water

MDPI

*Article*

# Influence on the Ecological Environment of the Groundwater Level Changes Based on Deep Learning

**Yu Zhou [1], Lili Zhang [1,2,*], Haoran Li [1], Yunsheng Yao [1], Yaowen Zhang [1,2] and Qixin Wang [3]**

[1] Institute of Ecology and the Environment, Institute of Disaster Prevention, Sanhe 065201, China; zy975656334@163.com (Y.Z.); 22661341@st.cidp.edu.cn (H.L.); yys@cidp.edu.cn (Y.Y.); zhangyaowen@cidp.edu.cn (Y.Z.)
[2] Hebei Key Laboratory of Resource and Environmental Disaster Mechanism and Risk Monitoring, Sanhe 065201, China
[3] Liaoning Water Conservancy and Hydropower Survey and Design Research Institute Co., Ltd., Shenyang 110003, China; wqx16881111@163.com
* Correspondence: zhanglili@cidp.edu.cn

**Abstract:** In recent years, frequent floods caused by heavy rainfall and persistent precipitation have greatly affected changes in groundwater levels. This has not only caused huge economic losses and human casualties, but also had a significant impact on the ecological environment. The aim of this study is to explore the effectiveness of the new method based on Long Short-Term Memory networks (LSTM) and its optimization model in groundwater level prediction compared with the traditional method, to evaluate the prediction accuracy of the different models, and to identify the main factors affecting the changes in groundwater level. Taking Chaoyang City in Liaoning Province as an example, four assessment indicators, $R^2$, MAE, RMSE, and MAPE, were used. The results of this study show that the optimized LSTM model outperforms both the traditional method and the underlying LSTM model in all assessment metrics, with the GWO-LSTM model performing the best. It was also found that high water-table anomalies are mainly caused by heavy rainfall or heavy storms. Changes in the water table can negatively affect the ecological environment such as vegetation growth, soil salinization, and geological hazards. The accurate prediction of groundwater levels is of significant scientific importance for the development of sustainable cities and communities, as well as the good health and well-being of human beings.

**Keywords:** deep learning; water level prediction; ecological environment; sustainable cities and communities; good health and well-being

check for updates

**Citation:** Zhou, Y.; Zhang, L.; Li, H.; Yao, Y.; Zhang, Y.; Wang, Q. Influence on the Ecological Environment of the Groundwater Level Changes Based on Deep Learning. *Water* **2024**, *16*, 3656. https://doi.org/10.3390/w16243656

Academic Editor: Paolo Fabbri

Received: 7 October 2024
Revised: 17 November 2024
Accepted: 17 December 2024
Published: 18 December 2024

**Correction Statement:** This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

**Copyright:** © 2024 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

Groundwater is a crucial source of drinking water and a strategic resource in China, playing a significant role in the country's sustainable development [1]. Groundwater serves as a fundamental resource for agricultural irrigation, drinking water supply, and industrial production, contributing notably to the advancement of local economies [2]. Due to the increasing depletion of groundwater resources, the prediction, management, and rational utilization of these resources have become vital tasks for China. Groundwater level changes are influenced by various factors, including meteorological conditions, topography, aquifer media, and human activities. These factors result in the dynamic behavior of groundwater levels, characterized by complexity, including seasonality, trends, lag effects, and randomness. A large rise in groundwater levels over a short period of time may have multiple ecological impacts, for example:

(1) Vegetation growth pressure: A rapid rise in groundwater level in a short period of time may lead to excessive soil saturation, reducing the air content in the soil and causing plant root hypoxia. Long-term root hypoxia can cause poor plant growth and

even death. In addition, the rise in groundwater level may also alter the original soil moisture balance, affecting the composition and structure of plant communities.

(2) Wetland expansion and its ecological effects: Wetland ecosystems typically rely on the maintenance of groundwater levels. When the groundwater level suddenly rises, it may lead to the expansion of wetland areas and change the hydrological conditions of wetland ecosystems. This change may increase the biodiversity of wetlands in the short term, but if the water level is too high, it may lead to the long-term waterlogging of wetlands, thereby changing their plant and animal community structure.

(3) Intensifying soil salinization: In areas with high salt content in groundwater, a short-term, sharp rise in water level may trigger soil salinization. As groundwater rises, salt is carried to the surface, especially in areas with strong evaporation, and this phenomenon becomes more severe. Soil salinization can damage the growth of crops and vegetation and reduce land productivity.

(4) Changes in surface water quality: The rise in groundwater level will increase the exchange between groundwater and surface water, which may lead to changes in the water quality of surface water. If groundwater contains a high number of dissolved substances, nutrients, or pollutants, these components may enter rivers, lakes, or wetlands, leading to water quality deterioration and ultimately affecting the health of aquatic ecosystems.

(5) Geological hazard risk: The rapid rise in groundwater level in a short period of time may increase the risk of geological hazards in certain areas, such as landslides, mudslides, and ground subsidence. These disasters not only damage natural ecosystems, but may also pose a threat to the human living environment.

(6) Changes in animal habitats: Many animals rely on specific water level conditions for habitat and reproduction. A sudden rise in groundwater level may inundate low-lying areas, alter animal habitats and foraging areas, or force them to migrate or adapt to new environmental conditions, which may lead to a decrease in the population or habitat quality of certain species.

Large increases in groundwater levels over short periods of time can have complex ecological and environmental effects, which may favor the restoration and expansion of certain ecosystems, but more often lead to unfavorable environmental problems. Accurately predicting groundwater level dynamics is, therefore, a complex, nonlinear problem requiring sophisticated mathematical and physical methods for processing [3]. As a result, groundwater level prediction has gradually become a focal point and a challenging issue in the field of hydrogeological research.

Traditional groundwater level-prediction models include conceptual models, physical models, numerical models, and statistical models. Among these, numerical simulation methods based on physical models are commonly used [4], such as MODFLOW and FEFLOW. Ning Ge et al. utilized MODFLOW software (Visual MODFLOW 2011.1) to establish a groundwater-flow model in an over-exploited area, conducting simulations and analyses of groundwater level predictions under various future scenarios. This provided valuable references for the management and evaluation of water resources in over-exploited regions [5]. Similarly, Wen Haiyan et al. developed a groundwater-flow model for the Beiliuhe Water Source Area to predict the groundwater-flow field and drawdown over the next 20 years, determining the optimal extraction rates for each well [6]. Yuan Guangkun used FEFLOW software (Visual MODFLOW 2011.1) to perform numerical simulations of groundwater level dynamics in the study area, predicting the maximum water levels under extreme conditions [7]. Liu Wenlu and colleagues constructed a numerical model of groundwater in the Beijing Plain, forecasting groundwater level rises and their impact, which serves as a reference for optimizing recharge schemes [8].

With the rapid development of artificial intelligence, neural network models have been widely applied in hydrological analysis and prediction [9]. Unlike traditional groundwater level predictions based on physical models, deep learning methods can model groundwater dynamics without requiring a deep understanding of the underlying physical processes.

By analyzing the relationships between input data, these methods capture the nonlinear connections between input and target variables and identify hidden patterns in time-series data [10]. Due to this capability, various deep learning models have been extensively employed in groundwater level-prediction modeling, including Artificial Neural Networks (ANNs) [11], Recurrent Neural Networks (RNNs) [12], Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [13], Graph Neural Networks (GNNs) [14], Convolutional Neural Networks (CNNs) [15], and Long Short-Term Memory networks (LSTM) [16]. Guoyan Xu et al. proposed a novel hydrological-prediction method combining ARIMA and RNNs, which improved the prediction accuracy of water level changes by analyzing the correlation between water levels and environmental factors and validated the model's effectiveness with real data [17]. Yue Zhang explored a flood-prediction approach that integrates Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, incorporating lag time preprocessing to enhance prediction accuracy. This study found that considering the time delay between upstream and downstream hydrological stations significantly improved the model's predictive performance [18]. Gao Linqing et al. introduced a multi-time-scale prediction model based on an Adaptive Inertia Weight Comprehensive Learning Particle Swarm-Optimized LSTM network. The optimized algorithm obtains the optimal parameters for the LSTM network, overcoming the limitations of traditional LSTM networks in parameter selection and prediction accuracy [19]. Additionally, Yang Xingyu et al. utilized a combined grid structure of Convolutional Neural Networks (CNNs) and LSTM networks to extract the temporal dependency of groundwater levels on meteorological factors. Their study concluded that the hybrid CNN-LSTM-ML model outperformed other models in both short-term and long-term prediction accuracy [20].

In summary, the evolution of research on water level prediction from traditional numerical-simulation prediction to machine learning and deep learning demonstrates the importance of technological advances to improve prediction accuracy and efficiency. This study predicts groundwater levels in Chaoyang City, Liaoning Province, using both traditional and deep learning techniques. By comparing the accuracy of traditional and deep learning prediction results, the optimal model is selected to predict groundwater levels under extreme weather conditions and to analyze the impact of groundwater level changes on the ecological environment. The traditional prediction method employs MODFLOW software, while the deep learning approach utilizes LSTM and gray wolf-optimized LSTM (GWO-LSTM) models. The results of this study are of great significance for mitigating the impact of groundwater level changes on the ecological environment under extreme weather conditions in the study area.

## 2. Materials and Methods

### 2.1. Study Area

Chaoyang City is located in the western part of Liaoning Province, China. The city is strategically positioned with access to both inland and coastal areas. It administers Shuangta District, Longcheng District, Beipiao City, Lingyuan City, Chaoyang County, Jianping County, and Kazuo County, covering a total area of 19,736 $km^2$. The region features diverse landforms, with interspersed plains, valleys, hills, and mountains, creating a landscape of alternating ravines and undulating hills. The main rivers in the area include the Qinglong River, Laoha River, Daling River, and Xiaoling River, with precipitation being the primary source of river recharge. The region has a north temperate continental monsoon climate, with cold and dry winters, hot and rainy summers, and strong seasonality, with an average temperature of 5.4–8.7 °C and annual precipitation of 450–580 mm, showing a spatial distribution pattern of low in the northwest and high in the southeast. The soils in the study area are mainly classified into three categories: brown soil, cinnamon soil, and meadow soil. The geographical location of this study area is shown in Figure 1.

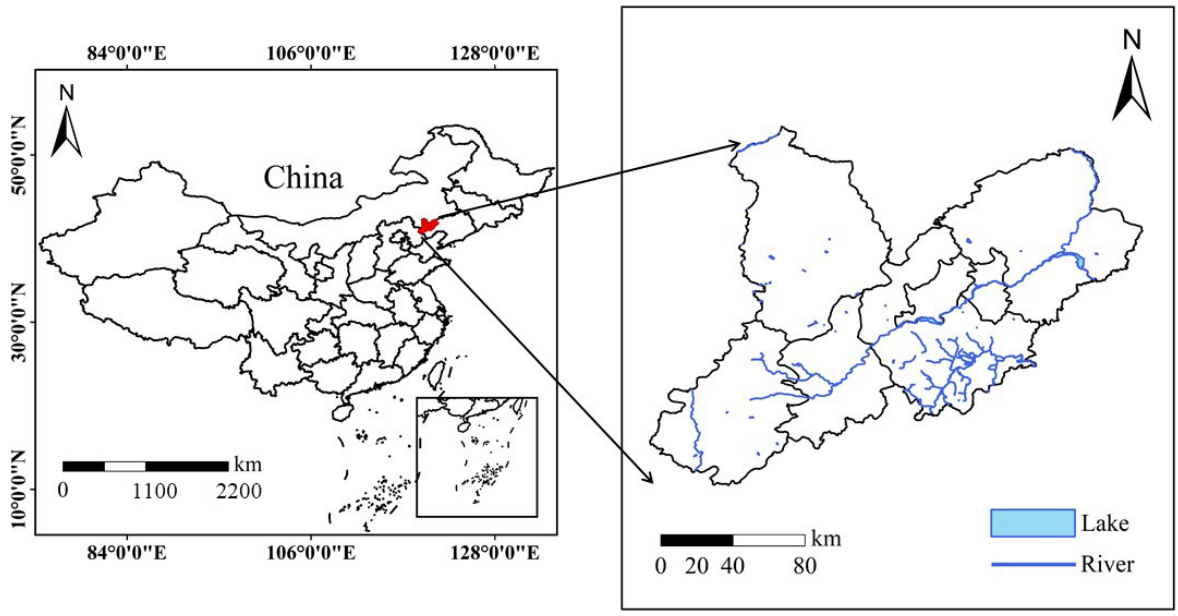



**Figure 1.** Geographical location map of the research area.

### *2.2. Data*

This study utilizes a dataset comprising groundwater level data from observation wells in Chaoyang City and a time series of rainfall data for the same area. The dataset spans from 2017 to 2019, providing a rich set of training and testing samples for the water level-prediction model. The groundwater level data from the observation wells record the daily water level changes in the region, serving as the primary target variable for the model's predictions. The rainfall data, which reflect precipitation levels across different areas, are key factors influencing groundwater levels. The groundwater level data are shown in Figure 2, and the rainfall data are shown in Figure 3.

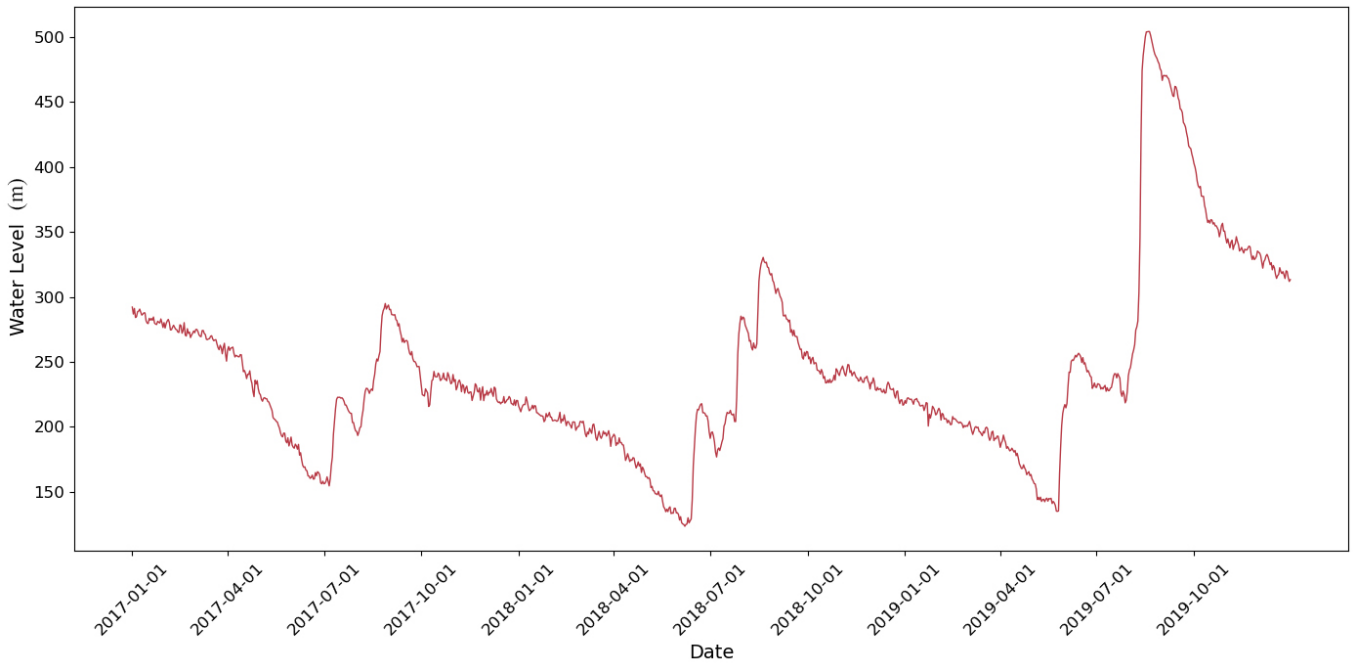



**Figure 2.** Water level changes in Chaoyang well.

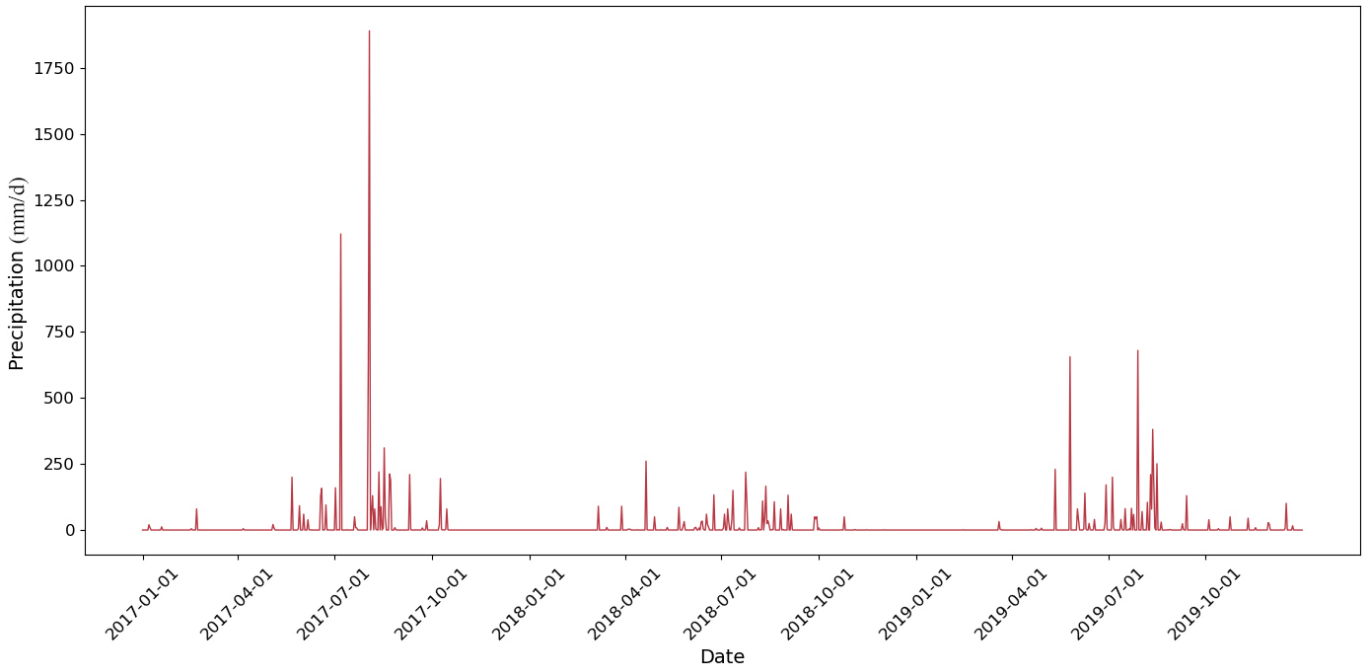



**Figure 3.** Chaoyang precipitation variability map.

During the data-processing phase, to ensure the model's effectiveness and consistency, the groundwater level and rainfall data were first normalized. By scaling all feature values to the range of 0–1, the impact of differing units and value ranges on model training was eliminated. This normalization allows the model to more fairly consider the importance of each feature, aiding in faster convergence and improving prediction accuracy.

### *2.3. Methodology*

#### 2.3.1. MODFLOW

Visual MODFLOW is developed by Waterloo Hydrogeology Inc. of Canada on the basis of the original MODFLOW software by applying modern visualization technology, and it is currently the most popular solute transport simulation and evaluation and 3D groundwater-flow-visualization standard professional software system. The software primarily consists of three components: planar and profile flow line-tracking analysis (Modpath), water flow evaluation (Modflow), and solute transport evaluation (MT3D). It features a highly interactive and user-friendly interface. The interface is designed with three interconnected yet relatively independent modules: the input module, the run module, and the output module [21]. Visual MODFLOW is suitable for evaluating safe groundwater supply, predicting inflow rates, and forecasting groundwater levels. Fan Yulu used MODFLOW to establish a groundwater-flow model for the windy beach area in northern Shaanxi to predict the groundwater dynamics in the coming years, and the correlation coefficients, $R^2$, were in the range of 0.76 to 0.99, with a better groundwater dynamics fit [22]. Luo Bin predicted the groundwater level in Xianyang City, revealing the intra- and inter-annual dynamics of the groundwater level in Xianyang City, with model correlation coefficients of $R^2$ ranging from 0.78 to 0.94 [23].

#### 2.3.2. LSTM (Long Short-Term Memory Model)

The Long Short-Term Memory network (LSTM) is a special type of Recurrent Neural Network (RNN) proposed by Hochreiter [24]. Designed to address the vanishing and exploding gradient problems encountered by traditional RNNs when processing long sequences, LSTM effectively manages the storage, updating, and utilization of information by introducing memory cells ($C_t$) and three key gating mechanisms: the forget gate ($f_t$), the

input gate ($i_t$), and the output gate ($h_t$). These mechanisms enable the network to capture long-term dependencies. The basic network framework of LSTM is illustrated in Figure 4.

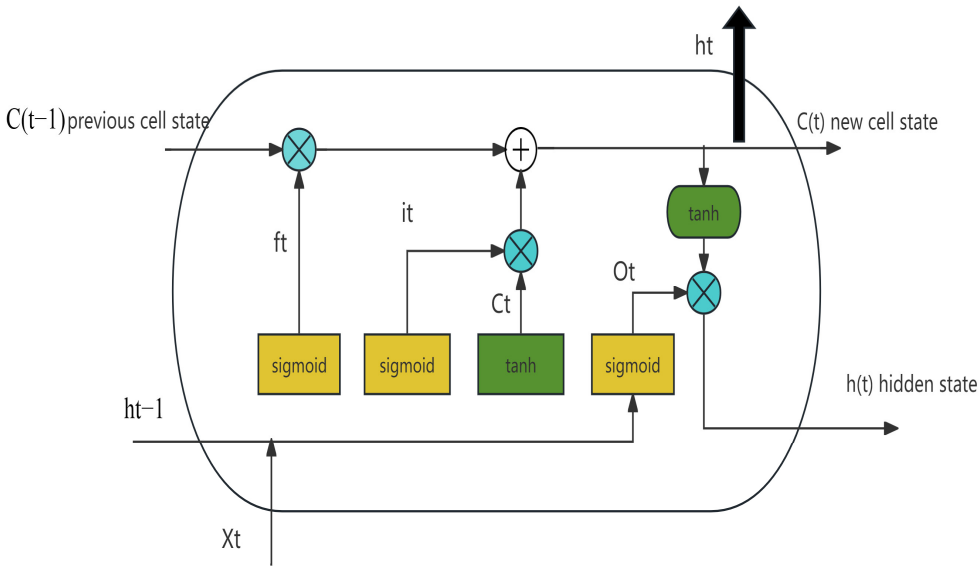



**Figure 4.** Basic network framework of LSTM.

The memory cell is the core component of the LSTM network, responsible for carrying crucial information across time steps in a sequence. Its update is governed by the following equations:

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{1}$$

In Formula (1): $C_t$ is the unit state of the current time step. $C_{t-1}$ is the unit state of the previous time step. $*$ represents dot multiplication, also known as Hadamard product. The updating of memory cells is a combination of old information controlled by the forget gate and new information controlled by the input gate.

The main function of an input gate is to control the flow of new information into memory cells. The input gate consists of two parts: the sigmoid layer, which determines which values need to be updated, and the tanh layer, which creates a new candidate vector to be added to the memory cells. The main formula is as follows:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i \tag{2}$$

$$\hat{C}_t = tanh(W_c * [h_{t-1}, x_t] + b_c \tag{3}$$

In Formulas (2) and (3), $i_t$ is the output of the input gate, $\hat{C}_t$ is the new candidate value that will be added to the cell state, $W_i$ and $W_c$ are the weight matrices of the input gate and candidate value vector, and $b_i$ and $b_c$ are the bias terms of the input gate and candidate value vector, respectively.

The forget gate controls the amount of information discarded from the unit state. It determines which information to retain by observing the current input and the previous hidden state:

$$f_t = \sigma\left(W_f * [h_{t-1}, x_t] + b_f\right) \tag{4}$$

In Formula (4): $f_t$ is the output of the forget gate at time step $t$, $W_f$ is the weight matrix of the forget gate, and $h_{t-1}$ is the hidden state of the previous time step. $x_t$ is the input for the current time step. $b_f$ is the bias term of the forget gate, and σ is the sigmoid activation function, which outputs a number between 0 and 1 indicating how many cell states from the previous time step should be retained. The specific parameters of the model are shown in Table 1.

**Table 1.** LSTM model parameters.

| Model Name | Input Dimensions | Output Dimensions | Key Layer Composition | Optimizer |
|---|---|---|---|---|
| LSTM | (epoch, 6.5) | (epoch, 1) | 2× LSTM, 1× Dense | TensorFlow Adam with learning rate 0.0001 |

#### 2.3.3. FA-LSTM (Firefly Algorithm-Optimized LSTM)

The FA-LSTM model is a hybrid intelligent algorithm model that combines the firefly algorithm (FA) and Long Short-Term Memory (LSTM) networks. The optimization algorithm was proposed by Yang in 2009 [25]. The firefly optimization algorithm optimizes the LSTM network parameters through the global search capability, which improves the performance and prediction accuracy of the model when dealing with complex time-series data.

The firefly algorithm simulates the flashing attraction mechanism of fireflies in nature, where each firefly represents a potential solution in the solution space, and its brightness (fitness) is directly related to the accuracy of water level prediction. Fireflies attract each other based on brightness and move towards a better solution, guiding the optimization process of LSTM parameters through this mechanism. Formula (5) is the formula for updating the position of fireflies in FA:

$$x_i^{(t+1)} = x_i^{(t)} + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^{(t)} - x_i^{(t)}\right) + \alpha(rand - 0.5) \quad (5)$$

Used for iteratively updating the weights and biases of LSTM, where $\beta_0$ is the cardinality of attraction, $\gamma$ controls the rate at which attraction decreases with distance, and $\alpha$ adjusts the randomness of the search step size to effectively explore the solution space. The specific parameters of the model are shown in Table 2.

**Table 2.** FA-LSTM model parameters.

| Model Name | Number of Neurons | Dropout Ratio | Batch Size | Initial Randomness ($\alpha$) | Intensity Absorption Coefficient ($\gamma$) | Minimum Attractiveness ($\beta_0$) |
|---|---|---|---|---|---|---|
| FA-LSTM | 5–300 | 0.0001–0.99 | 2–256 | 0.5 | 1 | 0.20 |

#### 2.3.4. GWO-LSTM (Grey Wolf Optimizer)

The GWO-LSTM model is an innovative model that combines the grey wolf optimizer (GWO) algorithm with the Long Short-Term Memory (LSTM) network. The optimization algorithm was proposed by Mirjalili et al. in 2014 [26]. It utilizes the group collaborative search capability of GWO and the powerful time-series data-processing capability of LSTM, aiming to improve the accuracy and efficiency of predicting complex time-series data, especially data with long-term dependencies and nonlinear features.

The grey wolf optimization algorithm is a swarm intelligence optimization algorithm that simulates the hunting behavior of grey wolf groups. In nature, grey wolves are known for their highly organized and strategic hunting behavior, and the GWO algorithm is based on this behavior pattern. Individuals in the grey wolf pack are divided into four levels: $\alpha$ (leader), $\beta$ (deputy leader), $\delta$ (subordinate), and $\omega$ (observer).

In the GWO-LSTM model, the GWO algorithm is used to optimize the parameters of the LSTM network, including network weights and bias terms. The main formula of the GWO algorithm reflects the mathematical model of grey wolf hunting behavior, mainly including the behavior of wolf packs surrounding, tracking, and attacking prey. Formulas (6)–(9) are the main formulas:

$$\vec{X}(t+1) = \vec{X}_{prey} - \vec{A}\cdot\vec{D} \quad (6)$$

Among them, $\vec{X}(t+1)$ represents the position of the wolf in the next iteration, $\vec{X}_{prey}$ is the position of the prey, and $\vec{A}$ and $\vec{D}$ respectively represent the coefficient vector and distance vector between the wolf and the prey.

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_{prey} - \vec{X} \right| \tag{7}$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r_1} - \vec{a} \tag{8}$$

$$\vec{C} = 2 \cdot \vec{r_2} \tag{9}$$

Among them, $\vec{a}$, $\vec{r_1}$, and $\vec{r_2}$ dynamically decrease with iteration, and $\vec{a}$ linearly decreases from 2 to 0. The specific parameters of the model are shown in Table 3.

**Table 3.** GWO-LSTM model parameters.

| Model Name | Number of Neurons | Dropout Ratio | Batch Size | Input Dimensions | Output Dimensions |
|---|---|---|---|---|---|
| GWO-LSTM | 1–300 | 0.0001–0.99 | 2–256 | (epoch, 10) | (epoch, 1) |

## 3. Research Results and Discussion

In this study, both the MODFLOW software and Long Short-Term Memory (LSTM) models were employed to predict the water levels in Chaoyang City. Additionally, the impact of two different optimization algorithms on the predictive performance of the LSTM network was examined, specifically the firefly algorithm (FA-LSTM) and the gray wolf optimization algorithm (GWO-LSTM). Compared to the traditional LSTM model, these optimization algorithms significantly enhanced the accuracy of the models in time-series prediction tasks. The goal was to identify the optimal model among these enhanced versions to achieve more accurate water level predictions.

### *3.1. MODFLOW Prediction Results*

Based on the collected hydrogeological data, the hydrogeological conditions of the study area were conceptualized, and key parameters such as permeability coefficient, specific yield, rainfall, recharge, and evaporation were determined. A groundwater-flow model for Chaoyang City was then established using these parameters to predict water levels. The generalization of the hydrogeological model takes into account the internal structure of the aquifer, the hydraulic characteristics of the aquifer, the treatment of the boundaries of the study area, and the treatment of the source and sink terms, as shown below:

(1) Generalization of the internal structure of the aquifer

According to the hydrogeological conditions of the study area, the aquifer is regarded as a single submersible aquifer. The bottom plate of the Upper Pleistocene strata is taken as the bottom plate of the aquifer. According to the type, lithology, thickness, and hydraulic conductivity characteristics of the aquifer, the model is generalized to a non-homogeneous isotropic aquifer, which can be locally regarded as homogeneous.

(2) Generalization of hydraulic characteristics of the aquifer

The groundwater level in the study area is subject to certain changes due to the influence of dry and abundant water periods, and the water flow is unsteady, but, in general, the regional groundwater is a laminar movement, and the groundwater seepage conforms to Darcy's law, which can be regarded as an unsteady two-dimensional planar flow.

(3) Boundary processing of the study area

According to the distribution of observation wells (holes) at the boundary of the study area, the time-series function provided by MODFLOW should be used to define the boundary as a 'given head boundary' that changes with time, and it can also be approximated and generalized as a water-isolated boundary of the shallow groundwater system (controlled by the topography, with a short runoff and fast alternation), and the boundary conditions should be determined according to the specific hydrogeological conditions. The boundary conditions depend on the specific hydrogeological conditions, and the value of the water level at the boundary point is determined according to the hydrogeological conditions of the study area and the data from the long-term observation wells of the boundary groundwater level.

(4) Treatment of source and sink items

The source term is mainly considered to have atmospheric precipitation infiltration recharge. Atmospheric precipitation infiltration recharge is partitioned according to the intensity of precipitation infiltration (the riverbed part is not included in the partition), and the intensity of precipitation infiltration is calculated according to the coefficient of precipitation infiltration and the amount of precipitation measured by the survey to calculate the amount of precipitation recharge per unit area.

The main considerations of the sinks are evaporation, agricultural water extraction, industrial water extraction, and groundwater recharge to rivers. Evaporation is calculated according to evaporation intensity; agricultural water extraction includes water for domestic use in villages, which is measured according to population and water use quotas; agricultural irrigation water is zoned according to extraction intensity, which is calculated according to crop area and irrigation quotas; industrial water extraction is calculated according to the amount of extraction surveyed; and groundwater recharge to rivers is treated in the same way as infiltration of rivers for recharge.

The hydrogeological parameters selected for the model are permeability coefficient, water supply degree, and porosity, and the initial values of the above parameters are mainly given according to the hydrogeological tests in the survey. The initial values of the model parameters and the values of the parameters after parameter adjustment are shown in Table 4.

**Table 4.** The value of each parameter before and after the transfer.

| | **Conductivity** | **Specific Yield** | **Porosity** |
|---|---|---|---|
| value of a parameter before modulation | 100 | 0.18 | 0.50 |
| value of a parameter after modulation | 130 | 0.15 | 0.50 |

The period from 20 March 2017, to 22 August 2018, was selected as the model-validation period. Representative observation wells were chosen for the comparison between the calculated water levels and the measured values. The comparison results between the calculated and observed water levels are presented in Figures 5 and 6. In order to make the model-identification results more intuitive, two quantitative metrics for evaluating the accuracy of the model were used, i.e., the root-mean-square error (RMSE) and the correlation coefficient ($R^2$) between the simulated and measured values of the model, as shown in Table 5.

As shown in the above graphs, from the graph of the validation results in March, it can be seen that the maximum error occurs at point G4, which is 9.13 m, and the minimum error occurs at point G6, which is 6.68 m. From the graph of the validation results in August, it can be seen that the maximum error occurs at point G6, which is 0.35 m, and the minimum error occurs at point G2, which is 0.07 m. Combined with the simulation results of validation periods, the correlation coefficient ($R^2$) of the model ranges from 0.78 to 0.95,

indicating that the simulated values of monitoring well-water level have a good correlation with the actual values.

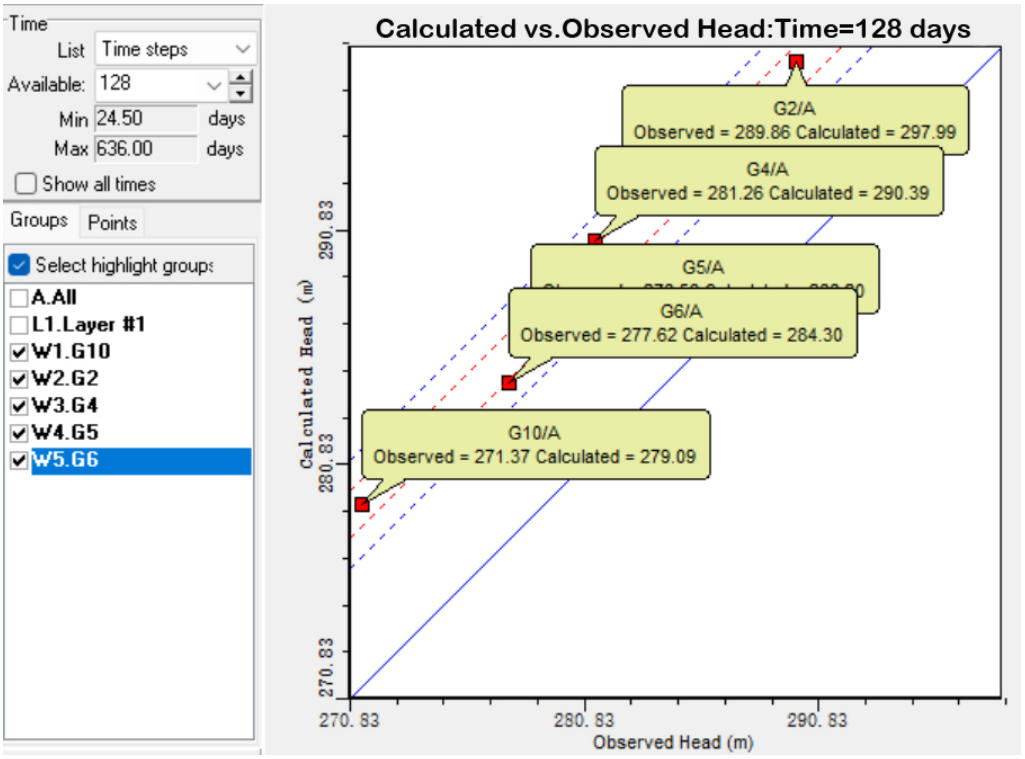



**Figure 5.** Comparison between observed water level and calculated water level on 20 March 2017.

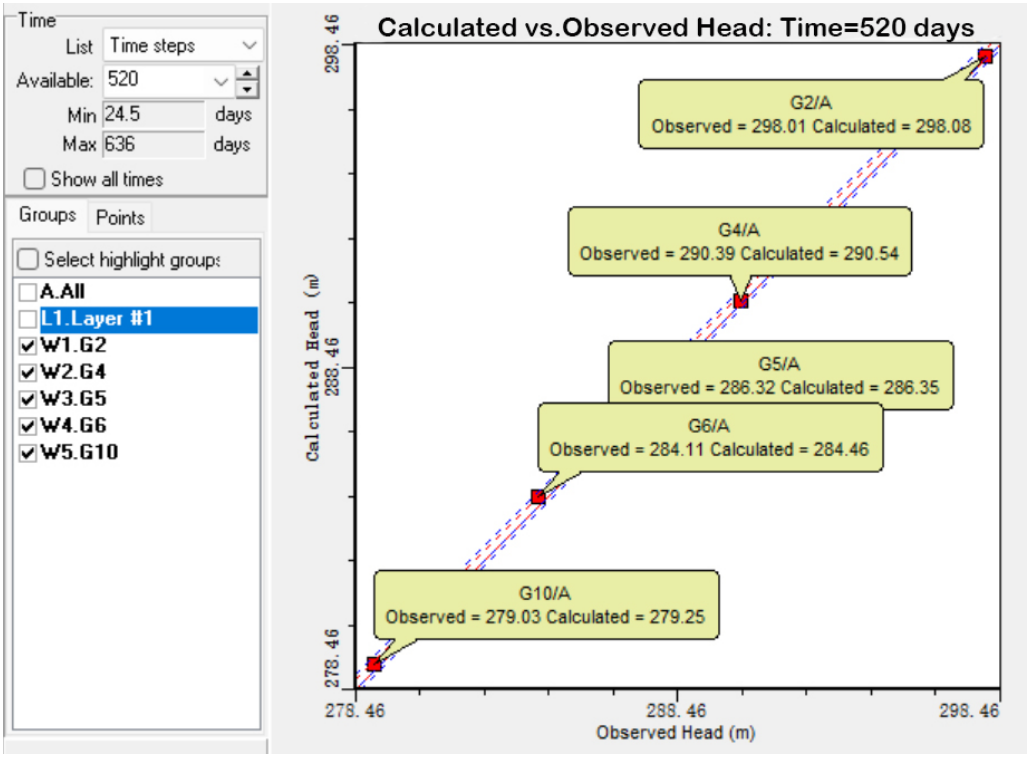



**Figure 6.** Comparison between observed water level and calculated water level on 22 August 2018.

**Table 5.** Root mean square error and correlation coefficient for each observation well.

| Observation Well | RMSE | $R^2$ |
|---|---|---|
| G2 | 6.37 | 0.84 |
| G4 | 7.28 | 0.78 |
| G5 | 6.16 | 0.95 |
| G6 | 5.56 | 0.91 |
| G10 | 5.87 | 0.87 |

### *3.2. Prediction Results of Deep Learning Models*

The groundwater level-prediction results of three models, namely, the Long Short-Term Memory network (LSTM), firefly algorithm (FA-LSTM), and gray wolf optimization algorithm (GWO-LSTM), are shown in Figure 7. The average performance indicators of the model are shown in Table 6.

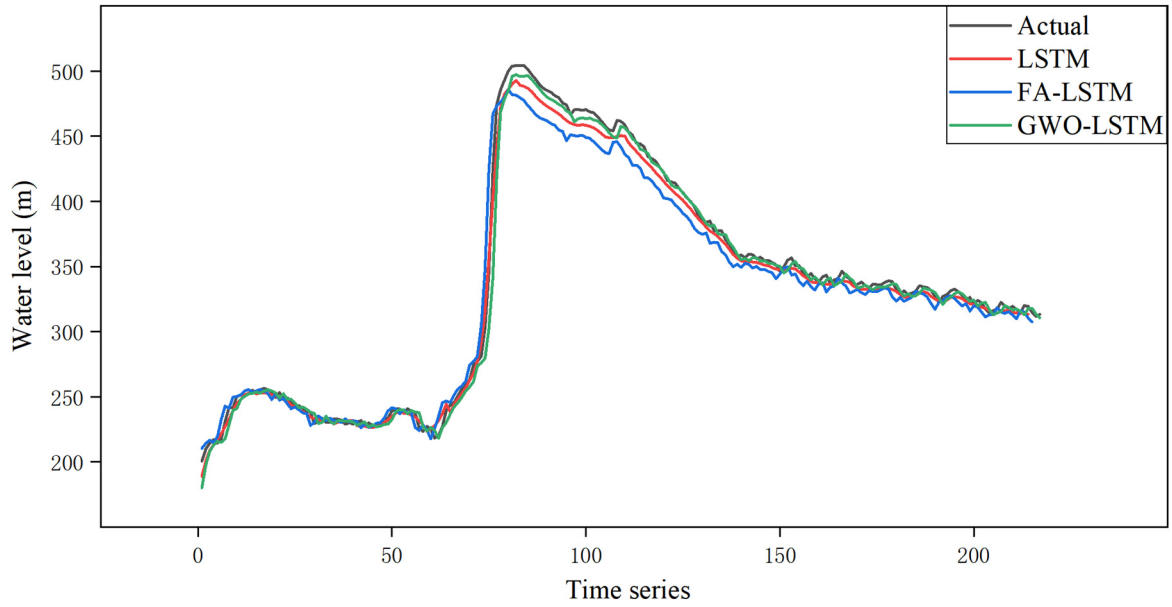



**Figure 7.** Deep learning model result graph.

**Table 6.** Model average indicators.

| Model Evaluation Indicators | $R^2$ | MAE | RMSE | MAPE |
|---|---|---|---|---|
| LSTM | 0.9421 | 8.5282 | 20.4631 | 2.45% |
| FA-LSTM | 0.9810 | 7.8515 | 11.7396 | 2.08% |
| GWO-LSTM | 0.9891 | 4.4729 | 8.9476 | 1.35% |

Figure 7 clearly demonstrates that the LSTM model is unable to effectively capture the rapidly fluctuating water level information, resulting in a significant discrepancy in the predicted high water level. The LSTM model optimized by the firefly algorithm demonstrates enhanced capability in capturing rapidly fluctuating water level information; however, a lag persists, and the prediction of the high water level remains suboptimal. The LSTM model optimized by the grey wolf optimization algorithm demonstrates enhanced precision in both the prediction of high water levels and the capture of rapidly fluctuating water levels.

By comparing and analyzing the data in Table 6, it was found that the optimized model significantly improved performance compared to the base-model LSTM. The FA-LSTM model adopts the brightness-attraction principle of the firefly algorithm to guide

parameter optimization, achieving good prediction accuracy. The $R^2$ value is 0.9810, and the MAE, RMSE, and MAPE values are reduced to 7.8515, 11.7396, and 0.0208, respectively. The GWO-LSTM model achieved the best performance among the three models by adjusting the parameters of social class and hunting strategy through the grey wolf optimization algorithm, with an $R^2$ value of 0.9891, and other performances also reached the optimal level.

Using the FA-LSTM model for water level prediction in the study area can effectively capture the nonlinear characteristics and long-term dependencies of water level changes. The global optimization capability of FA significantly improves the optimization efficiency of LSTM parameters, thereby enhancing the accuracy and reliability of water level prediction.

The GWO algorithm can guide the LSTM network parameters to gradually approach the global optimal solution, thereby improving the accuracy of the model's prediction of time-series data. In the application case of water level prediction in the research area, the GWO-LSTM model effectively captures the complex nonlinear patterns and long-term dependencies of water level changes by precisely adjusting the LSTM parameters, thereby significantly improving the prediction performance.

### *3.3. Discussion*

When comparing the traditional groundwater numerical-simulation prediction model (MODFLOW) and the deep learning model (e.g., LSTM), we note that when the groundwater-flow numerical model is repeatedly adjusted according to the hydrogeological data, the prediction results with higher accuracy can be obtained. At this time, the prediction accuracy of the traditional groundwater-flow numerical model is slightly better than that of the LSTM model, but there is still room for improvement, and the traditional model still has a better application prospect in some cases; the optimization of the LSTM model by the firefly optimization algorithm (FA-LSTM) and the gray wolf optimization algorithm (GWO-LSTM) significantly improves the prediction accuracy. The percentage improvement in model accuracy after optimization is shown in Table 7.

**Table 7.** Degree of improvement in model metrics after optimization.

| Improvement Degree | $R^2$ | MAE | RMSE | MAPE |
|---|---|---|---|---|
| LSTM | 0.9421 | 8.5282 | 20.4631 | 2.45% |
| FA-LSTM | 4.25% | 18.50% | 43.75% | 22.04% |
| GWO-LSTM | 4.99% | 47.55% | 56.27% | 44.89% |

From the table we can see that the firefly-optimized LSTM model improves $R^2$ by 4.25% compared to the before-optimization $R^2$, and MAE, RMSE, and MAPE are reduced by 18.50%, 43.75%, and 22.04% respectively; the gray wolf-optimized LSTM model improves $R^2$ by 4.99% compared to the before-optimization $R^2$, and MAE, RMSE, and MAPE are reduced by 47.55%, 56.27%, and 44.89%. It can be seen that the deep learning method demonstrates high accuracy in groundwater level prediction, showing its potential in groundwater level prediction. The high water level anomalies in the prediction results may be caused by heavy rainfall, and these anomalous high water levels may bring about environmental problems, so the optimized LSTM model can be subsequently used to mark and predict extreme events to prevent the adverse impacts of high water levels on ecological environments and to provide a scientific basis for sustainable development and human health.

## 4. Conclusions

This study compared the accuracy of traditional methods (such as MODFLOW), Long Short-Term Memory Networks (LSTM), LSTM models for groundwater level prediction after firefly optimization (FA) and grey wolf optimization (GWO), with the aim to explore

and evaluate the performance and effectiveness of these models. It was found that deep learning models have good accuracy in groundwater level prediction. Due to the fact that many areas in the real study do not have complete information on hydrogeological parameters, and the establishment of a good hydrogeological model requires complete and accurate data, the groundwater level prediction using MODFLOW does not always have good prediction results. In deep learning, to predict groundwater level, only rainfall and groundwater level data are needed, which greatly reduces the difficulty of data collection, so deep learning has great potential in the field of predicting groundwater level.

This methodology highlights the ability of deep learning techniques to deal with complex spatio-temporal sequence data, especially the role of optimization algorithms in enhancing the optimization of model parameters, among others. Specifically, the results reveal the significant improvement of the LSTM model after GWO (gray wolf optimization) and FA (firefly optimization) compared to the LSTM model in terms of the metrics such as R2, MAE, RMSE, and MAPE, which stems from the optimization algorithm's ability to dynamically capture the optimized parameters of the underlying LSTM model. In particular, the grey wolf-optimized LSTM model demonstrates optimal prediction accuracy and model-generalization ability through the global-optimization capability of the optimization algorithm combined with the global-search capability of the wolf pack.

In summary, this study demonstrates the value of deep learning models, especially the LSTM model after gray wolf optimization (GWO), in groundwater level prediction, and the optimized model not only can effectively capture the non-linear change characteristics of groundwater level, but also has a strong generalization ability. Future research can further explore the integration of more optimization algorithms with deep learning models and verify their applicability to different regions and datasets. This study provides a new method and technical reference for groundwater level prediction, which is of great significance for water resource management and environmental protection.

**Author Contributions:** Y.Z. (Yu Zhou): Data collection, water level prediction, writing initial drafts, conceptual and formal analysis. L.Z.: Supervision, review, editing, and funding acquisition. H.L.: Conceptualization, methodology, and software. Y.Z. (Yaowen Zhang): Formal analysis. Y.Y.: Funding acquisition and data. Q.W.: review. All authors have read and agreed to the published version of the manuscript.

**Funding:** Hebei Key Laboratory of Resource and Environmental Disaster Mechanism and Risk Monitoring (Grant No. FZ248107). This study is financially supported by the National Natural Science Foundation of China (Granted Nos. 41702264 and 42174177).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to our data involves confidentiality, please contact the corresponding author for specific data.

**Acknowledgments:** Thank you to the China Meteorological Administration for providing rainfall data and the China Earthquake Administration for providing well-water level data. Thank you for the funding support from Lili Zhang and Yunsheng Yao.

**Conflicts of Interest:** Author Qixin Wang was employed by the company Liaoning Water Conservancy and Hydropower Survey and Design Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Liu, Y.F.; Chen, H.W.; Zhou, H.; Li, Z.Z. A groundwater level prediction system based on MIDBO-BP. *Comput. Knowl. Technol.* **2024**, *21*, 10–14. Available online: https://oversea.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&filename=DNZS202421003&dbname=CJFDLAST2024 (accessed on 5 October 2024).
2. Tian, L.X.; Qin, H.Q. Multivariable LSTM model-based groundwater level prediction. *J. Arid Land Resour. Environ.* **2024**, *9*, 138–146. Available online: https://oversea.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&filename=GHZH202409015&dbname=CJFDLAST2024 (accessed on 5 October 2024).

3. Wang, W.X.; Wang, X.Q.; Cai, Y.M.; Xuan, W.; Li, W.P.; An, Y.H. Determination of Groundwater Recharge and Discharge Layer by Stratified Groundwater Monitoring Data—A Case Study of Zhangye Basin. *Ground Water* **2019**, *6*, 34–36+44. [CrossRef]
4. Dong, Y.H.; Fu, Y.M.; Wang, L.H.; Wang, J.; Zhang, Q.; Zong, Z.H.; Zhou, Z.C. Regional groundwater flow pattern in Beishan, Hexi Corridor and Qilian Mountain. *Bull. Geol. Sci. Technol.* **2022**, *1*, 79–89. [CrossRef]
5. Ning, G.; Su, D.Y.; Gao, C.J.; Su, A.; Lu, D.X. Numerical simulation and predictive analysis of shallow groundwater in overexploited areas. *J. China Inst. Water Resour. Hydropower Res.* **2024**, *3*, 283–297. [CrossRef]
6. Wen, H.Y.; Wu, H.N. Groundwater Level Prediction of the Beiliuheng Water Source Area in Fengnan District Based on Visual MODFLOW. *Henan Sci. Technol.* **2021**, *20*, 13–15.
7. Yuan, G.K. Groundwater dynamic simulation and maximum water level prediction of Erqi business district in Wuhan. *J. Chongqing Univ.* **2020**, *9*, 109–117.
8. Liu, W.L.; Hao, L.Z.; Yang, Y.; Fan, F.C.; Wang, T.; Li, W.J.; Jin, Y. Simulation and prediction of the impact of ecological water supplement of Chaobai River on groundwater. *Beijing Water* **2022**, *1*, 35–38. [CrossRef]
9. Ardana, P.D.H.; Redana, I.W.; Yekti, M.I.; Simpen, I.N. Prediction of Groundwater Level in The Shallow Aquifer Using Artificial Neural Network Approach. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *837*, 012010. [CrossRef]
10. Chen, H.Y.; Vojinovic, Z.; Lo, W.; Lee, J.W. Groundwater Level Prediction with Deep Learning Methods. *Water* **2023**, *15*, 3118. [CrossRef]
11. Azizi, E.; Yosefvand, F.; Yaghoubi, B.; Izadbakhsh, M.A.; Shabanlou, S. Prediction of groundwater level using GMDH artificial neural network based on climate change scenarios. *Appl. Water Sci.* **2024**, *14*, 77. [CrossRef]
12. Nourani, V.; Khodkar, K.; Paknezhad, N.J.; Laux, P. Deep learning-based uncertainty quantification of groundwater level predictions. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 3081–3107. [CrossRef]
13. Navale, V.; Mhaske, S. Artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) model for Forecasting groundwater level in the Pravara River Basin, India. *Model. Earth Syst. Environ.* **2023**, *9*, 2663–2676. [CrossRef]
14. Bai, T.; Tahmasebi, P. Graph neural network for groundwater level forecasting. *J. Hydrol.* **2023**, *616*, 128792. [CrossRef]
15. Pan, M.; Zhou, H.; Cao, J.; Liu, Y.; Hao, J.; Li, S.; Chen, C.H. Water level prediction model based on GRU and CNN. *IEEE Access* **2020**, *8*, 60090–60100. [CrossRef]
16. Li, H.; Zhang, L.; Zhang, Y.; Yao, Y.; Wang, R.; Dai, Y. Water-Level Prediction Analysis for the Three Gorges Reservoir Area Based on a Hybrid Model of LSTM and Its Variants. *Water* **2024**, *16*, 1227. [CrossRef]
17. Xu, G.; Cheng, Y.; Liu, F.; Ping, P.; Sun, J. A water level prediction model based on ARIMA-RNN. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019; pp. 221–226. [CrossRef]
18. Zhang, Y.; Zhou, Z.; Van Griensven Thé, J.; Yang, S.X.; Gharabaghi, B. Flood Forecasting Using Hybrid LSTM and GRU Models with Lag Time Preprocessing. *Water* **2023**, *15*, 3982. [CrossRef]
19. Gao, L.; Ha, D.; Ma, L.; Chen, J. The prediction model of water level in front of the check gate of the LSTM neural network based on AIW-CLPSO. *J. Comb. Optim.* **2024**, *47*, 5. [CrossRef]
20. Yang, X.; Zhang, Z. A CNN-LSTM model based on a meta-learning algorithm to predict groundwater level in the middle and lower reaches of the Heihe River, China. *Water* **2021**, *14*, 2377. [CrossRef]
21. Zhao, X.N. Introduction to the Application of Visual Modflow in Groundwater Numerical Simulation Research. *Technol. Innov. Appl.* **2017**, *16*, 21–22.
22. Fan, Y.L. Characterization and Numerical Simulation of Groundwater Dynamics in the Wuding River Basin of Windy Beach Area in Northern Shan Xi Province. Master's Thesis, Xi'an University of Technology, Xi'an, China, 2024. [CrossRef]
23. Luo, B. Dynamic Change Characteristics and Numerical Simulation of Groundwater Resources in Xian Yang City. Master's Thesis, Xi'an University of Technology, Xi'an, China, 2021. [CrossRef]
24. Hochreiter, S. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
25. Yang, X.S. Firefly algorithms for multimodal optimization. In *International Symposium on Stochastic Algorithms*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–178.
26. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]


**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.