safety warning study, predictive research on the vibration responses of the station is conducted. The ARIMA-FCM-ANFIS-WOA hybrid prediction model is utilized to predict the vibration trends of displacement, velocity, acceleration and the first principal stress of the extreme point of the concrete structure in the pumping house. Figure 13 displays the original, training and predicted values of the corresponding vibration responses of the concrete structure, where *E* represents the absolute error between the original value and the fitted or predicted value. For the training phase, 640 datasets between 6.0 and 9.2 s are selected, accounting for 80% of the total samples. Additionally, 160 datasets between 9.2 and 10.0 s are chosen as test samples. The results indicate that the hybrid model, incorporating WOA, exhibits excellent prediction performance in vibration curves characterized by strong stationarity and regularity. These vibration prediction results serve as the foundational data for the safety evaluation study in the subsequent section.



Figure 13. Prediction results of the concrete structure of pumping station (ARIMA–FCM–ANFIS–WOA model). (a) Displacement; (b) velocity; (c) acceleration; (d) the first principal stress.

4. The Safety Early warning Study of the Pumping Station Based on the D-S Evidence Theory

Structural safety prediction is based on the variation pattern of vibration response. It estimates the changing trends of historical monitoring data, providing reasonable references for pumping station vibration response status and future trends to maintenance personnel. However, it usually cannot evaluate the goodness or badness of these variable changes in terms of value significance. On the other hand, warning utilizes useful information from the abundant data in pumping station operation to define safety warning indicators and evaluation criteria. It constructs a safety warning system for the pumping station, evaluates the predictive values from a value perspective and provides an interval for objective safety judgment and decision making by decision makers. Therefore, safety prediction provides a data foundation for safety warning, while safety warning represents a higher-level safety prediction that can support decision making and operations.

Considering the pumping station's complex structural components, multiple vibration sources and uncertainty of measured data, this study proposes a safety early warning model based on D-S evidence theory for the pumping station. The evidence set is established based on the vibration response prediction results of the pumping station, specifically focusing on the responses of the extreme points. The vibration data are analyzed using the D-S information fusion method, with vibration displacement, velocity, acceleration and stress being chosen as the fusion indicators. The operational status of the pumping station is evaluated, and early warnings are issued accordingly.

4.1. Evaluation Criteria for Vibration Response of the Pumping Station

According to the vibration control standard of the pumping station, the maximum permissible vibration displacement for concrete is 0.20 mm, the maximum permissible vibration acceleration is 1.0 m/s². The stress control value for the concrete structure is 17.5 MPa, and for the metal structure, it is 175 MPa. Based on relevant literature [49], the criteria for evaluating the vibration response of pumping stations using D-S evidence theory are defined and presented in Table 4. The limit values for extremely unsafe level IV, unsafe level III, relatively safe level II and safe level I are 90%, 80~90%, 70~80% and 70% of the allowable value, respectively.

. .	Displacement (µm)	Velocity (mm/s)	• • • • • • • • • • • • • • • • • • • •	Stress (MPa)		
Level			Acceleration (m/s ²)	Concrete Structure	Metal Structure	
Level I	<140	<3.5	<0.7	<12.3	<122.5	
Level II	140~160	3.5~4.0	0.7~0.8	12.3~14.0	122.5~140.0	
Level III	160~180	$4.0 \sim 4.5$	0.8~0.9	14.0~15.8	140.0~157.5	
Level IV	>180	>4.5	0.9	15.8	157.5	

Table 4. Evaluation criterion of the vibration responses of pumping station based on D-S evidence theory.

4.2. Identification Framework for Vibration Safety Warning of Pumping Station

The D-S evidence theory combines multiple pieces of evidence to reduce system uncertainty and determines to which subset of Θ an event belongs. Its essence lies in synthesizing the basic probability distribution function for multiple pieces of evidence. In this research, the identification framework for vibration safety includes the subsets A_1 , A_2 , A_3 , A_4 , A_5 and A_u within the set Θ . The evidence set Θ consists of L1, L2, L3, L4 and L5, which correspond to extreme response point data for displacement, velocity, acceleration, first principal stress and effective stress, respectively. The basic probability distribution function, m_1 , m_2 , m_3 , m_4 and m_5 , represents the supporting probability set for each level within Θ . Figure 14 depicts the D-S evidence theory matrix, where each line represents the support probability of the corresponding evidence for different operation levels of the pumping station, with a sum value of 1. Column *j* indicates the support probability for a

Θ	m	A_1	A_2	A_3	A_4	A_s	A_u		The sum of
L1	m_1	$m_1(A_1)$	$m_1(A_2)$	$m_1(A_3)$	$m_1(A_4)$	$m_1(A_s)$	$m_1(A_u)$		the basic
L2	<i>m</i> ₂	$m_2(A_1)$	$m_2(A_2)$	$m_2(A_3)$	$m_2(A_4)$	$m_2(A_s)$	$m_2(A_u)$	\Rightarrow	distribution
L3	<i>m</i> ₃	$m_3(A_1)$	$m_3(A_2)$	$m_3(A_3)$	$m_3(A_4)$	$m_3(A_s)$	$m_3(A_u)$		function of measuring
L4	<i>m</i> ₄	$m_4(A_1)$	$m_4(A_2)$	$m_4(A_3)$	$m_4(A_4)$	$m_4(A_s)$	$m_4(A_u)$		point L2 is
L5	<i>m</i> ₅	$m_5(A_1)$	$m_5(A_2)$	$m_5(A_3)$	$m_5(A_4)$	$m_5(A_s)$	$m_5(A_u)$		
Si	Support probability of all evidences classified as A_2								

specific level of pumping station operation. A high value indicates a high probability for this level.

Figure 14. Schematic diagram of D-S evidence theory matrix.

As depicted in Figure 13, the maximum vibration amplitudes of displacement, velocity, acceleration and the first principal stress of the concrete structure in the pumping station are $1.75 \,\mu\text{m}$, $0.04 \,\text{mm/s}$, $4.15 \,\text{mm/s}^2$ and $0.13 \,\text{MPa}$, respectively. The effective stress of the blade is 24.80 MPa. The basic probability distribution value of each piece of evidence is calculated and normalized according to Equations (38) and (40), as shown in Table 5.

Table 5. Basic probability table showing data fusion of monitoring points in the pumping station.

т	A_1	A_2	A_3	A_4	A_s	A_u
m_1	0.40	0.05	0.05	0.05	0.40	0.05
m_2	0.40	0.05	0.05	0.05	0.40	0.05
m_3	0.40	0.05	0.05	0.05	0.40	0.05
m_{41}	0.40	0.05	0.05	0.05	0.40	0.05
m ₄₂	0.41	0.05	0.05	0.05	0.39	0.05

4.3. Multi-Source Information Fusion Results

Information fusion research is conducted to determine the operational status of the pumping station. This process involves two levels of fusion: data-level fusion and decision-level fusion. In data-level fusion, each evaluation index datum represents evidence of the pumping station's safety. By fusing the evaluation index data of the same type, comprehensive evidence for that type can be obtained. In decision-level fusion, the fusion results from data-level fusion, which are obtained from different types of evaluation indicators, serve as evidence for the overall safety of the pumping station. This evidence is further fused to obtain the final results, which represent the comprehensive evaluation of the pumping station's safety.

The fused basic probability distribution after the fusion of m_{41} and m_{42} is $m_4\{A_1, A_2, A_3, A_4, A_s, A_u\} = \{0.6919, 0.0600, 0.0107, 0.0107, 0.2230, 0.0036\}$. The vibration data of measurement points L1, L2, and L3, as well as the stress data after the first-level fusion, were then subjected to second-level fusion. The final result of the information fusion is $M\{A_1, A_2, A_3, A_4, A_s, A_u\} = \{0.9462, 0.0240, 0.0000, 0.0000, 0.0297, 0.0000\}$, the belief measure $Bel\{A_1, A_2, A_3, A_4, A_s, A_u\} = \{0.9462, 0.0240, 0.0000, 0.0000, 1.0000, 0.0000\}$ and the plausibility measure $Pl\{A_1, A_2, A_3, A_4, A_s, A_u\} = \{0.9759, 0.0537, 0.0000, 0.0000, 1.0000, 0.0000\}$.



Table 6. Basic probability calculation table of monitoring data after D-S evidence fusion.

Figure 15. Support probability distribution of data fusion of monitoring point of the pumping station at (**a**) the original status; (**b**) the first-level fusion; (**c**) the second-level fusion.

Bel(*A*) represents the degree of trust in the proposition *A* being true, while *Pl*(*A*) represents the degree of not opposing the proposition *A*. Table 6 shows that *Bel*(A_1) = 0.9462, indicating that the trust interval for the safe operation status of the pumping station is [0.9462, 1.0000]. It demonstrates that the pumping station's safety status under design operating conditions is very good. Based on the upper and lower limits of the trust interval, the uncertainty interval is only 0.0538, indirectly proving the high reliability of the D-S evidence theory for the evaluation of the pumping station's safety status. *Bel*(A_s) = *Pl*(A_s) = 1, meaning that the supporting evidence interval for the pumping station's safety warning model reaches 1, indicating that the pumping station's operating status is safe. Considering the complexity and fuzziness of the safety influencing factors of the pumping station, the D-S evidence method integrates different types of evidence information, including the displacement, velocity, acceleration and stress indicators, of the pumping station's vibration response. It quantitatively displays the probabilities and degrees of trust of each proposition, providing reliable references for the evaluation and decision making of the pumping station's operation status.

5. Conclusions

This research focuses on methods of prediction and safety early warning for vibration responses of the pumping station. Due to the limitations of single model prediction, this research proposes a hybrid prediction method based on ARIMA–ANFIS–WOA for predicting vibration responses in pumping stations. The performance of the developed models was studied based on the effective stress vibration data of the blades. The D-S evidence theory was employed to establish an early warning model for the vibration response of the pumping station, allowing for quantitative evaluation of its operation status. The main conclusions are as follows:

- (1) Vibration prediction research was performed using the ARIMA model. The model order was determined based on the minimum sum of *AIC* and *BIC* information. The prediction results for effective stress indicate a good fit for most time points, but there are cases where the curve abruptly changes, leading to relatively large prediction errors.
- (2) The prediction study involved using various fuzzy structures for the combined ANFIS prediction model. GA was employed to optimize the important model parameters, including the time lag, the number and type of the membership function in the

GP–ANFIS model, the influence radius in the SC–ANFIS model and the weighted exponent in the FCM–ANFIS model. The prediction results demonstrate that the FCM–ANFIS model has better accuracy and efficiency compared to the GP–ANFIS model and SC–ANFIS model.

- (3) In the research on vibration prediction using the hybrid model, the weight coefficients were derived to integrate intermediate results using the WOA algorithm. The research findings indicate that the ARIMA–FCM–ANFIS–WOA model has higher prediction accuracy compared to the hybrid model using GA and PSO algorithms. The hybrid model exhibits higher accuracy than the single model ARIMA and the combined model ANFIS. The vibration responses of the concrete structure and metal unit comply with the vibration standard and do not exceed the specified amplitude in the vibration specifications.
- (4) A safety early warning model was developed using the D-S evidence theory to assess the safety of the pumping station. Displacement, velocity, acceleration and stress indicators were selected as the fusion indicators for analyzing the vibration data. The results indicate a confidence interval of [0.9462, 1.0000], suggesting excellent pump operation status in the design condition. The D-S evidence method provides a quantitative display of the probability and confidence of each proposition, making it highly credible for evaluating and making decisions regarding the operating state of the pumping station.
- (5) The proposed framework for predicting and providing warning for the vibration responses can be applied to the real-time operation and management of pumping stations. The prediction model exhibits strong generalizability and computational efficiency. However, there is room for improvement in predictive performance, particularly in areas with significant data fluctuations. Comparing the proposed models with existing benchmark datasets or alternative methods will also be considered in our future research to support the validity and significance of the predictions. In addition, the safe operation of pumping stations is influenced by numerous factors that interact in a complex manner. The models developed for pumping station safety warnings in this research do not account for all these factors and are relatively simplistic. Further research should be conducted to incorporate multiple factors and develop pumping station safety warning models that rely on more accurate predictions.

Author Contributions: Conceptualization, S.W., L.Z. and G.Y.; data curation, S.W. and G.Y.; formal analysis, S.W. and G.Y.; funding acquisition, S.W. and L.Z.; investigation, S.W.; supervision, L.Z.; validation, S.W. and G.Y.; writing—original draft, S.W.; writing—review and editing, S.W., L.Z. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Jiangsu Provincial Transportation Technology Plan Project (Grant No. 2020QD28), Jiangsu Funding Program for Excellent Postdoctoral Talent (Grant No. 2022ZB188) and the National Key Research and Development Plan of China (Grant No. 2017YFC0404903).

Data Availability Statement: The data are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Mirjavadi, S.S.; Afshari, B.M.; Shafiei, N.; Rabby, S.; Kazemi, M. Effect of temperature and porosity on the vibration behavior of two-dimensional functionally graded micro-scale Timoshenko beam. *J. Vib. Control* **2018**, *24*, 4211–4225. [CrossRef]
- Forsat, M. Investigating nonlinear vibrations of higher-order hyper-elastic beams using the Hamiltonian method. *Acta Mech.* 2020, 231, 125–138. [CrossRef]
- 3. Lian, J.; Zhang, H.; Wang, H. Prediction of vibration response of powerhouse structures by means of artificial neural network method. *J. Hydraul. Eng.* **2007**, *3*, 361–364. (In Chinese)
- 4. Miao, Z.; Ma, Z.; Wang, Y.; Zhi, B. Vibration Response Prediction of Powerhouse Structure Based on Artificial Neural Network. *Water Resour. Power* **2010**, *28*, 88–90. (In Chinese)

- 5. Xu, G.; Han, W.; Wang, H.; Zhang, H. Study on vibration responses of powerhouse structures based on FOA-GRNN. J. Hydroelectr. Eng. 2014, 33, 187–191. (In Chinese)
- 6. Xu, G.; Han, W.; Wang, H. Vibration response prediction of a powerhouse structure based on SSPSO-GRNN. *J. Vib. Shock* **2015**, *34*, 104–109. (In Chinese)
- Wang, H.; Mao, L.; Lian, J. Structural vibration prediction for a hydropower house based on RVM method. J. Vib. Shock 2015, 34, 23–27. (In Chinese)
- Liu, D.; Du, Z. The Prediction of Vibration Response of Hydropower House Based on IBA-RBF. *China Rural Water Hydropower* 2020, *8*, 249–253. (In Chinese)
- 9. Song, Z.; Geng, D.; Su, C.; Liu, Y. Vibration prediction of a hydro-power house base on IFA-BPNN. *J. Vib. Shock* 2017, *36*, 64–69. (In Chinese)
- Shi, J.; Guo, J.; Zheng, S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renew. Sustain. Energy Rev.* 2012, 16, 3471–3480. [CrossRef]
- Garg, N.; Soni, K.; Saxena, T.; Maji, S. Applications of AutoRegressive Integrated Moving Average (ARIMA) approach in time-series prediction of traffic noise pollution. *Noise Control Eng. J.* 2015, 63, 182–194. [CrossRef]
- Milan, S.G.; Roozbahani, A.; Azar, N.A.; Javadi, S. Development of adaptive neuro fuzzy inference system—Evolutionary algorithms hybrid models (ANFIS-EA) for prediction of optimal groundwater exploitation. J. Hydrol. 2021, 598, 126258. [CrossRef]
- 13. Tran, D.S.; Songmene, V.; Ngo, A.D. Regression and ANFIS-based models for predicting of surface roughness and thrust force during drilling of biocomposites. *Neural Comput. Appl.* **2021**, *33*, 11721–11738. [CrossRef]
- 14. Sharifi, H.; Roozbahani, A.; Shahdany, S.M.H. Evaluating the Performance of Agricultural Water Distribution Systems Using FIS, ANN and ANFIS Intelligent Models. *Water Resour. Manag.* **2021**, *35*, 1797–1816. [CrossRef]
- 15. Armstrong, J.S. Combining forecasts: The end of the beginning or the beginning of the end? *Int. J. Forecast.* **1989**, *5*, 585–588. [CrossRef]
- 16. Wei, B.; Liu, B.; Yuan, D.; Mao, Y.; Yao, S. Spatiotemporal hybrid model for concrete arch dam deformation monitoring considering chaotic effect of residual series. *Eng. Struct.* **2021**, *228*, 111488. [CrossRef]
- 17. Liu, B.; Wei, B.; Li, H.; Mao, Y. Multipoint hybrid model for RCC arch dam displacement health monitoring considering construction interface and its seepage. *Appl. Math. Model.* **2022**, *110*, 674–697. [CrossRef]
- 18. Wen, Z.; Zhou, R.; Su, H. MR and stacked GRUs neural network combined model and its application for deformation prediction of concrete dam. *Expert Syst. Appl.* **2022**, 201, 117272. [CrossRef]
- Jain, A.; Kumar, A.M. Hybrid neural network models for hydrologic time series forecasting. *Appl. Soft Comput.* 2007, 7, 585–592. [CrossRef]
- Ibrahim, K.S.M.H.; Huang, Y.F.; Ahmed, A.N.; Koo, C.H.; El-Shafie, A. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alex. Eng. J.* 2022, *61*, 279–303. [CrossRef]
- 21. Humphrey, G.B.; Gibbs, M.S.; Dandy, G.C.; Maier, H.R. A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. *J. Hydrol.* **2016**, *540*, 623–640. [CrossRef]
- Samantaray, S.; Sahoo, A.; Mishra, S.S. Chapter 37—Flood forecasting using novel ANFIS-WOA approach in Mahanadi river basin, India. In *Current Directions in Water Scarcity Research*; Zakwan, M., Wahid, A., Niazkar, M., Chatterjee, U., Eds.; Elsevier: Amsterdam, The Netherlands, 2022.
- 23. Luo, H.; Zhou, P.; Shu, L.; Mou, J.; Zheng, H.; Jiang, C.; Wang, Y. Energy Performance Curves Prediction of Centrifugal Pumps Based on Constrained PSO-SVR Model. *Energies* **2022**, *15*, 3309. [CrossRef]
- 24. Huang, R.; Zhang, Z.; Zhang, W.; Mou, J.; Zhou, P.; Wang, Y. Energy performance prediction of the centrifugal pumps by using a hybrid neural network. *Energy* 2020, *213*, 119005. [CrossRef]
- 25. Zhang, J.; Yang, Z. A Study on Reservoir Dam Defects and Breaches in China; Science Press: Beijing, China, 2014. (In Chinese)
- Sang, L.; Wang, J.C.; Sui, J.; Dziedzic, M. A New Approach for Dam Safety Assessment Using the Extended Cloud Model. Water Resour. Manag. 2022, 36, 5785–5798. [CrossRef]
- 27. He, G.; Chai, J.; Qin, Y.; Xu, Z.; Li, S. Coupled Model of Variable Fuzzy Sets and the Analytic Hierarchy Process and its Application to the Social and Environmental Impact Evaluation of Dam Breaks. *Water Resour. Manag.* **2020**, *34*, 2677–2697. [CrossRef]
- Yang, X.; Yuan, C.; Hou, M.; Zhou, C.; Ju, Y.; Qi, F. Law and Early Warning of Vertical Sluice Cluster Displacements in Soft Coastal Soil. KSCE J. Civ. Eng. 2023, 27, 698–711. [CrossRef]
- 29. Dou, Z.; Xu, X.; Lin, Y.; Zhou, R. Application of D-S Evidence Fusion Method in the Fault Detection of Temperature Sensor. *Math. Probl. Eng.* **2014**, 2014, 395057. [CrossRef]
- Zhao, Y.; Jia, R.; Liu, C. An Evaluation Method of Underwater Ocean Environment Safety Situation Based on D-S Evidence Theory. *Adv. Meteorol.* 2015, 207656. [CrossRef]
- Yang, M.; Jia, L.; Gao, T.; He, Y.; Gui, B.; Zhang, T. Cloud Platform Credibility Assessment System Based on D-S Theory and Blockchain Technology. Wirel. Commun. Mob. Comput. 2022, 2022, 5866570. [CrossRef]
- Chen, A.; Tang, X.; Cheng, B.; He, J. Multi-source monitoring information fusion method for dam health diagnosis based on Wasserstein distance. *Inf. Sci.* 2023, 632, 378–389. [CrossRef]
- Xu, C.; Zhang, H.; Peng, D.; Yu, Y.; Xu, C.; Zhang, H. Study of Fault Diagnosis of Integrate of D-S Evidence Theory Based on Neural Network for Turbine. *Energy Procedia* 2012, 16, 2027–2032. [CrossRef]
- 34. Box, G.E.P.; Jenkins, G.M. Time Series Analysis: Forecasting and Control (Revised Ed.); Holden-Day: San Francisco, CA, USA, 1976.

- 35. Jang, J.-S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man Cybern.* 1993, 23, 665–685. [CrossRef]
- Schmittner, C.; Kosleck, S.; Hennig, J. A Phase-Amplitude Iteration Scheme for the Optimization of Deterministic Wave Sequences. In Proceedings of the ASME 2009 28th International Conference on Ocean, Offshore and Arctic Engineering, Honolulu, HI, USA, 31 May–5 June 2009; pp. 653–660. [CrossRef]
- 37. Chiu, S.L. Fuzzy Model Identification Based on Cluster Estimation. J. Intell. Fuzzy Syst. 1994, 2, 267–278. [CrossRef]
- Zhang, T.J.; Chen, D.; Sun, J. Research on Neural Network Model Based on Subtraction Clustering and Its Applications. *Phys. Procedia* 2012, 25, 1642–1647.
- Dunn, J.C. A Graph Theoretic Analysis of Pattern Classification via Tamura's Fuzzy Relation. IEEE Trans. Syst. Man Cybern. 1974, SMC-4, 310–313. [CrossRef]
- 40. Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms; Plenum Press: New York, NY, USA, 1981.
- Mirnezami, S.V.; Hassan-Beygi, S.R.; Banakar, A.; Ghobadian, B. Modelling total weighted vibration of a trailer seat pulled by a two-wheel tractor consumed diesel-biodiesel fuel blends using ANFIS methodology. *Neural Comput. Appl.* 2017, 28, 1197–1206. [CrossRef]
- 42. Yang, H.; Hasanipanah, M.; Tahir, M.M.; Bui, D.T. Intelligent Prediction of Blasting-Induced Ground Vibration Using ANFIS Optimized by GA and PSO. *Nat. Resour. Res.* **2020**, *29*, 739–750. [CrossRef]
- 43. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. Adv. Eng. Softw. 2016, 95, 51–67. [CrossRef]
- 44. Zhang, X.; Liu, Z.; Miao, Q.; Wang, L. Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time-frequency atom dictionary. *Mech. Syst. Signal Process.* **2018**, *107*, 29–42. [CrossRef]
- 45. Fan, Q.; Yu, F.; Xuan, M. Transformer fault diagnosis method based on improved whale optimization algorithm to optimize support vector machine. *Energy Rep.* 2021, *7*, 856–866. [CrossRef]
- Samantaray, S.; Sahoo, A.; Agnihotri, A. Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River Basin. *Methodsx* 2023, 10, 102060. [CrossRef] [PubMed]
- Samantaray, S.; Sahoo, A.; Paul, S.; Ghose, D.K. Prediction of Bed-Load Sediment Using Newly Developed Support-Vector Machine Techniques. J. Irrig. Drain. Eng. 2022, 148, 04022034. [CrossRef]
- 48. Samantaray, S.; Biswakalyani, C.; Singh, D.K.; Sahoo, A.; Satapathy, D.P. Prediction of groundwater fluctuation based on hybrid ANFIS-GWO approach in arid Watershed, India. *Soft Comput.* **2022**, *26*, 5251–5273. [CrossRef]
- 49. Zhou, J.; Sun, B.; Huang, Y. Real-time online monitoring technology of large sector lock gate with medium and high water head. *Port Waterw. Eng.* **2021**, *579*, 99–103. (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.