The Relationship of Different COD Concentration, Sediment Pollutant Content with Hydrodynamic in Black Bloom Water Based on Stepwise Regression Method

Yu Shan1,2,3, Lie Hong4, Bing-Jie Lu1,2,3, Jun Chen1,2,3, Jie Jin1,2,3, Ke Wu1,2,3*, Xinwu Gao1, Yonglian Song2 and Jun Liu5

1 School of Biology, Food and Environment, Hefei University, Hefei 230601, China
2 Anhui Key Laboratory of Sewage Purification and Eco-restoration Materials, Hefei Anhui. 230088, China
3 Anhui Provincial Collaborative Innovation Center for Environmental Pollution Prevention and Ecological Restoration, Hefei Anhui. 230061, China
4 Engineering and Technology Center for Solid Waste to Energy of Anhui Province, Hefei University, Hefei, 230601, China
5 Anhui Zhonghuan Environmental Protection Technology Co., Ltd., Hefei Anhui. 230022
E-mail: 610113863@qq.com; wuke@hfuu.edu.cn

Abstract. To analyze the relationship of different COD concentration, sediment pollutants with hydrodynamic. Taking a polluted river as the research object, the data of flow velocity, water temperature, COD concentration, sediment OM content, sediment TN content and sediment TP content at nine sampling points were analyzed by stepwise regression and SPSS 25 data analysis software. Construct linear regression model of COD concentration Y, river flow velocity X1, water temperature X2, sediment OM content X3, sediment TN content X4, sediment TP content X5. The regression equation Y=106.172-64.304X4 + 0.065X5 with correlation coefficient R=0.954 and determination coefficient R²=0.91 was obtained by stepwise regression method. The results showed that the concentration of COD in water was correlated with the content of nitrogen and phosphorus in river sediments, and negatively correlated with the content of TN and positively correlated with the content of TP. The model can be used to estimate COD load of water body by analyzing the nitrogen and phosphorus content in river sediments.

1. Introduction
The quality of rivers had an important impact on its ecological value. As one of the important indicators for evaluating water quality, COD had an important impact on the quality of water quality [1-3]. River polluted was the result of its own meteorological and hydrodynamic, geographical environment, river sediments, etc, where the flow rate had a certain impact on the nutrient release of the river sediment [4]. In the process of simulating the change of water quality, the migration and transformation law of nitrogen and phosphorus in water was an important parameter [5-6]. There was a migration and transformation law of nitrogen and phosphorus at the interface between sediment and...
water [7-9], and there was also a correlation between the change of COD and the content of nitrogen and phosphorus in water [10-11]. Therefore, there might be a correlation between the content of pollutants in river sediments and hydrodynamic.

The main reason for eutrophication of river was that the high level of nutrient content of nitrogen and phosphorus, which increase the amount of algae and other plankton and organic matter [12], and destroy the ecological balance of water to make water quality deterioration. COD can be used to indicate the amount of organic matter in water, and may be has a correlation of the content of nitrogen and phosphorus in water. However, the nitrogen and phosphorus substances in the water were closely related to the composition of sediments. Studies shown that there is a correlation between different concentrations of Fe$^{2+}$, S$^{2-}$ with hydrodynamic, DO in black bloom water [13]. The source of COD concentration in water was related to human activities. Population density, livestock and poultry breeding and agricultural planting were all the influencing factors of COD emission and river load [14-18]. The analysis of the release law of nutrients in sediments and the correlation between nutrients and water pollutants can predict the trend of water pollution and provide technical support for the treatment of water pollution [19-20]. Therefore, river flow rate, water temperature, sediment OM, TN in sediment, TP content in sediment and COD concentration were selected. Multiple linear regression analysis and correlation were judged to find out the relevant factors affecting the concentration of COD, and it had a certain effect on estimating and predicting the trend of water COD concentration.

2. Materials and Methods

2.1. Sample Collection
From the upstream to the downstream of the river, nine sample sections were set, one sample point per section. Use the transparent acrylic water sample collector to collect the water sample, put it into the incubator and bring it back to the laboratory. Use the sediment sampler to collect the sediment sample (about 15cm deep), put it into the plastic bag and bring it back to the laboratory, then dry it at room temperature to remove the residue of animals, plants and small stones.

2.2. Experimental Methods
The water sample stored at low temperature was measured by rapid digestion spectrophotometry. The OM content of the sediment was determined by the burning reduction method, the TN content was determined by Kjeldahl method, and the alkali fusion-molybdenum antimony spectrophotometric method was used to determine the TP, the data obtained would be multivariate stepwise regression analysis using SPSS 25 data analysis software.

3. Results and Discussion

3.1. Experimental Data
The measured flow velocity (m·s$^{-1}$) and temperature (°C) of the sample section were set to $X_1$ and $X_2$, respectively. The measured data of the organic matter content (g·kg$^{-1}$), TN (g·kg$^{-1}$), and TP (mg·kg$^{-1}$) were set to $X_3$, $X_4$, and $X_5$, respectively. The water sample COD concentration was set to $Y$, as shown in table 1 below.

3.2. Model Establishment
Select river flow rate (m·s$^{-1}$) $X_1$, temperature (°C) $X_2$, sediment organic matter content (g·kg$^{-1}$) $X_3$, TN (g·kg$^{-1}$) $X_4$, TP (mg·kg$^{-1}$)$X_5$ and COD concentration $Y$, imported into SPSS 25 software for multivariate stepwise linear regression analysis, set the regression equation to

$$Y=a_0+a_1X_1+a_2X_2+a_3X_3+a_4X_4+a_5X_5$$

It can be seen from table 2 that stepwise linear regression enters and eliminates variables at each step. This stepwise regression analysis produces two models in two steps, the first step of river flow
rate (m·s⁻¹) $X_1$, temperature (°C) $X_2$, sediment OM content (g·kg⁻¹) $X_3$ were rejected, sediment TN (g·kg⁻¹) $X_4$ into the model, second step river flow rate (m·s⁻¹) $X_1$, temperature (°C) $X_2$, sediment OM content (g·kg⁻¹) $X_3$ were removed, and sediment TP (mg·kg⁻¹) $X_5$ entered the model and the analysis was completed. In summary, the final equation was

$$Y = a_0 + a_4 X_4 + a_5 X_5$$

### Table 1. Data of each monitoring section.

| Sample point | $X_1$ (Flow rate) (m·s⁻¹) | $X_2$ (Temperature) (°C) | $X_3$ (Organic matter content) (g·kg⁻¹) | $X_4$ (TN content) (g·kg⁻¹) | $X_5$ (TP content) (mg·kg⁻¹) | $Y$ (COD content) (mg·L⁻¹) |
|--------------|--------------------------|--------------------------|--------------------------------------|----------------------------|-----------------------------|---------------------------|
| 1            | 0.0150                   | 6.1                      | 31.8                                 | 0.91                       | 100.81                      | 56                        |
| 2            | 0.0262                   | 6                        | 48.3                                 | 1.41                       | 258.89                      | 31                        |
| 3            | 0.0165                   | 6.2                      | 44.3                                 | 1.53                       | 190.38                      | 16                        |
| 4            | 0.0157                   | 6.3                      | 42.2                                 | 1.51                       | 168.35                      | 21                        |
| 5            | 0.0159                   | 6.6                      | 35.6                                 | 1.46                       | 284.99                      | 26                        |
| 6            | 0.0171                   | 6.5                      | 41.2                                 | 1.57                       | 331.72                      | 26                        |
| 7            | 0.0157                   | 6.4                      | 51.1                                 | 1.64                       | 375.22                      | 31                        |
| 8            | 0.0154                   | 6.3                      | 46.2                                 | 1.56                       | 250.41                      | 21                        |
| 9            | 0.0152                   | 6.5                      | 51.8                                 | 1.54                       | 152.12                      | 21                        |

### Table 2. Excluded variables.

| Model | Enter Beta | t | Significant | Partial correlation | Collinear statistics | Tolerance | VIF | Minimum tolerance |
|-------|------------|---|-------------|---------------------|----------------------|-----------|-----|------------------|
|       |            |   |             |                     |                      |           |     |                  |
| 1     | X₁         | 0.027*   | 0.137       | 0.895               | 0.056                | 1         | 1   | 1                |
|       | X₂         | 0.067*   | 0.289       | 0.782               | 0.117                | 0.742     | 1.347 | 0.742            |
|       | X₃         | 0.136*   | 0.491       | 0.641               | 0.197                | 0.5       | 1.198 | 0.5              |
| 2     | X₅         | 0.496*   | 3.151       | 0.02                | 0.789                | 0.604     | 1.655 | 0.604            |
|       | X₁         | -0.073*  | -0.543      | 0.611               | -0.236               | 0.938     | 1.066 | 0.567            |
|       | X₂         | 0.016*   | 0.103       | 0.922               | 0.046                | 0.733     | 1.365 | 0.525            |
|       | X₃         | 0.274*   | 1.81        | 0.13                | 0.629                | 0.475     | 2.107 | 0.32             |

Note: * Dependent variable: COD concentration; * Dependent in model: (constant), sediment TN content; * Dependent in c model: (constant), sediment TN content, sediment TP content.

### 3.3 Model Analysis

Table 3 shown the results of analysis of variance. It can be seen that the sum of squared residuals from model 1 to model 2 decreased from 262.31 to 98.823. The sum of squared residuals was an important indicator of equation fit. The smaller the sum of squares of residual errors, the better the correlation between independent variables and dependent variables. The significance test statistic F of the multivariate stepwise regression equation increased from 22.355 to 30.393, which passed the significance test, indicating that the regression coefficient of the model was not significantly zero, and the significance $P < 0.05$, indicating that the regression coefficient of at least one independent variable was less than 0.05. The established statistical model was meaningful.

In table 4, the R, $R^2$ and adjusted $R^2$ indicators were close to 1, the better the equation fitting effect. It can be seen from table 4 that the R, $R^2$ and adjusted $R^2$ indicators risen from the 0.873, 0.762 and 0.727 of model 1 to model 2 0.954, 0.91 and 0.88, indicating that after two-step regression analysis, model 2 had a better fit than model 1 equation. Model 2 Durbin-Watson $wd = 2.168$, indicating that
the model variables were not sequence related.

| Model | Sum of square | Degree of freedom | Mean square | F       | Significant |
|-------|---------------|-------------------|-------------|---------|-------------|
| Return | 837.69        | 1                 | 837.69      | 22.355  | 0.002b      |
| 1 Residual | 262.31 | 7                 | 37.473      |         |             |
| Total  | 1100          | 8                 |             |         |             |
| Return | 1001.177      | 2                 | 500.589     | 30.393  | 0.001c      |
| 2 Residual | 98.823  | 6                 | 16.47       |         |             |
| Total  | 1100          | 8                 |             |         |             |

Note: a Dependent variable: COD concentration; b Predictor: (constant), sediment TN content; c Predictor: (constant), sediment TN content, sediment TP content.

| Model | R | R² | Adjusted R² | Standard estimated error | Durbin-Watson |
|-------|---|----|-------------|--------------------------|---------------|
| 1     | 0.873a | 0.762 | 0.727 | 6.122 | 2.168 |
| 2     | 0.954b | 0.91 | 0.88 | 4.058 | 1.655 |

Note: a Predictor: (constant), sediment TN content; b Predictor: (constant), sediment TN content, sediment TP content; c Dependent variable: COD concentration.

Table 5 given the partial regression coefficients and their respective standard deviations for each item in the regression model and the t-test results for whether each parameter was equal to zero. The variables of models 1 and 2 were significantly less than 0.05, indicating that the predictor was significantly correlated with the dependent variable. The model 1 variable tolerance was 1, VIF=1, the model 2 variable tolerance was 0.604, VIF=1.655. It can be seen from tables 5–6 that there was no collinear problem between the equation independent variables.

| Model | Unnormalized coefficient | Standardization coefficient | t  | Significant |
|-------|--------------------------|-----------------------------|----|-------------|
|       | B | Standard error | Beta |                  | |
| 1 (constant) | 96.777 | 14.759 | 6.557 | 0 |
| X₄ | -47.372 | 10.019 | -0.873 | -4.728 | 0.002 |
| 2 (constant) | 106.172 | 10.229 | 10.38 | 0 |
| X₄ | -64.304 | 8.544 | -1.185 | -7.526 | 0 |
| X₅ | 0.065 | 0.021 | 0.496 | 3.151 | 0.02 |

Note: a Dependent variable: COD concentration.

| Model | Dimension | Eigenvalues | Conditional indicator | Variance ratio (constant) | X₄ | X₅ |
|-------|-----------|-------------|----------------------|--------------------------|----|----|
| 1     | 1         | 1.99        | 1                    | 0                         | 0  | 0  |
| 2     | 2         | 0.01        | 14.396               | 1                         | 1  | 1  |
|       | 1         | 2.929       | 1                    | 0                         | 0  | 0  |
|       | 0.014     | 6.776       | 0.08                 | 0.01                      | 0.69 |
|       | 0.007     | 20.414      | 0.92                 | 0.99                      | 0.3 |

Note: a Dependent variable: COD.
In summary, the final fitting equation was 

\[ y = 106.172 - 64.304 x_4 + 0.065 x_5 \]

Figure 1 was a histogram of the normalized residuals, showing the normality of the data. Figure 2 was the normal p-p plot of the regression standardized residuals. The scatters of each observation were basically linear, indicating that the equation was meaningful.

4. Conclusion

Among the five variables of flow rate, water temperature, OM content of sediment, TN content of sediment and TP content of sediment, the variable with significant correlation with COD concentration was the TN and TP content in the sediment. Through multivariate stepwise regression analysis, it can be found that the concentration of COD in water was negatively correlated with TN content in river sediments, and the TP content was positively correlated. Therefore, it was possible to estimate and predict the water COD concentration by studying the trend of nitrogen and phosphorus content in the sediment.

Acknowledgments

This research was supported by National Special Item on Water Resource and Environment (2017ZX07603-003), Anhui Provincial Central Committee Guides Local Science and Technology Development Project (YDZX20183400004146), Opening Project of Anhui Key Laboratory (MJ-ZK201701). Financial Support Project for Top-notch Talents in the Universities of Anhui (gxgnfx2018030). Hefei Independent Innovation Policy ‘Borrowing and Subsidizing’ Fund Project (J2018G22). Talent Research Foundation of Hefei University(16-17RC04). Innovation Team for Manucipal Solid Waste Treatment of Anhui.

Reference

[1] Kawabe M and Kawabe M 1997 J of Oceanography. 53 443-453
[2] Zhang L, Shi Y, Zhang J L, Liu Z C, Qin Y W and Chi M H 2017 Res of Env. Sci. 30 041-1049
[3] Liu L N, Ma C Z, Zhang J T, He Z S, Huo S L and Xi B D 2019 Res of Env. Sci. 32 1108-1116
[4] Zhong X Y, Wang C H, Yu C R, Wen L and Duan P Y 2017 Acta Scientiae Circumstantiae. 8 2862-2869
[5] Hu, Li C H, Ye C, Wang J, Wei W W and Deng Y 2019 Res of Env. Sci. 1-21
[6] Liu D H, Yu J H, Zhong J C, Zhong W H and Fan C X 2016 China Env. Sci. 36 125-132
[7] Chen Y C and Tang L 2005 J of Yunnan Agricultural University 20 527-533
[8] Cheng X J, Guo Z R, Liu G and Li B 2014 *J of Tropical Oceanography* 33 77-84
[9] Li D P and Huang Y 2010 *Env. Chem.* 29 1075-1078
[10] Yang Wan-Ling, Lai Zi-N, Zeng Y Y, Shuai F M, Li H Y and Wang C 2017 *Ecol. Env. Sci.* 26 643-648
[11] Zhang X F, Lin Y S, Yu F, Xu Y G, Hu L and Li M 2006 *Jiangsu Env. Technol.* 5 1-3
[12] Chen Hong, Liu Qing, Pan J X, Wang S and Wang Z Z 2019 *Acta Ecologica Sinica.* 39 173-184
[13] Wang Y L, Wang L and Hua Z L 2018 *China. Env. Sci.* 38 627-633
[14] Liao Y S, Zhou M N, Li D Q and Guo T L 2013 *Env. Sci.* 34 3019-3024
[15] Liu M X, Zhou J G, Huang X, Xie K J, Li J, Lv D Q and Yu-Yuan Li 2017 *Res of Agric Modernization.* 38 168-175
[16] Gong D L, Hong X, Zeng G J, Wang Y, Zuo S M, Liu X L and Wu J S 2017 *Journal of Ecol. Rural Env.* 33 509-518
[17] Yin C, Liu M, Sun F Y, Li C L and Xiang W N 2017 *Chinese J of Applied Ecol.* 27 911-919
[18] Ji F, Wang G X, Han R M, Li S Y, Ma J and Cao X 2017 *Chinese J of Env Engineering.* 10 3462-3470
[19] Xie A L, Xu F, Xiang L, Xu B, Lin L C and Wang C L 2017 *J of Hohai Univ (Natural Sciences)* 45 391-397
[20] Wang Qian, Wu Y D, Ding Q L, Zhang Q H and Wang X C 2017 *China Env. Sci.* 37 2699-2707