Simulation study on Multi-parameter Detection method of polluted Water quality based on Sponge City

Wei Zhong, Xinfei Zhao*
School of Management, Tianjin University of Technology, China
*Corresponding author’s e-mail: 819753100@qq.com

Abstract. With the accelerated pace of urbanization, the construction of sponge cities is particularly essential for maintaining the ecological environment. Given the traditional multi-parameter detection of heavily polluted water quality in cities, there are many problems, such as long completion time, low accuracy, high cost. A multi-parameter detection method of water quality based on the random forest classification algorithm is proposed. Through the analysis of urban heavily contaminated water quality, the probability synthesis matrix of heavily polluted water quality index is obtained by combining the water quality fluctuation probability of a single parameter and the dynamic correlation coefficient of multi-parameters. The membership function of the fluctuation of the treated water quality index is obtained, and the weak correlation among the water quality indexes is removed by the obtained membership function, and the strong relationship characteristic quantity of the water quality index is obtained, which is input into the random forest classifier to complete the multi-parameter detection of river water quality.

1. Introduction
Different from the traditional grey infrastructure, "Sponge City" is a living water ecological infrastructure. It not only maintains the essential environmental processes of urban rainfall and waterlogging regulation and storage, water source protection and conservation, groundwater recharge, rain and pollution purification, habitat restoration, and soil purification, but also can be scientifically laid out and operated on the ground in space. However, with the continuous progress of urbanization, the discharge of pollutants from urban work and life brings severe pollution to the river, and river pollution is becoming more serious [1]. According to the monitoring data of urban heavily polluted water quality by relevant regulatory departments, the relevant indicators of water quality have seriously exceeded the standard. The water environment has been severely polluted [2-3]. For the traditional water quality detection methods, there are many problems, such as long detection time, low accuracy, high cost, and so on. Given this situation, the study of an efficient and accurate river quality monitoring method has become an urgent problem to be solved in today's society [6]. To solve the above issues, a multi-parameter detection method of water quality based on the random forest classification algorithm is proposed. The experimental results show that the proposed method has the advantages of short completion time, high accuracy, and low cost.

2. The multi-parameter Detection method of Water quality in Sponge City
The technical route of domestic "sponge city" construction has only 12 words "source emission reduction, process control, system control". In contrast, the leading engineering and technical measures are more concise, with only six words "infiltration, lag, storage, net, use, and discharge."
2.1. Obtaining probability synthesis matrix of river water quality index

The multi-parameter index of the water quality of urban heavily polluted rivers can show different forms of fluctuation in a specific time range, because a large number of pollutants are discharged into the river due to heavy urban pollution, and the abnormal water quality causes the water quality index to fluctuate to a certain extent.

For the time series of a single parameter index of water quality, water quality fluctuation refers to the degree of difference between the monitored value and the reference value in a certain period. The primary probability distribution function of Dmurs evidence theory is used to determine the fluctuation probability of each index.

\[
m(\text{Abnormal}) = \frac{1 - \exp(-D/\sigma)}{1 + \exp(-D/\sigma)}
\]

\[
m(\text{Normal}) = 1 - m(\text{Abnormal})
\]

In the formula, \(m(\text{Abnormal})\) represents the probability of a fluctuation of the water quality index of the current urban heavily polluted river, \(D\) represents the short-term root mean square of the monitored water quality index, and \(\sigma\) represents the short-term RMS average value of the water quality of the urban heavily polluted river when the water quality does not fluctuate. When the fluctuation of river water quality index time series is not visible or there is no fluctuation, \(m(\text{Abnormal}) = 0\); when the change of river water quality index time series is pronounced, \(m(\text{Abnormal}) = 1\).

The dynamic time warping method is introduced into the time series analysis of water quality indexes. This method has better stability for the synchronization of water quality index time series, and the allowed unsynchronized points can be calculated accordingly so that there are stretching and shifting time series on the time axis. This method can be used to effectively measure the approximation between different water quality index time series. Assuming that \(\gamma\) represents electrical conductivity, the distance matrix of the water quality index is obtained according to the Euclidean distance between two points of water quality index.

\[
d(p, q) = \left(X_i^p, X_i^q\right)^2
\]

The essence of the dynamic time regulation method is to find a matching pattern with the shortest distance among all water quality index matching points.

\[
DTW(X_i, X_j) = \min \sum_{n=1}^{k} d_n / k
\]
Based on equation (3), the dynamic distance is obtained when the maximum allowable time deviation is zero.

$$DTW \left( X_t^{t_0}, X_j^{t_0} \right) = \min_{0 \leq p, q \leq t} \left( DTW \left( X_t^{t_0-p}, X_j^{t_0-q} \right) \right)$$  \hspace{1cm} (4)

The synthesis rule is the primary probability fusion method to calculate the time series of different water quality indexes under the same detection framework. In the water quality detection, under the fluctuation of the normal water quality index, the correlation of each parameter is small. When the river pollution is more serious, the water quality fluctuates obviously, and the relationship of each water quality index parameter is more excellent. The correlation coefficient between the two index parameters is multiplied by the dynamic time warping method, and the probability synthesis matrix of the water quality index is obtained.

$$P = \left[ p(X_t^{t_0})p(X_j^{t_0}) \right]$$  \hspace{1cm} (5)

2.2. Parameter detection method based on the random forest classification algorithm

According to the dynamic time regulation method, the process of multi-parameter optimization of river water quality is locally optimal rather than global optimal. According to this situation, the method can satisfy the exchange law.

$$R(i, j) = \max \left( R(i, j), R(j, i) \right)$$  \hspace{1cm} (6)

The characteristic quantity of water quality index is input into a random forest classifier to classify the multi-parameter test samples of heavily polluted river water quality. The random forest algorithm is an improved decision tree method. The classifier method mainly has two characteristics, information gain, and Gini coefficient. According to the principle of maximizing information gain and minimizing the Gini coefficient, the information entropy of each node of the river water quality parameter is calculated by information gain.

$$Info(K, S) = \sum_{K=1}^{n} \frac{|K_i|}{|K|} Info(K_i)$$  \hspace{1cm} (7)

If the Gini coefficient is used to measure the impurity of a multi-parameter sample of water quality, the expression is

$$Gini(K) = 1 - \sum_{i=1}^{m} P_i^2$$  \hspace{1cm} (8)

Based on formula (7) and formula (8), formula (9) is used to classify the multi-parameter test samples of heavily polluted river water quality.

$$c(x) = \arg \max_c \left( \frac{1}{M} \sum_{m=1}^{M} I(h_m(x) = c) \right)$$  \hspace{1cm} (9)

3. Proved by experiment and simulation

To verify the comprehensive effectiveness of the proposed water quality multi-parameter detection method based on random forest classification algorithm, an experiment needs to be carried out. The experimental environment is Intel-Core3-5304G, operating system Windows8, to compare the proposed method with the multivariable water quality parameter time series data detection method and the spectral on-line water quality detection method based on baseline correction and principal component analysis. The proposed method is compared with the time series data detection method based on multivariable water quality parameters and the spectral on-line water quality detection method based on baseline correction and principal component analysis ((s)). The experimental results are shown in Table 1. In Table 1, W represents the number and unit of water quality samples, g, F represents the test completion time, the group is seconds, s represents the proposed method. Q represents the detection method based on time series data of multivariable water quality parameters; Y represents the spectral on-line water quality detection method based on baseline correction and principal component analysis.
The analysis of Table 1 shows that with the continuous increase in the number of water quality samples, the detection completion time of the three methods also increases. When the amount of water quality samples increases from 25 to 90, the detection completion time of the proposed method increases by 29s, and the detection completion time of the method based on multivariable water quality parameters time series data increases by 61s. The completion time of spectral on-line water quality detection method based on baseline correction and principal component analysis is increased by 91s. The experimental results show that the proposed method has a small increase in detection time for the continuous increase in the amount of data, which is better than the other two ways and has higher detection efficiency.

According to the analysis of figure 2, with the continuous increase in the number of water quality samples, the corresponding detection accuracy of the three methods varies in varying degrees. When the amount of water quality samples is increased from 0 to 700, the detection accuracy of the proposed method has been fluctuating up and down in the range of 92% to 82%, and the fluctuation is relatively smooth. The detection accuracy of the detection method based on time series data of multivariable water quality parameters has been fluctuating up and down in the range of 85% to 51% and varies greatly. The detection accuracy of spectral on-line water quality detection method based on baseline correction and principal component analysis has been shifting in the range of 57% to 29%. The experimental results show that the proposed method can effectively ensure the stability of the test results and has good application value.

Table 2 shows that the cost consumption of the three methods increases with the continuous increase in the number of samples. When the amount of water quality samples is 18, the detection cost of the
proposed method is 101 yuan. The detection cost of the detection method based on multivariable water quality parameter time series data is 156 yuan, and the detection cost of spectral on-line water quality detection method based on baseline correction and principal component analysis is 498 yuan. Compared with the detection cost, the proposed method can save 55 yuan compared with the time series data detection method based on multivariable water quality parameters, and 397 yuan compared with the spectral on-line water quality detection method based on baseline correction and principal component analysis. When the number of water quality samples is 58, the detection cost of the proposed method is 197 yuan, the detection cost of the technique based on multivariable water quality parameter time series data is 471 yuan, and the detection cost of spectral on-line water quality detection method based on baseline correction and principal component analysis is 1432 yuan. Compared with the detection cost, the proposed method can save 274 yuan compared with the detection method based on multivariable water quality parameters time series data. Compared with the spectral on-line water quality detection method based on baseline correction and principal component analysis, it can save 1235 yuan. To sum up, the detection cost of the proposed method is the lowest, followed by the time-series data detection method based on multivariable water quality parameters. In contrast, the spectral on-line water quality detection method based on baseline correction and principal component analysis has the highest cost, which is not conducive to practical application.

4. Concluding remarks
Given the shortcomings of the current multi-parameter detection methods of heavily polluted water quality in sponge cities, a multi-parameter detection method of water quality based on the random forest classification algorithm is proposed. This method is better than the current detection methods, has higher detection accuracy, and has lower detection time and cost so that it can be used in the field of urban water quality monitoring.

References
[1] Chen Yi, Liu Ying, Chen Xingru, Liu Ruijun. Simulation evaluation method of pesticide residue pollution based on visual analysis [J]. Computer Simulation, 2017 Rd 34 (10): 347-351.
[2] Gan Liping, Wu Yuan, Shi Rujie, Zhang Jing, Zhou Jiangcui. Detection of surface runoff indicator microorganisms and water quality of the Zhaoxi River in the Three Gorges Reservoir area [J]. Journal of Southwest Agriculture, 2017, 30 (03): 692-696.
[3] Li Gong, Guo Hui, Zhao Hui, Zhao Jianlong, Jia Chunping, Jin Qinghui. Development of portable water quality testing instruments based on Android [J]. Journal of Environmental Engineering, 2016 jade 10 (07): 3973-3976.
[4] Mao Yingchi, Qi Hai, Ji Qing, Wang Longbao. MnurTAEDA: abnormal event detection algorithm for time series data of multivariable water quality parameters [J]. Computer applications, 2017, 37 (01): 138-144.
[5] Guo Bingbing, Hou Dibo, Jin Yu, Yin Hang, Huang Pingjie, Zhang Guangxin, Zhang Hongjian. UV-visible spectrum on-line water quality anomaly detection method based on baseline correction and principal component analysis [J]. Spectroscopy and Spectral Analysis, 2017 focus 37 (05): 1460-1465.
[6] Xie Bin, Zhang Shuo, Li Li, Zhang Hu. Characteristics of phytoplankton community structure and its relationship with water quality parameters in marine pastures of Haizhou Bay [J]. Journal of Environmental Science, 2017, 37 (01): 121-129.
[7] Wang Lucia, Chen Chaohua, Wang Tao, Xiang Wei. Design and implementation of a chromatographic peak searching algorithm in water quality detection [J]. Computer Engineering and Application, 2017, 53 (11): 237-240,259.
[8] Wu Dezao, Wei Biao, Feng Peng, Tang Bin, Liu Juan. UV-vis spectral denoising algorithm for water quality detection based on two-dimensional reorganization and dynamic pane [J]. Spectroscopy and Spectral Analysis, 2016 Jing 36 (04): 1044-1050.
[9] Li Kui, Zhang Xuhui, Li Xiaoli, Wang Xin, Liu Jiyun, Duan Zhiquan. This paper analyzes the problems and countermeasures of all levels of disease control and water quality testing laboratories in Yunnan Province from the results of quality control assessment [J]. Modern Preventive Medicine, 2016 pr 43 (15): 2767-2769.