

## Article

# Water Quality Prediction Method Based on Multi-Source Transfer Learning for Water Environmental IoT System

Jian Zhou <sup>1,2,\*</sup> , Jian Wang <sup>1,2</sup>, Yang Chen <sup>1,2</sup>, Xin Li <sup>1,2</sup> and Yong Xie <sup>1,2</sup> 

<sup>1</sup> College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; b18030412@njupt.edu.cn (J.W.); chenyang\_njupt@126.com (Y.C.); xinli@njupt.edu.cn (X.L.); yongxie@njupt.edu.cn (Y.X.)

<sup>2</sup> Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China

\* Correspondence: zhoujian@njupt.edu.cn; Tel.: +86-189-0518-2929

**Abstract:** Water environmental Internet of Things (IoT) system, which is composed of multiple monitoring points equipped with various water quality IoT devices, provides the possibility for accurate water quality prediction. In the same water area, water flows and exchanges between multiple monitoring points, resulting in an adjacency effect in the water quality information. However, traditional water quality prediction methods only use the water quality information of one monitoring point, ignoring the information of nearby monitoring points. In this paper, we propose a water quality prediction method based on multi-source transfer learning for a water environmental IoT system, in order to effectively use the water quality information of nearby monitoring points to improve the prediction accuracy. First, a water quality prediction framework based on multi-source transfer learning is constructed. Specifically, the common features in water quality samples of multiple nearby monitoring points and target monitoring points are extracted and then aligned. According to the aligned features of water quality samples, the water quality prediction models based on an echo state network at multiple nearby monitoring points are established with distributed computing, and then the prediction results of distributed water quality prediction models are integrated. Second, the prediction parameters of multi-source transfer learning are optimized. Specifically, the back propagates population deviation based on multiple iterations, reducing the feature alignment bias and the model alignment bias to improve the prediction accuracy. Finally, the proposed method is applied in the actual water quality dataset of Hong Kong. The experimental results demonstrate that the proposed method can make full use of the water quality information of multiple nearby monitoring points to train several water quality prediction models and reduce the prediction bias.

**Keywords:** water quality prediction; multi-source transfer learning; echo state network; adjacency effect; distributed computing; environmental IoT system



**Citation:** Zhou, J.; Wang, J.; Chen, Y.; Li, X.; Xie, Y. Water Quality Prediction Method Based on Multi-Source Transfer Learning for Water Environmental IoT System. *Sensors* **2021**, *21*, 7271. <https://doi.org/10.3390/s21217271>

Academic Editor: Hisashi Satoh

Received: 4 September 2021

Accepted: 27 October 2021

Published: 1 November 2021

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



**Copyright:** © 2021 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

As an important part of the natural environment, water environment plays a vital role in human life. With the rapid development of industry, the discharge of industrial wastewater has increased day by day [1], leading to the deterioration of water environment, and water environment protection is facing severe challenges. Accurate water quality prediction is the basis for water environment protection. The monitoring points are equipped with various water quality Internet of Things (IoT) devices to build the water environmental IoT system [2], which can collect water quality information in real time, making the prediction of accurate water quality possible.

Traditional methods of water quality prediction can be classified into three types: regression analysis, grey systems, and neural networks [3]. Water quality prediction method based on regression analysis is derived from mathematical statistics. It determines the relationship between the dependent variable and the independent variable through the analysis of statistical data, and calculates the correlation coefficient through a certain

algorithm, thereby constructing a regression equation to predict water quality information. Ratko et al. [4] proposed a water temperature prediction method based on Gaussian process regression to predict the daily average water temperature of the river. Anja et al. [5] proposed a water quality prediction method based on partial least square regression analysis to predict the water quality information of mining wastewater. Mohammad et al. [6] proposed a prediction method based on M5 model tree and multiple adaptive regression to predict the daily river flow. Water quality prediction method based on grey systems regards the water environment system as a grey system. After that, a strong regular series for water quality prediction is generated by identifying the relationships of system factors. Zhang et al. [7] constructed a grey prediction model to predict the chemical oxygen of industrial wastewater. Yang et al. [8] constructed a GM (1,1) model to predict the water quality information of the lake. Xue et al. [9] constructed a grey prediction model to predict the mineralization of groundwater. Xiao et al. [10] applied grey theory to construct a model to predict the affecting factors of water bloom. Water quality prediction method based on neural networks forms an adaptive nonlinear system through the connection of neurons, using the neural networks to adaptively learn the trend of water quality information. With the emergence of cloud computing, edge computing and other technologies [11–13], neural networks requiring complex computation were gradually applied to water quality prediction. Dawood et al. [14] constructed an artificial neural network to predict the water quality information. Zhou et al. [15] proposed a water quality prediction method based on improved grey relational analysis and long-short term memory (LSTM) neural network to predict the dissolved oxygen. Dong et al. [16] proposed a water quality prediction method based on Savitzky-Golay and LSTM to predict the water quality information. Hu et al. [17] constructed a deep LSTM to predict pH and water temperature. Considering the temporality and the nonlinearity of water quality information, neural networks have more advantages and better prediction performance than the other two types of methods but require a large number of training samples. If the target monitoring point has too few training samples, the accuracy of water quality prediction will be reduced.

Water flows and exchanges between multiple monitoring points [18] in the same water area result adjacency effect in their water quality information. The prediction accuracy of neural networks can be improved if the adjacency effect is used for neural networks. Traditional transfer learning methods, such as transfer component analysis (TCA) [19], are usually used for single-source transfer, which can transfer the features of water quality samples from a single nearby monitoring point to a target monitoring point [20]. However, TCA does not consider the bias between the features of water quality samples of multiple nearby monitoring points, which makes it not applicable for the transfer of water quality samples of multiple nearby monitoring points. Target monitoring points are often surrounded with multiple nearby monitoring points in practice. Compared with traditional transfer learning methods, which can only effectively use one source domain, multi-source transfer learning (MSTL) [21,22] can make full use of multiple source domains. Therefore, we proposed a water quality prediction framework based on MSTL, effectively using the water quality information of multiple nearby monitoring points with distributed computing.

The water quality information changes periodically along with time, so it has the nature of temporality. By using the temporality, the accuracy of water quality prediction can be effectively improved. Echo state network (ESN) [23], as an improved model of recurrent neural network (RNN) [24], retains the information left at the last moment through the internal connections of reservoir, which can effectively use the temporality. Moreover, ESN only needs to use the linear regression algorithm to train the output weights, which can solve the problem of slow convergence speed of traditional RNN. Therefore, we establish the distributed water quality prediction models based on ESN at multiple nearby monitoring points in the framework, effectively using the temporality of water quality information.

Bias [25] exists not only in the feature alignment of nearby monitoring points and target monitoring points, but also in the model alignment of the water quality prediction models at multiple nearby monitoring points. Therefore, we optimize the prediction parameters of MSTL to improve the prediction accuracy of the models.

In this paper, we propose a water quality prediction method based on MSTL, for the purpose of making full use of the adjacency effect of water quality information. The contributions of this paper are listed as follows.

- (1) We construct a water quality prediction framework based on MSTL. In particular, the common features of water quality samples of multiple nearby monitoring points and the target monitoring point are extracted and then aligned. Afterwards, according to the aligned features of water quality samples, the water quality prediction models based on ESN at multiple nearby monitoring points are established with distributed computing, and then the prediction results of distributed water quality prediction models are integrated. This framework successfully solves the problem of an insufficient number of training samples of the target monitoring point.
- (2) We optimize the prediction parameters of MSTL. In particular, the back propagates the population deviation based on multiple iterations and can reduce the feature alignment bias and the model alignment bias to improve the prediction accuracy of the models.
- (3) We perform experiments in the actual water quality dataset of Hong Kong. The experimental results demonstrate that the proposed method can train multiple water quality prediction models by using the adjacency effect, and thus reduce the prediction bias and improve the prediction accuracy compared with other similar methods.

The rest of this paper is organized as follows. Section 2 gives the details of the proposed method, including the water quality prediction framework based on MSTL, the prediction parameters optimization of MSTL, and the overall process. Section 3 gives the experimental results and analyses. Section 4 is the summary of this paper.

## 2. Methods

### 2.1. Water Quality Prediction Framework Based on MSTL

We construct a water quality prediction framework based on MSTL, as shown in Figure 1. First, we use the feature extraction network based on the residual network [26] to extract the water quality features of nearby monitoring points and the target monitoring point into the same feature space, to obtain the common features of water quality samples of nearby monitoring points and the target monitoring point. Second, we use the feature alignment networks based on a bottleneck layer [27] to align the common features of water quality samples in the same feature space, to obtain the aligned features. Third, we establish the water quality prediction model based on ESN at every nearby monitoring point with distributed computing and predict the water quality information at the next moment according to the aligned features of water quality samples. Finally, we integrate the results of distributed water quality prediction models to reduce the prediction bias.

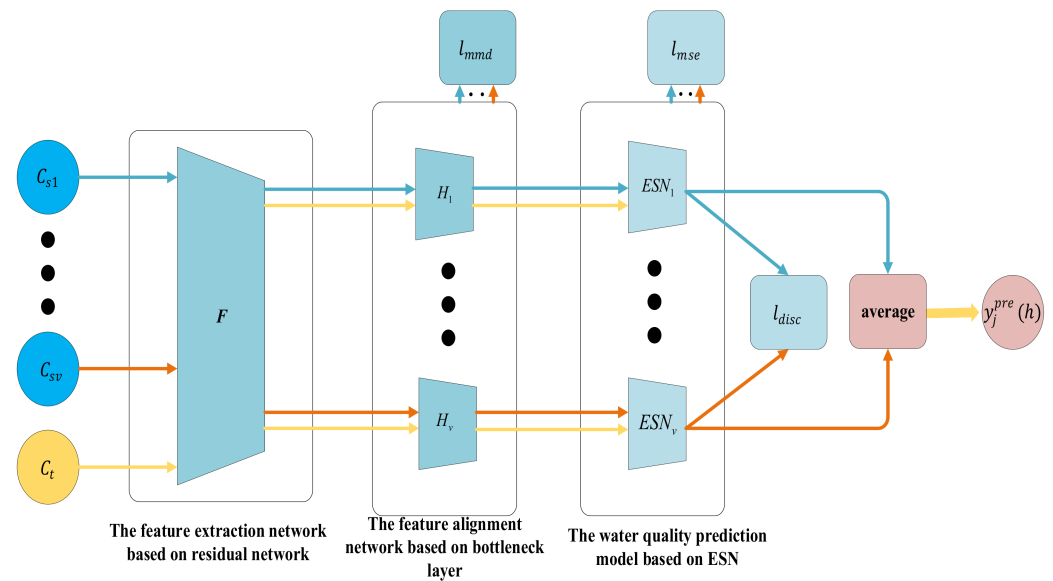


Figure 1. Water quality prediction framework based on MSTL.

If there are  $v$  nearby monitoring points around the target monitoring point, respectively construct the features of water quality samples of the  $j$ -th nearby monitoring point and the target monitoring point at the previous  $n$  moments as

$$C_{sj} = \{c_{sj}(d+1), c_{sj}(d+2), \dots, c_{sj}(h), \dots, c_{sj}(n)\} \quad (1)$$

$$C_t = \{c_t(d+1), c_t(d+2), \dots, c_t(h), \dots, c_t(n)\} \quad (2)$$

where  $c_{sj}(h) = \{x_{sj}(h-1), x_{sj}(h-2), \dots, x_{sj}(h-d)\}$  and  $c_t(h) = \{x_t(h-1), x_t(h-2), \dots, x_t(h-d)\}$  represent the features of water quality samples of the  $j$ -th nearby monitoring point and the target monitoring point at the  $h$ -th moment, respectively. In particular,  $d$  represents the size of the sliding window,  $x_{sj}(h-1)$  and  $x_t(h-1)$  represent the water quality information of the  $j$ -th nearby monitoring point and the target monitoring point at the  $(h-1)$ -th moment, respectively.

First, we construct the feature extraction network based on residual network (F), for the purpose of extracting the common features of water quality samples of  $v$  nearby monitoring points and the target monitoring point. The structure of this network is shown in Figure 2.

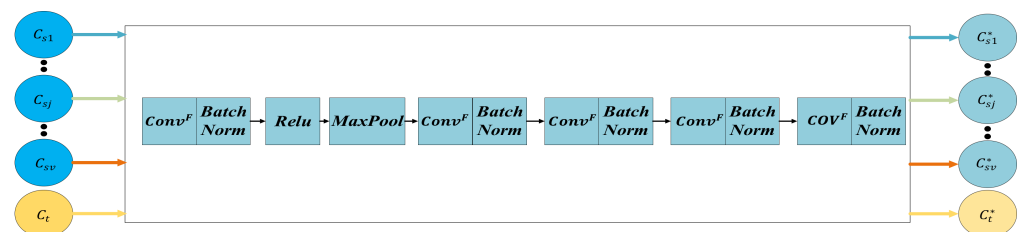


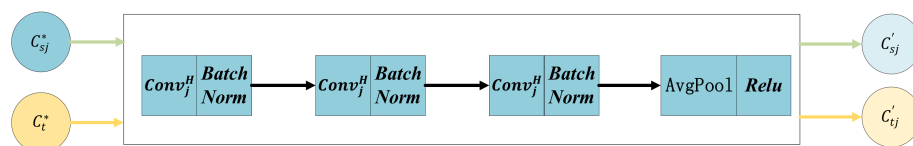
Figure 2. Structure of feature extraction network based on residual network.

In Figure 2,  $Conv^F$  is the convolution kernel,  $BatchNorm$  is the normalization algorithm,  $Rule$  is the activation function, and  $MaxPool$  is the max pooling layer. The features of water quality samples extracted from the  $j$ -th nearby monitoring point and the target monitoring point are respectively  $C_{sj}^*$  and  $C_t^*$ , and they are calculated by

$$C_{sj}^* = F(C_{sj}) \quad (3)$$

$$C_t^* = F(C_t) \quad (4)$$

Second, we construct the feature alignment networks based on the bottleneck layer ( $\{H_1, H_2, \dots, H_j, \dots, H_v\}$ ) at  $v$  nearby monitoring points, for the purpose of aligning the common features extracted from the nearby monitoring point with the features extracted from the target monitoring point. In particular,  $H_j$  is the feature alignment network at the  $j$ -th nearby monitoring point, and its structure is shown in Figure 3.



**Figure 3.** Structure of the feature alignment network based on the bottleneck layer at the  $j$ -th nearby monitoring point.

In Figure 3,  $Conv_j^H$  is the convolution kernel,  $Rule$  is the activation function, and  $AvgPool$  is the average pooling layer. The aligned features of water quality sample of the  $j$ -th nearby monitoring point and the target monitoring point are respectively  $C'_{sj}$  and  $C'_{tj}$ , and they are calculated by

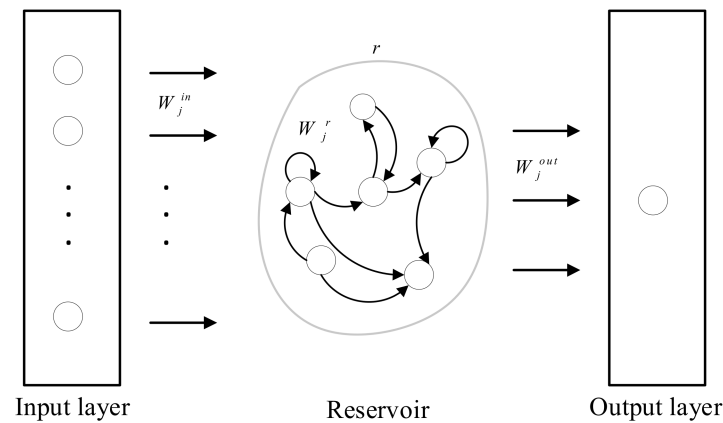
$$C'_{sj} = H_j(C_{sj}^*) \quad (5)$$

$$C'_{tj} = H_j(C_t^*) \quad (6)$$

After aligning the common features of water quality samples, construct the water quality sample sets of the  $j$ -th nearby monitoring point and target monitoring point as  $U_{sj} = \{u_{sj}(d+1), u_{sj}(d+2), \dots, u_{sj}(h), \dots, u_{sj}(n)\}$  and  $U_{tj} = \{u_{tj}(d+1), u_{tj}(d+2), \dots, u_{tj}(h), \dots, u_{tj}(m)\}$ , respectively. In particular,  $u_{sj}(h) = \{c'_{sj}(h), y_{sj}(h)\}$  and  $u_{tj}(h) = \{c'_{tj}(h), y_t(h)\}$  respectively represent the water quality samples of the  $j$ -th nearby monitoring point and the target monitoring point at the  $h$ -th moment, where  $c'_{sj}(h)$  and  $c'_{tj}(h)$  are the aligned feature of water quality sample of the  $j$ -th nearby monitoring point and the target monitoring point at the  $h$ -th moment.  $y_{sj}(h)$  and  $y_t(h)$  are the real water quality information of the  $j$ -th nearby monitoring point and the target monitoring point at the  $h$ -th moment.

We combine  $U_{sj}$  and  $U_{tj}$  to obtain the water quality sample set  $U_j^{train} = U_{sj} \cup U_{tj} = \{u_j^{train}(d+1), u_j^{train}(d+2), \dots, u_j^{train}(h), \dots, u_j^{train}(m+n-d)\}$ , where  $u_j^{train}(h) = \{c_j^{total}(h), y_j^{total}(h)\}$  is the water quality sample at the  $h$ -th moment,  $c_j^{total}(h)$  is the feature of water quality sample at the  $h$ -th moment, and  $y_j^{total}(h)$  is the real water quality information at the  $h$ -th moment.  $U_j^{train}$  will be used to train the following water quality prediction model.

Afterwards, we construct the water quality prediction models based on ESN at  $v$  nearby monitoring points ( $\{ESN_1, ESN_2, \dots, ESN_j, \dots, ESN_v\}$ ), where  $ESN_j$  is the distributed water quality prediction model at the  $j$ -th nearby monitoring point, and its structure is shown in Figure 4. The model consists of an input layer with  $d$  neurons, a reservoir with  $r$  neurons, and an output layer with one neuron. Besides, the input of the model is the feature of water quality sample at the  $h$ -th moment ( $c_j^{total}(h)$ ), and the output is the predicted water quality information at the  $h$ -th moment ( $y_j^{pre}(h)$ ).



**Figure 4.** Structure of water quality prediction model based on ESN.

The calculation of the water quality prediction model based on ESN at the  $j$ -th nearby monitoring point is as

$$s_j(h) = \text{Tanh}\left(W_j^r s_j(h-d) + W_j^{in} c_j^{total}(h)\right) \quad (7)$$

$$y_j^{pre}(h) = W_j^{out} s_j(h) \quad (8)$$

where  $\text{Tanh}$  is the activation function, and  $s_j(h)$  is the internal state vector of the reservoir.  $W_j^{in}$  is the input layer weight,  $W_j^r$  is the reservoir weight, and  $W_j^{out}$  is the output weight. In particular,  $W_j^{out}$  is trained by the ridge regression algorithm [28] according to  $U_j^{train}$ ,  $W_j^r$  and is scaled by

$$W_j^r = \alpha \frac{1}{|\rho|} W_0 \quad (9)$$

where  $\alpha$  is the scaling range and  $0 < \alpha < 1$ .  $\rho$  is the spectral radius of  $W_j^r$ , and  $W_0$  is a sparse matrix which is randomly generated.

Finally, we integrate the prediction results of distributed water quality prediction models at multiple nearby monitoring points to obtain the final prediction result ( $y^{pre}(h)$ ) by using the arithmetic average.  $y^{pre}(h)$  is calculated by

$$y^{pre}(h) = \frac{1}{v} \sum_{j=1}^v y_j^{pre}(h) \quad (10)$$

## 2.2. Prediction Parameters Optimization of MSTL

We optimize the prediction parameters of MSTL to reduce the feature alignment bias between nearby monitoring points and the target monitoring point, and to minimize the model alignment bias between the water quality prediction models. Specifically, to minimize the overall bias ( $l_{total}$ ),  $CONV^F$ ,  $CONV_j^H$  and  $W_j^{out}$  are updated by the stochastic gradient descent (SGD), since  $W_j^{out}$  affects the prediction results of the water quality prediction model,  $CONV^F$  and  $CONV_j^H$  affect the aligned features obtained by the feature alignment networks. The smaller  $l_{total}$  is, the better the prediction accuracy is.  $l_{total}$  is calculated by

$$l_{total} = l_{mse} + \lambda(l_{mmd} + l_{disc}) \quad (11)$$

where  $\lambda$  is the trade-off parameter, which is used to measure the importance of  $l_{mmd}$  and  $l_{disc}$ .  $\lambda$  is calculated by

$$\lambda = \frac{2}{1 + e^{\frac{10i}{iter}}} - 1 \quad (12)$$

where  $iter$  is the total number of iterations, and  $i$  is the current number of iterations.

$l_{mse}$  is the model prediction bias of the water quality prediction models at the  $v$  nearby monitoring points. The smaller  $l_{mse}$  is, the smaller the model prediction bias is.  $l_{mse}$  is calculated by

$$l_{mse} = \frac{1}{v} \sum_{j=1}^v Mse(y_j^{pre}(h), y^{true}(h)) \quad (13)$$

where  $y_j^{pre}(h)$  is the predicted water quality information of the prediction model at the  $j$ -th nearby monitoring point,  $y^{true}$  is the real water quality information, and  $Mse$  is the mean square error function.

$l_{mmd}$  is the feature alignment bias between nearby monitoring points and the target monitoring point. The smaller  $l_{mmd}$  is, the smaller the feature alignment bias is.  $l_{mmd}$  is calculated by

$$l_{mmd} = \frac{1}{v} \sum_{j=1}^v MMD(C'_{sj}(h), C'_{tj}(h)) \quad (14)$$

where  $MMD$  is the maximum mean discrepancy function [29], which is used to measure the distance between the aligned features of water quality samples of nearby monitoring points and the target monitoring point after mapping to the same feature space.

$l_{disc}$  is the model alignment bias between the water quality prediction models at multiple nearby monitoring points. The smaller the  $l_{disc}$  is, the smaller the model alignment bias is.  $l_{disc}$  is calculated by

$$l_{disc} = \frac{2}{(v-1)v} \sum_{j=1}^{v-1} \sum_{i=j+1}^v (|y_i^{pre}(h) - y_j^{pre}(h)|) \quad (15)$$

### 2.3. Process of Water Quality Prediction Method Based on MSTL

The overall process of the water quality prediction method based on MSTL is summarized in Figure 5. The specific steps are as follows:

Step 1: Use the feature extraction network based on residual network to extract the common features of water quality samples of the  $j$ -th nearby monitoring point and target monitoring point ( $C_{sj}^*$  and  $C_t^*$ ).

Step 2: Use the feature alignment network based on the bottleneck layer of the  $j$ -th nearby monitoring point to align the common features of water quality samples of the  $j$ -th nearby monitoring point and the target monitoring point ( $C'_{sj}$  and  $C'_{tj}$ ).

Step 3: Construct the training set  $U_j^{train}$ , and train the distributed water quality prediction model based on ESN at the  $j$ -th nearby monitoring point.

Step 4: Repeat Steps 1–3 to obtain the common features and the aligned features of water quality samples of all nearby monitoring points, and train distributed water quality prediction models based on ESN at all nearby monitoring points.

Step 5: Calculate the overall bias ( $l_{total}$ ) according to the aligned features of water quality samples of all nearby monitoring points and the prediction results of distributed water quality prediction models based on ESN.

Step 6: Judge whether  $l_{total}$  meets the accuracy requirements. If the requirements are met, go to step 8. Otherwise, go to step 7.

Step 7: For every nearby monitoring point, update  $CONV^F$ ,  $CONV_j^H$ , and  $W_j^{out}$  through multiple iterations and back-propagating  $l_{total}$ . After that, go to step 1.

Step 8: At the  $j$ -th nearby monitoring point, input the water quality information of the previous  $d$  moments of the current time of the target monitoring point into the optimized water quality prediction framework based on MSTL, then obtain the prediction result through distributed computing. In the same way, the prediction results of distributed water quality prediction models of all nearby monitoring points are obtained.

Step 9: Integrate the prediction results of distributed water quality prediction models at all nearby monitoring points to obtain the final prediction result ( $y^{pre}(h)$ ).

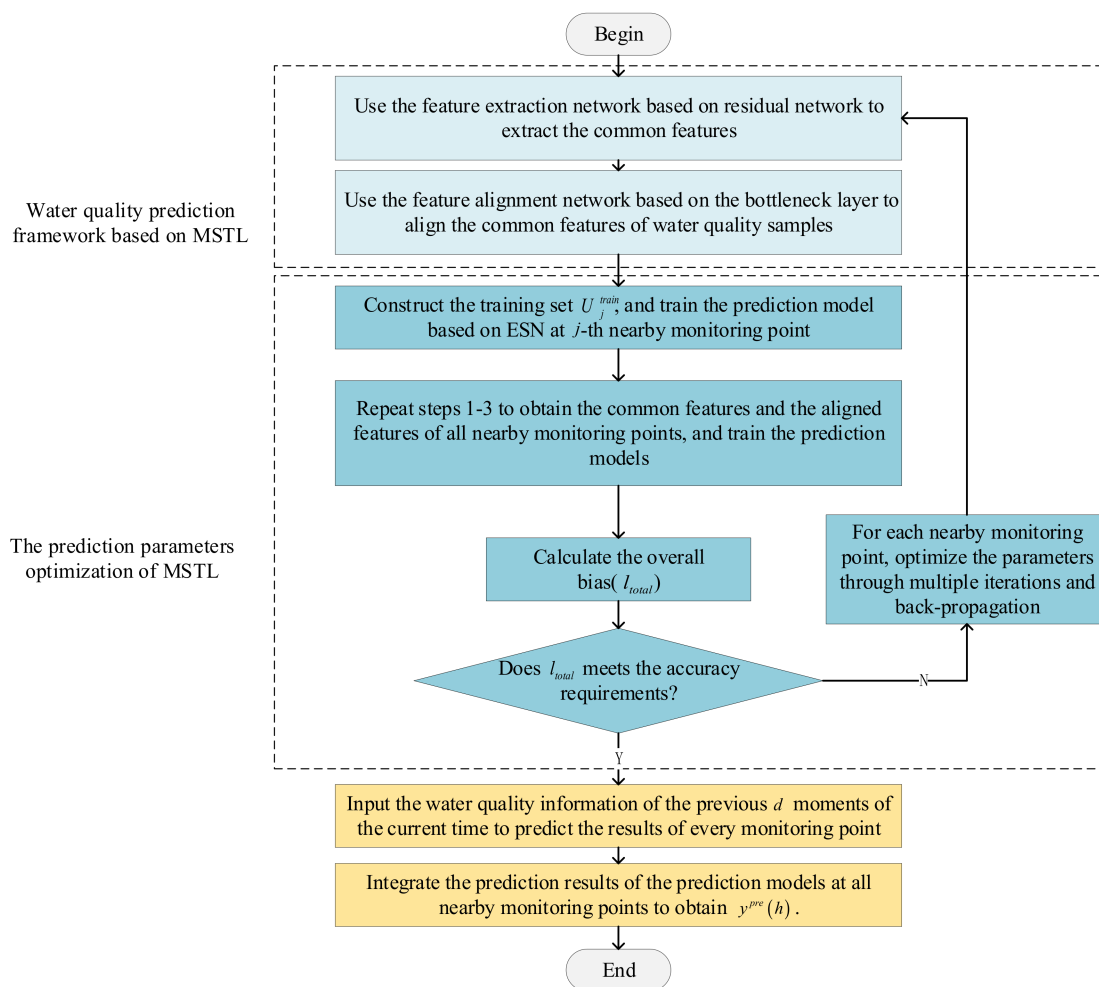


Figure 5. Overall process of water quality prediction method based on MSTL.

### 3. Experimental Results and Analyses

The proposed method is implemented by Python and Torch. First, we describe the specific dataset of the experiments. Second, we select the prediction parameters of MSTL. Afterwards, MSTL is compared with other transfer methods. Finally, we compare ESN with other prediction models.

We set 20% samples of the target monitoring point as the test sample set and 20% as the validation sample set. Thus, the training sample set is composed of the remaining 60% samples of the target monitoring point and the samples transferred from nearby monitoring points. The mean squared error (MSE) is chosen as the indicator measuring the prediction bias. The smaller MSE is, the smaller the prediction bias is. Specifically, MSE is calculated by

$$MSE = \frac{1}{q} \sum_{t=1}^q (y^{true}(t) - y^{pre}(t))^2 \quad (16)$$

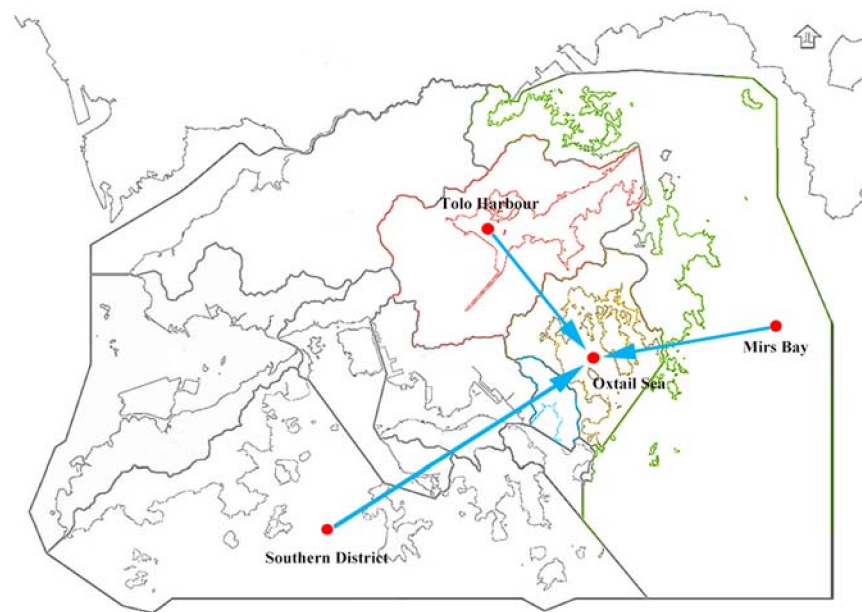
where  $q$  is the number of samples,  $y^{true}$  is the real water quality information, and  $y^{pre}$  is the predicted water quality information.

#### 3.1. Datasets

We performed two experiments. In the first experiment, we set Oxtail Sea as the target monitoring point. Oxtail Sea has only 3193 pieces of water quality information, which is slightly insufficient. The spatial location of monitoring points in the first experiment is shown in Figure 6. Tolo Harbour, Mirs Bay and Southern District are close to Oxtail Sea in



water area, and they have the adjacency effect. As a result, we consider these three locations as nearby monitoring points. Afterwards, we use the framework based on MSTL to align the features of water quality samples of these three nearby monitoring points with Oxtail Sea, and then use the samples of these three nearby monitoring points to train three water quality prediction models. Among them, Tolo Harbour has 3192 pieces of water quality information, Mirs Bay has only 758 pieces of water quality information, and Southern District has 4467 pieces of water quality information. The water quality indicators of Oxtail Sea include dissolved oxygen (DO), phosphate, water temperature (WT), and nitrite. The data of these indicators are collected by the sensors and transmitted back approximately every fifteen days. The purpose of the experiment is to predict the DO of Oxtail Sea at the next moment.



**Figure 6.** Spatial location of Oxtail Sea and nearby monitoring points.

In the second experiment, we set the Western Buffer District as the target monitoring point. Western Buffer District has only 1440 pieces of water quality information, which is also slightly insufficient. The spatial location of monitoring points in the second experiment is shown in Figure 7. Figure 7 shows the Northwestern District, the Southern District and the Victoria Harbour are close to Western Buffer District in water area and they have the adjacency effect. As a result, we consider these three locations as nearby monitoring points. The experimental procedure is the same as the first experiment. Among them, the Northwestern District has only 1630 pieces of water quality information, the Southern District has 4467 pieces of water quality information, and the Victoria Harbour has 4158 pieces of water quality information. The water quality indicators and the prediction purpose of the Western Buffer District are the same as the first experiment.

### 3.2. Parameters Selection

In order to improve the prediction accuracy, we select the parameters including the size of sliding window ( $d$ ), the size of reservoir ( $r$ ) and the size of spectral radius ( $\rho$ ) in the water quality prediction models based on ESN. In the experiment of Oxtail Sea,  $l_{total}$  converges when the number of iterations is 300. Table 1 shows the prediction results of distributed water quality prediction models based on ESN with different parameters in Oxtail Sea. When  $d = 3$ ,  $r = 500$ , and  $\rho = 0.7$ , the prediction result of Oxtail Sea is the best ( $l_{total} = 0.085$ ,  $MSE = 0.0060$ ). Similarly, in the experiment of the Western Buffer District,  $l_{total}$  converges also when the number of iterations is 300. Table 2 shows the prediction results of distributed water quality prediction models based on ESN with

different parameters in the Western Buffer District. When  $d = 3$ ,  $r = 500$ , and  $\rho = 0.6$ , the prediction result of the Western Buffer District is the best ( $l_{total} = 0.0111$ ,  $MSE = 0.0106$ ). The optimal parameters mentioned above are used in the subsequent experiments.

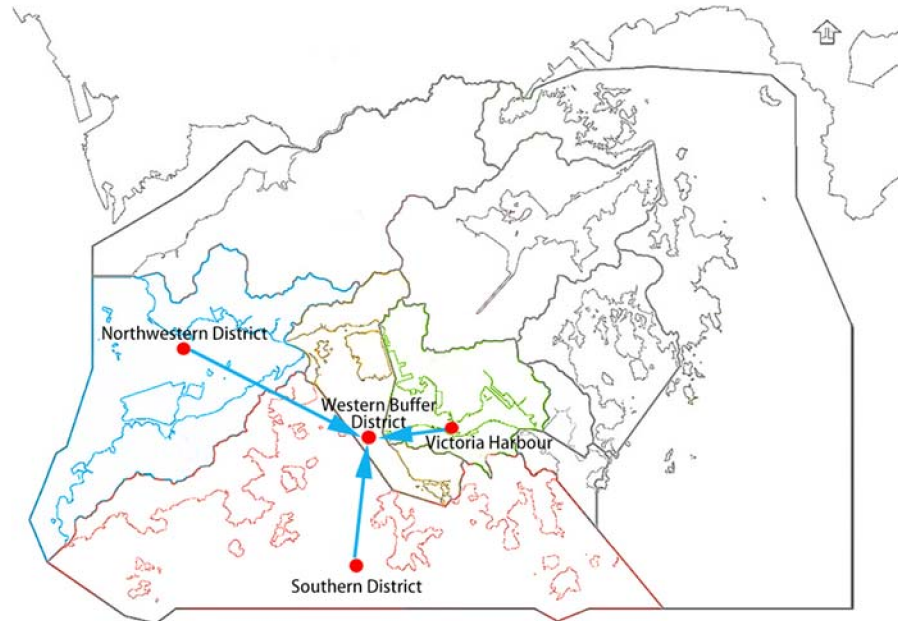


Figure 7. Spatial location of Western Buffer District and nearby monitoring points.

Table 1. Prediction results with different parameters in Oxtail Sea. The bold numbers are the best parameters.

$d$	$r$	$\rho$	$l_{total}$	$MSE$
2	250	0.6	0.092	0.0062
	250	0.7	0.093	0.0064
	250	0.8	0.092	0.0062
	250	0.9	0.091	0.0061
	500	0.6	0.092	0.0063
	500	0.7	0.093	0.0062
	500	0.8	0.092	0.0061
	500	0.9	0.095	0.0065
	750	0.6	0.092	0.0062
	750	0.7	0.093	0.0063
3	250	0.6	0.089	0.0063
	250	0.7	0.088	0.0062
	250	0.8	0.088	0.0062
	250	0.9	0.087	0.0062
	500	0.6	0.087	0.0061
	500	<b>0.7</b>	<b>0.085</b>	<b>0.0060</b>
	500	0.8	0.088	0.0063
	500	0.9	0.086	0.0061
	750	0.6	0.087	0.0062
	750	0.7	0.089	0.0063
750	0.8	0.089	0.0064	
750	0.9	0.088	0.0063	

**Table 1.** *Cont.*

$d$	$r$	$\rho$	$l_{total}$	$MSE$
4	250	0.6	0.093	0.0066
	250	0.7	0.089	0.0062
	250	0.8	0.092	0.0063
	250	0.9	0.090	0.0061
	500	0.6	0.092	0.0063
	500	0.7	0.094	0.0065
	500	0.8	0.093	0.0062
	500	0.9	0.092	0.0061
	750	0.6	0.092	0.0062
	750	0.7	0.093	0.0063
	750	0.8	0.094	0.0064
	750	0.9	0.089	0.0062

**Table 2.** Prediction results with different parameters in the Western Buffer District. The bold numbers are the best parameters.

$d$	$r$	$\rho$	$l_{total}$	$MSE$
2	250	0.6	0.0113	0.0108
	250	0.7	0.0114	0.0108
	250	0.8	0.0112	0.0107
	250	0.9	0.0115	0.0110
	500	0.6	0.0114	0.0108
	500	0.7	0.0114	0.0109
	500	0.8	0.0115	0.0111
	500	0.9	0.0114	0.0109
	750	0.6	0.0115	0.0110
	750	0.7	0.0116	0.0109
	750	0.8	0.0115	0.0108
	750	0.9	0.0115	0.0109
3	250	0.6	0.0116	0.0111
	250	0.7	0.0116	0.0110
	250	0.8	0.0115	0.0109
	250	0.9	0.0115	0.0108
	<b>500</b>	<b>0.6</b>	<b>0.0111</b>	<b>0.0106</b>
	500	0.7	0.0113	0.0108
	500	0.8	0.0114	0.0109
	500	0.9	0.0115	0.0109
	750	0.6	0.0115	0.0110
	750	0.7	0.0114	0.0109
	750	0.8	0.0114	0.0108
	750	0.9	0.0115	0.0109
4	250	0.6	0.0113	0.0107
	250	0.7	0.0114	0.0108
	250	0.8	0.0114	0.0108
	250	0.9	0.0115	0.0110
	500	0.6	0.0116	0.0112
	500	0.7	0.0115	0.0111
	500	0.8	0.0114	0.0109
	500	0.9	0.0113	0.0108
	750	0.6	0.0114	0.0110
	750	0.7	0.0115	0.0111
	750	0.8	0.0114	0.0109
	750	0.9	0.0113	0.0108

### 3.3. Comparison of Transfer Methods

We compare MSTL with non-expansion, TCA [20] and the joint class proportion and optimal transport (JCPOT) [30]. In particular, non-expansion uses only the water quality information of the target monitoring point. TCA transfers the features of water quality samples from a single nearby monitoring point to the target monitoring point, and selects the high-quality samples based on similarity and time sequence. JCPOT predicts the water quality information by using the optimal transport to correct and align the feature alignment bias between multiple source domains and target domain. MSTL extracts and aligns the features of water quality samples of multiple nearby monitoring points and the target monitoring point and trains the model through the aligned samples. The prediction results of different transfer methods in the Oxtail Sea and the Western Buffer District locations are shown in Table 3.

**Table 3.** Prediction results of different transfer methods in the Oxtail Sea and the Western Buffer District.

Monitoring Points	Method	Nearby Monitoring Point	MSE
Oxtail Sea	non-expansion	non	0.0167
	TCA	Tolo Harbour	0.0118
		Mirs Bay	0.0144
	JCPOT	Southern District	0.0120
	JCPOT	Tolo Harbour, Mirs Bay, Southern District	0.0133
	MSTL	Tolo Harbour, Mirs Bay, Southern District	0.0048
Western Buffer District	non-expansion	non	0.0128
	TCA	Northwestern District	0.0120
		Southern District	0.0125
	JCPOT	Victoria Harbour	0.0118
	JCPOT	Northwestern District, Southern District, Victoria Harbour	0.0125
	MSTL	Northwestern District, Southern District, Victoria Harbour	0.0104

From Table 3, we can observe that the prediction bias of MSTL is lower than that of non-expansion either in the Oxtail Sea or in the Western Buffer District. Besides, the prediction bias of MSTL is lower than that of TCA and JCPOT, because TCA can only use the water quality information of a single nearby monitoring point and JCPOT does not consider the effect of the feature alignment bias between different source domains. The prediction results show that MSTL can effectively use the water quality information of nearby monitoring points to train multiple water quality prediction models, which can reduce the model prediction bias and improve the prediction accuracy.

### 3.4. Comparison of Prediction Models

In the water quality prediction framework based on MSTL, we compare the water quality prediction models based on ESN with the water quality prediction models based on back propagation (BP) network, and the water quality prediction models based on gated recurrent unit (GRU) network. Like ESN, both BP and GRU have only one hidden layer. As a widely used basic neural network, BP has the advantages of simple structure and small calculation. As an improvement of LSTM, GRU adds a gating mechanism to make it have a memory ability. Compared with BP, the training of GRU is more complex. Partial prediction results of different prediction models in the Oxtail Sea and the Western Buffer District are shown in Figures 8 and 9, respectively. The comparisons of different water quality prediction models in terms of prediction bias and training time are shown in Figures 10 and 11.

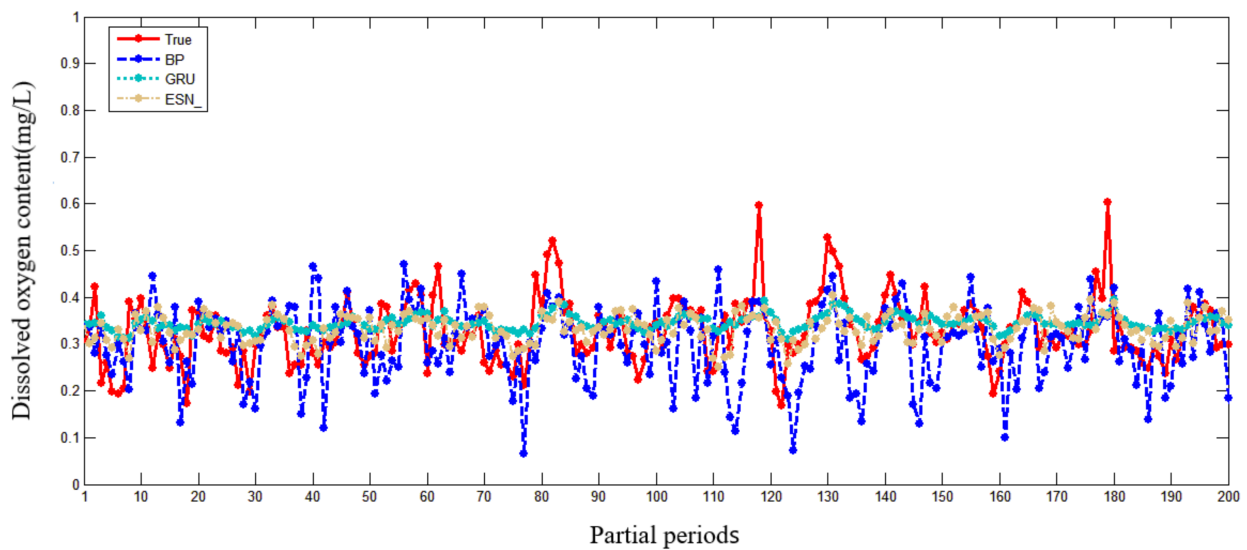


Figure 8. Partial prediction results of different prediction models in Oxtail Sea.

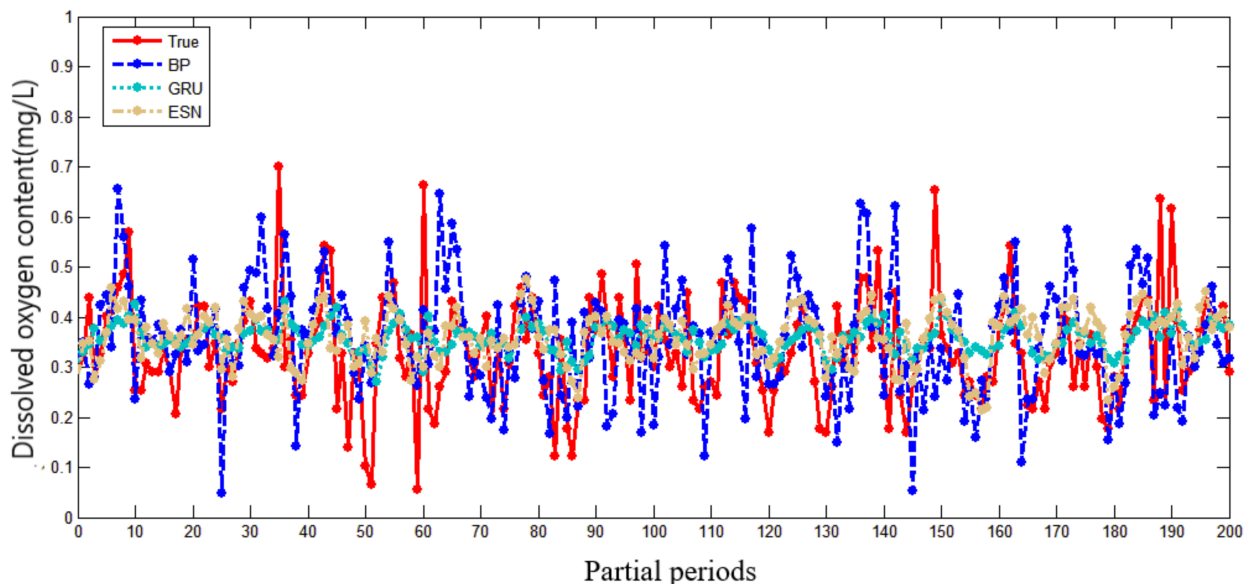


Figure 9. Partial prediction results of different prediction models in Western Buffer District.

Figures 8 and 9 show that the accuracy of BP is poor, and the prediction results fluctuate greatly. The prediction results of GRU and ESN are close when the data fluctuate slightly. Overall, ESN has better prediction ability than that of GRU either in the peak or in the valley part of the data. Figures 10 and 11 show that ESN has the smallest prediction bias and the shortest training time in the Oxtail Sea or the Western Buffer District, because ESN has a special reservoir structure and use only a simple linear regression algorithm for training.

To further illustrate the prediction accuracy of the proposed method, Figure 12 gives the box-plot comparison of the predicted water quality information and the real water quality information in the Oxtail Sea and the Western Buffer District. As seen in the figure, there exists a nearly uniform presentation through the observations of the measures, including the upper and lower quartiles, the upper and lower bound, the median and the outliers.

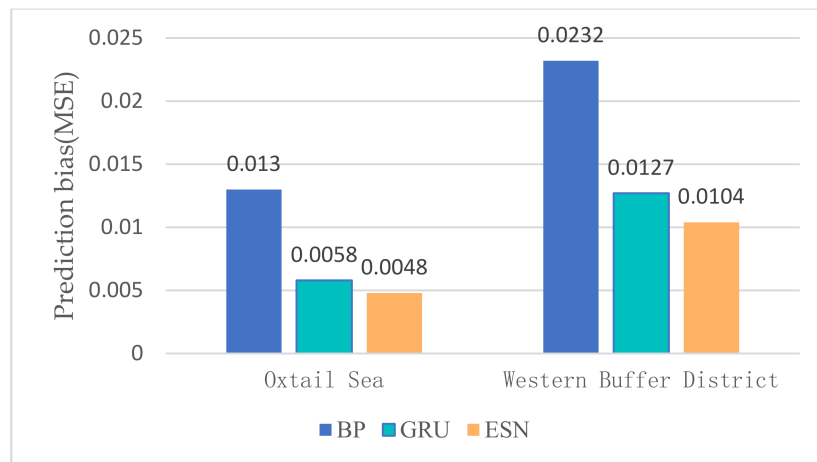


Figure 10. Comparison of prediction bias of different water quality prediction models.

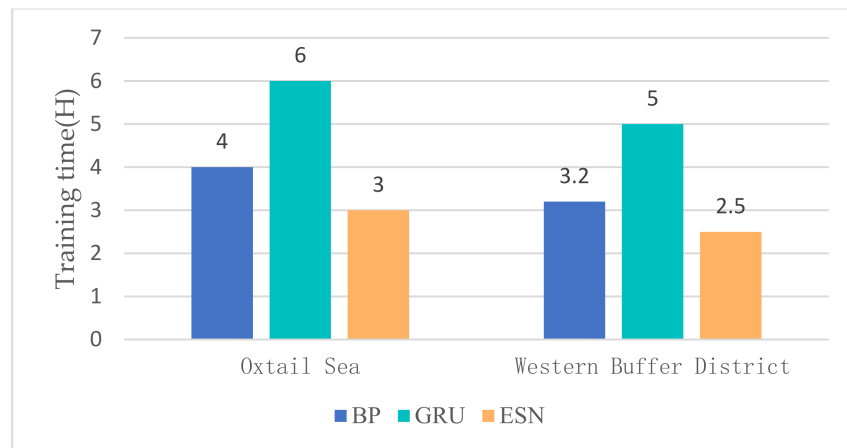


Figure 11. Comparison of training time of different water quality prediction models.

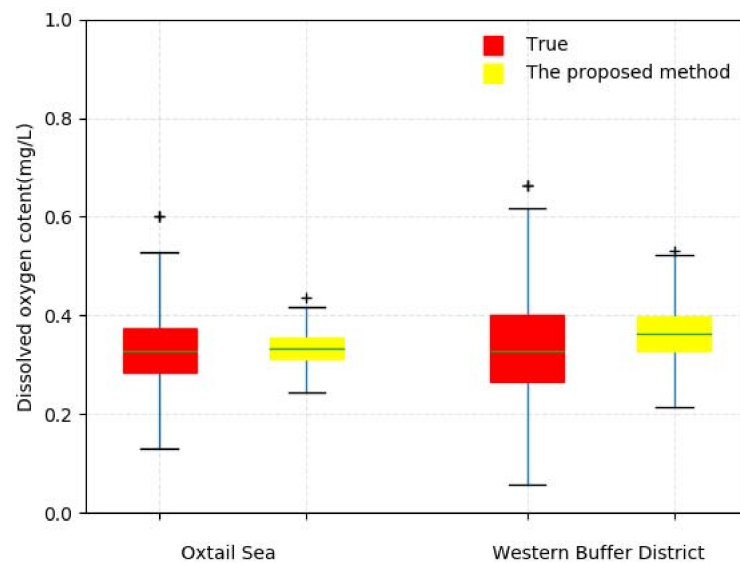


Figure 12. Box-plot comparison of predicted water quality information and real water quality information.

#### 4. Conclusions

Water environmental IoT system, which can collect water quality information in real time, provides the possibility for accurate water quality prediction. In this paper,

we propose a water quality prediction method based on MSTL for water environmental IoT system, to effectively use the water quality information of nearby monitoring points, and then improve the prediction accuracy of water quality. First, a water quality prediction framework based on MSTL is constructed, which establishes multiple water quality prediction models based on ESN at multiple nearby monitoring points with distributed computing. Second, the water quality prediction parameters of MSTL are optimized. Specifically, the back propagates population deviation based on multiple iterations reducing the feature alignment bias and the model alignment bias. Finally, the proposed method is compared with other similar methods in the actual water quality dataset of Hong Kong. The experimental results demonstrate that the proposed method can effectively align the features of water quality samples of multiple nearby monitoring points through MSTL and use the aligned samples of multiple nearby monitoring points to train multiple water quality prediction models, which can effectively reduce the prediction bias. It should be noted that the same type of sensors needs to be used at different monitoring points to collect the data of the same water quality indicators, so that the prediction models at different monitoring points have the same input parameter. In the following work, we will study to break through this limitation.

**Author Contributions:** Conceptualization, J.Z. and J.W.; Methodology, J.Z.; Software, Y.C.; Formal Analysis, X.L.; Writing—Original Draft Preparation, J.W.; Writing—Review & Editing, Y.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China (No. 61972210, 61873131, 61802206, 61872191, 61803212), Natural Science Foundation of Jiangsu Province (No. BK20211272) and the 1311 Talent Program of Nanjing University of Posts and Telecommunications.

**Acknowledgments:** The authors wish to thank the environmental protection department of Hong Kong for providing water quality dataset.

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

## References

1. Li, X.; Sha, J.; Wang, Z.L. Chlorophyll-A prediction of lakes with different water quality patterns in China based on hybrid neural networks. *Water* **2017**, *9*, 524. [[CrossRef](#)]
2. Pasika, S.; Gandla, S.T. Smart water quality monitoring system with cost-effective using IoT. *Heliyon* **2020**, *6*, 1–9. [[CrossRef](#)]
3. Liu, S.Y.; Tai, H.J.; Ding, Q.S.; Li, D.L.; Xu, L.Q.; Wei, Y.G. A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Math. Comput. Model.* **2013**, *58*, 458–465. [[CrossRef](#)]
4. Ratko, G. Stream water temperature prediction based on gaussian process regression. *Expert Syst. Appl.* **2013**, *40*, 7407–7414.
5. Anja, D.P.; Tertius, H.; Fethi, A. Quantifying and predicting the water quality associated with land cover change: A case study of the blesbok spruit catchment, South Africa. *Water* **2014**, *6*, 2946–2968.
6. Rezaie-Balf, M.; Kim, S.; Fallah, H.; Alaghmand, S. Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in Iran and South Korea. *J. Hydrol.* **2019**, *572*, 470–485. [[CrossRef](#)]
7. Zhang, F.; Xue, H.F.; Ma, X.M.; Wang, H.N. Grey Prediction model for the chemical oxygen demand emissions in industrial waste water: An empirical analysis of china. *Glob. Congr. Manuf. Manag.* **2017**, *174*, 827–834. [[CrossRef](#)]
8. Yang, L.; Qun, C. Poyang lake water quality model for dynamic prediction. In Proceedings of the International Conference on Computational and Information Sciences, Chongqing, China, 17–19 August 2012; pp. 1214–1216.
9. Xue, D.; Sudan, G.; Tong, L. Study on the prediction of mineralization degree of groundwater based on grey prediction model. In Proceedings of the Conference on Earth and Environmental Science, Pangkal Pinang, Indonesia, 3–4 September 2019; pp. 1–9.
10. Xiao, M.; Li, W.M.; Liu, D.F.; Xie, S.; Liu, X.Q. Study on the trend prediction of water bloom change in the backwater area of the tributaries of the Xiangxi River in the Three Gorges Reservoir area based on multiple optimization gray models. *Acta Sci. Circumstantiae* **2017**, *37*, 1153–1161.
11. Zhou, J.L.; Sun, J.; Zhang, M.Y.; Ma, Y. Dependable scheduling for real-time workflows on cyber-physical cloud systems. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7820–7829. [[CrossRef](#)]
12. Wang, T.; Lu, Y.C.; Wang, J.H.; Dai, H.N.; Zheng, X.; Jia, W.J. EIHPD: Edge-intelligent hierarchical dynamic pricing based on cloud-edge-client collaboration for IoT systems. *IEEE Trans. Comput.* **2021**, *70*, 1285–1298. [[CrossRef](#)]
13. Wu, Z.B.; Sun, J.; Zhang, Y.; Zhu, Y.Q.; Li, J.; Plaza, A.; Benediktsson, J.A.; Wei, Z.H. Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures. *IEEE Trans. Cybern.* **2020**, *51*, 3588–3601. [[CrossRef](#)] [[PubMed](#)]

14. Dawood, T.; Elwakil, E.; Novoa, H.M.; Delgado, J.F.G. Toward urban sustainability and clean potable water: Prediction of water quality via artificial neural networks. *J. Clean. Prod.* **2020**, *291*, 1–12.
15. Zhou, J.; Wang, Y.Y.; Xiao, F.; Wang, Y.Y.; Sun, L.J. Water quality prediction method based on IGRA and LSTM. *Water* **2018**, *10*, 1148. [[CrossRef](#)]
16. Dong, Q.X.; Lin, Y.Z.; Bi, J.; Yuan, H.T. An integrated deep neural network approach for large-scale water quality time series prediction. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Bari, Italy, 6–9 October 2019; pp. 3537–3542.
17. Hu, Z.H.; Zhang, Y.R.; Zhao, Y.C.; Xie, M.S.; Zhong, J.Z.; Tu, Z.G.; Liu, J.T. A water quality prediction method based on the Deep LSTM network considering correlation in smart mariculture. *Sensors* **2019**, *19*, 1420. [[CrossRef](#)]
18. Lee, J.H. Determination of optimal water quality monitoring points in sewer systems using entropy theory. *Entropy* **2013**, *15*, 3419–3434. [[CrossRef](#)]
19. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2010**, *22*, 199–210. [[CrossRef](#)] [[PubMed](#)]
20. Zhou, J.; Chen, Y.; Xiao, F.; Yan, X.Y.; Sun, L.J. Water quality prediction method based on transfer learning and echo state network. *J. Circuits Syst. Comput.* **2021**, 2150262. [[CrossRef](#)]
21. Sun, S.; Shi, H.; Wu, Y. A survey of multi-source domain adaptation. *Inf. Fusion* **2015**, *24*, 84–92. [[CrossRef](#)]
22. Yao, Y.; Doretto, G. Boosting for transfer learning with multiple sources. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 13–18.
23. Jaeger, H. The “echo state” approach to analysing and training recurrent neural networks. *Natl. Res. Cent. Inf. Technol.* **2001**, *148*, 1–47.
24. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [[CrossRef](#)]
25. Jun, M.; Knutti, R.; Nychka, D.W. Spatial analysis to quantify numerical model bias and dependence. *J. Am. Stat. Assoc.* **2008**, *103*, 934–947. [[CrossRef](#)]
26. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Sza, V.; Chen, Y.H.; Yang, T.J.; Emer, J.S. Efficient processing of deep neural networks: A tutorial and survey. *Proc. IEEE* **2017**, *105*, 2295–2329. [[CrossRef](#)]
28. Jaeger, H.; Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **2004**, *304*, 78–80. [[CrossRef](#)]
29. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Scholkopf, B.; Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
30. Redko, I.; Courty, N.; Flamary, R.; Tuia, D. Optimal transport for multi-source domain adaptation under target shift. In Proceedings of the Conference on Artificial Intelligence and Statistics, Okinawa, Japan, 16–18 April 2019; pp. 1–10.