Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 3, No. 2, 2024

Research on Flood Prediction and Risk Assessment Based on Multiple Models and Entropy Weight Method

Yupeng Shi, Peimeng Wu, Yu Qin

School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China

Abstract: The global water cycle under the combined influence of climate warming and economic development is the focus of international attention. Warming can cause melting glaciers and higher sea levels, and trigger extreme weather, increasing damage from flooding. Flood prediction and risk assessment have become the key to disaster prevention and mitigation work, requiring the influence of multiple factors. For this phenomenon, this paper establishes a mathematical analysis model to make reasonable planning and decisions to prevent economic losses and casualties caused by flood disasters. In this paper, we calculate the Spearman correlation coefficient using SPSS and train it using a classification algorithm. The CSV was divided into different categories, the weights of different indicators were obtained according to the entropy weight analysis, and the seven heaviest indicators were visualized with bar graphs. It was concluded that factors such as monsoon intensity, topographic drainage, and river management were unrelated to urbanization and agricultural practices. Then the flood prediction model based on entropy weight topography was preliminarily constructed and verified. By comparing the prediction results of the test set, the model's accuracy can be improved by drawing the error, and according to the requirements of the problem, five indicators are selected to build multiple linear regression and flood probability models. The flood prediction model is based on mathematical, physical principles and historical data, taking into account meteorological factors (such as rainfall, rainfall intensity, etc.), hydrology (such as river level, flow), terrain, soil, etc., to simulate and predict the future flood situation in a specific area. These models can help relevant departments and personnel to prepare for flood control in advance, formulate emergency measures, and optimize water resources management, to reduce the loss of life and property and social and economic impact caused by floods.

Keywords: Correlation, Multiple linear regression, Entropy weight, Topological value.

1. Introduction

With the development of society, urbanization, population growth, erosion, and frequent occurrence, flood is one of the most destructive natural disasters, it can quickly sweep through a large area, endangering people's lives, bringing huge property casualties, seriously threatening the sustainable development of social economy and the stability of the ecological environment. In this scenario, flood prediction can safely protect human life[1]. By preventing floods, reducing casualties in disasters, and reducing public safety and property losses. By preventing floods, the unrest caused by social unrest and disasters can be reduced. Besides, it can help protect the environment and ecosystems.

This paper will visualize the data, test whether the data have pathological values, and use the SPSS software. Then, the data were divided into different categories using a classification algorithm. In this paper, the weights of entropy weights were used to obtain different indicators, and the seven indicators with the largest proportion were visualized by the bar graph. Then, the flood prediction model is preliminarily established according to the entropic right tremor method. Finally, a multiple linear regression probability model was developed by dividing the test and training set by 8 to 2.

2. Definition and function of the flood forecast model

2.1. Classification

Common flood prediction models are as follows:

2.1.1. Hydrological model

(1) Empirical statistical model: it is generally only applicable to smaller water conservancy systems, but for larger systems, its application will become very complex. In this case, the operating rules and water systems may not show the desired feedback effect, resulting in the whole system reaching only suboptimal control[2].The empirical statistical model is predictions based on historical data and statistical relationships, and the method is relatively simple intuitive, and easy to understand and apply. However, for non-stable flood events, the accuracy of the empirical statistical model may be difficult to guarantee due to the complex and variable influencing factors. In particular, when encountering special situations such as extreme climate or abnormal rainfall, the prediction results of the model may have large errors.

(2) Concept hydrological model: Based on the conceptual description of the hydrological process, the watershed is taken as a system, and the rainfall runoff process is divided into several interrelated sub-processes, such as the new river model. It considers multiple hydrologic processes, such as evaporation, infiltration, and runoff. The conceptual hydrological model has a physical basis and the characteristics of a statistical regression model. This makes the model in this paper more comprehensive and accurate in describing the hydrological processes. However, the paper needs detailed watershed parameters, such as topographic characteristics, soil type, vegetation distribution, etc. However, these parameters are difficult to measure and quantify in many watersheds of the article, which increases the difficulty and uncertainty of model application.

2.1.2. Hydrodynamic model

(1) One-dimensional hydrodynamic model: Influenced by natural meteorological conditions, watershed characteristics, and human activities, the runoff sequence often shows non-linearity and randomness[3].Based on DEM, land use, and

soil type data, a prediction model was established, optimized the parameters using particle swarm optimization algorithm, and the accuracy of the model by simulating other floods[4]. During the simulation process, the calculation is relatively simple and more efficient because it considers only the change of water flow in a single direction. However, the one-dimensional model is calculated based on the sectional data, which cannot reflect the plane shape of the channel. This limits the applicability of the model for simulating large areas or complex topography

(2) Two-dimensional hydrodynamic model: considering the movement of water flow in two directions on the plane, it means that the inundation of complex terrain (such as floodplain, estuary, etc.) can be simulated more accurately. The model can provide richer computational information, such as flood arrival time, flooding range, submerged water depth, and flooding duration, which contributes to a more comprehensive understanding of the effects of flooding. However, at the same time, the model needs to consider the flow movement in both directions of the plane, so the calculation process is relatively complex and long-consuming

2.1.3. The model that is based on the AI

(1) Artificial neural network model: For non-linear problems, the artificial neural network is a common intelligent solution, among which a multi-layer perceptron neural network is one of the most popular neural networks in hydrological applications[5]. The learning and training of the artificial neural network model in this paper does not require a large amount of topographic and engineering data but only uses the existing hydrological observation data. This reduces the difficulty and cost of data collection. Although the learning and training of artificial neural networks does not require a large amount of topographic and engineering data, it still requires a large amount of high-quality hydrological observation data for training. If the data is insufficient or the quality is low, the model may be unable to fully learn the evolution rules of a flood, thus affecting the accuracy of prediction.

(2) Deep learning models: Deep learning architectures, such as convolutional neural networks (CNN) and long-term and short-term memory networks (LSTM), have also been applied to flood prediction. The long and short-term memory neural network (LSTM) in artificial intelligence methods has been studied a lot. Compared with other intelligent methods, the LSTM neural network has the characteristics of easy-to-build, adaptive, and real-time learning. To improve the generalization and fitting ability of LSTM, many scholars have conducted signal decomposition of non-stationary runoff data and optimization of model hyperparameters combined with an optimization algorithm[6].

2.2. Application Scenario

The application scenarios of the flood prediction model mainly include the following aspects:

2.2.1. Flood control and disaster reduction

Real-time monitoring of hydrological data, using the model to quickly predict the scale and arrival time of the flood, and timely release of early warning information to the possible affected areas, so that residents can be prepared to avoid risk in advance. Assist relevant departments in formulating emergency rescue plans, such as determining evacuation areas, setting up personnel resettlement sites, deploying relief materials, etc.

2.2.2. Water Resources Management Field

According to the flood prediction results, the water storage and discharge of reservoirs should be reasonably arranged, sufficient flood control and water storage capacity should be stored, the utilization of water resources in the non-flood period should be optimized, and the distribution and utilization of upstream and downstream water resources should be coordinated to avoid the waste and excessive development of water resources.

2.2.3. Urban planning and infrastructure construction

Predict the possible waterlogging in cities under different rainfall intensities, provide the basis for the planning, transformation, and optimization of drainage pipe network, and ensure smooth urban drainage. At the same time, when planning new buildings, transportation facilities, and other important infrastructure, the flood prediction model should be adopted to avoid the areas prone to flood or take corresponding flood control and protection measures.

2.2.4. Insurance and Finance Field

Insurance companies can evaluate the flood risk in different areas according to the flood prediction model, to set the insurance rate more scientifically and reduce their own risk. Financial institutions should judge the feasibility and potential risk of the project based on the results of the flood prediction model.

2.3. Functions

2.3.1. Data processing and analysis

Meteorological data (rainfall, temperature, etc.), hydrology (water level, flow rate, etc.), terrain, soil, and other types of data can be collected and processed. Analyze the internal connections between these data, such as the relationship between rainfall and the rate of river water level rise.

2.3.2. Simulation and Prediction

Repeat the evolution process of historical floods, including the rise of floods, and the formation and extinction of flood peaks. According to the existing data and historical law, predict the possibility of flood occurrence, flood peak flow, and flood range in the future.

2.3.3. Decision Assistance

Provide a decision-making basis for the planning and implementation of flood control works (such as dam building, river dredging, etc.). Help to determine the reasonable deployment plan of personnel and materials (such as sandbags, emergency rescue equipment, etc.). In different flood situations.

2.4. Significance

The flood prediction model in this paper can allow residents to transfer, avoid risks, and reduce the risk of casualties. Protection measures should be taken in advance for important facilities (such as hospitals, schools, electric power facilities, etc.). To reduce the economic losses. Avoid blind flood control, and carry out targeted protection and emergency rescue work in key areas according to the forecast results. Make emergency preparations in advance, shorten the emergency response time, and improve the overall efficiency of flood control and disaster relief.

Through the prediction and management of floods, the relationship between flood control and water resource utilization can be better coordinated to realize the sustainable utilization of water resources. Reasonable flood management can reduce the damage of floods to the ecological environment and is conducive to maintaining the health and stability of rivers, wetlands, and other ecosystems.

3. Data source and preliminary analysis

The data for this study are from www.apmcm. And org, the following table 1 is the table of data visualization processing we performed.

Table 1: Data visualization and processing table

	meaning	standard error	The lower limit of the 95% confidence interval of the mean	The upper limit of the 95% confidence interval of the mean	deflection	kurtosis	least value	crest value
Monsoon intensity	4.92199	2.05618	4.91806	4.92593	0.44378	0.33948	0	16
Terrain drainage	4.92655	2.09364	4.92254	4.93056	0.45931	0.24079	0	18
River management	4.95477	2.07178	4.95081	4.95874	0.42651	0.22504	0	16
Take off the forest	4.94226	2.05173	4.93834	4.94619	0.43405	0.27451	0	17
urbanization	4.94294	2.08314	4.93896	4.94693	0.44157	0.24651	0	17
climate change	4.93418	2.05792	4.93024	4.93812	0.42919	0.24091	0	17
Dam quality	4.95565	2.08215	4.95166	4.95963	0.44093	0.26	0	16
silt up	4.92695	2.06553	4.923	4.9309	0.44842	0.28606	0	16
agrotechnical measures	4.94264	2.06799	4.93868	4.9466	0.41987	0.21142	0	16
corrode	4.94846	2.08312	4.94447	4.95244	0.4641	0.26866	0	18
Ineffective disaster prevention	4.9448	2.07814	4.94082	4.94877	0.44353	0.20702	0	16
drainage system	4.94725	2.0725	4.94329	4.95122	0.44123	0.2942	0	17
Coastal vulnerability	4.95406	2.08907	4.95006	4.95806	0.44103	0.24233	0	17
hill-creep	4.93076	2.07777	4.92679	4.93474	0.42571	0.18736	0	16
watershed	4.92891	2.08235	4.92492	4.93289	0.45074	0.23699	0	16
Infrastructure deterioration	4.92619	2.06453	4.92223	4.93014	0.44363	0.24503	0	17
Population score	4.92796	2.07451	4.92399	4.93193	0.45197	0.25418	0	18
Wetland loss	4.95033	2.06888	4.94637	4.95429	0.43751	0.23524	0	18
Insufficient planning	4.94067	2.08169	4.93668	4.94465	0.45571	0.25027	0	16
Policy factors	4.93932	2.09032	4.93532	4.94332	0.43749	0.19864	0	16

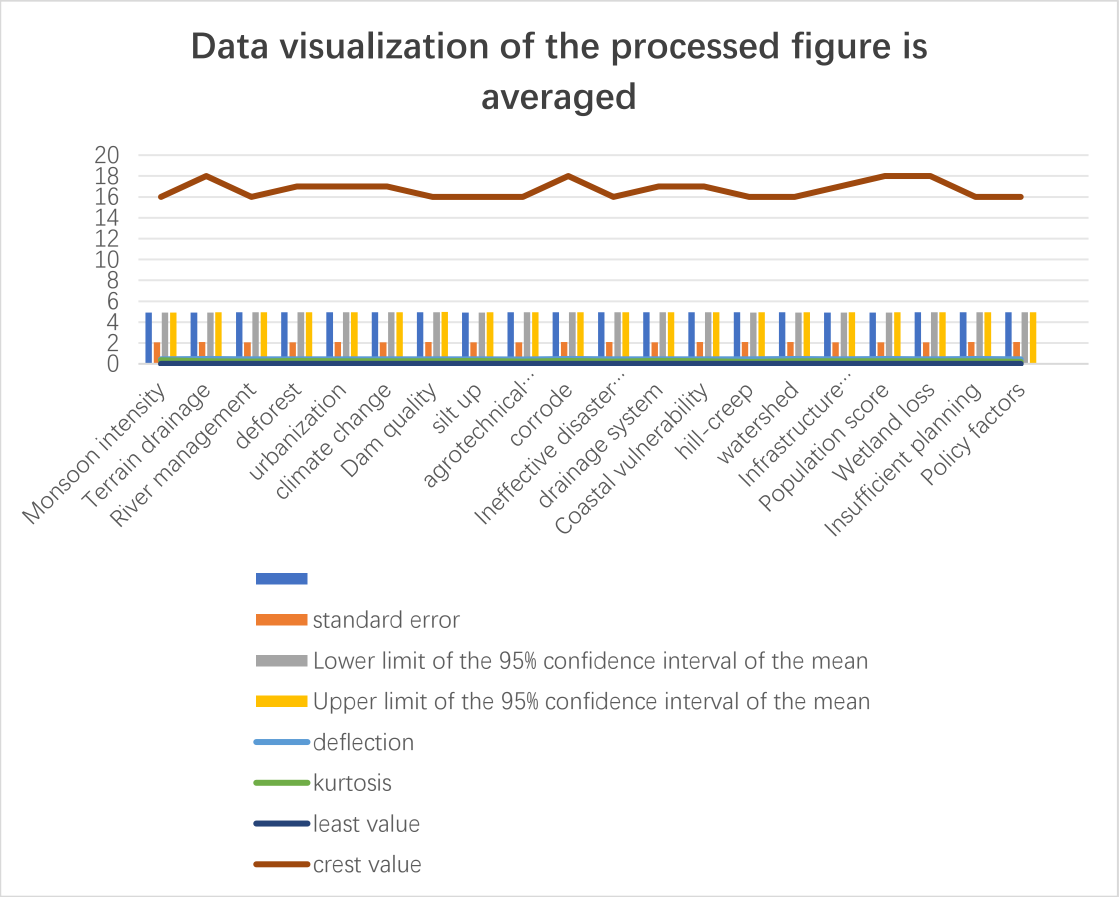

Figure 1: Data visialiazation

The occurrence of the flood was closely related to monsoon intensity, topographic drainage, river management, deforestation, climate change, dam quality, siltation, ineffective disaster prevention, landslide, deterioration of infrastructure, and population growth, which was calculated by the Spearman correlation coefficient of ssps software. To accurately predict the probability of flooding, the data set was split into a training set and a test set, with 30% of the data used as the test set and 70% as the training set. The model of linear regression, multilayer percepeptron, decision tree, random forest linear regression, and multilayer perceptron integration were selected to predict the probability of flood occurrence[7]. The specific operation is to first convert each variable to the corresponding rank, and then calculate the correlation coefficient, which is bounded as 0.175, and the correlation coefficient greater than 0.175 is considered closely related. However, the occurrence of flooding has little correlation with urbanization, agricultural practices, erosion, drainage systems, coastal vulnerability, watershed, wetland loss, inadequate planning, and policy factors.

Table 2: Spearman correlation analysis table

And Spearmans correlation coefficient	flood probability
Monsoon intensity	0.181
Terrain drainage	0.180
River management	0.180
Take off the forest	0.175
urbanization	0.173
climate change	0.176
Dam quality	0.178
silt up	0.178
agrotechnical measures	0.172
corrode	0.171
Ineffective disaster prevention	0.175
drainage system	0.171
Coastal vulnerability	0.171
hill-creep	0.176
watershed	0.173
Infrastructure deterioration	0.183
Population score	0.175
Wetland loss	0.173
Insufficient planning	0.174
Policy factors	0.173

There are many causes of flooding. In terms of the monsoon intensity, when the monsoon is strong, the precipitation is abundant and the rainfall increases significantly. Once it exceeds the carrying capacity of the river, it causes flooding. Terrain and drainage is also important. Flat terrain or low-lying areas with poor drainage, prone to flooding, steep upstream rivers, mountainous or hilly areas due to a short time of rainstorm, flood may also occur. In addition, poor river management, such as river blockage, siltation, illegal sand mining, barrage construction, and other unreasonable river projects, will reduce the flood discharge capacity of the river channel, change the natural state of the river, and affect the drainage. Deforestation destroys surface vegetation, reduces vegetation cover, and weakens soil water retention capacity, making it easier for rainwater to form surface runoff, increasing the risk of flood formation and flooding. The impact of climate change should not be ignored. Warming conditions leading to melting glaciers, increased river numbers, leading to flooding, and increased frequency and intensity of extreme weather events also increase the probability of flooding. If the quality of the dam is not up to standard, there will be cracks, leaks, and even breaks of the dam during the design or construction process, causing downstream floodwater, and poor maintenance and management of the dam will also lead to the failure of the dam in the flood. In terms of sedimentation and ineffective disaster prevention measures, severe sedimentation of a reservoir or lake will reduce storage capacity, reduce capacity, and increase flood risk. The construction and aging of flood control facilities may also make them lose their flood control function. Landslides can block rivers and form a barrier lake, causing major flooding, as well as aging or damage to infrastructure such as bridges and levees. In addition, population growth leads to the acceleration of urbanization, and urban expansion occupies the natural areas such as the drainage system

4. Multiple exponential analysis of topology based on the entropy weight method

In a post-hoc study, the articles included river management, coastal vulnerability, catchment, inadequate planning, population scores, infrastructure deterioration, urbanization as high risk indicators, topographic drainage, deforestation, climate change, agricultural practices, erosion, landslides, and wetland losses as moderate risk indicators. Monsoon intensity. Dam quality, siltation, disaster prevention, and drainage systems are low-risk indicators.

Next, this article used the entropy weight topology method to calculate the weights of different indicators, as shown in the figure below. The entropy weight method is a multi-index weight determination method, that is often used in decision evaluation or multi-factor decisions. It is based on information entropy theory and determines the weight of each index in the comprehensive evaluation by calculating the information entropy value of each index.

The specific steps are described as follows:

1. Calculate the information entropy value of each index: For several given indicators, the information entropy of each index is calculated first to reflect its contribution and dispersion in the overall evaluation. The formula for calculating the information entropy is:

$HX)=EL1(X)J=E[-In(P(X))]$.

2. Calculate the weight of each index: the weight is calculated according to the information entropy value of each index. The greater the information entropy value, the greater the weight of the index.

3. Normalization treatment: The calculated weights are normalized to ensure that the sum of each weight is 1, which facilitates the subsequent weight allocation and comprehensive evaluation.

The entropy weight method is applicable where the treatment exponents have a certain correlation or cannot be directly quantified. Through the calculation of information entropy, the importance of each index can be objectively reflected in the comprehensive evaluation, to realize the reasonable weight allocation.

The TOPSIS (preference sorting technique similar to an ideal solution) method is a multi-attribute decision method for evaluating the advantages of a candidate solution or project. It is based on the concept of geometric distance, evaluates the relative advantage of each candidate solution and the distance between the ideal solution and the negative ideal solution, and finally determines an optimal solution.

The specific steps are described as follows:

1. Determine the decision matrix: First, all candidates are scored according to each evaluation index to form a decision matrix. Each row represents a candidate scheme, and each column represents an evaluation index.

2. Normalization: Each index is normalized and converted to a dimensionless index value. Normalization is often referred to as normalization, which often requires data normalization to eliminate the effects of different dimensions between variables and speed up model training[8].

Ideal solution (A *): For each index, determine the maximum (income index) or minimum (cost index).

Negative ideal solution (A-): For each index, the minimum value (income index) or the maximum value (cost index) will be determined.

4. Calculate the distance to the ideal solution: calculate the Euclidean distance or other distance measure between each candidate solution and the ideal solution and the negative ideal solution.

5. Comprehensive score calculation: the comprehensive score is calculated according to the distance between the candidate scheme the ideal scheme and the negative ideal solution.

The TOPSIS method can effectively help decision makers to optimize under multiple evaluation indicators,

The importance of each indicator and the actual performance of the candidate scheme are considered to make a reasonable and scientific decision making.

The following table shows the results of the article weight analysis:

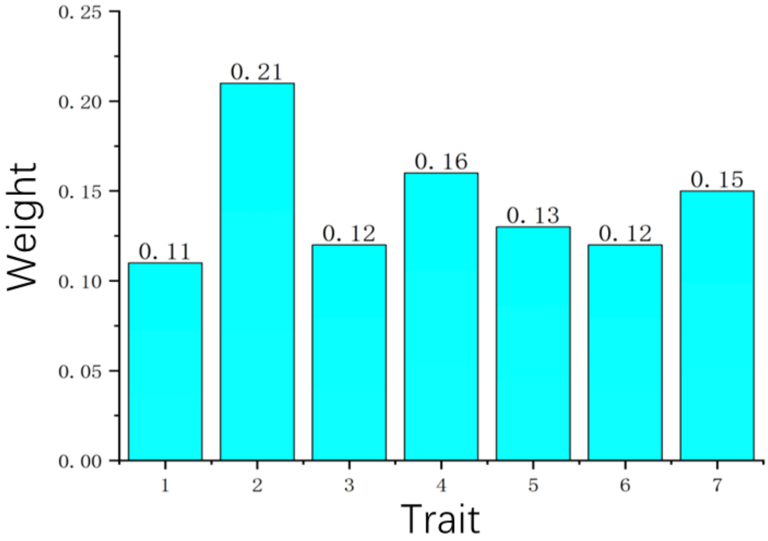

Figure 1. Weight analysis of the 26 indicator features

5. A multivariate-based linear regression model was combined with a sensitivity analysis

To further conduct early warning assessment of different flood risks, we further constructed a data model based on multiple linear regression. Here are the model and sensitivity analyses.

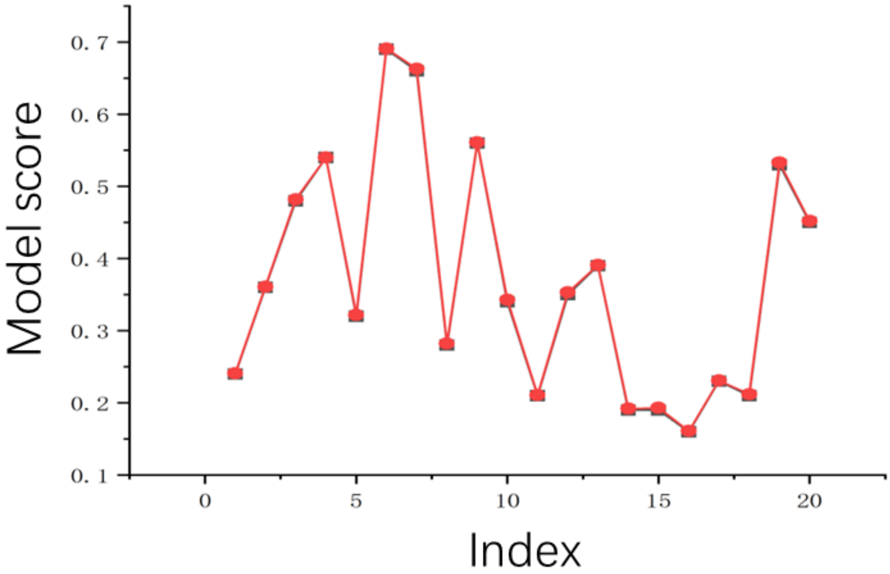

Figure 2 Sensitivity analysis

As can be seen from the figure above, the sensitivity of the model construction is consistent with the requirements, and the basic error is very small.

6. Conclusion

This study mainly uses various data and algorithms to analyze and predict flood disasters. It starts with visual inspection and Spearman correlation analysis of abnormalities in the data, revealing a strong correlation between monsoon intensity, population growth, and weaker associations with urbanization. A classification algorithm was used to classify the risk level, and the most significant flood indicators were identified using the entropy weight method. The flood prediction model was established using entropy topology and verified by testing and training data. The predictive power of the model was further verified by a Q-Q plot, confirming a normal distribution. The study highlights the importance of rich data sources, improved data quality, well-established algorithms, and enhanced computational power for flood prediction. It also highlights the practical application of adaptation and success based on regional characteristics.

References

[1] Cheng Nan, Gu Xianhui, Ren Mingxuan, et al. Global forecast of future flood evolution and extreme floods [J / OL]. Journal of Wuhan University (Science Edition), 1-10 [2024-10-04]

[2] Li Fa-xiao Li. Study on prediction and control of nonlinear model in flood forecasting [J]. Water Conservancy Technology and Economy, 2022,28 (02): 36-39.

[3] Cui Yabo, Luo Qingyuan, Liu Lina. A flood prediction algorithm based on improved nonlinear autoregressive networks. Journal of Shenyang University of Science and Technology, 2023,45 (01): 84-89.

[4] Liu Yaqian. Reservoir flood prediction scheme based on the Liuxi River model [J]. Yunnan Hydropower Power Generation, 2024,40 (03): 12-16.

[5] Xu Yueping, Zhou Xinlei, Wang Ruotong, etc. Research on flood runoff prediction based on integrated learning and deep learning. Changjiang, 1-9 September 13,2024.

[6] Zhang Shu. Flood disaster prediction and risk assessment based on data analysis and machine learning [J]. Information Technology and Informatization, 2024, (08): 189-194.

[7] Zhou Xia, Zhou Feng. Prediction model of non-stationary extreme runoff based on VMD-NGO-LSTM [J]. Peoples Pearl River, 2024,45 (06): 127-137.

[8] Wang Jun, Zhang Yuhang, Cui Yunye, et al. Study on flood prediction based on the CEEMDAN-LightGBM model [J]. Peoples Yellow River, 2024,46 (09): 99-105.