



Article Predicting Groundwater Indicator Concentration Based on Long Short-Term Memory Neural Network: A Case Study

Chao Liu¹, Mingshuang Xu¹, Yufeng Liu², Xuefei Li¹, Zonglin Pang¹ and Sheng Miao^{2,*}

- ¹ School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao 266033, China
- ² School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266033, China
- * Correspondence: smiao@qut.edu.cn

Abstract: Prediction of groundwater quality is an essential step for sustainable utilization of water resources. Most of the related research in the study area focuses on water distribution and rational utilization of resources but lacks results on groundwater quality prediction. Therefore, this paper introduces a prediction model of groundwater quality based on a long short-term memory (LSTM) neural network. Based on groundwater monitoring data from October 2000 to October 2014, five indicators were screened as research objects: TDS, fluoride, nitrate, phosphate, and metasilicate. Considering the seasonality of water quality time series data, the LSTM neural network model was used to predict the groundwater index concentrations in the dry and rainy periods. The results suggest the model has high accuracy and can be used to predict groundwater quality. The mean absolute errors (MAEs) of these parameters are, respectively, 0.21, 0.20, 0.17, 0.17, and 0.20. The root mean square errors (RMSEs) are 0.31, 0.29, 0.28, 0.27, and 0.31, respectively. People can be given early warnings and take measures according to the forecast situation. It provides a reference for groundwater management and sustainable utilization in the study area in the future and also provides a new idea for coastal cities with similar hydrogeological conditions.

Keywords: groundwater quality; deep learning; predictive modeling; LSTM

1. Introduction

Urbanization and population growth have increased the consumption of water resources [1]. Water resources are more important for water-scarce areas. With the economic development in recent years, the problems of groundwater pollution and over-exploitation have become more significant. As an important component of water resources, groundwater plays an irreplaceable role in social development, especially in arid and semi-arid regions [2]. The quality of groundwater directly affects the living standard of residents and agricultural development [3]. Many scholars have evaluated the groundwater quality in Nordic and Baltic countries, Korea, and Beijing [4–6]. Scholars assessed the quality, groundwater chemical characteristics, and quantity of groundwater by different methods. Affected by geology, pollution indicators, and human factors, the states of groundwater are different. Groundwater resources should be developed and utilized sustainably, avoiding pollution and overuse as much as possible [7].

The long short-term memory (LSTM) neural network is a recurrent neural network (RNN). LSTM is designed to solve the dependency problem of the general neural network [8]. This neural network is widely used in processing the sequence data, such as monitoring data. The regular prediction method are deterministic model, based on the relationship between influencing factors and groundwater quality [9]. Groundwater quality prediction can provide the basis for environmental protection and sustainable development. Different from the traditional prediction methods, this paper combines LSTM to predict the concentration of groundwater indicators.



Citation: Liu, C.; Xu, M.; Liu, Y.; Li, X.; Pang, Z.; Miao, S. Predicting Groundwater Indicator Concentration Based on Long Short-Term Memory Neural Network: A Case Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15612. https://doi.org/10.3390/ ijerph192315612

Academic Editors: Huiping Zeng, Yan Han and Yanan Cai

Received: 6 October 2022 Accepted: 17 November 2022 Published: 24 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). This paper, based on rainfall and indicator concentration monitoring data, established the LSTM to predict groundwater quality. By collecting monitoring data and monthly rainfall data from October 2000 to October 2014 of Wendeng, the subdistrict of Weihai, the characteristic of groundwater was analyzed. According to the characteristics and groundwater utilization, total dissolved solids (TDS), fluoride, nitrate, phosphate, and metasilicate were screened as characteristic factors. LSTM is established, and the five indicators' concentrations can be predicted for the dry and rainy periods. Considering the effect of the five indicators on groundwater quality comprehensively, the coastal groundwater quality and changing trend can be assessed and predicted.

The major contributions of this paper are expressed as follows: (1) According to the characteristics of coastal areas, the LSTM established in this paper is used to predict groundwater quality. The basic data used to establish LSTM are previous long-term monitoring data, which overcomes reliance on real-time monitoring data. (2) Combined with prediction data, the changing trend of quality is realized, and protection measures can be taken in advance. (3) The conclusions of this paper can be applied to coastal cities with similar hydrogeological conditions. It provides a basis for the rational use of groundwater resources and the realization of sustainable development.

2. Related Works

In arid areas, rainfall is sparse and evaporation rates are high, so groundwater is an important local source of drinking water, for agricultural, industrial, and domestic purposes [10]. Groundwater has less evapotranspiration and is less sensitive to contamination. The challenges to global groundwater supply over the past decade have been enormous. On the one hand, the quantity of groundwater is decreasing [11]. Population growth, urbanization, and excessive extraction of groundwater lead to the water level declining severely [12]. Irrigation using groundwater also decreases the available quantity [13]. On the other hand, the quality of groundwater is worse and worse. Human activities affect groundwater quality [14]. Due to the imperfect sewage treatment facilities, wastewater would pollute the surrounding surface water systems and soil [15,16]. The discharge of domestic wastewater and the excessive use of chemical fertilizers aggravate the pollution of groundwater.

In order to evaluate the pollution degree of the environment, the concentration of factors that caused pollution must be identified [17–19]. The groundwater quality index (GWQI) and irrigation water quality index (IWQI) can evaluate groundwater quality for drinking and irrigation purposes [20]. To consider the geographical elements, seepage of surface water pollutants, and other influencing factors, the geographical information system (GIS)-based analytic hierarchy process (AHP) model is applied to predict groundwater quality [21,22].

With the application of computing techniques, different neural network models were developed to evaluate the quantity and quality of groundwater [23,24]. Artificial intelligence (AI) models take advantage of processing nonlinearity data, such as river water quality [25]. Machine learning (ML) models including quantile regression forest (QRF), random forest (RF), radial support vector machine (SVM), stochastic gradient boosting (GBM), and gradient boosting machines are applied to predict water quality (WQ) [26]. Based on the small-scale catchment of Klang River, the novel H_2O deep learning (DL) and RF models prediction of river WQI classification is better [27]. Combined with neural network models, the groundwater quality and changing trend would be predicted in advance [28,29]. In particular, artificial neural network (ANN) models have been applied widely in water quality [30].

In the study area, the residents rely on groundwater for subsistence and agricultural needs [31]. However, due to excessive groundwater extraction and agricultural pollution, groundwater quality is under severe pressure [32]. The state and changing trend of groundwater quality would be revealed in advance so that protection measures will be taken. The ion concentration of groundwater is continuous and stable, and LSTM takes advantage

of processing sequence data. The ion concentrations can be predicted based on LSTM, providing reference basics for groundwater management.

3. Materials and Methods

3.1. Study Area

Wendeng District is bordered by land on the north, west, and east, and by the sea on the south. It is located on the Pacific plate subduction front. The regional tectonic plate is in the SuLu orogenic belt, which is located at the southern end of the Jiaonan-Weihai uplift and Weihai-Rongcheng uplift. From the perspective of stratigraphy, this area belongs to the Ludong stratigraphic division of North China. According to the rainfall and temperature information of the study area, it has a coastline of 156 km with an average annual temperature of 11.5 °C and an average annual precipitation of 762.2 mm. The latitude of the study area ranges from 36 degrees to 37 degrees in the northern hemisphere, with a temperate continental monsoon climate. The precipitation is unevenly distributed in the study area, with summer precipitation accounting for about 70% of that annually. The data studied in this paper are based on groundwater monitoring data from three wells. The locations of the study area, the indicator concentrations of well locations, and the borehole histograms are shown in Figure 1.





Figure 1. The basic information of study area and three wells.

3.2. Index Selection

Due to the shortage of freshwater resources in the study area, the population and industrial and agricultural production are more dependent on groundwater sources. In order to screen indicators on behalf of aspects influencing the quality of groundwater, this paper identifies the beneficial indicators and harmful indicators of water quality.

This paper screened the following indicators to predict and analyze: TDS, fluoride, nitrate, phosphate, and metasilicate. Firstly, dental fluorosis is endemic in around volcanic areas, due to the high fluoride content in daily drinking water. Scholars have analyzed the correlative relationship between dental fluorosis and high fluoride [33]. Secondly, the TDS is screened owing to the agricultural purpose of groundwater in the study area. The study area is coastal, the groundwater depth is shallow, with strong evaporation, TDS will accumulate in the topsoil causing soil salinization [34]. Thirdly, nitrate pollution is caused by the overuse of agricultural nitrogen fertilizers, leaching from municipal landfills, and leakage from industrial wastewater pipelines [35]. According to the International Agency for Cancer Research (IARC), nitrate and nitrite are probable carcinogens [36]. Fourth, according to the news related to the study area, pollution of farmland with superphosphate has occurred. Nitrate and phosphate can assess the impact of urban development and fertilizer use on groundwater quality. Finally, metasilicate is the only positive factor for groundwater quality. According to relevant regulations, when the content of metasilicate is not less than 25 mg/L, it is metasilicate-rich mineral water. As a natural nutritional supplement, metasilicate can help soften blood vessels and promote bone development [37].

3.3. Data Sources and Monitoring Methods

In order to demonstrate the seasonal distribution pattern of rainfall and the distribution pattern of dry and rainy periods, the four seasons were used as the time dimension to count the proportion of rainfall in each season to the annual rainfall from 2000 to 2014, and then to account for the proportion of rainfall during the dry and rainy periods. The seasonal and annual variation patterns of precipitation directly determine whether an accurate ion concentration prediction model can be built. Precipitation data are obtained from the long-term monitoring well clusters. Groundwater sampling was carried out once a month according to the operation specification in the Technical Specification for Groundwater Environmental Monitoring (HJ/T164-2004). After determining the groundwater level, washing the well, and stabilizing the parameters, a 500 mL polyethylene sampling bottle was cleaned 2–3 times using the collected water samples and sealed. The ion concentration detection method and standard are shown in Table 1.

Table 1. Ion concentration de	etection methods and stand	lards.
--------------------------------------	----------------------------	--------

No.	Indicators	Standards	Measurement Method
1	TDS	"Standard examination methods for drinking water—Organoleptic and physical parameters"(GB/T5750.4-2006)	Gravimetric method
2	Fluoride	"Water Quality-Determination of Fluoride-Ion Selective Elec-trode Method"(GB7484-87)	Ion selective electrode method
3	Nitrate	"Water quality—Determination of nitrate-nitrogen—Ultraviolet spectrophotometry" (HJ/T346-2007)	Ultraviolet spectro- photometry
4	Phosphate	"Standard examination methods for drinking water—Nonmental parameters" (GB/T 5750.5-2006)	Molybdenum blue spectrophotometric method
5	Metasilicate	"Drinking natural mineral water test method" (GB8538-2016)	Molybdosilicate blue photometry

5 of 14

3.4. Data Preparation

The data are normalized to eliminate the effect of dimensionality between different metrics. Normalized data make the prediction model training results converge to the optimal solution more quickly. Kolmogorov–Smirnov (KS test) and Shapiro–Wilk (SW test) tests are performed by SPSS software. The results are shown in Table 2, the data are less than 0.05, indicating that the data are not a normal distribution. A more suitable normalization method is the min-max normalization. Min-max normalization is characterized by compressing the data interval between [0,1] without changing the characteristics of the data distribution. It is applicable to data sets determined by maximum and minimum values. The calculation formula is as follows:

$$c' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where *x* is the data to be normalized, x' is the data from normalized, x_{\min} are the minimum value of data, x_{\max} is the maximum value of data.

Table 2. KS test and SW test results of P values.

Methods	TDS	Fluoride	Nitrate	Phosphate	Metasilicate
Kolmogorov–Sminov (KS test)	0.030	0.000	0.000	0.000	0.013
Shapiro–Wilk (SW test)	0.000	0.000	0.000	0.000	0.039

A total of 501 data sets were used in this study, taking into account that water quality time series data are generally seasonal, non-linear, and fuzzy, the water quality time series data are divided into rainy period and dry period data. During May to October of each year is the rainy period, and During November to April of the following year is the dry period.

Considering the impact of water quality by this low value, the prediction of the future dry 6 months of the indicator concentration should be based on the past 12 months of monitoring data. In addition, the impact of water quality by rainfall should also be based on the same month of the previous year or years of water quality monitoring data for prediction, so this selection is for the same month for the data of the past two years. Therefore, the input layer of the model is selected for the 12 months before the prediction month and the same months of the previous two years of water quality monitoring data, the output is the water quality data for a month in the dry or rainy season. The monthly data on water quality include five dimensions: TDS, fluoride, nitrate, phosphate, and metasilicate. Therefore, a single sample in the data set is composed of 5-dimensional water quality data of the predicted month.

According to the monitoring data, the three wells are close and the changing trend of indicator concentrations is similar. Therefore, the training set and the test set are established based on the three wells' monitoring data. After preprocessing the rainfall and monitoring data, 80% of the monitoring data set is randomly screened as the training set and the other 20% as the test set.

3.5. Long Short-Term Memory Neural Network

The LSTM model is a gated neural network, and the existence of a "gate" can realize selective memory of information. It is composed of a Sigmoid neural network layer and dot multiplication operation. When the output is 0, no information is passed at this time. If the output is 1, all the information can pass. LSTM model is shown in Figure 2.

$$f_t = \sigma \Big(W_f[h_{t-1}, x_t] + b_f \Big)$$
⁽²⁾

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$C'_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * C'_t \tag{5}$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = O_t * \tanh(C_t) \tag{7}$$

The operating mechanism of LSTM is that with the continuous replacement of data, the model remembers the effective information, forgets the invalid information, and constantly updates the weight status. Each hidden layer of the LSTM model contains three gates (forgetting gate, input gate, and output gate) and a corresponding cell state (C). Firstly, forgetting information is realized by the forgetting gate (F). When the new data x_t and the output data h_{t-1} of the last moment enter the model, the forgetting gate processes the data with the help of the Sigmoid function controls the degree of data forgetting and then updates the weight of forgetting gate. The next step is to reach the input gate (I), as shown in Equation (3), the information is processed again with the aim of control that can be stored in the cell state. The degree to which the current calculated state is updated to the cell state can be seen. The cell state at the last moment is multiplied by f_t to represent the part to be forgotten, and the new candidate value is represented by $i_t * C_t'$, C_t' to create a new vector of alternative values for a *tanh* shaped network layer, as in Equation (4). The updated cell state C_t is shown in Equation (5). The output gate (O) is then reached, which is used to determine the output content, as shown in Equation (6). The Sigmoid function determines the cell state information to be output, and then the *tanh* function is used to specify the value between -1 and 1. The tanh adjusted value is then multiplied by the output value as in Equation (7) so that the output value is determined by the model. W_f , W_i , W_c , and W_o , represent the corresponding weights, b represents the corresponding offset term, σ represents the Sigmoid function, and *Tanh* represents the hyperbolic tangent activation function.



Figure 2. LSTM model structure.

The LSTM model needs to go through two stages before it is put into use: the training stage and the testing stage. The core goal of the training phase is to find the optimal weight. Assuming that there is a certain time sequence data, the data at this time are taken as the target value, the continuous data before this time is used to train the model, and each weight is constantly updated. The training phase ends when the target value is met. In the test phase, the target value is not set, and the error between the output data at the next moment and the real value is checked under the condition of the optimal weight in the current stage. If the error is small, the test phase is over, and the model in this state can be used for this data prediction.

As shown in the training section of Figure 2, the water quality data of TDS, fluoride, nitrate, phosphate, and metasilicate for 14 months are applied in the model used in this

paper, and the corresponding month's (t_15) output water quality data are used to train model. The input of the LSTM model is set to 14×5 , which corresponds to the five-dimension water quality data received in 14 months. The output is set to 5, which is responsible for the output of the predicted value of five water quality data.

The purpose of LSTM model training is to find a set of optimal model parameters that minimize errors. The water quality prediction model in this paper should be evaluated from the overall performance, hoping that the overall forecast data will be more accurate. Therefore, mean square error (MSE) is selected as the loss function of training. MSE reflects the degree of prediction error through the average sum of the squares of the difference between the predicted data y' and the real measured data y. Compared with the first-order error loss function MAE, the result is closer to the real situation, which is more sensitive to abnormal outliers and has higher requirements for the overall performance of the model. Finally, the test set is applied to the trained model to get the prediction result of the final model.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (y - y')^2$$
(8)

where *y* is the measured value, y' are output values.

Taking into account the difference in groundwater quality between the rainy and dry periods, groundwater quality in the rainy and dry periods was predicted with the help of long short-term memory networks based on the time series of water sample data in the study area.

Applying the model to groundwater concentration indicators in the study area requires determining the step size, the number of implied layers, and the number of data dimensions. The concentration of groundwater indicators in the study area is influenced by the amount of rainfall. The annual variation trend is stable. In addition, the variation trends are similar in the dry and rainy periods. Combined with the monitoring data in the study area, the step size of the prediction model is 14.

The prediction accuracy of the model is low, when the number of hidden layers is too small, while the number of hidden layers is too large, it will overfit. After repeated debugging, the final number of hidden layers is one. Combined with the groundwater state of the study area, this paper screened five indicators for prediction, from the perspective of health (TDS and fluoride), agricultural pollution (nitrate and phosphate), and positive index (metasilicate). The number of dimensions of the prediction model is five.

4. Results and Discussion

4.1. Rainfall Data

The climate type of the study area is a temperate continental monsoon climate. The southeast monsoon blowing from the tropical ocean brings abundant rainfall, there is more rainfall in summer and autumn, while there is less rainfall in winter and spring. The proportion of dry and rainy periods to the total annual rainfall is about 20% and 80%, respectively. The rainfall data are presented in Table 3. Precipitation alternates significantly between rainy and dry periods, with stable water distribution and similar annual distribution. From 2000 to 2014, the proportion of rainfall in rainy and dry periods is essentially constant every year. The stability of seasonal and annual variation is the basis for the concentration prediction model.

No	1#		2#		3#	3#	
rear -	Dry Period	Wet Period	Dry Period	Wet Period	Dry Period	Wet Period	
2000	0.20	0.80	0.21	0.79	0.20	0.80	
2001	0.24	0.96	0.20	0.80	0.24	0.76	
2002	0.16	0.84	0.17	0.83	0.17	0.84	
2003	0.21	0.79	0.19	0.81	0.22	0.79	
2004	0.15	0.84	0.15	0.85	0.15	0.84	
2005	0.16	0.84	0.22	0.78	0.16	0.84	
2006	0.14	0.86	0.10	0.89	0.13	0.86	
2007	0.05	0.95	0.07	0.93	0.05	0.95	
2008	0.29	0.72	0.27	0.73	0.28	0.72	
2009	0.21	0.79	0.18	0.82	0.21	0.79	
2010	0.08	0.92	0.11	0.89	0.07	0.92	
2011	0.27	0.73	0.28	0.72	0.27	0.73	
2012	0.20	0.80	0.21	0.78	0.20	0.80	
2013	0.25	0.75	0.30	0.70	0.25	0.75	
2014	0.26	0.74	0.30	0.71	0.26	0.74	
Average	0.19	0.82	0.20	0.80	0.19	0.81	

Table 3. The proportion of rainfall in each wet and dry period from 2000 to 2014.

4.2. Monitoring Results of Groundwater

Affected by surface water infiltrating, the concentrations of groundwater indicators are changing during both rainy and dry periods. The statistics of groundwater indicator concentrations and the longitude and latitude of wells in rainy period are shown in Table 4 and those in dry period are shown in Table 5.

Names of Wells	Geographical Location	Statistical Indicators	TDS (mg/L)	Fluoride (mg/L)	Nitrate (mg/L)	Phosphate (mg/L)	Metasilicate (mg/L)
1	121.831 °E /36.985 °N	Avg Std. CV (%)	421.44 81.02 19.22	93.64 51.05 54.52	0.17 0.12 73.40	0.04 0.03 64.75	30.86 5.11 16.55
2	121.882 °E /37.063 °N	Avg Std. CV (%)	918.52 351.12 38.23	139.60 126.21 90.41	0.18 0.11 63.05	0.09 0.08 94.91	23.75 6.83 28.74
3	121.886 °E /37.093 °N	Avg Std. CV (%)	698.09 196.32 28.12	100.42 60.81 60.55	0.25 0.09 34.51	0.30 0.19 62.23	21.53 7.37 34.23

Table 4. Statistics of groundwater indicators concentration in the rainy period.

Abbreviations used in this table include: Avg: average; Std. = standard deviation: CV: coefficient of variation (%).

Due to the three wells being close, the changing trends of the indicators are similar. According to Tables 4 and 5, the average indicator concentration in the dry period is higher than that in the rainy period. Standard deviation and coefficient of variation would measure the statistical dispersion of data, the data in the dry period are higher than that in the rainy period, indicating that the indicator concentrations are more similar in the rainy period.

Names of Wells	Longitute/ Latitude	Statistical Indicators	TDS (mg/L)	Fluoride (mg/L)	Nitrate (mg/L)	Phosphate (mg/L)	Metasilicate (mg/L)
1	121.831 °E /36.985 °N	Avg Std. CV (%)	516.90 195.35 37.79	132.22 101.35 76.65	0.25 0.11 44.30	0.04 0.01 40.54	34.84 2.41 6.91
2	121.882 °E /37.063 °N	Avg Std. CV (%)	934.70 359.17 38.43	186.93 135.35 72.41	0.14 0.14 100.56	0.18 0.09 47.17	26.68 4.29 16.08
3	121.886 °E /37.093 °N	Avg Std. CV (%)	764.94 81.64 10.67	138.57 38.90 28.07	0.28 0.21 77.26	0.28 0.14 48.46	26.68 4.71 17.67

Table 5. Statistics of groundwater indicators in the dry period.

Abbreviations used in this table include: Avg: average; Std.: standard deviation; CV: coefficient of variation (%).

4.3. Modeling Result

In order to train and validate the proposed LSTM model, this paper applied Tensor-Flow 2.6 and Python 3.8 to implement the experiments, Tensorflow is one of the most popular machine learning frameworks available today, it can flexibly create complex topological networks and execute the environment for debugging. The training epoch was set to 50 to achieve a convergent model. Considering that large training epochs can cause undesirable overfitting, this paper employed a model with 200 epochs, as shown in Figure 3. The historical graph of the loss function shows that in the rainy and dry periods, the training set rapidly declines and then slowly converges, while the validation set rapidly declines and then remains stable, and the model converges without overfitting. The models were trained on the two-core Intel(R) Xeon(R) Silver 4210R CPU and NVIDIA GeForce RTX3090 GPU server.

The predicted values of TDS, fluoride, nitrate, phosphate, and metasilicate in the rainy season are shown in Figure 4. The predicted values of TDS, fluoride, nitrate, phosphate, and metasilicate in the dry period are shown in Figure 5. In the experiment of this paper, MAE is used to measure the error between the real value and the predicted value of the two models. MAE reflects the degree of prediction deviation through the absolute value of the difference between the predicted data and the measured data. This result can reflect the real deviation of five water quality metrics and has realistic evaluation significance. The errors are shown in Table 6.

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |y - y'|$$
(9)

Table 6. MAE of prediction results based on LSTM.

Indicators (mg/L)	MAE (Rainy Period)	MAE (Dry Period)
TDS	89.45	76.35
Fluoride	0.09	0.06
Nitrate	10.32	5.47
Phosphate	0.23	0.17
Metasilicate	4.21	2.78



Figure 3. LSTM model training results: (a) rainy period; (b) dry period.



Figure 4. Prediction values and true values for the five indicator concentrations in the rainy period.





Figure 5. Prediction values and true values for the five indicator concentrations in the dry period.

According to Table 5, the fitting degree of the five indicators is good, the training of the model for the rainy and dry period has been completed, and the output value of the model can predict the actual concentration. As shown in Figures 4 and 5, the predicted values have similar trends to the true values. In the study area, the true values of the five indicator concentrations changed stably, the concentration in the dry period is higher than that in the wet season. The prediction can reflect the law of change. If the prediction values of TDS are high, it is necessary to pay attention to whether there is a demand for farmland irrigation in the future. Corresponding measures should be taken in time to avoid soil caking caused by slightly saline or saline water, which affects crop yields. If nitrate and phosphate concentrations are predicted to be high, an early warning should be made to note whether the local farming season is approaching and proper planning should be made to avoid excessive use of agricultural fertilizer. When the concentration of metasilicate increases at a certain stable water source point, as a reference, the water source point can be considered as the drinking water source for optimal utilization of the water source.

The model established in this paper can predict groundwater indicator concentration. Compared with traditional methods, such as the MODFLOW, this model does not rely on geological data, and it is simple to carry out. When the indicator concentration data for one consecutive month are available, the model can predict the concentration of the next six months using input data, and the model can continuously output data by circulating input data. The protection measures can be taken in advance, and the predicted values are set as a reference for groundwater protection. Therefore, the model can guide practical work effectively and contribute to the predictability of groundwater quality changes.

4.4. Application of the Concentration Prediction

In this paper, the LSTM neural network model is established and applied to predict the concentrations of five indicators in three specialized observation wells. The prediction values are based on the changing principles of previous indicators' concentrations, indicating that the model could predict the indicators' concentrations for the next period. The predicted changing trends of five indicators are shown in Figure 6. From November to February, the concentrations of TDS, nitrate, and metasilicate would fluctuate, and the maximum occurred in February. Due to the quantity of groundwater decreasing, there is less rainwater supply, leading to the concentration increase. The concentration would increase by various degrees from March to April. Compared with TDS and nitrate, the concentration of metasilicate is more stable. The concentrations of fluoride and phosphate would increase from November to March, and decrease from March to April.



Figure 6. The predicted concentration values of five indicators during the next period.

5. Conclusions

Groundwater, as a main water source, is important for residents' life. The quality of groundwater plays an irreplaceable role in social development. Groundwater sampling is difficult, causing the monitoring data to not to reflect the current state. Quality prediction

results could provide a reference to the administrative department. In this paper, Wendeng District was selected as a study area, and the five indicators were screened, including TDS, fluoride, nitrate, phosphate, and metasilicate. This paper proposed a prediction model based on LSTM, combined with the characteristics of the study area. Based on three wells' monitoring data over fifteen years, groundwater quality could be predicted. Affected by surface water infiltrating, the concentrations of groundwater indicators change during dry and rainy periods.

In this paper, prediction models of dry and rainy periods were established. MAE is used to measure the errors between the true values and the predicted values. According to the results of LSTM, MAE is low, indicating that the accuracy of the prediction model was high. This model could reflect the groundwater quality changes accurately. Applying the dry period prediction model, the concentration of five indicators was predicted. Prewarning system could be established based on prediction data, abnormal changes will be forecasted. Considering the present and prediction stats of groundwater quality comprehensively, protection measures could be taken in advance.

In future works, more indicators could be taken into consideration, in order to assess groundwater quality more comprehensively. Moreover, the indicators' concentrations are visible by Arcgis.

Author Contributions: Conceptualization, C.L., M.X.; Data acquisition and curation, writing—original draft, C.L.; Modeling and validation, Y.L.; Methodology, writing—review & editing, M.X., X.L., Z.P., S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

- 1. Song, M.; Tao, W.; Shang, Y.; Zhao, X. Spatiotemporal characteristics and influencing factors of China's urban water resource utilization efficiency from the perspective of sustainable development. *J. Clean. Prod.* **2022**, *338*, 130649. [CrossRef]
- Wang, W.; Zhang, Z.; Yin, L.; Duan, L.; Huang, J. Topical Collection: Groundwater recharge and discharge in arid and semi-arid areas of China. *Hydrogeol. J.* 2021, 29, 521–524. [CrossRef]
- Bouimouass, H.; Fakir, Y.; Tweed, S.; Sahraoui, H.; Leblanc, M.; Chehbouni, A. Traditional irrigation practices sustain groundwater quality in a semiarid piedmont. *Catena* 2022, 210, 105923. [CrossRef]
- 4. Kitterod, N.O.; Kvaerner, J.; Aagaard, P.; Arustien, J.; Biks, J.; Dagestad, A.; Gundersen, P.; Hansen, B.; Hjartarson, A.; Karro, E.; et al. Hydrogeology and groundwater quality in the Nordic and Baltic countries. *Hydrol. Res.* 2022, *53*, 958–982. [CrossRef]
- 5. Ha, K.; Lee, E.; An, H.; Kim, S.; Park, C.; Kim, G.B.; Ko, K.S. Evaluation of Seasonal Groundwater Quality Changes Associated with Groundwater Pumping and Level Fluctuations in an Agricultural Area, Korea. *Water* **2021**, *13*, 51. [CrossRef]
- 6. Zhang, Y.; Jia, R.; Wu, J.; Wang, H.; Luo, Z. Uncertain in WQI-based groundwater quality assessment methods: a case study in east of Beijing, China. *Environ. Earth Sci.* 2022, *81*, 1–16. [CrossRef]
- Aladejana, J.A.; Kahn, R.M.; Sentenac, P.; Hassan, I. Assessing the Impact of Climate Change on Groundwater Quality of the Shallow Coastal Aquifer of Eastern Dahomey Basin, Southwestern Nigeria. *Water* 2020, 12, 224. [CrossRef]
- 8. Yan, R.; Liao, J.; Yang, J.; Sun, W.; Nong, M.; Li, F. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* **2021**, *169*, 114513. [CrossRef]
- 9. Xie, W.; Ren, B.; Hursthouse, A.; Wang, Z.; Luo, X. Simulation of Manganese Transport in Groundwater Using Visual MODFLOW: a Case Study from Xiangtan Manganese Ore Area in Central China. *Pol. J. Environ. Stud.* **2021**, *30*, 1409–1420. [CrossRef]
- 10. Bierkens, M.F.P.; Wada, Y. Non-renewable groundwater use and groundwater depletion: A review. *Environ. Res. Lett.* **2019**, *14*, 063002. [CrossRef]
- 11. Ostad-Ali-Askari, K.; Shayannejad, M. Quantity and quality modelling of groundwater to manage water resources in Isfahan-Borkhar Aquifer. *Environ. Dev. Sustain.* **2021**, *23*, 15943–15959. [CrossRef]
- Haghnazar, H.; Johannesson, K.H.; Gonzalez-Pinzon, R.; Pourakbar, M.; Aghayani, E.; Rajabi, A.; Hashemi, A.A. Groundwater geochemistry, quality, and pollution of the largest lake basin in the Middle East: Comparison of PMF and PCA-MLR receptor models and application of the source-oriented HHRA approach. *Chemosphere* 2022, 288, 132489. [CrossRef]

- Koh, E.H.; Hyun, B.S.; Lee, E.; Kim, M.C.; Kang, B.R.; Park, W.B.; Jun, S.C. Crop Field Level Estimation of Nitrogen Input from Fertilizer Use in Jeju Island, South Korea: Management Methods to Prevent Groundwater NO3-N Contamination. *Water* 2021, 13, 2715. [CrossRef]
- 14. Wang, Z.; Li, Z.; Zhan, H.; Yang, S. Effect of long-term saline mulched drip irrigation on soil-groundwater environment in arid Northwest China. *Sci. Total. Environ.* **2022**, *820*, 153222. [CrossRef]
- Liu, G.; Ye, J.; Chen, Y.; Yang, X.; Gu, Y. Analysis of Water Pollution Causes and Control Countermeasures in Liaohe Estuary via Support Vector Machine Particle Swarm Optimization under Deep Learning. CMES Comput. Model. Eng. Sci. 2022, 130, 315–329. [CrossRef]
- El Mountassir, O.; Bahir, M.; Ouazar, D.; Chehbouni, A.; Carreira, P.M. Temporal and spatial assessment of groundwater contamination with nitrate using nitrate pollution index (NPI), groundwater pollution index (GPI), and GIS (case study: Essaouira basin, Morocco). *Environ. Sci. Pollut. Res.* 2022, 29, 17132–17149. [CrossRef] [PubMed]
- 17. Yafouz, A.; AlDahoul, N.; Birima, A.H.; Ahmed, A.N.; Sherif, M.; Sefelnasr, A.; Allawi, M.F.; Elshafie, A. Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alex. Eng. J.* **2022**, *61*, 4607–4622. [CrossRef]
- 18. Shi, L.; Zhang, H.; Xu, X.; Han, M.; Zuo, P. A balanced social LSTM for PM2.5 concentration prediction based on local spatiotemporal correlation. *Chemosphere* **2022**, *291*, 133124. [CrossRef]
- 19. Moghaddam, H.K.; Rajaei, A.; Rahimzadeh kivi, Z.; Moghaddam, H.K. Prediction of qualitative parameters concentration in the groundwater resources using the Bayesian approach. *Groundw. Sustain. Dev.* **2022**, *17*, 100758. [CrossRef]
- 20. Orellana-Macias, J.M.; Perles Rosello, M.J. Assessment of Risk and Social Impact on Groundwater Pollution by Nitrates. Implementation in the Gallocanta Groundwater Body (NE Spain). *Water* **2022**, *14*, 202. [CrossRef]
- 21. Cao, L.; Huang, L.; Zhao, P.; Liu, Y. Study on the Whole Process Simulation and Regulation Mechanism of Urban Green Ecological Rainwater Drainage System. *CMES Comput. Model. Eng. Sci.* 2021, 129, 787–803. [CrossRef]
- 22. Jenifer, M.A.; Jha, M.K.; Khatun, A. Assessing Multi-Criteria Decision Analysis Models for Predicting Groundwater Quality in a River Basin of South India. *Sustainability* **2021**, *13*, 6719. [CrossRef]
- Abba, I, S.; Hadi, S.J.; Sammen, S.S.; Salih, S.Q.; Abdulkadir, R.A.; Pham, Q.B.; Yaseen, Z.M. Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *J. Hydrol.* 2020, 587, 124974. [CrossRef]
- 24. Bayatvarkeshi, M.; Imteaz, M.A.; Kisi, O.; Zarei, M.; Yaseen, Z.M. Application of M5 model tree optimized with Excel Solver Platform for water quality parameter estimation. *Environ. Sci. Pollut. Res.* **2021**, *28*, 7347–7364. [CrossRef] [PubMed]
- Tiyasha.; Tung, T.M.; Yaseen, Z.M. A survey on river water quality modelling using artificial intelligence models: 2000–2020. J. Hydrol. 2020, 585, 124670. [CrossRef]
- 26. Al-Sulttani, A.O.; Al-Mukhtar, M.; Roomi, A.B.; Farooque, A.A.; Khedher, K.M.; Yaseen, Z.M. Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction. *IEEE Access* **2021**, *9*, 108527–108541. [CrossRef]
- 27. Tiyasha.; Tung, T.M.; Yaseen, Z.M. Deep Learning for Prediction of Water Quality Index Classification: Tropical Catchment Environmental Assessment. *Nat. Resour. Res.* 2021, *30*, 4235–4254. [CrossRef]
- 28. Deng, Y.; Zhou, X.; Shen, J.; Xiao, G.; Hong, H.; Lin, H.; Wu, F.; Liao, B.Q. New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloketones in tap water. *Sci. Total. Environ.* **2021**, 772, 145534. [CrossRef]
- Huang, X.; Gao, L.; Crosbie, R.S.; Zhang, N.; Fu, G.; Doble, R. Groundwater Recharge Prediction Using Linear Regression, Multi-Layer Perception Network, and Deep Learning. *Water* 2019, 11, 1879. [CrossRef]
- Vijay, S.; Kamaraj, K. Prediction of Water Quality Index in Drinking Water Distribution System Using Activation Functions Based Ann. Water Resour. Manag. 2021, 35, 535–553. [CrossRef]
- Liu, W.; Zhao, C.; Zhou, Y.; Xu, X. Modeling of Vapor-Liquid Equilibrium for Electrolyte Solutions Based on COSMO-RS Interaction. J. Chem. 2022, 2022, 1–13. [CrossRef]
- 32. Siarkos, I.; Sevastas, S.; Mallios, Z.; Theodossiou, N.; Ifadis, I. Investigating groundwater vulnerability variation under future abstraction scenarios to estimate optimal pumping reduction rates. *J. Hydrol.* **2021**, *598*, 126297. [CrossRef]
- Miranda-Rius, J.; Brunet-Llobet, L.; Lahor-Soler, E.; Mrina, O.; Mashala, E.I.; Mahande, M.J. Periodontal and dental conditions of a school population in a volcanic region of Tanzania with highly fluoridated community drinking water. *Afr. Health Sci.* 2020, 20, 476–487. [CrossRef]
- 34. Liu, W.; Xu, X.; Lu, F.; Cao, J.; Li, P.; Fu, T.; Chen, G.; Su, Q. Three-dimensional mapping of soil salinity in the southern coastal area of Laizhou Bay, China. *Land Degrad. Dev.* **2018**, *29*, 3772–3782. [CrossRef]
- 35. Hu, Y.; You, M.; Liu, G.; Dong, Z. Distribution and potential health risk of nitrate in centralized groundwater sources of Wanbei Plain, Central China. *Aqua Water Infrastruct. Ecosyst. Soc.* **2021**, *70*, 684–695. [CrossRef]
- Temkin, A.; Evans, S.; Manidis, T.; Campbell, C.; Naidenko, O.V. Exposure-based assessment and economic valuation of adverse birth outcomes and cancer risk due to nitrate in United States drinking water. *Environ. Res.* 2019, 176, 108442. [CrossRef]
- 37. Li, Y.; Bian, J.; Li, J.; Ma, Y.; Auguiano, J.H.H. Hydrochemistry and stable isotope indication of natural mineral water in Changbai Mountain, China. *J. Hydrol. Reg. Stud.* **2022**, *40*, 101047. [CrossRef]