



Article

Water-Quality Assessment and Pollution-Risk Early-Warning System Based on Web Crawler Technology and LSTM

Guoliang Guan ¹, Yonggui Wang ¹ , Ling Yang ¹, Jinzhao Yue ¹, Qiang Li ¹ , Jianyun Lin ² and Qiang Liu ^{3,*}

¹ Department of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China

² Ningbo Ligong Environment and Energy Technology Co., Ltd., Ningbo 315800, China

³ Sichuan Province Environmental Monitoring Station, Chengdu 610091, China

* Correspondence: futureliuqiang@163.com

Abstract: The openly released and measured data from automatic hydrological and water quality stations in China provide strong data support for water environmental protection management and scientific research. However, current public data on hydrology and water quality only provide real-time data through data tables in a shared page. To excavate the supporting effect of these data on water environmental protection, this paper designs a water-quality-prediction and pollution-risk early-warning system. In this system, crawler technology was used for data collection from public real-time data. Additionally, a modified long short-term memory (LSTM) was adopted to predict the water quality and provide an early warning for pollution risks. According to geographic information technology, this system can show the process of spatial and temporal variations of hydrology and water quality in China. At the same time, the current and future water quality of important monitoring sites can be quickly evaluated and predicted, together with the pollution-risk early warning. The data collected and the water-quality-prediction technique in the system can be shared and used for supporting hydrology and in water quality research and management.



Citation: Guan, G.; Wang, Y.; Yang, L.; Yue, J.; Li, Q.; Lin, J.; Liu, Q. Water-Quality Assessment and Pollution-Risk Early-Warning System Based on Web Crawler Technology and LSTM. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11818. <https://doi.org/10.3390/ijerph191811818>

Academic Editor: Yan Li

Received: 18 August 2022

Accepted: 14 September 2022

Published: 19 September 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/>).

Keywords: water quality evaluation; pollution risk; water-quality early-warning system; machine learning; web crawler; LSTM

1. Introduction

Environmental monitoring is a fundamental component of environmental protection and a key tool for advancing ecological development. China has had an established national surface-water monitoring network since 1991. Since 1999, it has begun to conduct pilot automatic monitoring of the surface-water quality in some river basins [1]. Currently, it has thousands of national surface-water environmental-quality-monitoring sections, forming a national surface-water environmental-quality-monitoring network, which established a monitoring mode. This combines the manual monitoring of collection and separation with the automatic monitoring of water quality, and discloses these real-time monitoring data [2]. It is of great significance to fully dig out the value of real-time monitoring data of national surface-water quality, which provides a better comprehension of the changing trends in China's water environment [3].

With the rapid development of the global network, a large amount of information and data has flooded into the Internet [4]. Understanding how to effectively acquire useful information from network public data has emerged as a main issue. Web crawler technology was created in order to extract the relevant data in batches from the vast amount of data on the Internet automatically, and it is currently widely used in search engines, large-scale data mining and analysis, artificial intelligence and machine learning dataset production, among other fields [5]. A lot of methods have been created for data crawling and gathering from the web, such as the autonomous crawling and analysis tool from the dark web [6], the query-based crawler [7] and the web-crawled literature [8]. It is feasible

to readily and swiftly obtain real-time monitoring data from the opened national surface-water quality data through the robust data-crawling, processing and analysis capabilities of a web crawler.

In reality, several researchers are now using the national surface-water-quality-monitoring dataset as a data source and a support for their work. For example, the evolution of water quality in Ma'an City [9], Zhujiang River [10] and Taizi River [11] have been analyzed by crawled data from the open dataset. In recent years, the spatial distribution of river-water quality and its affecting factors at a national level has been conducted based on these data [12]. The aforementioned research demonstrates that, although the national surface-water quality-monitoring dataset has been fully utilized in the assessment of surface-water environmental quality and the analysis of historical water-quality changes, they are rarely used in the field of surface-water quality prediction [13]. In reality, employing historical data to predict and analyze water quality has substantial practical implications for successfully reducing the likelihood of water contamination incidents, assisting with environmental management and protection, etc. Massive, high-frequency and real-time monitoring data should not only participate in the evaluation, assessment and ranking of surface-water environmental quality, but it should also be applied to the early warning and prediction of water environmental quality [14,15]. Water-quality assessment and prediction are the basis of water-pollution prevention, as well as control and environmental management. Meanwhile, the statistics on water quality in the entire nation and in various regions are currently missing from China's public datasets on water quality, and the visualization quality is subpar. On the other hand, forecasting capabilities are lacking in the current national data sharing system for water quality monitoring.

Numerical models are used extensively in water quality prediction and early warning [16,17]. However, numerical models need a lot of condition data, which adds a lot of labor. Artificial neural networks (ANNs) have the following characteristics: strong nonlinear mapping ability, high learning accuracy and strong robustness. They represent a better modeling method for water quality prediction nationwide. Among these, long short-term memory (LSTM) networks have a good performance in processing data with time series characteristics [18]. They address the issues of gradient explosions and the gradient disappearance of conventional recurrent networks; LSTM networks are presently used in many fields, such as trend predictions of COVID-19 [19], sudden change simulations in financial markets [20] and water quality prediction [21,22]. This research has verified that the LSTM model has better prediction accuracy than the traditional numerical models. The LSTM neural network has been utilized extensively in the field of water environment, although the majority of these networks are only employed in theoretical research projects and cannot be integrated with monitoring data for real-time prediction, which has no practical application. In fact, pollution control, environmental governance and environmental protection all benefit significantly from real-time prediction using monitoring data. As a result, it is necessary to build a real-time water-quality-prediction system with the help of the LSTM neural network.

Artificial neural network technology has been widely used in the field of environmental management and prevention. Developing a real-time monitoring and early-warning system by fusing it with water quality monitoring with the aid of the neural network's potent learning capabilities and the help of the powerful learning ability of the neural network is a popular research issue. In fact, a number of established water-quality-prediction and early-warning systems have surfaced both domestically and internationally [23]. However, the majority of these systems concentrate on water quality monitoring and early warnings in local water and economic breeding fields, rather than establishing a national water-quality-prediction and early-warning system. Therefore, this paper constructs a water-quality-prediction and early-warning system, using web crawler technology to obtain real-time monitoring data from national surface-water monitoring stations and combining the LSTM neural network model to provide water-quality-prediction and early-warning services. Comparatively speaking, this system makes full and effective use of

the national water-quality-monitoring data and establishes a relatively complete national water-quality-prediction and early-warning system, as well as a data sharing system, which provides help for the prevention and control of national water environmental pollution and environmental management.

2. Data and Methods

2.1. Data Sources

The hydrological and water-quality data mentioned in this article were collected from the national water and rain information network of the Ministry of Water Resources of the People's Republic of China (http://xxfb.mwr.cn/sq_djdh.html accessed on 16 September 2022) and the national surface-water quality monitoring and publishing system of the Ministry of Ecology and Environment of the People's Republic of China (<https://szzdjc.cnemc.cn:8070/GJZ/Business/Publish/Main.html> accessed on 16 September 2022).

2.2. The Philosophy of This System

The purpose of this system was to develop functions for data crawling, storage, water quality prediction and early warnings based on the public hydrology and water-quality dataset. One of the most significant features in this process was the ability to automatically obtain public hydrology and water-quality data through data collecting, cleaning and storing in real time. On this basis, the LSTM model could be constructed and trained to accomplish the function of water quality prediction and early warnings. The philosophy of this system is shown in Figure 1.

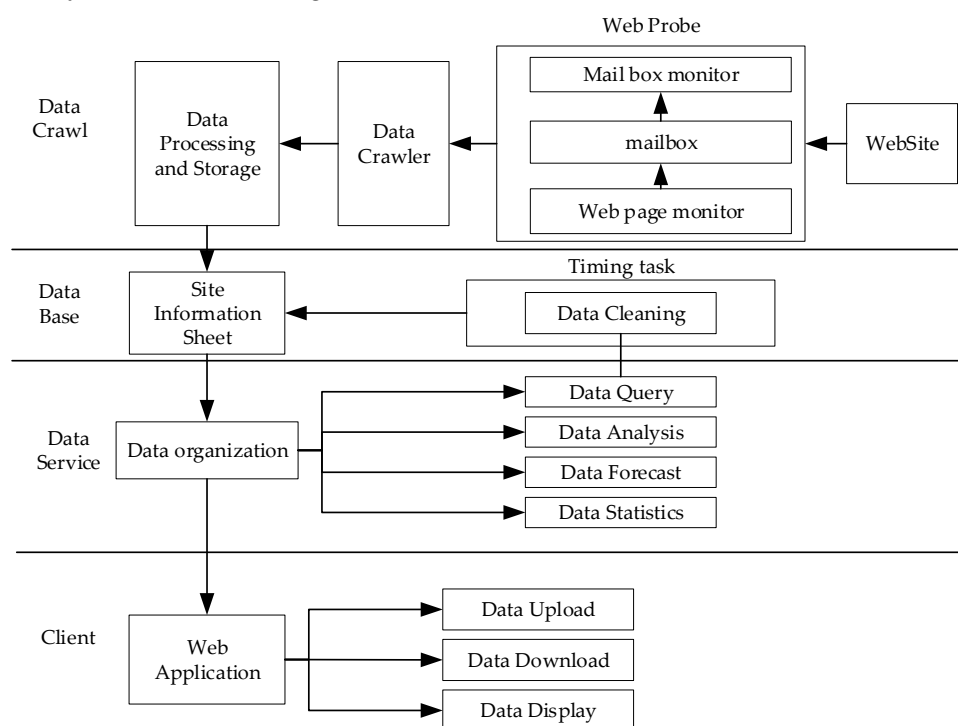


Figure 1. The philosophy of water-quality-prediction and early-warning system.

2.2.1. Data Crawling

A monitoring script was set up in the target website through the distill web monitor plug-in provided by the browser. Crawlers will be launched to acquire and store data automatically once the update of the target website is detected.

2.2.2. Database

In this system, we chose MongoDB as the storage database for data storage and maintenance functions, and to establish the data table shown in Table 1.

Table 1. The functions of tables used in the database.

ID	Tables in the Database	Functions
1	Hydrology and water quality station information table	Store the names and geographical coordinates of hydrology and water quality stations in China
2	Hydrology station table	Store the names of the hydrology sites crawled
3	Water quality site table	Store the names of the water quality sites crawled
4	Hydrological data table	Store the hydrological data crawled
5	Water quality data table	Store the water quality data crawled

Additionally, it also provided a data-cleaning service to remove redundant data and store new data.

2.2.3. Model Service

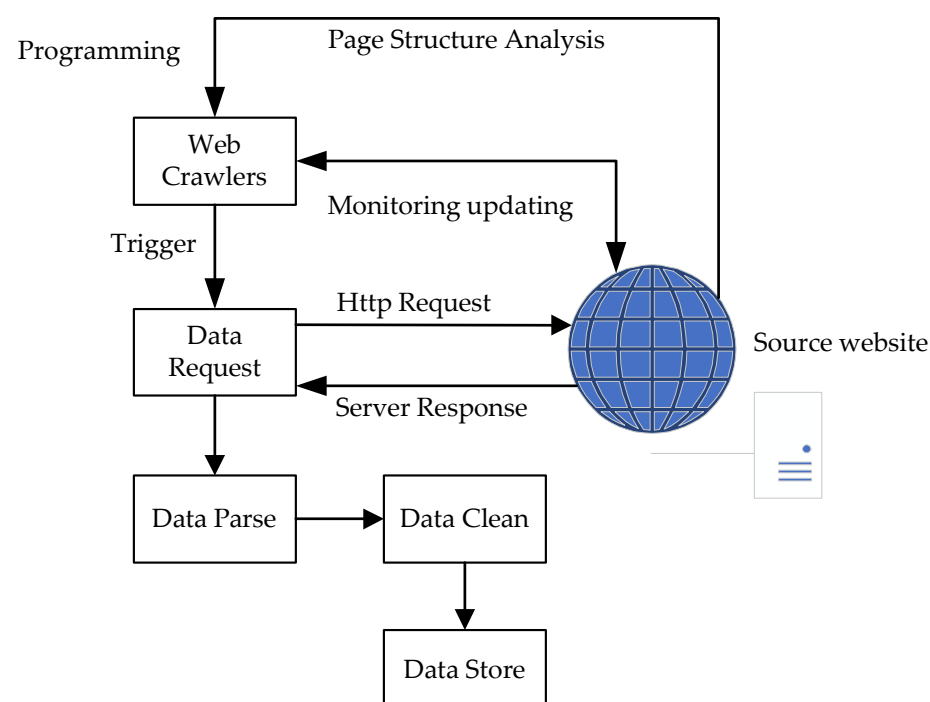
The data services section provided data manipulation functions. It was used to establish the connection session between the server and client, retrieve and analyze data from the database according to certain conditions, and return prediction results of water quality by using the LSTM neural network model.

2.2.4. Client

This module provided the functions of data interaction, including data upload, data download, rendering and producing displays through words, tables or charts. Users could log in to the system through a browser and make the related operations.

2.3. Process of Data Collecting by the Web Crawler

Web crawlers were used to periodically collect real-time monitoring data of water-quantity- and water-quality-monitoring stations. The process of the crawler is shown in Figure 2.

**Figure 2.** Process of data collecting by web crawler used in the system.

2.3.1. Website Monitoring and Crawler Starting

A monitor mechanism was created to monitor the content change of the national water and rain information network, and the national surface-water-quality-monitoring and publishing system through the Distill Web Monitor plug-in provided by the browser. An e-mail is sent to a specified mailbox if there are some data updates. The measurement time, used for data update judgment, of water quality data is monitored by Xpath syntax. The crawler starts once an update e-mail is captured and parsed by a mailbox-monitoring script.

2.3.2. Data Crawling

The URL of the target website is first parsed through the network data-catching tool after launching the crawler. Then, a POST request is sent to the object URL through the request library in Python. After that, the browser receives a response in JSON data from the object URL. Water quality data is obtained from the JSON data.

2.3.3. Data Cleaning and Storing

Data cleaning is essential before data is stored in a database. After obtaining the water quality data from the crawler module, the next step is to optimize them. The repeating data are deleted, as well as the null data. After that, the data are stored in MongoDB, which is accomplished by calling PyMongo through Python. Firstly, query the monitoring time of the site and estimate whether the current monitoring time is the same as that stored in the database table. If so, it can be considered that the data under the corresponding site was crawled and updated. The storing step will not be started. Otherwise, update the corresponding site data with the new crawled and cleaned data. Secondly, if the current traversal site is not recorded in the site data table, create corresponding site records in the site data table and store the data.

2.4. Model Services Development for Water Quality Assessment and Early Warnings Based on the LSTM Neural Network

The long short-term memory (LSTM) network is an improved and solvable special RNN model [24]. By introducing a new internal state and gate mechanism, LSTM networks have a good performance in solving the problems of gradient disappearance and long-term dependence, which makes LSTM have selective memory functions [25]. The construction process of the water-quality-prediction model for river pollution accidents based on LSTM is shown in Figure 3.

2.4.1. Dataset Organization and Analysis

The dataset contains water quality data from 1863 key monitored sections in China. The first step is to pretreat the obtained data. The input variables corresponding to the model output variables are determined through correlation analysis. After that, the dataset is divided into the training set, test set and validation set. Then, all of these data are standardized.

2.4.2. Model Developed by LSTM

The LSTM model is an advanced recurrent neural network (RNN) that includes specialized memory blocks for capturing multi-timestep relationships [26]. The model structure used in this system is shown in Figure 4.

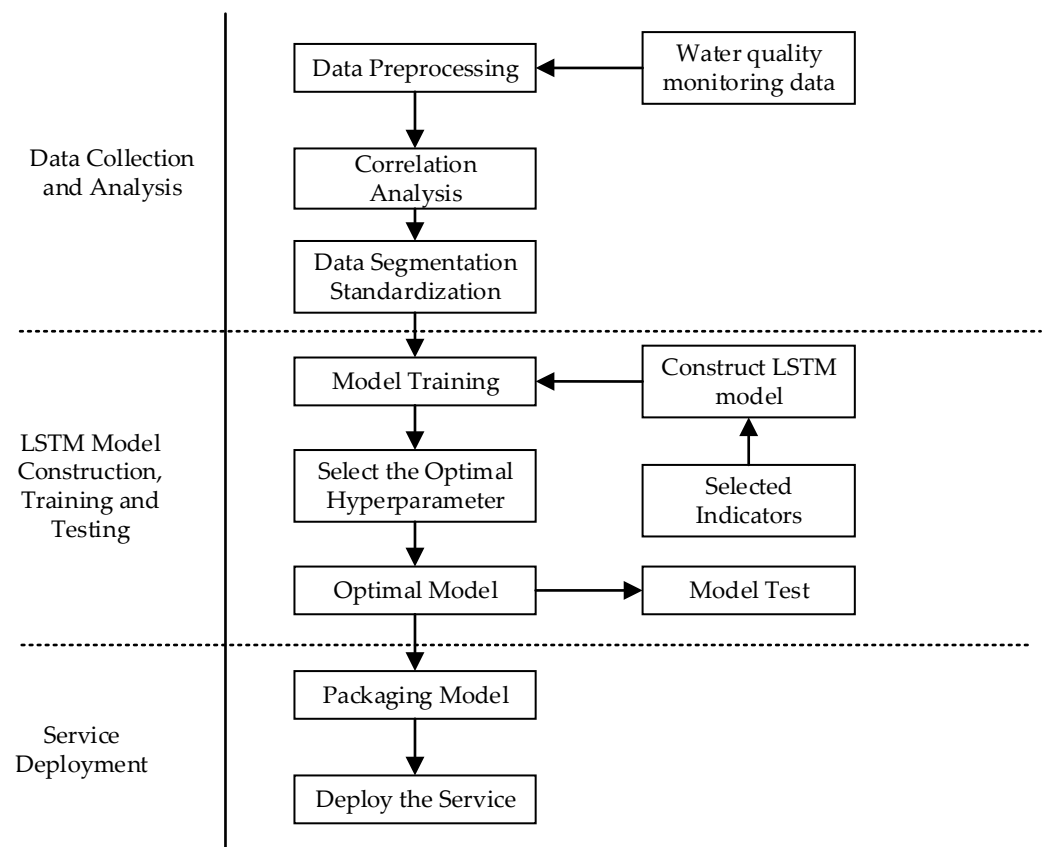


Figure 3. Model construction process of water quality prediction based on LSTM.

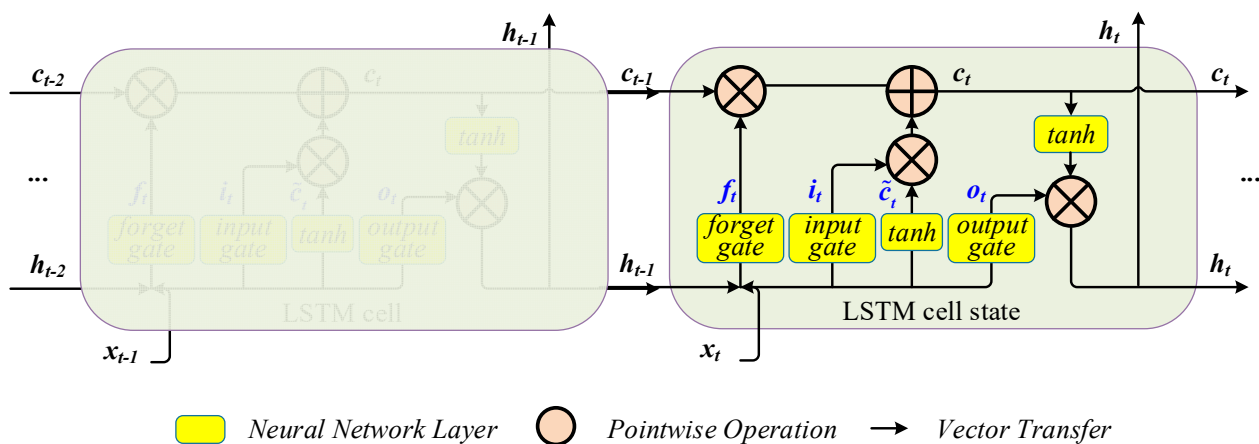


Figure 4. Structure of the LSTM model.

Where c_{t-1} is the cell state at time $t-1$, h_{t-1} is the output at time $t-1$, x_t are new inputs at time t , o_t is the value that the output gate produces to determine which parts of the cell state to output, $\tanh$ denotes the $\tanh$ function, $\tilde{c}_t$ is the candidate cell state and i_t is the coefficient of $\tilde{c}_t$.

The LSTM cell state is the memory space of the whole model, which changes over time. Three control gates, namely the neural network layer, are the core modules controlling what information is transferred. In the system, the LSTM model was built based on the TensorFlow framework. Water quality parameters, such as pH, ammonia nitrogen (NH_4^+), dissolved oxygen (DO), water temperature (WT), potassium permanganate index (COD_{Mn}) and turbidity (TU), are set for water environment evaluation. Future water quality parameters at time $t+1$ are simulated by LSTM using the historical value of water quality

parameters. The whole dataset in China is divided to 10 watersheds. Every watershed is divided into three parts: upstream, middle, and downstream sub-watershed. Every sub-watershed is modeled by a single LSTM model. Every water quality index has its own LSTM model. There are 210 developed LSTM models in the whole of China.

2.4.3. Model Training and Testing

After the model is built, the training set of these data is put into the LSTM model for model training. Optimal hyperparameters parameters are then determined according to the verification based on the model validation dataset. Finally, the test model accuracy performance of water quality prediction is made on the test set.

2.4.4. Model Services Deployment

After model testing, the adequate model is picked as a TensorFlow Server to the system as a model service.

2.5. Client Structures

In order to fulfill the need for a water-quality-assessment and pollution-risk early-warning system, the client module was designed to include a data visualization module, a water quality evaluation module and a water-quality pollution-risk early-warning module, as shown in Figure 5.

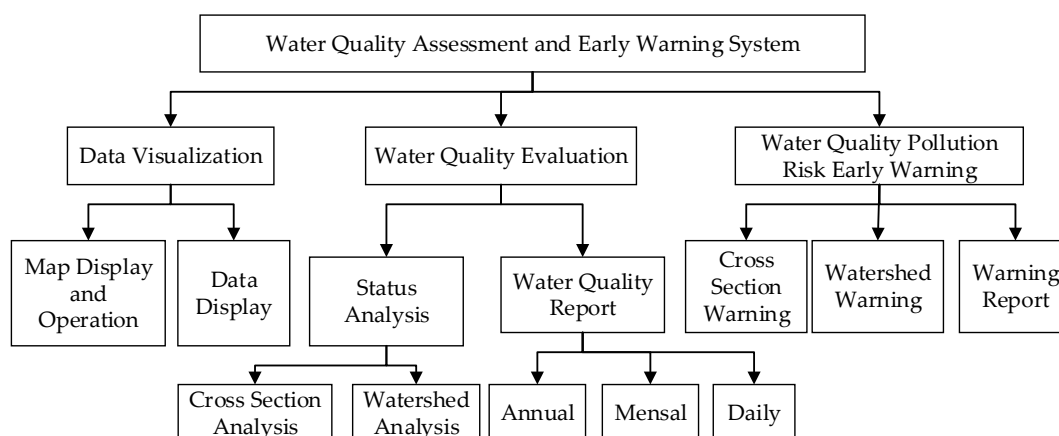


Figure 5. The submodules and functions designed in the client.

2.5.1. Data Visualization Module

This module is mainly used to visualize the spatial location information of water-quality-monitoring stations, data analysis and prediction results. Two important functions are designed in the data visualization module. For map display and operation, the GeoServer and webGIS are used to build a GIS map, which provides the function for map display, query and spatial data rendering. For data visualization, the time series data is visualized in the form of a table, chart and animation.

2.5.2. Water-Quality-Assessment Module

This module is mainly used to integrate and analyze the current water-quality-monitoring data and generate water quality reports within a certain time range, according to the historical monitoring data. The function of the current water quality analysis is developed. The water-quality-attainment-standard analysis, proportion of water quality categories analysis and the link relative ratio analysis are provided, which allows users to retrieve any site by watershed name, province name or cross-section name. Meanwhile, water quality reports can be made automatically according to the user's demands. The annual water quality report, monthly water quality report and customized definition report

can be made after determining the parameters, time range and evaluation methods selected by users.

2.5.3. Water-Quality-Prediction and Early-Warning Module

This module is mainly used to forecast the tendency of water quality in a specified cross-section or watershed through LSTM. An early warning is made based on the prediction result. The indicators of prediction can be selected from water quality indexes. The prediction module has two kinds of water-quality-prediction and early-warning modes.

Firstly, there is the automatic prediction and early warning: The water quality concentration and water quality grade for every water quality station, sub-watershed and watershed are made automatically for 24 h at intervals of one hour. The function is automatically started at 0:00 every day. The detail information of water quality trends can be obtained via the form of a table, chart, etc. The system raises a warning in time once the predicted indicator is over the standard. At the same time, a water quality report containing water quality station and watershed information whose water quality is unqualified is made and sent to the managers.

Secondly, there is the manual prediction and early warning: The system provides a user-defined prediction function. Based on user-specified water-quality parameters, stations and time periods, the LSTM model is launched. The prediction results, as well as early warning reports, are made.

3. Results and Discussion

3.1. Datasets That Were Crawled

3.1.1. The Spatial Distribution of Monitoring Stations

There are a total of 2591 hydrological monitoring stations and 3785 national surface-water quality stations obtained in the crawler. The distribution of the data sites crawled is shown in Figure 6.

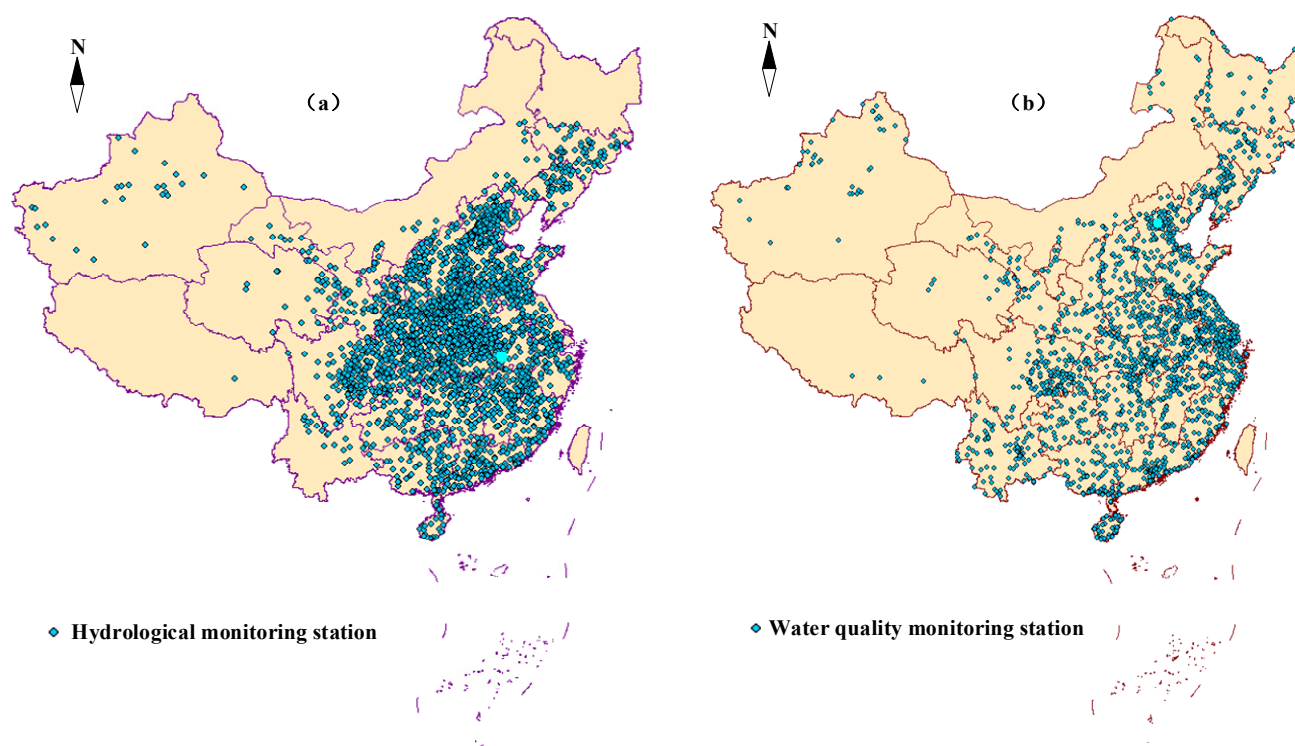


Figure 6. Distribution of hydrological monitoring stations (a) and water-quality-monitoring stations (b).

As can be seen in Figure 6, hydrological stations are mainly distributed in the eastern basin, and hydrological stations in the Yangtze River basin and the Yellow River basin are densely distributed. Hydrological stations are less distributed in the west than the east. Similar to hydrology stations, water quality stations are mainly distributed in the east of China, especially in the Yangtze River basin. On the whole, compared to most research, the location density of sites used for water quality prediction is better [12,27]. This means that the hydrological stations and water quality stations cover most of the water systems in China, and this reflects the characteristics of hydrology and water quality in China well.

Obviously, this dataset contains most of China's central, southern and eastern regions, and basically covers the main river basins in China. Features of hydrology data indicators and water quality data indicators obtained in the system are shown in Table 2.

Table 2. Features of the obtained data in the system.

Category	Variable	Indexes Name	Time Range	Units
Hydrological monitoring data	WL	Water level	Since 2005	m
	WWL	Warning water level		m
	Q	Flow		m ³ /s
Water-quality-monitoring data	Turb	Turbidity of water	Since 2000	NTU
	COD _{MN}	Permanganate index		mg/L
	NH ₄ ⁺	Ammonia nitrogen index		mg/L
	TP	Total phosphorus index		mg/L
	TN	Total nitrogen index		mg/L
	PH	Pondus Hydrogenii		-
	CHL	Chlorophyll α content		mg/L
	CA	Algal density of water		cells/L
	BOD ₅	Biochemical oxygen demand		mg/L
	DO	Dissolved oxygen		mg/L
	WT	Water temperature		°C
	WQC	Water quality classification		-

3.1.2. The Temporal Distribution of Monitoring Stations

Hydrological monitoring data from 2005 to today were obtained. These data were published at 8:00 and 14:00. Additionally, water quality created every four hours since 2000 was gained. Monitoring data from two stations was displayed as an example. A part of the hydrological monitoring data from the Taihu station and the water quality of Beijing's North Moat station are shown in Figure 7.

As shown in Figure 7a, the water stage, warning water state and quantification of flow will be released. It is shown that the water stage in the Taihu station was stable at around 3 to 4 m. For water quality, as is shown in Figure 7b, the concentration of PH, COD_{MN} and dissolved oxygen in the North Moat station was stable. Meanwhile, the turbidity had relatively large fluctuations from 17 September 2021 to 21 September 2021. These characteristics are similar to other published research [28,29]. Overall, the quality of these datasets is good enough to support water quality prediction.

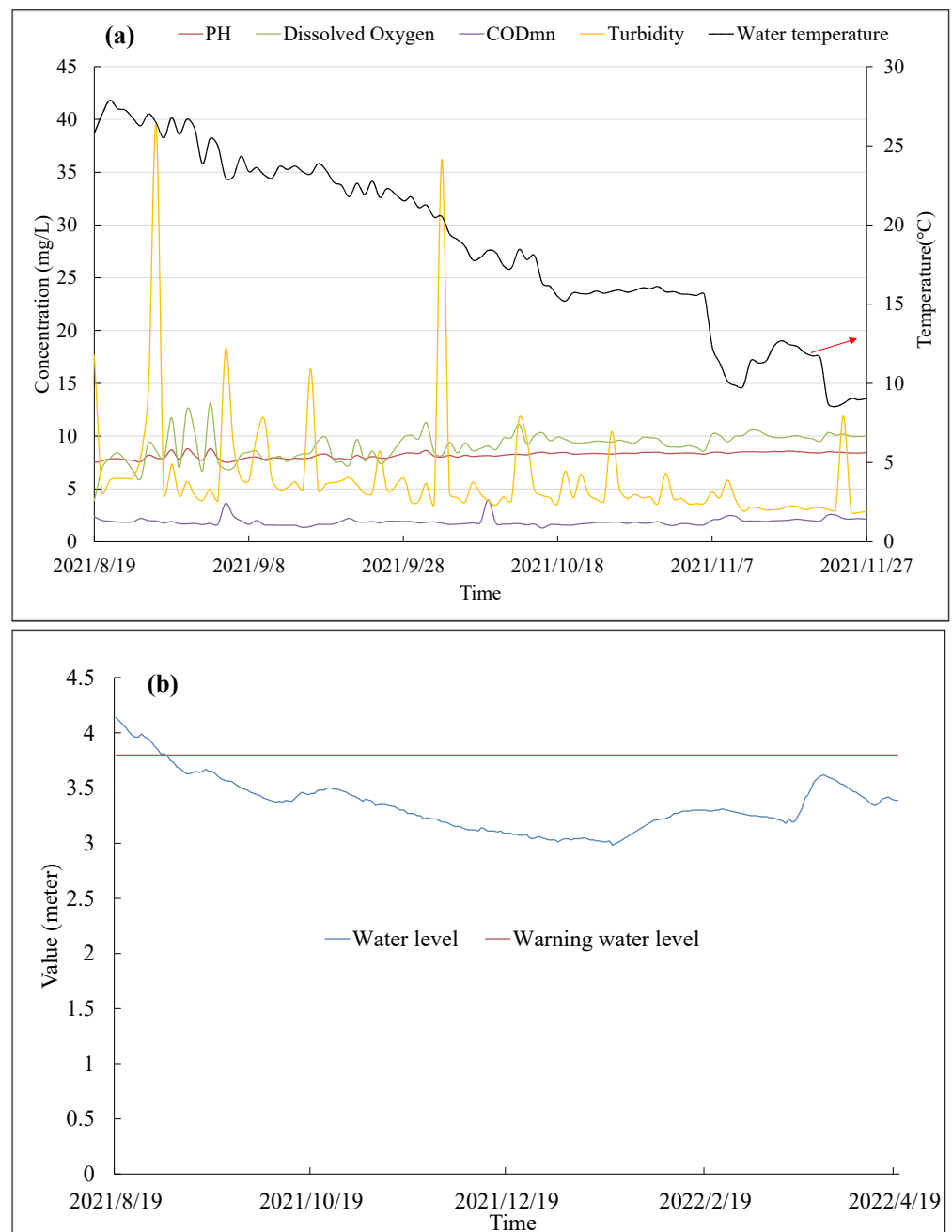


Figure 7. Part of the hydrological monitoring data at Taihu station (a) and water quality of North Moat station (b).

3.2. LSTM Model Training and Validation

LSTM models were trained and validated based on the water-quality-monitoring dataset of every sub-watershed. The Nash efficiency coefficient (NSE) of LSTM models used in this system for water quality prediction in the Haihe River basin can be seen in our published paper [26]. It verified that the LSTMs present are better predictors of BOD, COD_{Mn} , COD_{Cr} and TP (median Nash–Sutcliffe efficiency reaching 0.766, 0.835, 0.837, and 0.711, respectively) than of NH_4^+ , DO, and pH (median Nash–Sutcliffe efficiency of 0.638, 0.625, and 0.229). As an expert for PH (for the cause of large missing PH monitoring data), these models perform very well. Additionally, the LSTM model in the Yangtze River basin can be used as an example. In this dataset, we collected the water-quality-monitoring data of seven indicators in 22 sites from January 2003 to December 2018. The dataset was divided

into the training set, test set and validation set. Part of the results of LSTM developed for the Yangtze River basin are shown in Figure 8. CC stands for cross-correlation that indicates the effect of simulation, which can be calculated as follows:

$$CC(x, y, k) = CCF(x_t, y_{t-k}) \quad (1)$$

$$r(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$CS(x, y) = \frac{x^T y}{\|x\| \|y\|} \quad (3)$$

$$dCor(x, y) = \begin{cases} \sqrt{\frac{dCov(x, y)^2}{\sqrt{dCov(x)^2 dCov(y)^2}}} dCov(x)^2 dCov(y)^2 > 0 \\ 0 dCov(x)^2 dCov(y)^2 = 0 \end{cases} \quad (4)$$

$$MIC = \max \left\{ \frac{I(x, y)}{\log_2 \min(n_x, n_y)} \right\} \quad (5)$$

$$I(x, y) = H(x) + H(y) - H(x, y) \quad (6)$$

where CCF denotes the correlation metric formula (Equations (1)–(5)); k is the lag time ranging from 1 to 12; n is the number of variables for the simulation or measured data; x, y are two different water quality variables; $x_t = \{x_k, x_{k+1}, x_{k+2}, \dots, x_t\}$, $x_{t-k} = \{x_0, x_1, x_2, \dots, x_{t-k}\}$ and $y_{t-k} = \{y_0, y_1, y_2, \dots, y_{t-k}\}$; $x = \{x_1, x_2, \dots, x_n\}$ and $y = \{y_1, y_2, \dots, y_n\}$; $dCov(x)$ represents the Distance variance of x ; $dCov(x, y)$ denotes the Distance covariance between x and y ; $H(x)$ and $H(y)$ are the entropy of x and y , respectively, and $H(x, y)$ is their joint entropy; $B(n)$ is a coefficient of n ; $n_x n_y < B(n)$ and $B(n) = n^{0.6}$.

It is shown in Figure 8 that the CC value of the models of the validation set (part a) and test set (part b) is used for predicting ammonia nitrogen (NH_4^+), biochemical oxygen demand (BOD), permanganate index (COD_{Mn}), dissolved oxygen (DO), hydrogen ion concentration (PH), total phosphorus (TP) and water temperature (WT). In the test set, models used to predict PH, WT and TP are acceptable since most of their CC values are bigger than 0.65, and some models performed well ($CC > 0.75$). Additionally, the CCs for NH_4^+ and COD_{Mn} are close to 0.65. In the validation set, the CC of six indexes, COD_{Mn} , TP, NH_4^+ , WT, PH and DO, reach 0.65. Only the CCs for BOD are less than 0.65, which is around 0.4. The BOD is measured by the microbial electrode method and is only suitable for low concentrations of water. Meanwhile, the BOD in most of the test sections is relatively large, which makes the measured data fluctuate greatly and without obvious regularity. As a result, it is difficult to predict accurately.

Overall, these results are similar to or even better than the results of published research for monthly water quality predictions [30,31]. According to the previous research, the model is acceptable when the $CC \geq 0.65$ [32] and very good when the $CC \geq 0.75$ [26]. Accordingly, the model has a good performance in predicting water quality.

3.3. System Developments and Implementation

3.3.1. System Development Environment

Based on the B/S architecture and the model-template-view (MTV) framework for Django 1.8 and Pycharm 2021.3 professional software, the integrated development environment was built. MongoDB was set as the primary database to provide a data storage service. The TensorFlow framework was used to train, build and package the LSTM neural network model to be a TensorFlow Server.

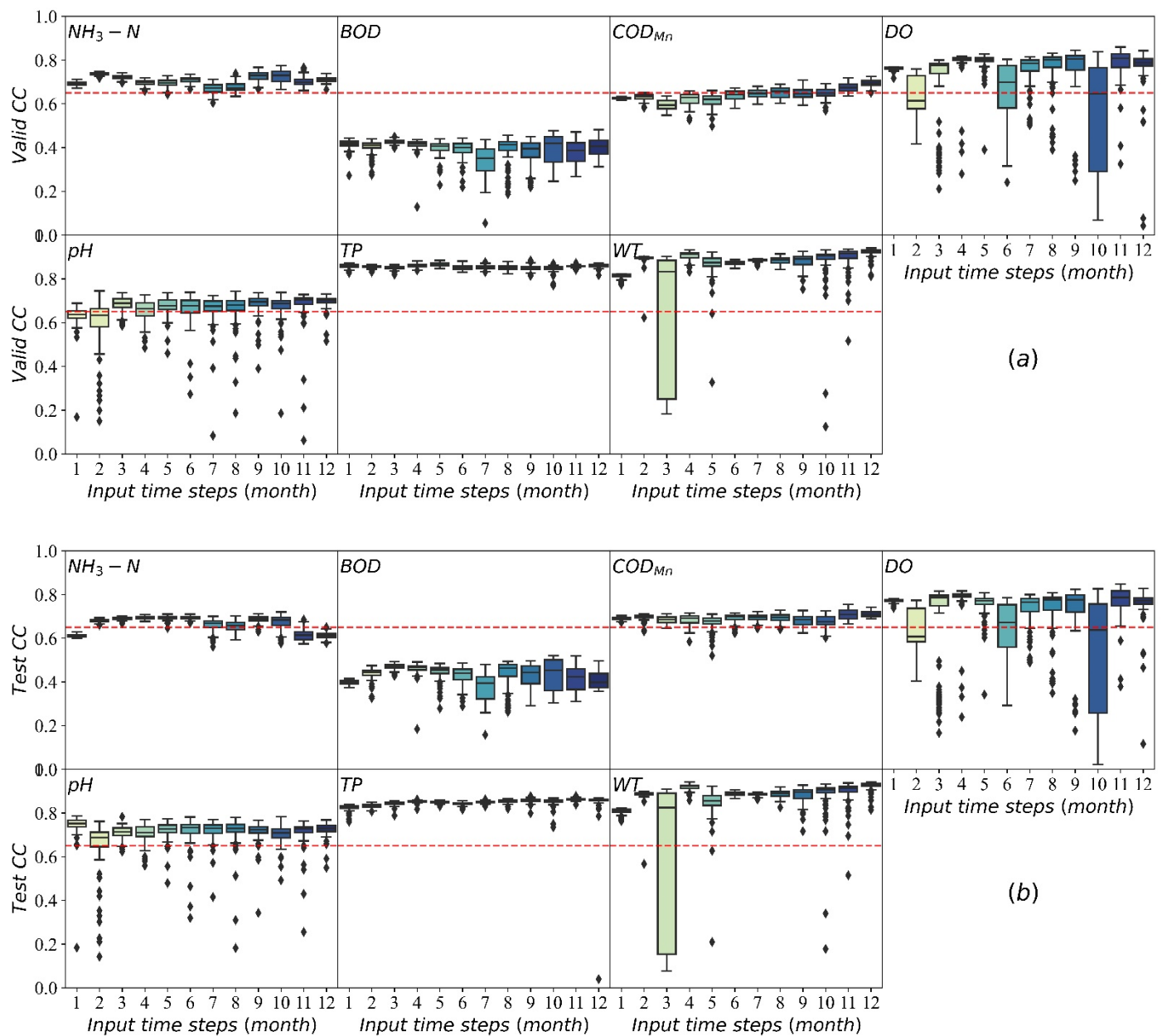


Figure 8. Distribution of cross-correlation (CC) values between predicted results and measured values for each water quality indicator predicted results in the validation set (a) and test set (b). The lower boundary of the box represents the minimum value. The upper boundary of the box represents the maximum value. The red dotted lines represent the threshold of acceptable model performance (CC = 0.65).

3.3.2. Functions of the System

These functions, such as water-environment-monitoring data visualization, water quality evaluation, water quality prediction and water-pollution-risk early warning were developed in the system (as shown in Figure 9a,b).

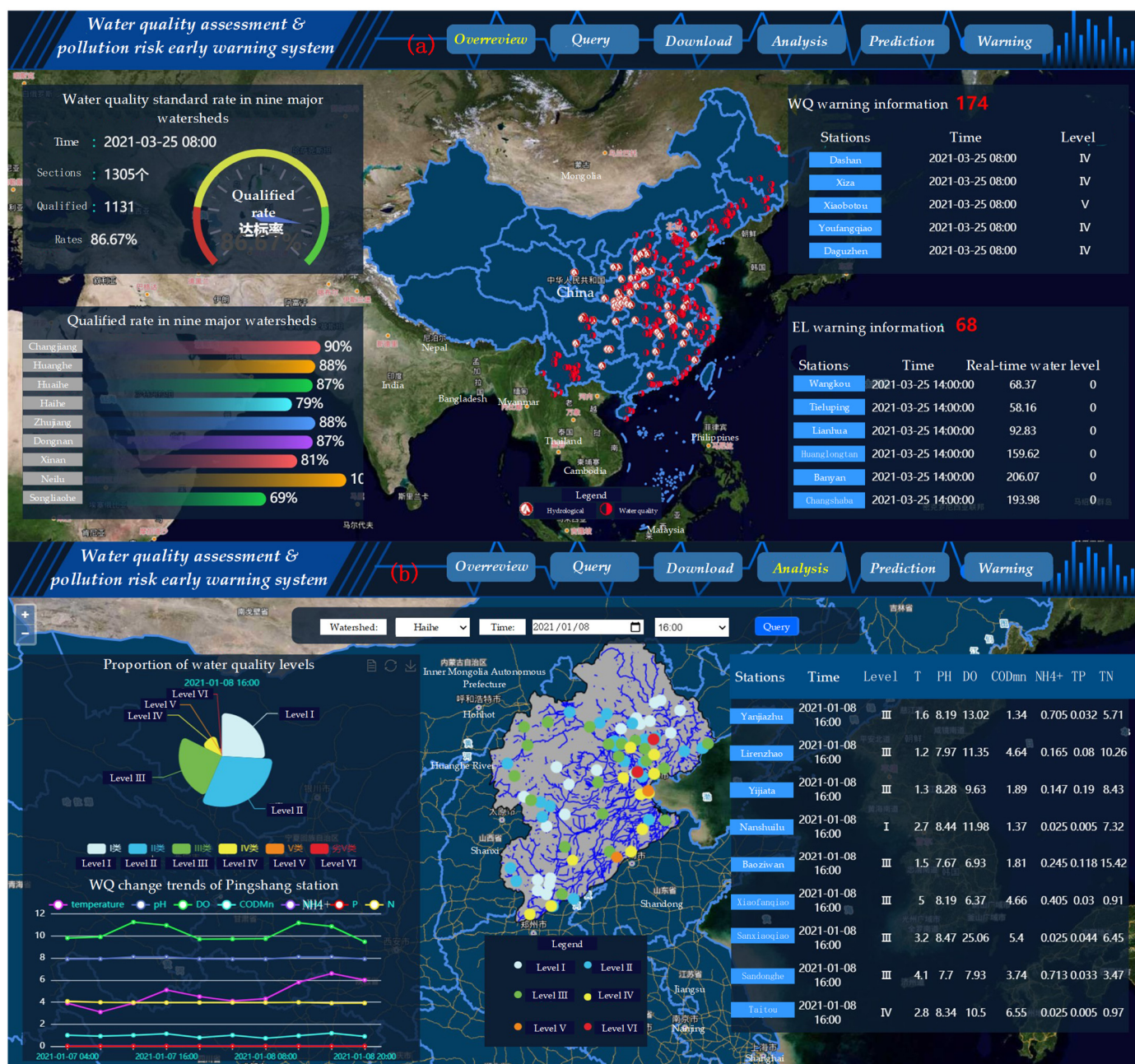


Figure 9. Functions of the water environment data sharing and analysis system ((a): overview of water-environment-monitoring data, (b): water quality analysis). Level VI of water quality means the water quality is worse than level V.

(1) Overview of water environment monitoring data

Figure 9a shows the visualization function of water-quality-monitoring data. The upper part of the interface integrates the system function area. In the GIS map, we chose the image map of the Tianditu map service to implement the function of basic map display (<https://www.tianditu.gov.cn/> accessed on 16 September 2020). In addition, the translucent and suspended window was used to display the hydrology water-quality-section alarm information. The water quality statistics were set to enhance interactivity with a GIS map.

(2) Water analysis with water quality evaluation and prediction

Figure 9b shows the general situation of all water quality sections in the Haihe River basin and the chart of trends in water quality indicators of specific sections that are predicted

the next day. The suspension toolbar integrates all options for the prediction of water quality of the watershed. It describes the summary of each cross-section and water quality trend, specifying the cross-section in the form of a table, chart, etc., which better reflects the hierarchical relationship between the watershed and section data.

3.4. Effect of a Case of Pollution Warning

The system automatically evaluates the water quality of all sites after prediction by the LSTM neural network model and renders a red point on the GIS map if the water quality is over the standard. As shown in Figure 10, the water quality (WQ) level of the Guojiatun station reached level IV on 26 March 2021 at 4:00, and then reached an inferior level V at 20:00. This means that there may be sudden pollution spills around this station. After the system detected the pollution, the alarm information of the WQ was displayed in the right column of the system. Meanwhile, a warning report of water pollution was generated and sent to the administrator's mailbox.



Figure 10. The water quality of pollution warning of prediction results.

In the early 1990s, the concept of an integrated energy and water quality management system (EWQMS) was developed to solve water quality, water supply and energy management problems, simultaneously. Since then, a lot of water quality management systems have been developed to display, monitor and assess water quality. From early displays of water-quality-monitoring data to water quality simulation and prediction with numerical models, water quality management systems have become more and more powerful [33,34]. Previously, we designed a water environment management system with flexible and extensible service-oriented architecture with the data center, system control center, model center and client center based on numerical models. This makes it possible for water quality forecasting in an automatic mode and a user-defined mode [35]. Meanwhile, as well as other water management systems, the previous systems make water quality simulations based on mechanism models. It needs preconcert development and a lot of basic condition data, such as topography data, geomorphology data, hydrology and water pollution source data. The system built in this paper can make water quality predictions rapidly, based on a small number of water-quality-concentration data.

4. Conclusions

This paper designed and developed a kind of water-quality-forecasting and warning system. The system can provide functions for data querying, presenting and analyzing,

as well as the retrieval and data sharing of historical data. The system can especially be used for water quality assessment and prediction based on LSTM models. It shows great significance for the water and environmental management department in obtaining water quality data and guarding water pollution. The water-quality-prediction and early-warning system constructed in this paper plays an important role in the promotion of environmental monitoring and monitoring work, the building of environmental-monitoring big data systems, and the promotion of ecological civilization construction and ecological environment-monitoring reform.

However, there are still some weaknesses in this system, such as imperfect data sharing, and a lack of verification of the accurate prediction and evaluation results. With the increase in monitoring data and the development of artificial neural networks, this system develops towards a faster, more professional, more reliable and lower-cost direction in the field of water quality monitoring and predicting.

Author Contributions: Data curation, J.L.; Methodology, Y.W.; Project administration, Q.L. (Qiang Liu); Software, Y.W. and J.Y.; Supervision, Q.L. (Qiang Liu); Writing—original draft, G.G.; Writing—review & editing, L.Y. and Q.L. (Qiang Li). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Open Research Fund Program of MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area (SZU51029202010).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available because of privacy concerns.

Conflicts of Interest: Author Qiang Liu has received research grants from Company Ningbo Ligong Environment and Energy Technology Co., Ltd., Author Jianyun Lin has been involved as a consultant and expert witness in Company Ningbo Ligong Environment and Energy Technology Co., Ltd.

References

1. Yin, Y.; Wang, H.; Gao, Y.; Li, X. Real-time monitoring and early warning of landslides at relocated Wushan Town, the Three Gorges Reservoir, China. *Landslides* **2010**, *7*, 339–349. [\[CrossRef\]](#)
2. Bownik, A.; Wlodkowic, D. Advances in real-time monitoring of water quality using automated analysis of animal behaviour. *Sci. Total Environ.* **2021**, *789*, 147796. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Pavai, G.; Geetha, T.V. Improving the freshness of the search engines by a probabilistic approach based incremental crawler. *Inf. Syst. Front.* **2017**, *19*, 1013–1028. [\[CrossRef\]](#)
4. Kumar, N.; Aggarwal, D. LEARNING-based Focused WEB Crawler. *IETE J. Res.* **2021**, 1–9. [\[CrossRef\]](#)
5. Velkumar, K.; Thendral, P. Web Crawler and Web Crawler Algorithms: A Perspective. *Int. J. Eng. Adv. Technol.* **2020**, *9*, 203–205. [\[CrossRef\]](#)
6. Hayes, D.R.; Cappa, F.; Cardon, J. A Framework for More Effective Dark Web Marketplace Investigations. *Information* **2018**, *9*, 186. [\[CrossRef\]](#)
7. Kumar, M.; Bindal, A.; Gautam, R.; Bhatia, R. Keyword query based focused Web crawler. *Procedia Comput. Sci.* **2018**, *125*, 584–590. [\[CrossRef\]](#)
8. Wang, Z.; Ren, H.; Zhang, A.; Zhuang, D. Spatiotemporal Hotspots of Study Areas in Research of Gastric Cancer in China Based on Web-Crawled Literature. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3997. [\[CrossRef\]](#)
9. Chen, Z.; Zhang, H.; Liao, M. Integration Multi-Model to Evaluate the Impact of Surface Water Quality on City Sustainability: A Case from Maanshan City in China. *Processes* **2019**, *7*, 25. [\[CrossRef\]](#)
10. Zhao, Y.; Song, Y.; Cui, J.; Gan, S.; Yang, X.; Wu, R.; Guo, P. Assessment of Water Quality Evolution in the Pearl River Estuary (South Guangzhou) from 2008 to 2017. *Water* **2020**, *12*, 59. [\[CrossRef\]](#)
11. Qin, G.; Liu, J.; Xu, S.; Wang, T. Water Quality Assessment and Pollution Source Apportionment in a Highly Regulated River of Northeast China. *Environ. Monit. Assess.* **2020**, *192*, 1–16. [\[CrossRef\]](#)
12. Huang, J.; Zhang, Y.; Bing, H.; Peng, J.; Dong, F.; Gao, J.; Arhonditsis, G.B. Characterizing the river water quality in China: Recent progress and on-going challenges. *Water Res.* **2021**, *201*, 117309. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Gordillo, G.; Morales-Hernández, M.; García-Navarro, P. Finite volume model for the simulation of 1D unsteady river flow and water quality based on the WASP. *J. Hydroinform.* **2020**, *22*, 327–345. [\[CrossRef\]](#)

14. Priya, S.K.; Shenbagalakshmi, G.; Revathi, T. Design of smart sensors for real time drinking water quality monitoring and contamination detection in water distributed mains. *Int. J. Eng. Technol.* **2018**, *7*, 47–51. [\[CrossRef\]](#)
15. Arndt, J.; Kirchner, J.S.; Jewell, K.S.; Schluesener, M.P.; Wick, A.; Ternes, T.A.; Duester, L. Making waves: Time for chemical surface water quality monitoring to catch up with its technical potential. *Water Res.* **2022**, *213*, 118168. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Ding, X.; Dong, X.; Hou, B.; Fan, G.; Zhang, X. Visual platform for water quality prediction and pre-warning of drinking water source area in the Three Gorges Reservoir Area. *J. Clean. Prod.* **2021**, *309*, 127398. [\[CrossRef\]](#)
17. Pyo, J.C.; Kwon, Y.S.; Ahn, J.-H.; Baek, S.-S.; Kwon, Y.-H.; Cho, K.H. Sensitivity Analysis and Optimization of a Radiative Transfer Numerical Model for Turbid Lake Water. *Remote Sens.* **2021**, *13*, 709. [\[CrossRef\]](#)
18. Hochreiter, S.; Unger Schmidhuber, J. Long Shortterm Memory. *Neural Comput.* **1997**, *9*, 17351780. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Chimmula, V.K.R.; Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* **2020**, *135*, 109864. [\[CrossRef\]](#)
20. Bukhari, A.H.; Raja, M.A.Z.; Sulaiman, M.; Islam, S.; Shoaib, M.; Kumam, P. Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. *IEEE Access* **2020**, *8*, 71326–71338. [\[CrossRef\]](#)
21. Liu, P.; Wang, J.; Sangaiah, A.K.; Xie, Y.; Yin, X. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability* **2019**, *11*, 2058. [\[CrossRef\]](#)
22. Xia, F.; Jiang, D.; Kong, L.; Zhou, Y.; Wei, J.; Ding, D.; Chen, Y.; Wang, G.; Deng, S. Prediction of Dichloroethene Concentration in the Groundwater of a Contaminated Site Using XGBoost and LSTM. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9374. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [\[CrossRef\]](#)
24. Lee, S.; Lee, D. Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1322. [\[CrossRef\]](#)
25. Akhter, F.; Siddiquei, H.R.; Alahi, E.E.; Jayasundera, K.P.; Mukhopadhyay, S.C. An IoT-Enabled Portable Water Quality Monitoring System With MWCNT/PDMS Multifunctional Sensor for Agricultural Applications. *IEEE Internet Things J.* **2022**, *9*, 14307–14316. [\[CrossRef\]](#)
26. Tung, T.M.; Yaseen, Z.M. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J. Hydrol.* **2020**, *585*, 124670. [\[CrossRef\]](#)
27. Ban, J.; Ling, B.; Huang, W.; Liu, X.; Peng, W.; Zhang, J. Spatiotemporal Variations in Water Flow and Quality in the Sanyang Wetland, China: Implications for Environmental Restoration. *Sustainability* **2021**, *13*, 4637. [\[CrossRef\]](#)
28. Jia, Z.; Chang, X.; Duan, T.; Wang, X.; Wei, T.; Li, Y. Water quality responses to rainfall and surrounding land uses in urban lakes. *J. Environ. Manag.* **2021**, *298*, 113514. [\[CrossRef\]](#)
29. Wang, Y.; Guo, Y.; Zhao, Y.; Wang, L.; Chen, Y.; Yang, L. Spatiotemporal heterogeneities and driving factors of water quality and trophic state of a typical urban shallow lake (Taihu, China). *Environ. Sci. Pollut. Res.* **2022**, *29*, 53831–53843. [\[CrossRef\]](#)
30. Li, L.; Jiang, P.; Xu, H.; Lin, G.; Guo, D.; Wu, H. Water quality prediction based on recurrent neural network and improved evidence theory: A case study of Qiantang River, China. *Environ. Sci. Pollut. Res.* **2019**, *26*, 19879–19896. [\[CrossRef\]](#)
31. Li, R. Water quality forecasting of Haihe River based on improved fuzzy time series model. *Desalination Water Treat.* **2018**, *106*, 285–291. [\[CrossRef\]](#)
32. Ritter, A.; Muñoz-Carpena, R. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *J. Hydrol.* **2013**, *480*, 33–45. [\[CrossRef\]](#)
33. Wang, X.; Yin, Z.-Y. Using GIS to assess the relationship between land use and water quality at a watershed level. *Environ. Int.* **1997**, *23*, 103–114. [\[CrossRef\]](#)
34. Cherchi, C.; Badruzzaman, M.; Oppenheimer, J.; Bros, C.M.; Jacangelo, J.G. Energy and water quality management systems for water utility's operations: A review. *J. Environ. Manag.* **2015**, *153*, 108–120. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Wang, Y.; Li, Q.; Zhang, W.; Hu, S.; Peng, H. The architecture and application of an automatic operational model system for basin scale water environment management and design making supporting. *J. Environ. Manag.* **2021**, *290*, 112577. [\[CrossRef\]](#) [\[PubMed\]](#)