

Article

Prediction of Tail Water Level under the Influence of Backwater Effect Based on Deep Learning Models: A Case Study in the Xiangjiaba Hydropower Station

Sen Zhang ^{1,2,3}, Shuai Xie ^{1,2,3}, Yongqiang Wang ^{1,2,3,*}, Yang Xu ⁴, Zheng Zhang ⁴ and Benjun Jia ^{4,*}

- ¹ Changjiang River Scientific Research Institute, Changjiang Water Resources Commission of the Ministry of Water Resources of China, Wuhan 430010, China; zhangsen_crsri@163.com (S.Z.); xieshuai@mail.crsri.cn (S.X.)
- ² Hubei Key Laboratory of Water Resources & Eco-Environmental Sciences, Changjiang River Scientific Research Institute, Wuhan 430010, China
- ³ Research Center on the Yangtze River Economic Belt Protection and Development Strategy, Wuhan 430010, China
- ⁴ Hubei Key Laboratory of Intelligent Yangtze and Hydroelectric Science, China Yangtze Power Co., Ltd., Yichang 443300, China; xu_yang@ctg.com.cn (Y.X.); zhangzhengno1@163.com (Z.Z.)
- * Correspondence: wangyq@mail.crsri.cn (Y.W.); hydrobj@163.com (B.J.)

Abstract: Accurate forecasting of the tail water level (TWL) is of great importance for the safe and economic operation of hydropower stations. The prediction accuracy is significantly influenced by the backwater effect of downstream tributaries and the operation of adjacent hydropower stations, but the explicit quantification method of the backwater effect is lacking. In this study, a deep-learning-model-based forecasting method for TWL predictions under the backwater effect is developed and applied in the Xiangjiaba (XJB) hydropower station, which is influenced by the backwater effect of downstream tributaries, including the Hengjiang River (HJR) and the Minjiang River (MJR). Firstly, the random forest algorithm was used to analyze the influence of HJR and MJR flows with different lag times on the TWL prediction error of the XJB hydropower station. The results show that the time lags of the backwater effect of HJR and MJR run offs on the TWL of the XJB are 5~7 h and 1~2 h, respectively. Then, the run off thresholds of the HJR and MJR for impacting the TWL of the XJB station are obtained through scenario comparison, and the results show that the run off thresholds of the HJR and the MJR are 700 m³/s and 7000 m³/s, respectively. Finally, based on the analysis of the time lag and the threshold of the backwater effect, a deep learning model (LSTM)-based TWL forecasting method is established and applied to predict the TWL of the XJB station. The results show that the forecasting model has a good predictive performance, with 98.22% of absolute errors less than 20 cm. The mean absolute error over the validation dataset is 5.27 cm, and the maximum absolute error is 63.35 cm. Compared with the LSTM-based prediction model without considering the backwater effect, the mean absolute error decreased by 31%, and the maximum absolute error decreased by 71%.

Keywords: tail water level prediction; backwater effect; LSTM; Xiangjiaba hydropower station



Citation: Zhang, S.; Xie, S.; Wang, Y.; Xu, Y.; Zhang, Z.; Jia, B. Prediction of Tail Water Level under the Influence of Backwater Effect Based on Deep Learning Models: A Case Study in the Xiangjiaba Hydropower Station. *Water* **2023**, *15*, 3854. <https://doi.org/10.3390/w15213854>

Academic Editors: Feng Huang and Carlos G. Ochoa

Received: 20 September 2023

Revised: 23 October 2023

Accepted: 24 October 2023

Published: 5 November 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/>).

1. Introduction

Hydropower stations are of great significance to flood control and drought resistance, water and energy security, and the comprehensive utilization and management of water resources. With the increase in the number of hydropower stations, the efficiency promotion of existing hydropower stations based on the premise of safe operation has been a growing concern in recent years. For most hydropower stations, the precise control of water levels is at the core of dispatching and operating hydropower stations and the base of safe and economic operation and management. Therefore, the accurate prediction of water levels has gained increasing attention as it provides valuable support for the precise operation of hydropower stations. In particular, the prediction of the tail water level (TWL) is of

great importance for the economic operation and management of a power station. The TWL directly impacts the power generation efficiency of a hydropower station [1]. Hence, accurate prediction of the TWL can significantly contribute to the operation planning of a hydropower station.

Current water level prediction methods can be divided into physics-driven models and data-driven models. Physical-driven models work by solving equations based on hydrodynamics equations or water balance equations [2]. The hydrodynamic model mainly uses the Saint-Venant equations and boundary condition data to calculate the river network. In order to improve the calculation feasibility, many assumptions are added, resulting in low calculation accuracy. With the deepening of people's understanding of the hydrodynamic mechanism and the development of computer technology, the hydrodynamic model is constantly improving, but its computational complexity is getting higher and higher [3]. However, it is still difficult to describe the physical mechanism accurately with equations when there is a propagation time problem between variables of the hydropower station or in more complex, special cases. Therefore, many hydropower stations use the combination of the physics-driven model and the statistical method in practical application to improve the calculation accuracy and efficiency. Through statistical calculations of historical operation data, the empirical curve describing the correlation of variables is summarized to calculate the variables that cannot be accurately described by the equation, such as the water level–storage capacity curve and the TWL–outbound flow rate curve [4].

A data-driven model refers to a model that summarizes patterns from historical operational data and applies them to future data forecasting [5]. The empirical curve is a simple data-driven model that summarizes the corresponding relationship between the combination of target variables and feature variables. A target variable can be calculated by finding a similar situation in the empirical curve table. However, the simple data-driven model cannot handle nonlinear relationships, which are ubiquitous in the research of water level prediction. Therefore, machine learning, such as artificial neural network (ANN), random forest (RF), and support vector machine (SVM), which can be immune to these limitations, is widely used in water level prediction research [6–9]. But, the machine learning methods still cannot adapt to the temporal dependencies in the water level prediction problem. The recurrent neural network (RNN), as a kind of deep learning model, can excavate temporal dependencies to improve the forecasting accuracy. As an improved RNN model, the LSTM model can effectively slow down the gradient disappearance or explosion that may occur in long sequence problems, which has attracted the attention of many scholars [10–12]. Shuofeng, L. [12] focus on the precipitation-only water level forecasting problem by using an LSTM-based hybrid model and try predicting the future water levels of all of the rivers in Japan by using simulated precipitation data from the database for policy decision making for future climate change (d4PDF). Zhang, Z. [13] use the maximal information coefficient and feature combination to select feature inputs, construct a CNN-LSTM neural network to predict the downstream water level of a reservoir, and then compare the method with four state-of-the-art prediction methods; they find that the designed method is also very competitive.

While there are numerous water level prediction methods available, the majority of them mainly focus on longer time scales, typically exceeding one day [14,15], and they primarily consider upstream water levels [16,17]. There are limited studies that specifically address tail water level (TWL) prediction and consider the downstream backwater effect, despite its obvious significance. The Xiangjiaba hydropower station (XJB), as China's third largest hydropower station and the fifth largest globally, plays a vital role in power generation, shipping, and ecological benefits. The downstream tributary's backwater effect on the XJB's TWL can reach approximately 3 m, significantly impacting the accuracy of tail water level predictions, thus posing a major operational challenge. Hence, the establishment of a TWL prediction model that accounts for the backwater effect becomes crucial in achieving accurate predictions under such conditions.

The structure of the paper is as follows. Section 2 will provide an overview of the research area and the research methodology employed. In Section 3, the research process will be analyzed. Section 4 will present the analysis of the model prediction results. Finally, the paper will conclude with a summary of the findings.

2. Case Study, Data, and Method

2.1. Case Study

The XJB's geographical location is shown in Figure 1. The dam site of the XJB is located at the canyon exit of the lower reaches of the Jinsha River [18,19]. The installed capacity of the XJB hydropower station is 6 million kW, the number of units is 8, the rated output of each unit is 750,000 kW, and the maximum capacity of each unit corresponds to the minimum head of 100 m. The regulation capacity of the reservoir is 900 million m³, the flood control capacity is 903 million m³, and the total capacity is 5.163 billion m³. Figure 1 shows that the upstream water level of the XJB is influenced by the operation of the Xiluodu Hydropower Station (XLD), while the downstream water level is affected by the backwater effect of the confluence of the Hengjiang River (HJR) and the Minjiang River (MJR). The complex boundary conditions make the prediction of water level by hydrodynamic method or water balance method prone to large errors [20].

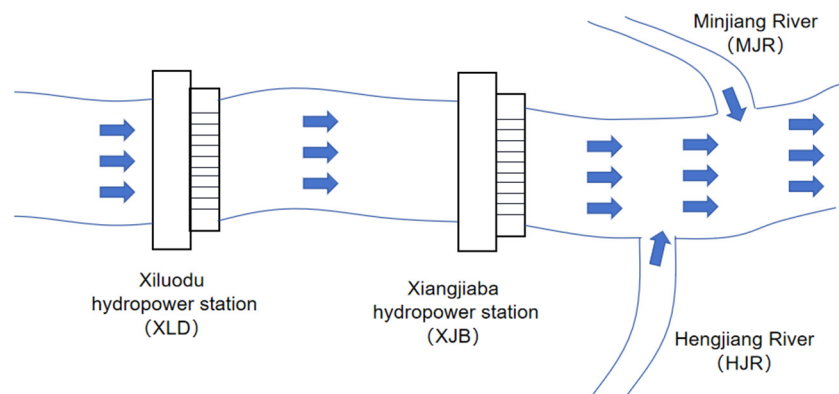


Figure 1. Schematic diagram of upstream and downstream of the XJB.

2.2. Data

This study uses the operation data of the XJB hydropower station from 2015 to 2020 for prediction and analysis. The content and classification of the data are shown in Table 1.

Table 1. Data type and time.

Data Type	Station	Time Series	Time Step
Water level	Upstream water level and tail water level of XJB	2015–2020	2 h, 1 h
Flow rate	XLD and XJB's outbound flow rate, inbound flow rate and abandon flow rate, HJR and MJR run off	2015–2020	1 h
Output	XLD, XJB plant, branch plant, and each unit's output	2015–2020	2 h, 1 h
Empirical curve	XLD's and XJB's water level–storage curve, head loss curve, unit flow rate curve, tail water level curve, etc.	/	/

2.3. Method

There are many factors affecting the XJB [6]. The most critical factor is the outbound flow of the XJB. The change trend of the XJB is consistent with the outbound flow, but there is a lag in time, as shown in Figure 2.

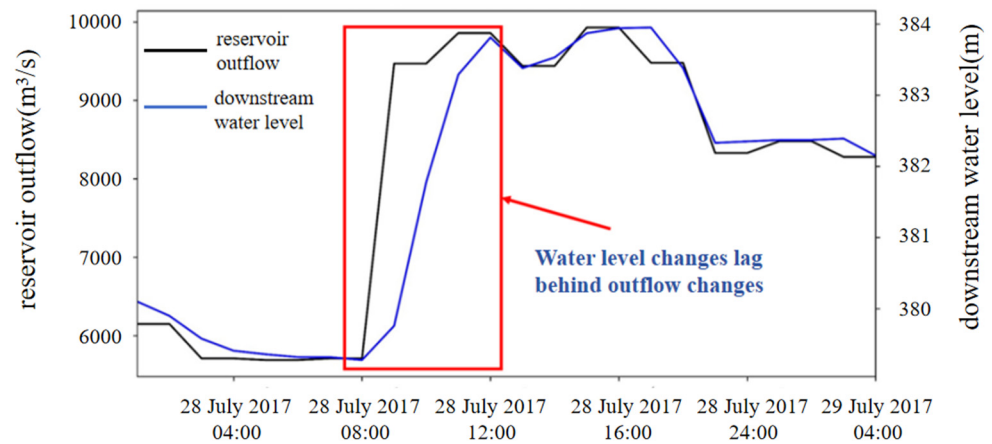


Figure 2. XJB outbound flow and tail water level change process chart.

The empirical curve of the outbound flow and the TWL is usually used to predict the TWL of the XJB, which is a characteristic curve based on historical data. The calculation method for determining the outbound flow involves assessing the water balance of the XJB reservoir, which relies on future predictions of the inbound flow rate and power station output planning. By utilizing the gross head of the hydropower station, the water consumption rate can be obtained through reference to the empirical curve of the water consumption rate versus the gross head. This rate is then multiplied by the planned output to determine both the power generation flow and the outbound flow rate. The computational process is illustrated in Figure 3. In the figure, red colour means key steps. In this study, the model developed based on the water balance principle is referred to as the water balance equation model (WBE model).

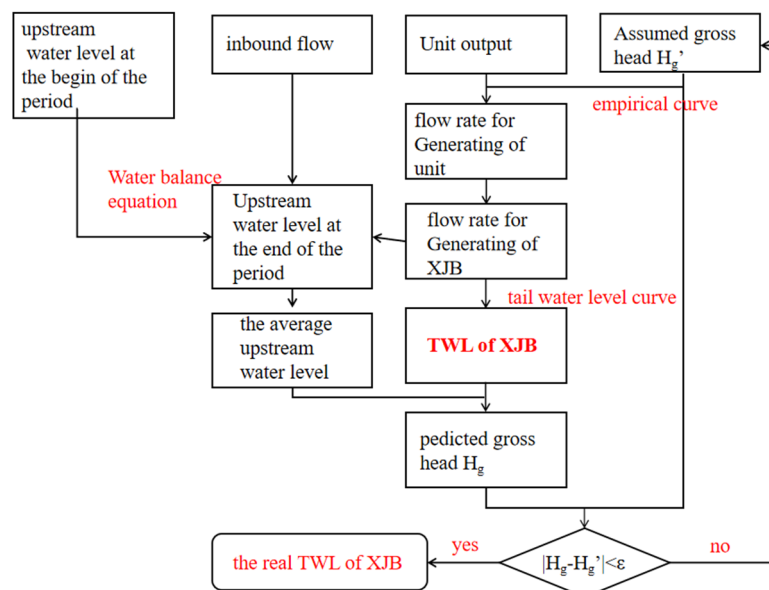


Figure 3. Calculating process of XJB using the water balance method.

Nevertheless, it is important to acknowledge that there are more intricate factors influencing the water level of the XJB hydropower station. This is illustrated in Figure 3. The

water balance method solely takes into account the variables related to the power station itself, which presents two noteworthy limitations: (1) It fails to consider the operating conditions of the upstream XLD, rainfall between the two hydropower stations, and the time lag effect mentioned earlier [21]. (2) The Hengjiang river flows in at 1.9 km downstream of the XJB and the Minjiang River flows in at 28 km downstream of the XJB, meaning that their backwater effects are not accounted for in the water balance method [22].

This study aims to analyze the time lag and the threshold of the backwater effect utilizing LSTM and machine learning techniques. On this basis, the LSTM prediction model for the TWL of the XJB considering the impact of the backwater effect is constructed.

To begin with, this study will analyze and discuss the relationship between the TWL of the XJB and relevant variables using the water balance method. Subsequently, a machine learning model will be constructed for predicting the water level in the TWL of the XJB. Additionally, the random forest model will be employed to analyze the characteristic relationship between the prediction error and the run offs of the Hengjiang River (HJR) and the Minjiang River (MJR) during different time periods. This analysis will help determine the time lag of the backwater effect. Furthermore, a deep learning water level prediction model for the TWL of the XJB will be established. Various scenarios will be categorized based on different run offs of the HJR and the MJR, and the backwater effect threshold will be derived by examining the prediction results. Finally, a water level prediction model for the TWL of the XJB that takes into consideration the impact of the backwater effect will be constructed. Finally, a water level prediction model for the TWL of the XJB considering the impact of the backwater effect will be constructed. The methodology roadmap is shown in Figure 4. In Figure 4, dotted box means the detailed steps of analyzing the influence of backwater effect and the blue box means key steps among them.

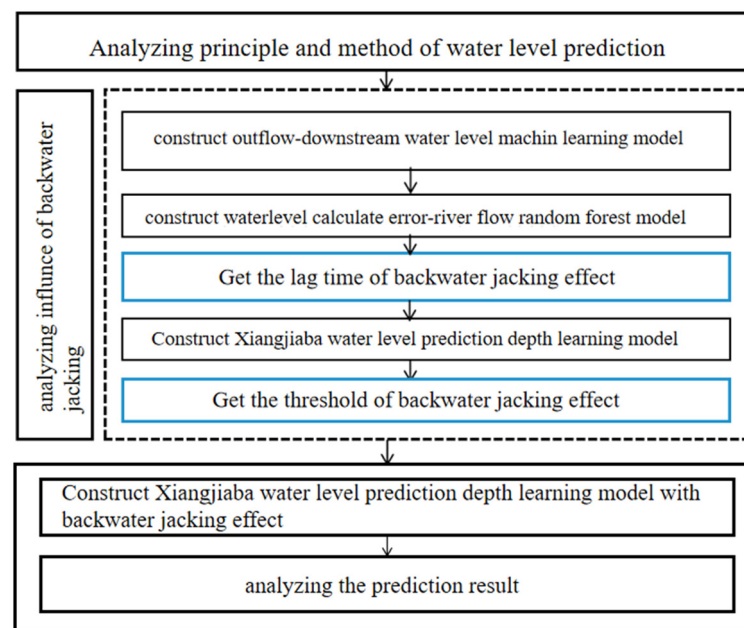


Figure 4. Flow chart of research ideas.

The selection of the LSTM (Long Short-Term Memory) algorithm to predict the TWL (Tide Water Level) of the XJB is motivated by the inherent time lag between the multiple variables involved in water level forecasting. The LSTM algorithm is renowned for its ability to mitigate the challenges of gradient disappearance or explosion, which are often encountered in long sequence problems. Thus, LSTM has demonstrated superior performance compared to traditional RNNs in addressing longer sequence prediction tasks. The LSTM model is shown in Figure 5.

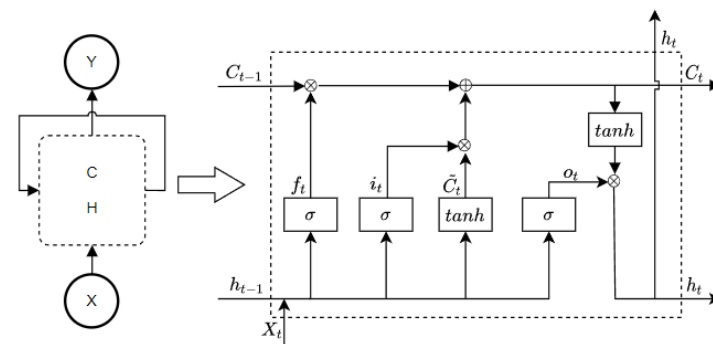


Figure 5. Schematic diagram of LSTM calculation steps.

LSTM implements the control of information state of memory unit through a forget gate and an input gate. The forget gate f_t determines which water level influence information of h_{t-1} and X_t should be discarded at time t . The input gate is used to determine how much X_t and h_{t-1} water level impact information needs to be passed to C_t at time t in order to update the information stored by C_{t-1} . LSTM uses the output gate to control the unit state C_t at the current time t and how much water level influence information in h_{t-1} and X_t should be output [23,24].

3. Analysis of Backwater Effect

3.1. Analysis Method of Backwater Effect

The analysis of time lag can be mainly divided into two primary steps. Initially, the linear regression model is employed to establish the corresponding relationship between the TWL of the XJB and the outbound flow rate; in this way, the prediction error between the calculated TWL of the XJB and the actual TWL is determined. Additionally, a relationship between the prediction error of the TWL and the run offs of the Hengjiang River (HJR) and the Minjiang River (MJR) before different hours is established. By utilizing the run offs data with different time lags as input variables and the water level prediction error as the output variable, the random forest model can be employed to calculate the significance of various input factors. This analysis aids in identifying the specific time lag from which the error predominantly originates, focusing on the run offs data of the correct time.

The threshold analysis builds upon the outcomes of the lag analysis. A 12 h prediction model without considering the backwater effect is established. This model allows for the categorization of different scenarios based on the varying run offs of the HJR and the MJR. By computing the water level prediction error under different scenarios, it becomes evident that scenarios with run offs exceeding the backwater effect threshold exhibit notably higher prediction errors compared to those under the threshold.

3.2. Time Lag Analysis of Backwater Effect

The random forest model employs a technique wherein the ordering of feature factors within the training set is randomly shuffled to analyze their individual impact on the resultant output variable. In the context of this study, a random forest model was constructed with the run offs of the HJR and the MJR at different time lags serving as input factors, while the TWL of the XJB error, calculated through machine learning methods, was set as the output. The model aimed to determine the characteristic weights between the run offs of the HJR and the MJR at various time lags on the water level error. In the analysis process, the calculation was repeated 30 times, and the influence degree of different time periods of the HJR's and the MJR's run offs on the prediction error was drawn as a box diagram, as in Figure 6. The y -axis in the figure represents different input factors. The 'MJR 6' represents the flow rate of the MJR six hours ago. Other input factors follow a similar pattern. On the x -axis, the decline in the model index resulting from the random disruption of each factor is depicted, thereby capturing the importance of the perturbed factor to the output variable.

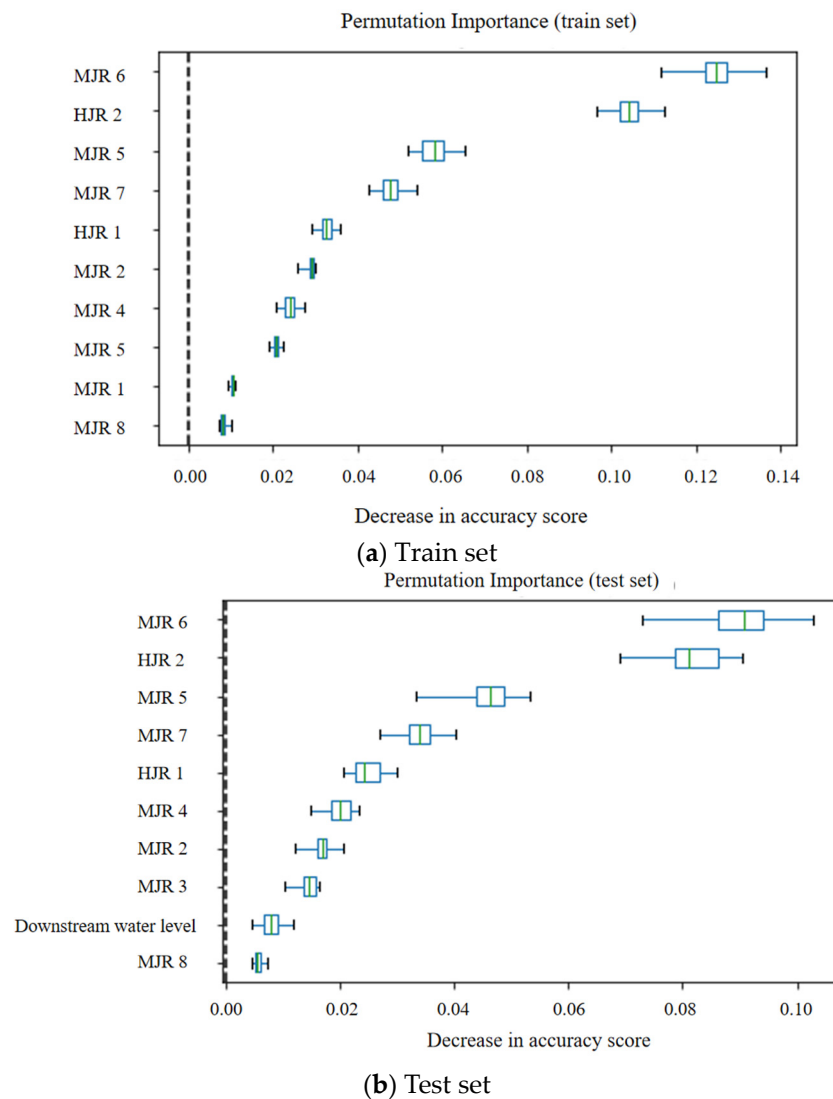


Figure 6. Characteristic importance of different input factors.

The presented figure demonstrates that both the test set and the training set exhibit identical distributions in terms of the calculation of characteristic importance. The top five factors of the two sets are the same, which are Minjiang river 6, Hengjiang river 2, Minjiang river 5, Minjiang river 7, and Hengjiang river 1. The results show that the lag time of the backwater effect of the MJR run off to the TWL of the XJB is about 5–7 h, and that of the HJR run off to the TWL of the XJB is approximately 1–2 h.

3.3. Threshold Analysis of Backwater Effect

The backwater effect threshold analysis method utilizes a deep learning model to predict the tail water level (TWL) of the XJB under different scenarios, and the model disregards the influence of the backwater effect. This analysis involves the segmentation of flow scenarios from the Hengjiang River (HJR) and the Minjiang River (MJR). Specifically, the analysis focuses on prediction and computation using data from the year 2019, and the scenarios are categorized based on predefined thresholds (700 m³/s for HJR run off and 7000 m³/s for MJR run off). These thresholds divide the scenarios into two types: Scenario 1, where either the HJR flow rate or the MJR flow rate exceeds their respective thresholds, and Scenario 2, where both flow rates are below their thresholds. Additionally, an overall scenario without threshold division is also considered. The 12 h prediction errors for three scenarios are visualized in Figure 7.

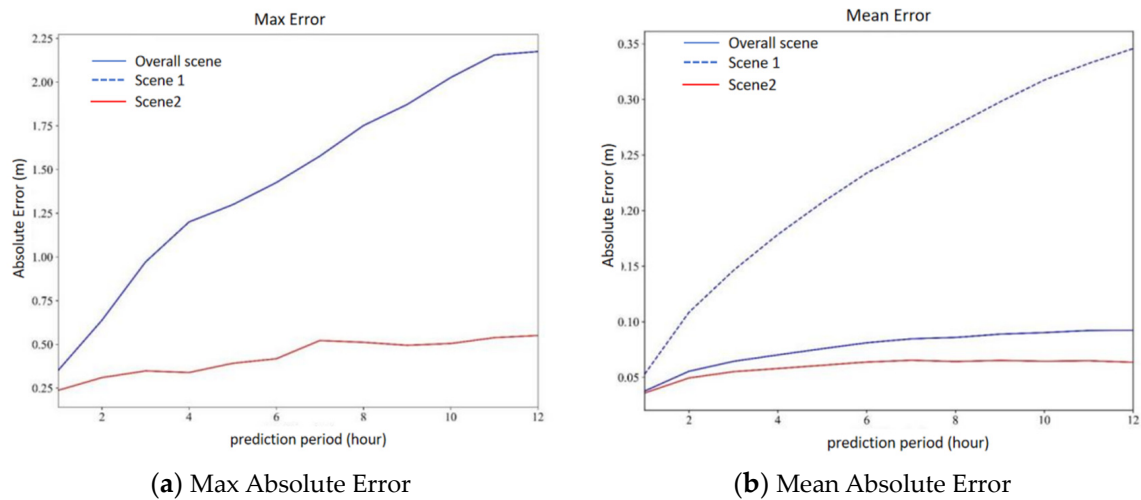


Figure 7. Max error and mean error of different scenarios.

The maximum prediction error of the overall scenario coincides with that of Scenario 1, while the average prediction error of Scenario 1 is much higher than that of the overall scenario and Scenario 2. The error distribution diagram of different scenarios is shown in Figure 8.

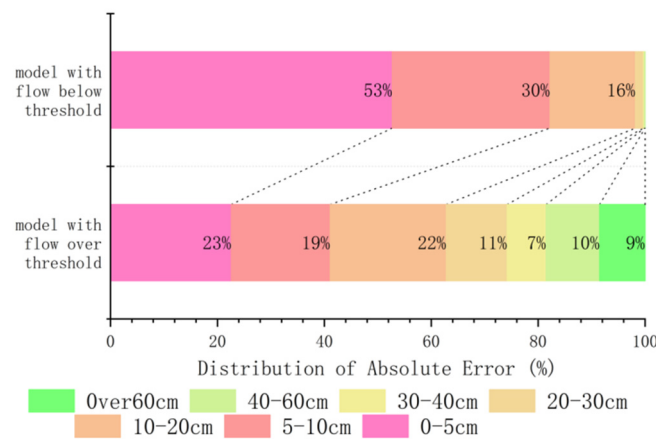


Figure 8. Distribution of absolute error of models with different flow rates.

From the distribution of errors, in Scenario 1, 18.49% of the errors are above 40 cm, while in Scenario 2, only 0.04% of the errors exceed 40 cm. Overall, using $700 \text{ m}^3/\text{s}$ as the threshold for the HJR flow rate and $7000 \text{ m}^3/\text{s}$ as the threshold for the MJR flow rate can effectively differentiate between forecast scenarios. When both flow rates are below the threshold, the maximum absolute prediction error is 55.02 cm, and the average prediction error is 5.93 cm. In comparison, it can be assumed that this scenario is minimally impacted by the backwater effect.

4. Analysis of Predictive Performance

4.1. Prediction Result of the WBE Model

Using the method described earlier, a Water Balance Error (WBE) model was formulated to assess the TWL of the XJB in 2020, employing a time step interval of two hours. The resultant absolute error measurements are presented in Figure 9, providing an informative illustration of the model’s performance.

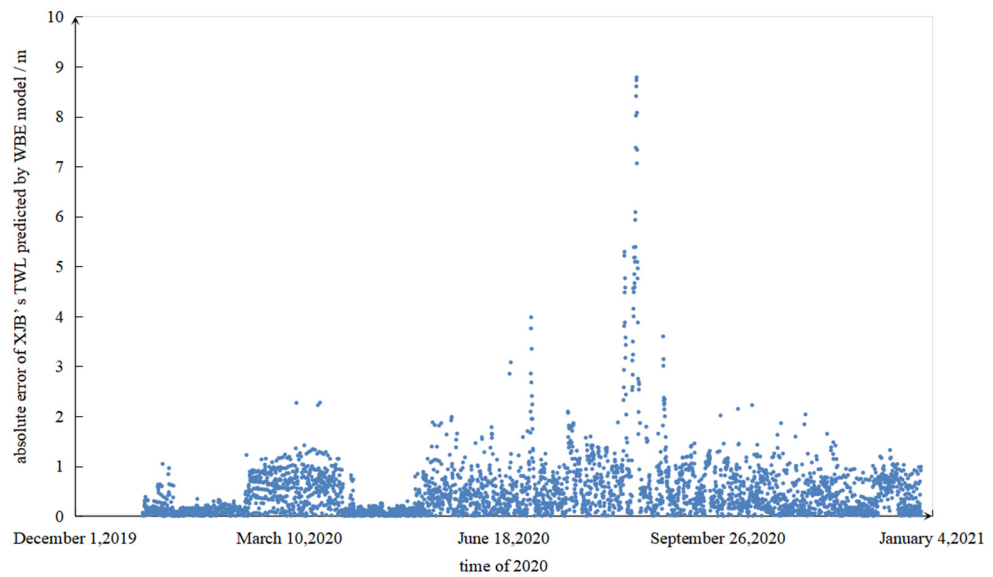


Figure 9. Annual distribution of absolute error of XJB’s TWL predicted by WBE model.

The evaluation of the results demonstrates a mean absolute error of 47.3 cm, with a maximum absolute error of 878.3 cm. Errors exceeding the 100 cm threshold constitute around 10% of the total errors. During the calculation process, it was observed that the error in the discharge flow exiting the XJB is less than 3%, indicating high accuracy. Consequently, the primary source of discrepancy in the estimation of the tail water level (TWL) of the XJB can be ascribed to two factors: the imprecision inherent in the TWL empirical curve and the backwater effect generated by the HJR and the MJR beneath the XJB dam.

4.2. Prediction Result of LSTM Model

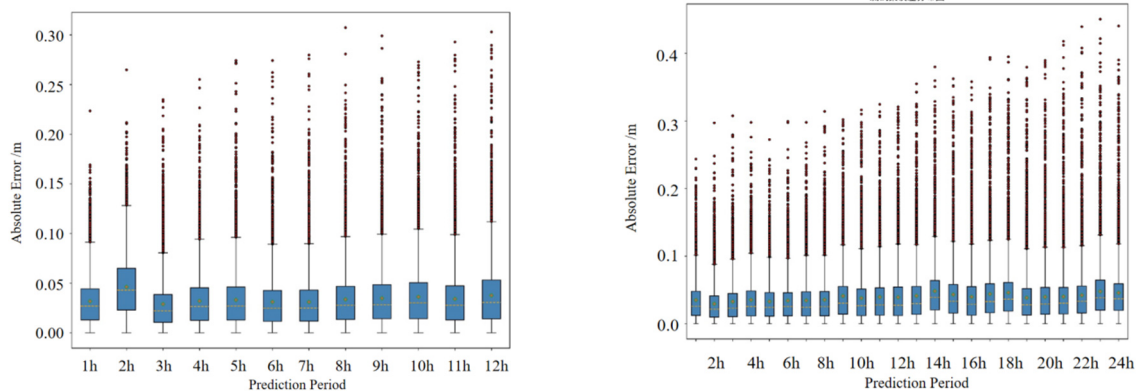
The model of the backwater effect threshold analysis does not incorporate the influence of rainfall within the stream direction interval, and it solely relies on the inflow of the XJB as the input parameter, which hampers its practical applicability. To address this limitation, a revised prediction model is proposed, wherein the inflow of the XJB is substituted with XLD output, XLD abandoned water, and rainfall within the stream direction interval. By adding these variables, the prediction model is rendered more suitable for practical applications. Notable characteristic variables considered in the model encompass the variables presented in Table 2.

Table 2. The content of characteristic variables considered in deep learning models.

Period	Characteristic Variable
Historical data	XJB upstream and downstream water levels, rainfall among XLD and XJB, Pengshan and MJR rainfall, HJR rainfall, HJR run off, MJR run off, XJB inflow, XJB output, XJB abandon flow
Future data	Rainfall among XLD and XJB, XLD and XJB output, XLD and Xiangjiaba abandon flow, HJR run off, MJR run off

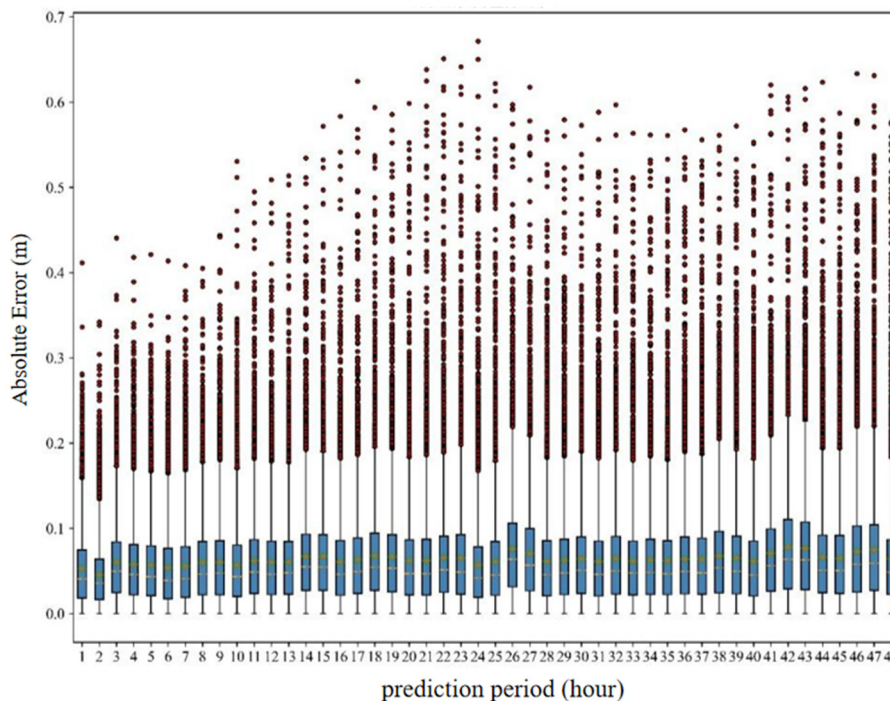
In the analysis, it is recognized that the forecast data used in the model contain inherent errors, and the impact of these errors on water level prediction remains unclear. Additionally, the backwater effect of the HJR and the MJR, as well as the conversion of rainfall into run off and inflow, entail a time lag. Consequently, when the forecast period is short, utilizing historical flow rate and rainfall data from the HJR, MJR, and XJB can enhance the predictive capabilities of the model and enhance its practicality. To compare the predictive performance of different forecast periods, three prediction models for the

XJB's TWL are constructed: a 12 h forecast, a 24 h forecast, and a 48 h forecast, each with a step of 1 h. The error statistics of different forecast periods are shown in Figure 10.



(a) Model error distribution for 12-h forecast period

(b) Model error distribution for 24-h forecast period



(c) Model error distribution for 48 h forecast period

Figure 10. Model error distribution for different hour forecast periods.

The error distribution of each model across different forecast periods shows a similar pattern, without any significant exceptions. Upon comparing the models, it is observed that the 12 h forecast model yields the most accurate predictions overall. It exhibits a mean absolute error of 3.43 cm and a maximum absolute error of 30.74 cm. These values are lower than the corresponding errors obtained from the previous 12 h of the 24 h forecast model, where the mean absolute error was 3.56 cm and the maximum absolute error was 32.47 cm. Similarly, the 24 h forecast model demonstrates a mean absolute error of 3.93 cm and a maximum absolute error of 45.07 cm, which are lower than the corresponding errors obtained from the previous 24 h of the 48 h forecast model, where the mean absolute error was 6.03 cm and the maximum absolute error was 67.14 cm. Therefore, in the final prediction, a comprehensive approach is taken by combining forecasts from three different models. Specifically, the first 12 h forecast is generated using the 12 h prediction model, the 13 to 24 h forecast is obtained from the 24 h prediction model, and the 25 to 48 h forecast is derived from the 48 h prediction model.

Based on the prediction models of different forecast periods, the overall error distribution within the 48 h forecast period is shown in Figure 11.

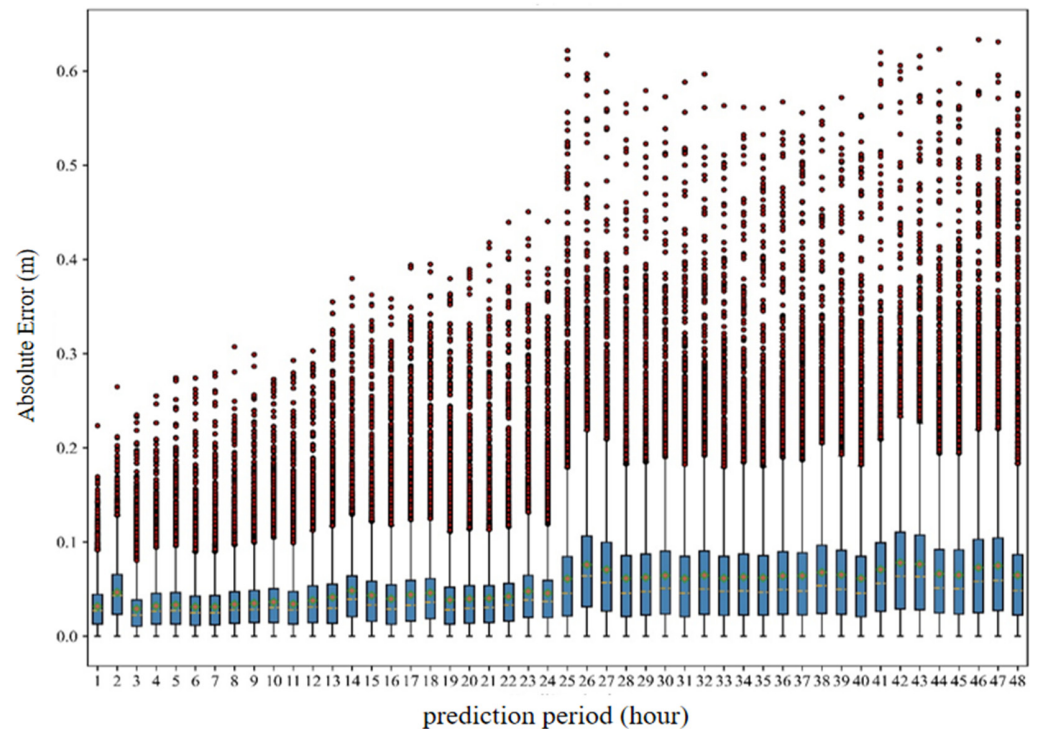


Figure 11. Overall prediction error.

From the graph, specifically, the error in the last 24 h of the forecast shows a noticeable increase compared to the first 24 h. The average absolute error for the initial 24 h is recorded as 3.87 cm, whereas it rises to 6.67 cm for the subsequent 24 h. Similarly, the maximum absolute error also displays an increase, with the value advancing from 45.07 cm in the first 24 h to 63.35 cm in the later period. Overall, when considering the entire range of predictions, the mean absolute error is calculated to be 5.27 cm.

4.3. Result Analysis

Based on the comparison between the WBE (water balance equation) model and the LSTM model and considering the backwater effect, it is evident that the LSTM model performs significantly better in terms of prediction accuracy. The mean absolute error of the LSTM model, which incorporates the backwater effect, is approximately 11%.

To analyze cases where the error exceeds 60 cm, Figure 12a illustrates the comparison between the predicted water level process and the measured water level process. When the prediction start time is delayed by one day, meaning that the prediction starts at 21:00 on 22 July 2019, the comparison between the predicted water level process and the measured process is shown in Figure 12b. It is observed that both the overall error and the maximum water level prediction error decrease. This indicates that utilizing rolling prediction can help mitigate the impact of prediction errors during practical applications.

Figure 13 represents the error distribution of the predicted results between LSTM_Model1, which is the LSTM prediction model without considering the backwater effect constructed in Section 3, and LSTM_Model2, which is the LSTM prediction model considering the backwater effect constructed in this section. The error distribution plot provides a visual comparison of the differences between the predicted values of the two models. It helps identify the variations in the errors present in their predictions. By analyzing this plot, we can understand the impact of incorporating the backwater effect on the accuracy of water level predictions.

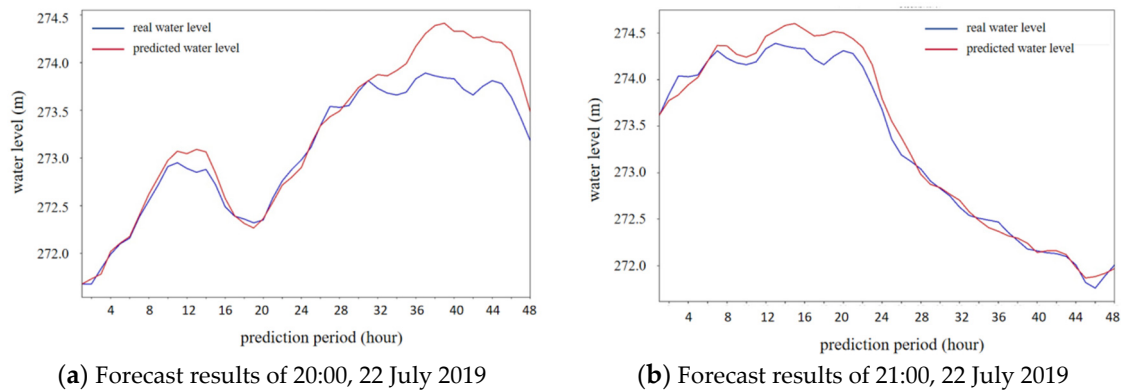


Figure 12. Case analysis of the XJB’s TWL prediction.

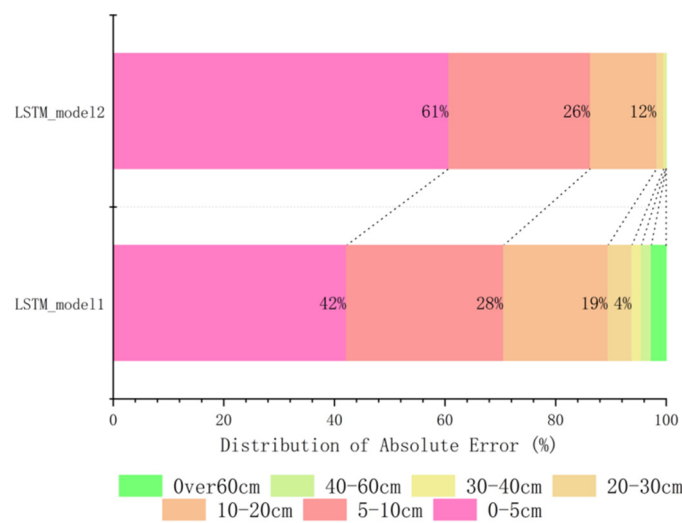


Figure 13. Comparison of prediction errors depending on whether the model considers the backwater effect.

The prediction error of the two models is shown in Table 3, and the maximum absolute error and mean absolute error have significantly decreased after considering the backwater effect. The mean absolute error was reduced from 7.66 cm to 5.272 cm, and the maximum absolute error was reduced from 217.52 cm to 63.35 cm.

Table 3. Error comparison depending on whether the model considers the backwater effect.

	Mean Absolute Error	Maximum Absolute Error
LSTM_Model1	7.66	217.52
LSTM_Model2	5.27	63.35

Indeed, deep learning models, such as the LSTM (Long Short-Term Memory) model, often have significant advantages in accurate predictions. The LSTM model excels in handling time series problems, and it can consider many characteristic variables that are challenging to include in traditional calculation-based models, like the WBE (water balance equation) model. Comparing LSTM_Model1 (without considering the backwater effect) and LSTM_Model2 (considering the backwater effect) helps demonstrate the correctness of analyzing the lag time and threshold values associated with the backwater effect.

The substantial reduction in the maximum absolute error indicates that the backwater effect is a significant contributing factor to large errors in predicting tail water levels (TWLs). By incorporating the backwater effect into the models, the stability and reliability of the

TWL predictions improve significantly. This highlights the importance of considering the backwater effect when aiming to achieve more accurate and robust predictions using the LSTM model.

5. Conclusions

The application of deep learning and machine learning provides flexibility, powerful computational power, and fitting capabilities in various domains. In addition to using these techniques for prediction based on input feature variables, they can also be employed for feature analysis. This paper aims to develop a deep learning model for predicting tail water level (TWL) considering the backwater effect. To accomplish this, it is necessary to identify the conditions that can lead to the backwater effect in the tributaries downstream of the dam. Subsequently, this information is incorporated as part of the flow variables in the deep learning model for training and prediction. The lag time of the backwater effect is analyzed using the random forest model. The threshold analysis involves dividing different scenarios based on the flow rate of downstream tributaries (the HJR and the MJR) of the XJB. By observing in which scenario the prediction error of the tail water level significantly increases, the threshold value can be determined.

After that, the TWL prediction model based on LSTM considering the backwater effect is constructed. Compared to the WBE model, the accuracy of the prediction results is improved by approximately 90%. Additionally, compared to other deep learning models that do not consider the backwater effect, the accuracy is increased by approximately 30%. The prediction error is controlled within 63 cm. This demonstrates that the application of deep learning methods and neural network models holds great promise in the field of water resources and provides a reliable alternative to traditional calculation approaches.

However, it is important to acknowledge that the model also has limitations. When the number of downstream tributaries of the dam increases, the number of scenarios required for backwater effect threshold analysis will also increase significantly. Each scenario necessitates model training and prediction, leading to a substantial increase in the time required for threshold analysis. Consequently, exploring a new method to determine the backwater effect threshold more quickly is an improvement direction of the model and the next goal of this study.

Author Contributions: Methodology, S.Z., S.X., Y.W. and Z.Z.; Software, S.Z. and S.X.; Investigation, S.X., Y.X. and Z.Z.; Resources, B.J.; Data curation, S.X. and Y.X.; Writing—original draft, S.Z.; Writing—review & editing, S.Z., S.X., Y.W., Z.Z. and B.J.; Visualization, S.Z. and S.X.; Supervision, Y.W. and B.J.; Project administration, Y.W. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2022YFC3202300), the National Natural Science Foundation of China (Nos. 42271044, 51979114, 42207526), the Fundamental Research Funds for Central Public Welfare Research Institutes (CKSF2021486), the Scientific Research Project of China Three Gorges Corporation (0704106, ZH21020001), and the Major science and technology project of the Ministry of Water Resources (SKS-2022128), the Natural Science Foundation of Hubei Province (No. 2022CFD164).

Data Availability Statement: The data is unavailable due to privacy restrictions.

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

References

1. Mossa, M.; Petrillo, A.; Chanson, H.; Yausda, Y.; Takahashi, M.; Ohtsu, I. Tailwater level effects on flow conditions at an abrupt drop. *J. Hydraul. Res.* **2005**, *43*, 217–224. [[CrossRef](#)]
2. Ahn, J.; Na, Y.; Park, W.S. Development of Two-Dimensional Inundation Modelling Process using MIKE21 Model. *KSCE J. Civ. Eng.* **2019**, *23*, 3968–3977. [[CrossRef](#)]
3. Zhu, S.; Hrnjica, B.; Ptak, M.; Choiński, A.; Sivakumar, B. Forecasting of water level in multiple temperate lakes using machine learning models. *J. Hydrol.* **2020**, *585*, 124819. [[CrossRef](#)]

4. Li, X.; Liu, B.; Wang, Y.; Yang, Y.; Liang, R.; Peng, F.; Xue, S.; Zhu, Z.; Li, K. The hydrodynamic and environmental characteristics of tributary bay influenced by backwater jacking and intrusion of main reservoir. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 5057–5076. [[CrossRef](#)]
5. Liang, Z. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *J. Hydrol.* **2020**, *581*, 124432. [[CrossRef](#)]
6. Choi, C.; Kim, J.; Han, H.; Han, D.; Kim, H.S. Development of water level prediction models using machine learning in wetlands: A case study of Upo wetland in South Korea. *Water* **2019**, *12*, 93. [[CrossRef](#)]
7. Wang, D.; Qi, X.; Wen, S.; Dan, Y. Robust nonlinear control and svm classifier based fault diagnosis for a water level process. *ICIC Express Lett.* **2015**, *9*, 767–774.
8. Haijiao, D.; Wengang, C. Prediction Model of River Water Level Based on LS-SVM. In Proceedings of the International Conference on Intelligent Computation Technology & Automation, Nanchang, China, 14–15 June 2015; IEEE Computer Society: Los Alamitos, CA, USA, 2015. [[CrossRef](#)]
9. Wang, H.; Song, L. Water Level Prediction of Rainwater Pipe Network Using an SVM-Based Machine Learning Method. *Int. J. Pattern Recognit. Artif. Intell.* **2019**, *34*, 2051002. [[CrossRef](#)]
10. Zhang, J.; Wang, X.; Zhao, C.; Bai, W.; Shen, J.; Li, Y.; Pan, Z.; Duan, Y. Application of cost-sensitive LSTM in water level prediction for nuclear reactor pressurizer—ScienceDirect. *Nucl. Eng. Technol.* **2020**, *52*, 1429–1435. [[CrossRef](#)]
11. Wunsch, A.; Liesch, T.; Broda, S. Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1671–1687. [[CrossRef](#)]
12. Shuofeng, L.; Puwen, L.; Koyamada, K. LSTM Based Hybrid Method for Basin Water Level Prediction by Using Precipitation Data. *J. Adv. Simul. Sci. Eng.* **2021**, *8*, 40–52. [[CrossRef](#)]
13. Zhang, Z.; Qin, H.; Yao, L.; Liu, Y.; Jiang, Z.; Feng, Z.; Ouyang, S.; Pei, S.; Zhou, J. Downstream Water Level Prediction of Reservoir based on Convolutional Neural Network and Long Short-Term Memory Network. *J. Water Resour. Plan. Manag.* **2021**, *147*, 04021060. [[CrossRef](#)]
14. Nourani, V. Reply to comment on ‘Nourani V, Mogaddam AA, Nadiri AO. An ANN-based model for spatiotemporal groundwater level forecasting. *Hydrological Processes* 2008, *22*, 5054–5066’. *Hydrol. Process.* **2010**, *24*, 370–371. [[CrossRef](#)]
15. Kochhar, A.; Singh, H.; Sahoo, S.; Litoria, P.K.; Pateriya, B. Prediction and forecast of pre-monsoon and post-monsoon groundwater level: Using deep learning and statistical modelling. *Model. Earth Syst. Environ.* **2021**, *8*, 2317–2329. [[CrossRef](#)]
16. Litrico, X.; Belaud, G.; Baume, J.P.; Ribot-Bruno, J. Hydraulic modeling of an automatic upstream water-level control gate. *J. Irrig. Drain. Eng.* **2005**, *131*, 176–189. [[CrossRef](#)]
17. Bazartseren, B.; Hildebrandt, G.; Holz, K.P. Short-term water level prediction using neural networks and neuro-fuzzy approach. *Neurocomputing* **2003**, *55*, 439–450. [[CrossRef](#)]
18. Yan, H.U. Reservoir sediment computation of Xiangjiaba hydropower plant. *Yangtze River* **2003**, *4*, 36–38+48.
19. Ma, W.; Nie, D.; Cao, Y. Key Technical Schemes for ±800 kV UHVDC Project from Xiangjiaba to Shanghai. *Power Syst. Technol.* **2007**, *38*, 34–38. [[CrossRef](#)]
20. Lu, P.; Zhou, J.; Wang, C.; Qiao, Q.; Mo, L. Short-term hydro generation scheduling of Xiluodu and Xiangjiaba cascade hydropower stations using improved binary-real coded bee colony optimization algorithm. *Energy Convers. Manag.* **2015**, *91*, 19–31. [[CrossRef](#)]
21. Yang, Y.; Zhang, M.; Sun, Z.; Han, J.; Li, H.; You, X. The relationship between water level change and river channel geometry adjustment in the downstream of the Three Gorges Dam. *J. Geogr. Sci.* **2018**, *28*, 19.
22. Cao, G.; Cai, Z. Research and Application of Optimized Model for Long-term Daily-operation of the Three Gorges & Gezhouba Cascade Power Stations. In *Proceedings of the International Symposium on Three Gorges Project and Water Resources Development and Protection of Yangtze River*; CTGPC: Wuhan, China, 2010.
23. Wang, B.; Liu, S.; Wang, B.; Wu, W.; Wang, J.; Shen, D. Multi-step ahead short-term predictions of storm surge level using CNN and LSTM network. *Acta Oceanol. Sin.* **2021**, *40*, 104–118. [[CrossRef](#)]
24. Jung, S.; Cho, H.; Kim, J.; Lee, J. Prediction of water level in a tidal river using a deep-learning based LSTM model. *J. Korea Water Resour. Assoc.* **2018**, *51*, 1207–1216.

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.