

Article

Spatio-Temporal Analysis of Marine Water Quality Data Based on Cross-Recurrence Plot (CRP) and Cross-Recurrence Quantitative Analysis (CRQA)

Zhigang Li ^{1,2}, Ting Sun ^{1,2,3}, Yu Wang ^{1,2,3}, Yujie Liu ^{1,2,3} and Xiaochuan Sun ^{1,2,*} 

¹ College of Artificial Intelligence, North China University of Science and Technology, Bohai Road, Tangshan 063210, China

² Hebei Key Laboratory of Industrial Perception, Tangshan 063210, China

³ Marine Ecological Restoration and Smart Ocean Engineering Research Center of Hebei Province, Qinhuangdao 066000, China

* Correspondence: sunxiaochuan@ncst.edu.cn

Abstract: In recent years, with the frequency of marine disasters, water quality has become an important environmental problem for researchers, and much effort has been put into the prediction of marine water quality. The temporal and spatial correlation of marine water quality parameters directly determines whether the marine time-series data prediction task can be completed efficiently. However, existing research has only focused on the correlation analysis of marine data in a certain area and has ignored the temporal and spatial characteristics of marine data in complex and changeable marine environments. Therefore, we constructed a spatio-temporal dynamic analysis model of marine water quality based on a cross-recurrence plot (CRP) and cross-recurrence quantitative analysis (CRQA). The time-series data of marine water quality were first mapped to high-dimensional space through phase space reconstruction, and then the dynamic relationship among various factors affecting water quality was visually displayed through CRP. Finally, their correlation was quantitatively explained by CRQA. The experimental results showed that our scheme demonstrated well the dynamic correlation of various factors affecting water quality in different locations, providing important data support for the spatio-temporal prediction of marine water quality.

Keywords: marine water quality; CRP; CRQA; dynamic correlation; spatio-temporal analysis



Citation: Li, Z.; Sun, T.; Wang, Y.; Liu, Y.; Sun, X. Spatio-Temporal Analysis of Marine Water Quality Data Based on Cross-Recurrence Plot (CRP) and Cross-Recurrence Quantitative Analysis (CRQA). *Entropy* **2023**, *25*, 689. <https://doi.org/10.3390/e25040689>

Academic Editor: Sotiris Kotsiantis

Received: 15 February 2023

Revised: 12 April 2023

Accepted: 14 April 2023

Published: 19 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the acceleration of industrialization, the discharge load of industrial, agricultural, and domestic sewage in coastal areas has increased, and the deterioration of marine water quality has worsened year by year. Red tide disasters and the eutrophication of water bodies occur frequently. The quality of marine water is not only important for the economy, but it also affects human life. Therefore, the analysis of marine water quality has become a hot topic for researchers in recent years. Industrialization, mining, pollution, and natural disasters affect water quality. They introduce or change various parameters of water, thus affecting whether water is suitable for human consumption or general use [1].

Because of the important role of the marine environment in social development, marine data analysis has always been a hot topic. Water quality is affected by many factors, such as meteorology, chemistry, and human activities, which leads to the nonlinearity, randomness, and delay of water quality parameters. Statistical learning and other methods are widely used in the analysis and modeling of marine data. Jiang et al. [2] used the parallel FP-growth algorithm to analyze the oxygen, temperature, phosphates, nitrates, and silicates in the ocean. Meanwhile, based on association rules, they analyzed the correlation between different data. Erni-Cassola et al. [3] used the meta-analysis method to detect increased marine plastic debris in the sea surface and explored the separation degree of plastic debris

in the water according to the density of microplastic polymers, so as to further determine whether the plastic debris sank under the sea. Lemenkova [4] used a statistical database embedded in Python for data analysis to study the interaction between environmental factors affecting the seafloor geomorphology of the Mariana Trench, established a geospatial data model, and studied the inhomogeneity of the seafloor structure. Deng et al. [5] put forward a water quality analysis framework based on the time-series data-mining method, determined the relationship between water quality in the main stream and tributaries of the Yangtze River and the variation rule of dissolved oxygen, and effectively mined valuable knowledge from historical time-series data of water quality. However, the above research methods did not pay attention to the changes in the data across time and space, ignoring the effects brought about by cross-regional factors. Additionally, they lacked an analysis of the internal dynamic characteristics and spatial characteristics of the data. It is necessary to accurately analyze the non-stationarity and spatial heterogeneity of marine data.

A CRP can mine the correlation characteristics of time-series data in space-time, process nonlinear and non-stationary high-dimensional data, and realize the visualization analysis of marine data information. A CRP detects the correlation between signals by mapping two signals to the same phase space. The details of dynamic changes between two signals can be intuitively explained by the classical structure of points, diagonals, and rectangles. In addition, the traditional analysis method of a CRP is CRQA, which describes the CRP of signals in different states through nonlinear characteristic quantities such as the recurrence rate, certainty rate, recurrence entropy, and stratification rate. Both have excellent applications in disease diagnosis, industrial fault identification, financial market analysis, intelligent oceans, and many other fields. Kanakambaran et al. [6] used a CRP to analyze signals captured by fiber sensors in order to improve the accuracy of partial discharge detection and localization. González et al. [7] used a CRP and CRQA to analyze systolic blood pressure and inter-beat intervals in healthy persons and nephrotic patients. Wang et al. [8] applied a CRP and CRQA to the time recording of current and voltage fluctuations in electrochemical noise analysis to identify dynamic characteristics and achieve efficient detection results. Li et al. [9] demonstrated the dependent behavior of chlorophyll a with various marine factors using a CRP and CRQA and quantified their dynamic similarity using recurrence analysis. Currently, the use of CRPs and CRQA in the marine field mainly focuses on a single area and lacks the exploration of the spatio-temporal dynamic characteristics of marine data. In view of the advantages of CRPs and CRQA in many fields and in the analysis of marine environmental data, in this study, we used them to analyze the dynamic correlation of marine water quality in time and space.

In order to solve the problem that marine water quality data analysis mainly focuses on a single region and ignores the spatio-temporal impact across regions, this paper proposes a spatio-temporal dynamic analysis model of marine water quality data based on CRPs and CRQA. Due to the complex and changeable characteristics of the marine environment, various data not only interact with each other, but also interact with factors in different locations. Considering the influence of time and space, the accurate analysis of marine data is the key premise for water quality prediction. Therefore, we selected eight major factors from three stations to analyze the dynamic correlation among different stations. Specifically, the scheme consisted of phase space reconstruction (PSR), spatio-temporal dynamic CRP analysis, and CRQA quantification. Firstly, a marine factor of one station and influencing factors of the other two stations were mapped to the same phase space, and then their dynamic correlation was displayed through a CRP. Some key indicators of CRQA are used to evaluate the correlation of multiple marine water quality parameters. In particular, mutual information entropy can be used to measure the dependence between two discrete variables, so mutual information entropy is commonly used as an index to evaluate the correlation degree of marine factors. Finally, we could screen out the key factors affecting water quality. The results indicated that the scheme proved the interaction between marine data of different stations and effectively selected important factors, providing data support

for the input of future marine prediction models. The key contributions of this work can be summarized as follows.

- This was the first attempt to use CRPs and CRQA to study the dynamic association between water quality factors at different stations, as well as to analyze the factors in time and space.
- The water quality time-series data were converted to high-dimensional phase space through phase space reconstruction. After the reconstruction parameters were determined, the dynamic correlation relationship between the water quality data of different stations was displayed through a CRP, and the correlation degree was quantitatively described by CRQA and mutual information entropy.
- The spatial correlation of different influencing factors of water quality was confirmed, providing more reliable data support for the input of marine prediction tasks.

Specifically, the framework of the model is shown in Figure 1.

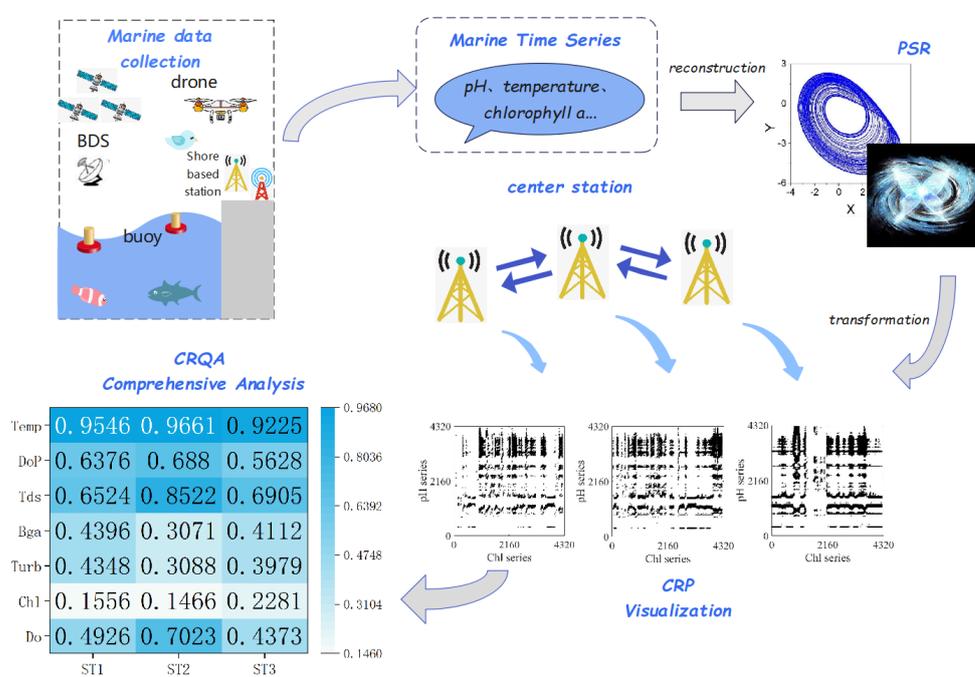


Figure 1. Marine water quality spatio-temporal analysis framework based on CRPs and CRQA.

The rest of this paper is structured as follows. Section 2 summarizes some relevant concepts in the correlation analysis of water quality factors. In Section 3, the experimental results from the CRP and CRQA investigation of the dynamic correlation between water quality factors at different stations are presented, and the correlation between the main water quality factors is discussed and explained. Finally, Section 4 provides conclusions and prospective research directions.

2. Materials and Methods

Herein, a new spatio-temporal recurrence analysis model for the influencing factors of marine water quality is introduced in detail. Firstly, the model is summarized, and then the main components of the model are introduced, including PSR, CRP, and CRQA.

2.1. Overview

In order to find the factors affecting the dynamic correlation of marine water quality in time and space and determine the main influencing factors of the quality of the water, we set up an analysis model based on CRPs and CRQA. According to the investigation and analysis, we selected eight kinds of water-quality-influencing factors from three stations

to analyze the spatio-temporal correlation between them. Specifically, our model consisted of the following three parts: PSR, CRP analysis, and CRQA quantification.

Taking pH as an example, firstly, the pH of the center station and the water-quality-influencing factors of the other three stations were mapped to the same phase space to calculate the reconstruction parameters. Secondly, the cross-recurrence matrix of the time series of the pH and another water quality factor was obtained through a CRP. Finally, some CRQA indicators were used to quantify the influence degree of each water quality factor on the pH. Based on the above experimental results, the temporal and spatial correlation between water quality factors was analyzed.

2.2. PSR Conversion

For chaotic time-series analysis, it is necessary to reconstruct the phase space formed by these sequence changes to explore the temporal dynamics of the system [10]. PSR can transform a one-dimensional time series to high-dimensional phase space. The reconstructed marine water quality time series could exhibit more nonlinear dynamic characteristics while preserving the continuity of the original series. PSR was considered as a prerequisite for CRPs in the task of the spatio-temporal correlation analysis of marine water quality factor data. Its purpose was to project the corresponding marine series into the high-dimensional phase space to obtain chaotic attractors. For the one-dimensional water quality time series $w(t) = w_1, w_2, \dots, w_i$ with a length of i , the reconstructed n -dimensional phase space could be expressed as follows:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} w_{1+t} & w_{1+2t} & \cdots & w_{1+(n-1)t} \\ w_{2+t} & w_{2+2t} & \cdots & w_{2+(n-1)t} \\ \vdots & \vdots & \vdots & \vdots \\ w_{N+t} & w_{N+2t} & \cdots & w_{N+(n-1)t} \end{bmatrix} \tag{1}$$

where $N = i + (n - 1)t$, t represents the delay time, and n represents the embedding dimension. In our paper, the mutual information method was considered to select the appropriate delay time, and the false nearest neighbors was used to select the suitable embedding dimension [11,12]. Note that unified reconstruction parameters should be selected so that two marine series can be reconstructed into the same phase space. A higher embedding dimension n and smaller delay time t were chosen in our scheme.

2.3. CRP Visualization

PSR is the first step in analyzing time series using a CRP [13]. After PSR, a CRP was used to detect the dynamic information of two marine water quality time series in the same phase space, and the data were visualized by two-dimensional graphics. The CRP was determined by a cross-recurrence matrix. For two reconstruction variables \vec{a} and \vec{b} , the matrix is defined as follows:

$$CR_{i,j}^{\vec{a};\vec{b}}(\varepsilon) = \Theta\left(\varepsilon - \|\vec{a}_i - \vec{b}_j\|\right), i = 1 \dots N, j = 1 \dots M \tag{2}$$

$$\Theta(e) = \begin{cases} 1, e \geq 0 \\ 0, e < 0 \end{cases} \tag{3}$$

where N and M denote the length of \vec{a}_i and \vec{b}_j , respectively; ε is the threshold that is 0.6% of the maximum phase space diameter [11]; and $\Theta(\bullet)$ represents the Heaviside function.

$CR_{i,j}$ is a two-dimensional matrix containing 1 and 0. When the distance between two sequences in the same phase space is less than ε , $CR_{i,j}$ is 1 and represented as a black point in the CRP; otherwise, $CR_{i,j}$ is 0 and represented as a white point in the CRP. Matching black and white dots with 1 and 0 were used to visualize similar behavior between two time series.

2.4. CRQA Quantification

In the analysis of marine time-series data, we used CRQA to quantify the frequency of similar changes between the time series of two water quality factors and then confirmed the dynamic spatial correlation of the time series of different water quality factors. CRQA indexes include mean diagonal line length (MDL), determinism (DET), laminarity (LAM), and recurrence rate (RR) [14–18].

RR, a metric for the density of recurrence points, exposes the probability that the time series of water quality factors remain similar under a certain delay. A high RR value indicates a high probability of similar states between both time series. It is determined by

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N CR_{i,j}(\varepsilon) \quad (4)$$

DET represents the ratio of diagonal structures. Usually, a high DET implies a deterministic process, while a low DET suggests a random process. In our case, it could evaluate the regularity and predictability of the interaction between two marine time series, expressed by

$$DET = \frac{\sum_{l=l_{\min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (5)$$

where l_{\min} is set to 2 [19], and $P(l)$ denotes the histogram of diagonals, defined as

$$P(l) = \sum_{i,j=1}^N (1 - CR_{i-1,j-1}) (1 - CR_{i+l,j+l}) \times \prod_{k=0}^{l-1} CR_{i+k,j+k} \quad (6)$$

LAM is the percentage of recurrence points comprising vertical structures. It hints at the steady states between two marine water quality time series, calculated as follows:

$$LAM = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (7)$$

where v_{\min} is set to 2, and $P(v)$ denotes the histogram of verticals, given by

$$P(v) = \sum_{i,j=1}^N (1 - CR_{i-1,j-1}) (1 - CR_{i,j+v}) \times \prod_{k=0}^{v-1} CR_{i,j+k} \quad (8)$$

MDL refers to the mean length of diagonal lines, and its value is the average time for which the time series of the marine water quality factors in the two groups are similar, given by

$$MDL = \frac{\sum_{l=l_{\min}}^N lP(l)}{\sum_{l=l_{\min}}^N P(l)} \quad (9)$$

2.5. Mutual Information Entropy

Information entropy is a measure of the uncertainty of random variable X in statistics. The higher the uncertainty of X, the greater the entropy. For a discrete random variable X,

its probability distribution is consistent with $p(x) = P(X = x)$, where $x \in X$. Information entropy is defined as follows:

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (10)$$

For two discrete random variables X and Y , the joint distribution probability is $p(X, Y)$, and the marginal distribution probability is $p(X)$, $p(Y)$, where $x \in X$ and $y \in Y$. Then, according to the definition of information entropy, the joint distribution entropy is

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (11)$$

Mutual information entropy is used to measure the degree to which one random variable X reduces the uncertainty of another random variable Y . The mutual information entropy between X and Y can be defined as

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (12)$$

3. Results

In this section, we present a thorough experimental evaluation of marine water quality within the framework of CRPs and CRQA. Firstly, we introduce the sources of the marine water quality data used in this study, followed by the determination of the PSR parameters. Secondly, a CRP was used to visualize the spatio-temporal characteristics of the marine water quality parameters at the three stations. Finally, some key CRQA indicators and mutual information entropy were used to quantify the degree of the temporal and spatial correlation of these parameters. We also explain the importance of the above research work for future marine water quality prediction tasks.

3.1. Datasets

The data used in this experiment came from a region in the Bohai Sea of China, and the data were collected by land-based stations and ocean buoys. A total of 4320 samples were collected from 22 July to 20 October 2021, with an interval of 30 min. After consulting the data, eight factors affecting the marine water quality were selected, including dissolved oxygen (Do), chlorophyll a (Chl), turbidity (Turb), blue-green algae (Bga), total dissolved solids (Tds), dissolved oxygen saturation (DoP), water temperature (Temp), and pH [20–25]. In order to explore the spatio-temporal characteristics of the marine water quality factors, the water quality factors were gathered from three adjacent stations. The locations and distances are shown in Figure 2. Specially, due to the large number of parameters calculated, the following example uses pH as the key indicator of marine water quality in order to illustrate the experimental process.



Figure 2. Geographical location of the three marine monitoring stations.

3.2. Parameter Determination

Before CRP was used for the dynamic correlation analysis of the marine water water quality data, one-dimensional time-series data of water quality had to be mapped to high-dimensional space, which was achieved through PSR. The delay time and embedding dimension had to be calculated. Taking pH as an example, the dynamic correlation between other water quality factors and the pH value was analyzed. Tables 1 and 2, respectively, show the embedding dimension and delay time of each water quality factor in the three stations. Table 3 shows the optimal matching of PSR parameters when pH and other water quality time series were reconstructed into the same high-dimensional space. A larger embedding dimension and a smaller delay time were generally considered the best matches.

Table 1. Determination of embedding dimension of water quality sequences at the three stations.

Station	Do	Chl	Turb	Bga	Tds	DoP	Temp	pH
ST1	5	8	6	5	6	5	5	6
ST2	4	9	8	8	7	6	5	6
ST3	4	7	15	17	7	5	6	5

Table 2. Determination of delay time of water quality sequences at the three stations.

Station	Do	Chl	Turb	Bga	Tds	DoP	Temp	pH
ST1	18	28	42	22	19	14	14	9
ST2	22	18	22	22	35	15	15	18
ST3	13	15	24	24	20	21	26	18

Table 3. Taking pH as an example, the best-match results of embedding dimension/delay time.

Station	Do	Chl	Turb	Bga	Tds	DoP	Temp	pH
ST1	6/9	8/9	6/9	6/9	6/9	6/9	6/9	6/9
ST2	6/9	9/9	8/9	8/9	7/9	6/9	6/9	6/9
ST3	6/9	7/9	15/9	17/9	7/9	6/9	6/9	6/9

3.3. Spatio-Temporal Visualization and Results

A CRP contains many classical structures to analyze the correlation of series, including recurrence points, diagonal lines, and vertical/horizontal diagonal distribution. Figures 3–5 show the correlation between the pH at ST1 and the other water quality factors at ST1, ST2,

and ST3. Taking pH as an example, we qualitatively determined the correlation between the water quality time-series data of the three stations and the pH of the central station, as shown in the figures below.

Figure 3 shows the correlation analysis between the pH and the other factors of the central station (ST1). It can be seen from the black recurrence points, rectangular structures, vertical horizontal lines, and other CRP structures in Figure 3 that all factors had a certain degree of correlation with the pH. Specifically, there are significantly more CRP structures in Figure 3d and Figure 3e,g than in Figure 3b,c. This indicates that the similarity of the recurrence behavior between Bga, Tds, and Temp and pH was stronger than that for Chl and Turb in ST1. The latter had a weak correlation with pH.

Figures 4 and 5 show the analysis results of the temporal and spatial correlation between the water quality factors at ST2 and ST3 and the pH at the central station (ST1), respectively. It can be clearly seen that the factors in Figures 4 and 5 also had similar characteristics to those in Figure 3. Specifically, there are a large number of rectangular structures in Figures 4e and 5e, and there are many horizontal line segments in Figures 4g and 5g. Nevertheless, in Figures 4b and 5b, there are few black recurrence points and large areas of blank space. This reflects a more synchronized state between the Tds and Temp at ST2 and ST3 and the pH at the central station. The pH and Chl at ST1 did not present obvious behavioral similarities. The above visualization results were consistent with those shown in Figure 3. By analyzing the recurrence behavior of factors among multiple stations, the temporal and spatial dynamic correlation characteristics of the marine water quality factors were verified.

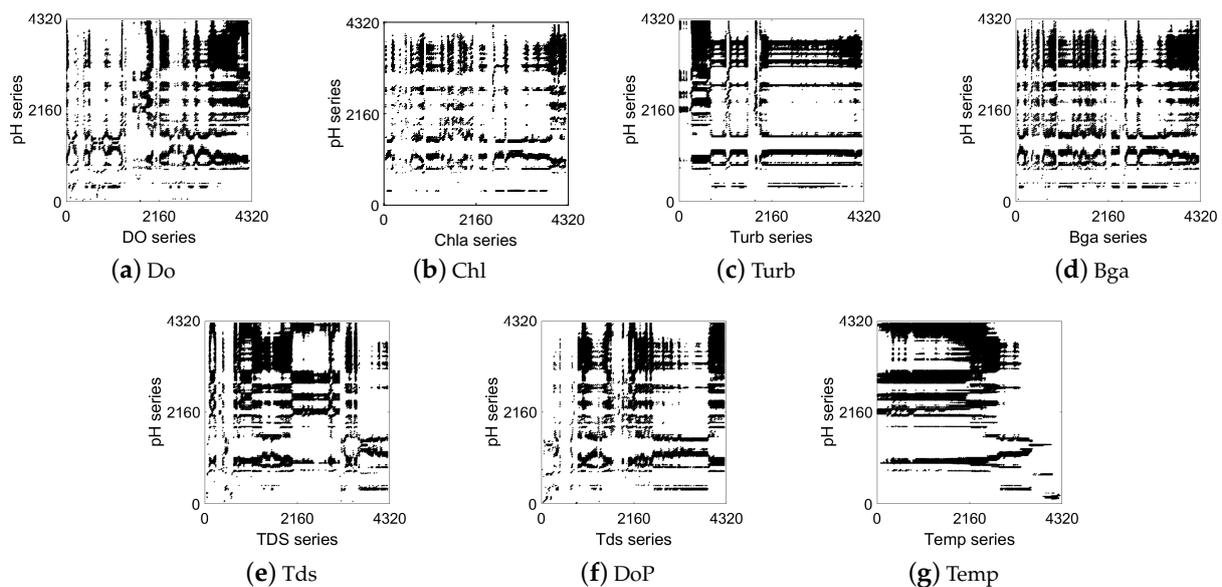


Figure 3. CRP visualization of correlation between marine factors and pH at ST1 station.

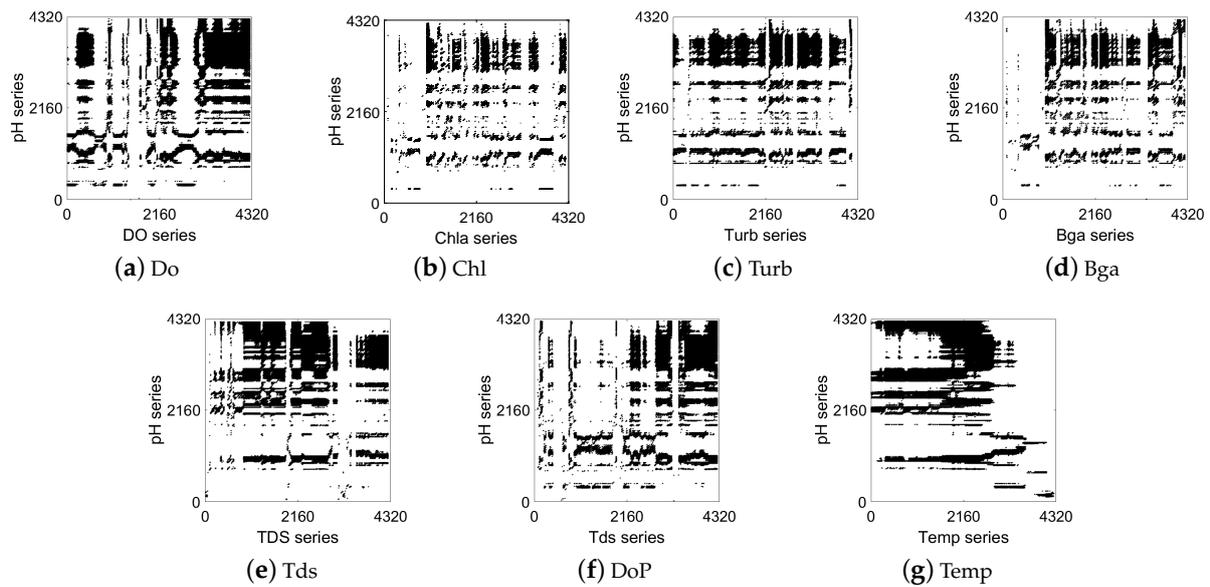


Figure 4. CRP visualization of correlation between marine factors at ST2 station and pH at ST1 station.

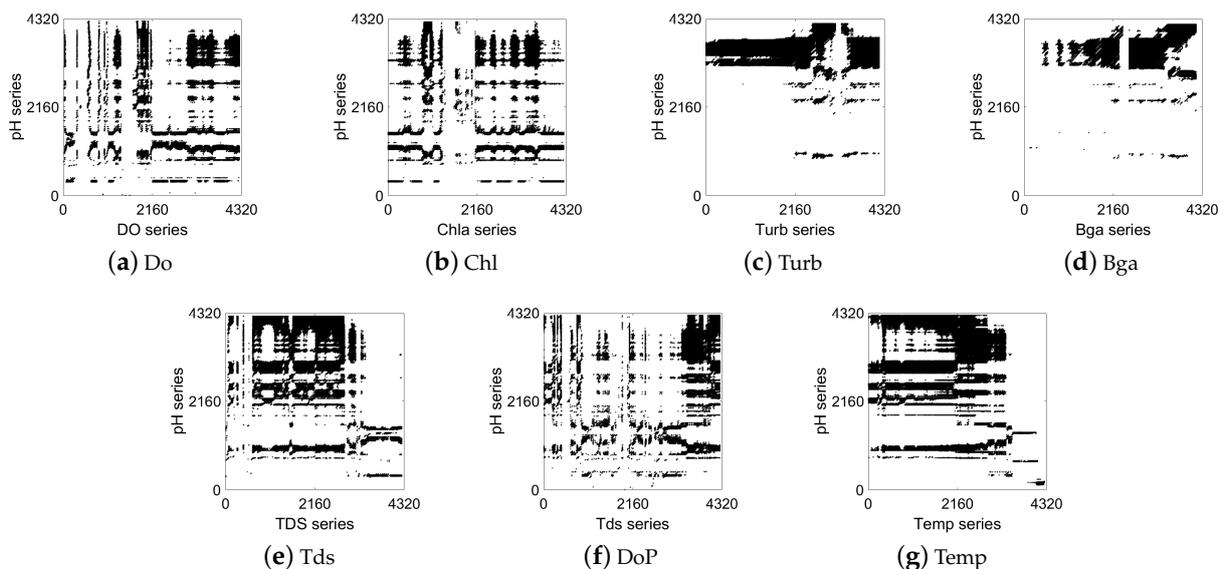


Figure 5. CRP visualization of correlation between marine factors at ST3 station and pH at ST1 station.

3.4. CRQA Analysis

As a quantitative analysis method of CRPs, CRQA includes several indicators to measure the influence degree of factors including RR, DET, MDL, LAM. In this study, we analyzed the influence degree of water quality factors through these four indicators and quantitatively evaluated the correlation of water quality factors by integrating the four indicators. Table 4 shows the quantification results of the four CRQA indexes for the correlation between seven water quality factors at three stations and pH at the central station.

RR indicates the probability that two water quality sequences had similar states. In the central station ST1, the RR value of Tds was 0.0727, which was significantly higher than that of the other water quality factors, indicating more similar recurrence states between the Tds and pH series. The RR value of Chl was only 0.0114, indicating a very weak correlation between the CHL and pH sequences. We also reached the same conclusion for the ST2 and ST3 stations, which indicated that the Tds at the ST2, ST3, and ST1 stations had a highly similar impact on the pH of the ST1 station. Similarly, Chl had the lowest RR values among

the three stations, indicating that there was no obvious correlation between Chl and pH when analyzing the impact of a single station or multiple stations.

As an important measure of the similarity of two time series, MDL reveals the average duration of similar states in the phase space of two time series. The values of Tds were 6.5991, 8.8260, and 7.0533, and the values of Temp were 9.0706, 7.3402, and 8.5231. These were still higher than the values for other water quality factors. The Chl MDL values were 3.4141, 3.6056, and 3.5846, representing the worst performance.

DET characterizes the determinacy of two sequences by calculating the ratio of the diagonal lines in the CRP. When two deterministic processes have similar phase space states, the corresponding DET value is larger. At the three stations, the values of Tds were 0.9355, 0.9513, and 0.9387, and the values of Temp were 0.9562, 0.9582, and 0.9531, respectively. The DET values of the two were significantly higher than those of the other factors, indicating that their states were more similar to the phase space of the pH sequence, while the findings for Chl were the opposite.

LAM represents the synchrony between sequences by quantifying the vertical/horizontal structures. From Table 4, it can be seen that the Tds and Temp values were the largest, while the Chl value was the smallest. This implied that Tds and Temp were more in sync with the pH sequences, and Chl was almost out of sync with pH.

Table 4. CRQA quantification results of associations between seven factors from three stations and pH at the central station.

Station	CRQA	Do	Chl	Turb	Bga	Tds	DoP	Temp	pH
ST1	MDL	5.2637	3.4141	5.2753	4.7075	6.5991	5.6745	9.0706	5.5753
	RR	0.0411	0.0114	0.0570	0.0386	0.0727	0.0452	0.0452	0.0602
	DET	0.8994	0.7758	0.8859	0.8684	0.9355	0.9063	0.9562	0.8940
	LAM	0.9281	0.8402	0.9087	0.9097	0.9446	0.9234	0.9574	0.9384
ST2	MDL	7.6584	3.6056	4.1526	3.9768	8.8260	0.7376	7.3402	9.1455
	RR	0.0848	0.0093	0.0339	0.0194	0.1029	0.0590	0.1082	0.1635
	DET	0.9398	0.7600	0.8166	0.8312	0.9513	0.9333	0.9582	0.9745
	LAM	0.9443	0.8580	0.8896	0.8874	0.9538	0.9417	0.9589	0.9866
ST3	MDL	4.7679	3.5846	5.8109	6.1439	7.0533	6.6590	8.5231	5.2118
	RR	0.0330	0.0218	0.0566	0.0317	0.0698	0.0401	0.1120	0.0456
	DET	0.8826	0.8180	0.8617	0.9069	0.9387	0.9222	0.9531	0.9236
	LAM	0.8992	0.8701	0.9225	0.9310	0.9501	0.9331	0.9553	0.9591

3.5. Quantitative Analysis

In Section 3.4, we found that under the four indexes, the water quality factors did not have exactly the same degree of influence on pH. For example, at ST2, the value of Turb was higher than that of Bga considering the influence of MDL, but the opposite was true at ST3. For LAM, the Do and DoP values also presented the same performance across stations. A single index cannot measure the effect of several water quality factors on another factor. Therefore, we added the value of mutual information entropy, normalized it with the four CRQA indexes, calculated the geometric mean value, and comprehensively measured the behavioral similarity among the water quality factors. By introducing mutual information entropy, the correlation between the data could be mined from the perspective of recurrence analysis and information theory, which made the scheme more convincing.

Figure 6 shows the comprehensive evaluation results of CRQA and mutual information entropy for the spatio-temporal correlation characteristics of the marine factors. The figure in the heat map is the combined value of the correlation between the two factors, ranging from 0 to 1. The larger the value, the stronger the temporal and spatial correlation between the two factors. At the same time, the colors reflect the degree of correlation along with the values. The darker the color, the stronger the relationship between the two factors. Taking Figure 6h as an example, considering their impact on the pH of the central station, the top four factors at ST1 were Temp, Tds, DoP, and Turb. At ST2, they were Temp, Tds, Do, DoP. At ST3, they were Temp, Tds, DoP, and Turb. In the case of interference by various factors, we could assume that the high-correlation factors of the three stations were basically the same. In the pH prediction task, we could select the water quality factors of the target station and the high-association factor data of the adjacent stations as the input. High-correlation factors for other water quality factors can also be found in Figure 6. In marine water quality multi-task prediction, one can conduct a comprehensive evaluation of water quality by predicting various factors affecting water quality at the same time, so that the prediction results are more convincing. In addition, the CRQA comprehensive measurement values corresponding to the marine water quality factors in Figure 6 are presented in Table A1.

Furthermore, in order to verify the scheme for multiple data lengths, we chose 2000 and 3000 pieces of data to draw a comparison with the analysis results of the total 4320 pieces of data. Figure 7 shows the correlation between the pH sequences for the three data lengths and other factors. As can be seen from the values and color shades in this figure, in the experiments with data lengths of 2000 and 3000, the correlation between the pH sequence and Tds and Temp was still large, while that for Chl was small. This was almost consistent with the results for 4320 pieces of data.

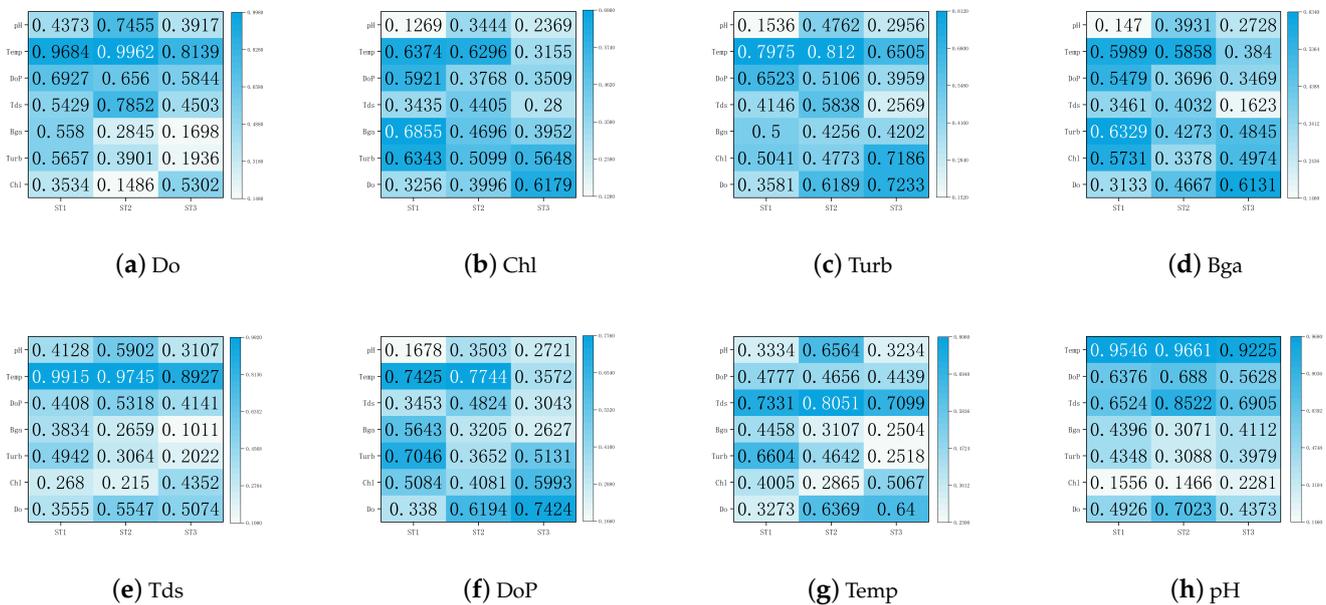


Figure 6. CRQA indicators and mutual information entropy comprehensive measurement results of the associations among marine factors.

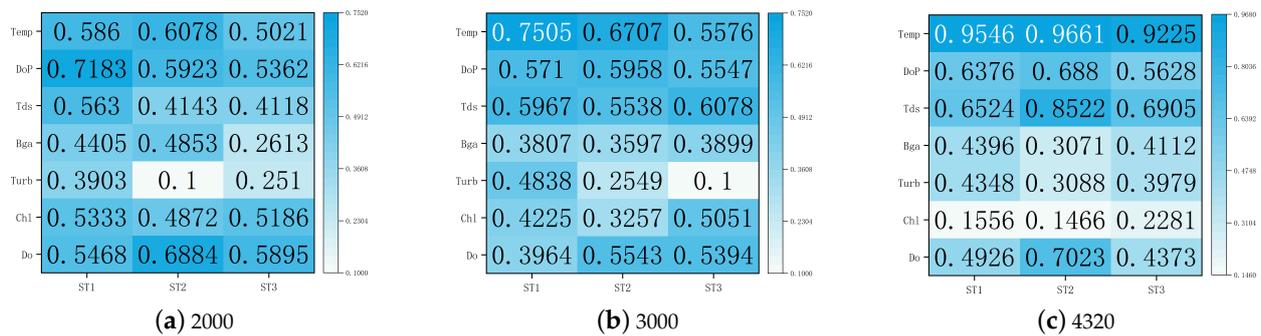


Figure 7. Comprehensive measurement results of correlation between pH sequences and other factors for different data lengths.

4. Conclusions

In this paper, we developed a spatio-temporal analysis model of marine water quality data based on CRPs and CRQA. Furthermore, mutual information entropy was introduced as one of the evaluation indexes. In contrast to single-station analysis, this model could analyze the spatio-temporal dynamic characteristics of water quality factors at multiple stations. Accordingly, we could obtain better prior data as the input for subsequent marine water quality forecasting tasks. Through spatio-temporal analysis, it could be concluded that certain variation factors of the target station were largely affected by other factors at adjacent stations. In future marine water quality multi-tasking prediction work, we will give full consideration to the effects of other stations and consider comprehensive data changes caused by many factors, so as to improve the accuracy of prediction and provide a more novel research scheme for marine prediction work.

Author Contributions: Conceptualization, Z.L. and X.S.; methodology, Y.L.; software, T.S.; validation, T.S. and Y.W.; formal analysis, Z.L.; investigation, Z.L. and Y.L.; data curation, Z.L.; writing—original draft preparation, T.S. and X.S.; writing—review and editing, T.S. and X.S.; visualization, Y.W.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of Hebei Education Department, grant ZD2021088; the open fund project of the Marine Ecological Restoration and Smart Ocean Engineering Research Center of Hebei Province, grant HBMESO2315; and the project Intelligent Early Warning Prediction and Demonstration Application of Pollutants from Land to Sea, grant 21373302D.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets for this study are available upon request from the corresponding author.

Acknowledgments: We thank the editorial board and all reviewers for their professional advice to improve this work.

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

Appendix A

Table A1. CRQA indicators and mutual information entropy comprehensive measurement results of the associations among 8 factors from 3 stations, in contrast to the central station factors.

Station	Contrast (from ST1)	Do	Chl	Turb	Bga	Tds	DoP	Temp	pH
ST1	pH	0.4926	0.1556	0.4348	0.4396	0.6524	0.6376	0.9546	0.1688
	Temp	0.3273	0.4005	0.6604	0.4458	0.7331	0.4777	0.9510	0.3334
	DoP	0.3380	0.5084	0.7046	0.5643	0.3453	0.4731	0.7425	0.1678
	Tds	0.3555	0.2680	0.4942	0.3834	0.4703	0.4408	0.9915	0.4128
	Bga	0.3133	0.5731	0.6329	0.4096	0.3461	0.5479	0.5989	0.1470
	Turb	0.3581	0.5041	0.6588	0.5000	0.4146	0.6523	0.7975	0.1536
	Chl	0.3256	0.3536	0.6343	0.6855	0.3435	0.5921	0.6374	0.1269
	Do	0.2996	0.3534	0.5657	0.5580	0.5429	0.6927	0.9684	0.4373
ST2	Ph	0.7023	0.1466	0.3088	0.3071	0.8522	0.6880	0.9661	0.3852
	Temp	0.6369	0.2865	0.4642	0.3107	0.8051	0.4656	0.9120	0.6564
	DoP	0.6194	0.4081	0.3652	0.3205	0.4824	0.3296	0.7744	0.3503
	Tds	0.5547	0.2150	0.3064	0.2659	0.4849	0.5318	0.9745	0.5902
	Bga	0.4667	0.3378	0.4273	0.1976	0.4032	0.3696	0.5858	0.3931
	Turb	0.6189	0.4773	0.3026	0.4256	0.5838	0.5106	0.8120	0.4762
	Chl	0.3996	0.2243	0.5099	0.4696	0.4405	0.3768	0.6296	0.3444
	Do	0.4799	0.1486	0.3901	0.2845	0.7852	0.6560	0.9962	0.7455
ST3	Ph	0.4373	0.2281	0.3979	0.4112	0.6905	0.5628	0.9225	0.2228
	Temp	0.6400	0.5067	0.2518	0.2504	0.7099	0.4439	0.7983	0.3234
	DoP	0.7424	0.5993	0.5131	0.2627	0.3043	0.2690	0.3572	0.2721
	Tds	0.5074	0.4352	0.2022	0.1011	0.3837	0.4141	0.8927	0.3107
	Bga	0.6131	0.4974	0.4845	0.2732	0.1623	0.3469	0.3840	0.2728
	Turb	0.7233	0.7186	0.5131	0.4202	0.2569	0.3959	0.6505	0.2956
	Chl	0.6179	0.3711	0.5648	0.3952	0.2800	0.3509	0.3155	0.2369
	Do	0.4511	0.5302	0.1936	0.1698	0.4503	0.5844	0.8139	0.3917

References

- Ajayi, O.O.; Bagula, A.B.; Maluleke, H.C.; Gaffoor, Z.; Jovanovic, N.; Pietersen, K.C. Waternet: A network for monitoring and assessing water quality for drinking and irrigation purposes. *IEEE Access* **2022**, *10*, 48318–48337. [\[CrossRef\]](#)
- Jiang, Y.; Zhao, M.; Hu, C. A parallel FP-growth algorithm on World Ocean Atlas data with multi-core CPU. *J. Supercomput.* **2019**, *75*, 732–745. [\[CrossRef\]](#)
- Erni-Cassola, G.; Zadjelovic, V.; Gibson, M.I.; Christie-Oleza, J.A. Distribution of plastic polymer types in the marine environment; A meta-analysis. *J. Hazard. Mater.* **2019**, *369*, 691–698. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lemenkova, P. Computing and plotting correlograms by Python and R libraries for correlation analysis of the environmental data in marine geomorphology. *Jeomorfol. Arařt. Derg.* **2019**, *3*, 1–16. [\[CrossRef\]](#)
- Deng, W.; Wang, G. A novel water quality data analysis framework based on time-series data mining. *J. Environ. Manag.* **2017**, *196*, 365–375. [\[CrossRef\]](#)
- Kanakambaran, S.; Sarathi, R.; Srinivasan, B. Identification and localization of partial discharge in transformer insulation adopting cross recurrence plot analysis of acoustic signals detected using fiber bragg gratings. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 1773–1780. [\[CrossRef\]](#)
- González, G.H.; Infante, O.; Martínez-García, P.; Pérez-Grovas, H.; Saavedra, N.; Caviedes, A.; Becerra, B.; Lerma, C. Dynamical interaction between heart rate and blood pressure of end-stage renal disease patients evaluated by cross recurrence plot diagonal analysis. *J. Appl. Physiol.* **2020**, *128*, 189–196. [\[CrossRef\]](#)

8. Liu, W.; Wang, D.; Chen, X.; Wang, C.; Liu, H. Recurrence plot-based dynamic analysis on electrochemical noise of the evolutive corrosion process. *Corros. Sci.* **2017**, *124*, 93–102. [[CrossRef](#)]
9. Li, Z.; Cai, D.; Wang, J. Machine learning based dynamic correlation on marine environmental data using cross-recurrence strategy. *IEEE Access* **2019**, *7*, 185121–185130. [[CrossRef](#)]
10. Shuai, H.; Wang, S.; Liu, Q. Phase space reconstruction driven spatio-temporal feature learning for dynamic facial expression recognition. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1466–1476. [[CrossRef](#)]
11. Li, Y.; Cai, D.; Wang, J.; Sun, X.; Li, Z.; Zhang, H.; Wang, N. Recurrence behavior statistics of blast furnace gas sensor data in industrial internet of things. *IEEE Internet Things J.* **2020**, *7*, 5666–5676. [[CrossRef](#)]
12. Xu, Z.; Zhong, L.; Zhang, A. Phase space reconstruction-based concepter network for time series prediction. *IEEE Access* **2019**, *7*, 163172–163179. [[CrossRef](#)]
13. Fan, G.F.; Peng, L.L.; Hong, W.C. Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *Appl. Energy* **2018**, *224*, 13–33. [[CrossRef](#)]
14. Coco, M.I.; Mønster, D.; Leonardi, G. Unidimensional and multidimensional methods for recurrence quantification analysis with CRQA. *arXiv* **2020**, arXiv:2006.01954. [[CrossRef](#)]
15. Yan, B.; Chan, P.W.; Li, Q. Dynamic analysis of meteorological time series in Hong Kong: A nonlinear perspective. *Int. J. Climatol.* **2021**, *41*, 4920–4932. [[CrossRef](#)]
16. Mukherjee, S.; Ray, R.; Samanta, R. Nonlinearity and chaos in wireless network traffic. *Chaos Solitons Fractals* **2017**, *96*, 23–29. [[CrossRef](#)]
17. Amiri, A.; Samet, H.; Ghanbari, T. Recurrence plots based method for detecting series arc faults in photovoltaic systems. *IEEE Trans. Ind. Electron.* **2021**, *69*, 6308–6315. [[CrossRef](#)]
18. Yang, D.; Ren, W.X.; Hu, Y.D. Selection of optimal threshold to construct recurrence plot for structural operational vibration measurements. *J. Sound Vib.* **2015**, *349*, 361–374. [[CrossRef](#)]
19. Martín-González, S.; Navarro-Mesa, J.L.; Juliá-Serdá, G.; Ramírez-Ávila, G.M.; Ravelo-García, A.G. Improving the understanding of sleep apnea characterization using recurrence quantification analysis by defining overall acceptable values for the dimensionality of the system, the delay, and the distance threshold. *PLoS ONE* **2018**, *13*, e0194462. [[CrossRef](#)]
20. Ashikur, M.; Rupom, R.; Sazzad, M. A remote sensing approach to ascertain spatial and temporal variations of seawater quality parameters in the coastal area of bay of Bengal, Bangladesh. *Remote. Sens. Appl. Soc. Environ.* **2021**, *23*, 100593. [[CrossRef](#)]
21. Deswati, D.; Safni, S.; Khairiyah, K.; Yani, E.; Yusuf, Y.; Pardi, H. Biofloc technology: Water quality (ph, temperature, do, cod, bod) in a flood & drain aquaponic system. *Int. J. Environ. Anal. Chem.* **2022**, *102*, 6835–6844. [[CrossRef](#)]
22. Elkiran, G.; Nourani, V.; Abba, S. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J. Hydrol.* **2019**, *577*, 123962. [[CrossRef](#)]
23. Najafzadeh, M.; Ghaemi, A.; Emamgholizadeh, S. Prediction of water quality parameters using evolutionary computing-based formulations. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 6377–6396. [[CrossRef](#)]
24. Xu, Z.; Shen, J.; Qu, Y.; Chen, H.; Zhou, X.; Hong, H.; Sun, H.; Lin, H.; Deng, W.; Wu, F. Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. *Chemosphere* **2022**, *286*, 131586. [[CrossRef](#)] [[PubMed](#)]
25. Susanti, N.D.; Sagita, D.; Apriyanto, I.F.; Anggara, C.E.W.; Darmajana, D.A.; Rahayuningtyas, A. Design and implementation of water quality monitoring system (temperature, ph, tds) in aquaculture using iot at low cost. In Proceedings of the 6th International Conference of Food, Agriculture, and Natural Resource (IC-FANRES 2021), Tangerang, Indonesia, 4–5 August 2021; pp. 7–11. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.