

## Analysis of water quality using multivariate statistical methods in Duliujian River, China

Xuewei Sun<sup>1, a</sup>, Xiaoqian Liang<sup>1, b</sup>, Tousheng Huang<sup>1, c</sup>,  
Huayong Zhang<sup>1, d, \*</sup>, Hai Huang<sup>1, e</sup>

<sup>1</sup> Research Center for Engineering Ecology and Nonlinear Science, North China Electric Power University, Beijing 102206, P.R. China

<sup>a</sup> Xuewei\_Sun@ncepu.edu.cn, <sup>b</sup> liangxiaoqian1949@126.com, <sup>c</sup> tous\_huang@ncepu.edu.cn,

<sup>d</sup> rceens@ncepu.edu.cn (corresponding author), <sup>e</sup> huanghai@ncepu.edu.cn

**Keywords:** water quality; principle component analysis; cluster analysis.

**Abstract.** Water quality assessments are essential for providing information to water resource management processes. The objective of this study was to investigate the internal correlation among water parameters, to analyze main water contamination and to identify the most polluted section along a seagoing river. Five water quality parameters (WT, EC,  $\text{NH}_4^+\text{-N}$ ,  $\text{NO}_3^-\text{-N}$ ,  $\text{NO}_2^-\text{-N}$ ) were chosen to assess the water quality using multivariate statistical methods (Spearman's correlation, PCA, CA). PCA results showed 2 PCs can explain 78% of the total variation. The main water contamination was related to anthropogenic activities (domestic wastewater, industrial effluents and livestock operations) as well as natural condition (seawater intrusion). CA results gave 3 clusters by analyzing similarities of each section, indicating that the middle-downstream was the most polluted and the last section was mainly influenced by seawater intrusion.

### Introduction

Water is the most significant natural resource for human society. Clean and safe water plays an important role not only on human healthy and factory producing, but also on environmental system. Thus, it is necessary to obtain information from river to analyze which section along the river is the most polluted and what is the most important water contamination. Many studies suggest that anthropogenic activities such as industrial effluents, agricultural waste water and domestic sewage systems are generally the significant factors in water degeneration [1, 2, 3]. Furthermore, natural condition, especially the sea water intrusion in coastal region, is not negligible [4]. Since multivariate statistical techniques have accurate calculation process and analyzable results, it is often used in identifying water contaminations and categorizing spatial similarity group [5, 6].

In this study, water quality of a seagoing river was evaluated by using multivariate statistical techniques including Spearman's rank correlation, Principal components analysis (PCA) and Cluster analysis (CA). The objective of this study was to investigate the internal correlation among water parameters, to analyze main water contamination, and to identify the most polluted section along this seagoing river.

### Material and methods

**Study area.** The Duliujian River, a tributary of the Haihe River in Tianjin City, flows from the conjunction of the Daqing River and the Ziya River to the Bohai Bay. The Duliujian River region is

located in warm temperate zone with semi-humid continental monsoon climate. The rainfall is concentrated from June to September; with an annual mean rainfall 571mm. Monthly mean temperature ranges from -3.4°C in January to 26.8°C in July. The investigated river goes through the urban population living quarters, so the up-middle stream flow is mainly polluted by domestic wastewater and intensive agricultural activities. The river crosses the city industrial park, and receives effluents produced by various industries and may be polluted by point and non-point sources.

**Sampling strategy and laboratory analysis.** In order to monitor spatial changes in water quality, 15 sampling sections were set in the Duliujian River and each sampling section has three sampling sites. The samples were collected 0.5m below the water surface using a polymethyl methacrylate sampler on December 4th to 14th, 2016 and were stored in a 1.5 L polyethylene plastic bottle pre-cleaned with detergent. The monitored water quality parameters, analytical methods and its units were summarized in Table 1. The water temperature (WT) and electrical conductivity (EC) of each water sample were measured using a YSI ProPlus in situ. Ammonia nitrogen, nitrate and nitrite were analyzed in laboratory under conducting by standard methods [7].

Table 1 The water quality parameters, abbreviation and their analytical method in this study.

Parameters	Abbreviation	Analytical method	Unit
Water temperature	WT	YSI ProPlus	°C
Electrical	EC	YSI ProPlus	μs/cm
Ammonia nitrogen	NH <sub>4</sub> <sup>+</sup> -N	Nessler's reagent spectrophotometric	mg/l
Nitrate	NO <sub>3</sub> <sup>-</sup> -N	Gas phase molecular Absorption	mg/l
Nitrite	NO <sub>2</sub> <sup>-</sup> -N	Gas phase molecular Absorption	mg/l

**Data processing and statistical analyses.** In order to evaluate water quality at different sampling sites, multivariate analyses of the water quality dataset were performed through non-parametric Spearman's rank correlation, Principal components analysis (PCA) and Cluster analysis (CA). All data calculations and statistical analyses were performed in IBM SPSS 19.

## Results and discussion

**Water quality evaluation.** Spearman's rank correlation coefficients (r) showed the possible relationships between physicochemical parameters from 45 sampling sites. Table 2 presented the values of Spearman's correlation coefficient, where the  $r > 0.50$ ,  $p < 0.05$  indicated a significant correlations [6]. The recorded WT had a significant positive correlation with ammonia and nitrite, the coefficient were 0.32, 0.50. It was significant negative correlated with EC ( $r = -0.31$ ). The recorded EC had a negative correlation with WT which probably indicated ions in water became compounds by high temperature. The negative relation is similar with other reports [1,9].

The investigated NH<sub>4</sub><sup>+</sup>-N was positively related with WT ( $r = 0.32$ ), while negatively related with EC ( $r = -0.59$ ). There were not many agricultural activities going on in the area during invested period (December, 2016). Thus, variation of NH<sub>4</sub><sup>+</sup>-N may be caused by urban domestic waste and industrial effluents. NO<sub>3</sub><sup>-</sup>-N concentration had a negative correlation with EC (-0.38) and EC had a

negative correlation with WT (-0.31), which indicated the correlation between EC and  $\text{NO}_3^-$ -N established by WT. In addition, the measured  $\text{NO}_3^-$ -N concentrations showed a significant positive correlation with WT ( $r=0.50$ ), which was possibly a result of potential biological assimilation and denitrification processes.  $\text{NO}_2^-$ -N in the study showed no correlation with other parameters. It was likely due to the low temperature influencing nitrification processes. Thus, the  $\text{NO}_2^-$ -N concentration was too low to get a proportion in DIN.

Table 2 Spearman's rank correlation coefficients.

Parameter	WT	EC	$\text{NH}_4^+$ -N	$\text{NO}_3^-$ -N	$\text{NO}_2^-$ -N
WT	1.00				
EC	-0.31*	1.00			
$\text{NH}_4^+$ -N	0.32*	<b>-0.59**</b>	1.00		
$\text{NO}_3^-$ -N	<b>0.50**</b>	-0.38*	0.28	1.00	
$\text{NO}_2^-$ -N	0.07	-0.14	-0.07	-0.25	1.00

\*\* Correlation is significant at  $p<0.01$  level; \* Correlation is significant at  $p<0.05$  level.

**Source identification.** PCA was employed to identify the main indexes that spatially affect water quality. Before performing the PCA, suitability of the variables was tested using the Kaiser-Meyer-Olkin ( $>0.5$ ) and Barlett's sphericity tests ( $p<0.05$ ) [10]. In the study, the Barlett's sphericity test result was  $p=0.00$  with KMO 0.570, which meant these parameters were satisfactory for PCA. As the Fig. 1 showed, the sorted eigenvalues decreased with the increased PC number. Two principal factors were selected since their eigenvalues were greater than 1 (Kaiser Normalization), which explained 78% of the total variance in the water dataset.

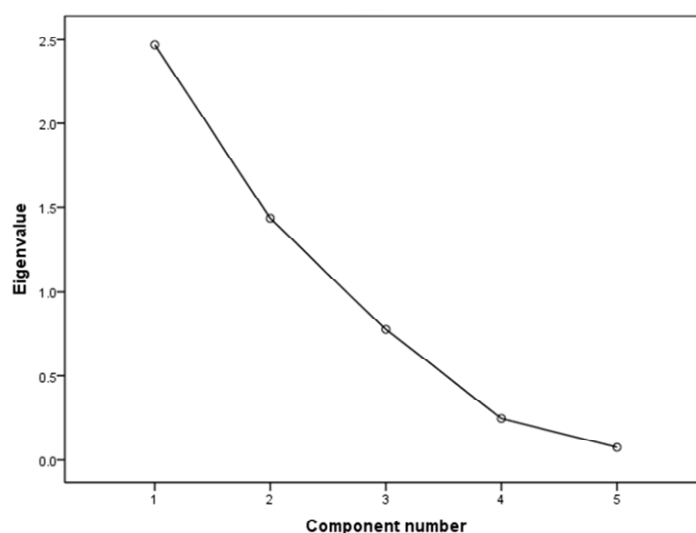


Fig. 1 Scree plot of the eigenvalues of water quality parameters

PCA result was presented in Table 3. The first component (PC1) explained 49.37% of the total variance and was positively and largely influenced by WT, EC and  $\text{NO}_3^-$ -N. It was weakly negative attributed by  $\text{NH}_4^+$ -N and  $\text{NO}_2^-$ -N. Since there were not many agricultural activities in winter,

$\text{NO}_3^-$ -N was not possibly derived from agricultural activities such as nitrogen fertilizers. Anthropogenic activities (industrial effluents) and biogenic (denitrification function) are mainly attributed to the factor. The second component (PC2), accounting 28.71% of the total variances, showed high positive loadings of  $\text{NH}_4^+$ -N and  $\text{NO}_2^-$ -N. This factor including nutrient variables could be attributed to anthropogenic activities such as livestock operations, domestic sources, as well as municipal point-source discharges.

Table 3 Principal component analysis (PCA) results in the Duliujian River

Parameter	PC1	PC2
WT	0.96	-0.01
EC	0.93	-0.26
$\text{NH}_4^+$ -N	-0.07	0.86
$\text{NO}_3^-$ -N	0.83	0.38
$\text{NO}_2^-$ -N	-0.01	0.70
Eigenvalue	2.47	1.44
Total variance [%]	49.37	28.71
Cumulate [%]	49.37	78.08

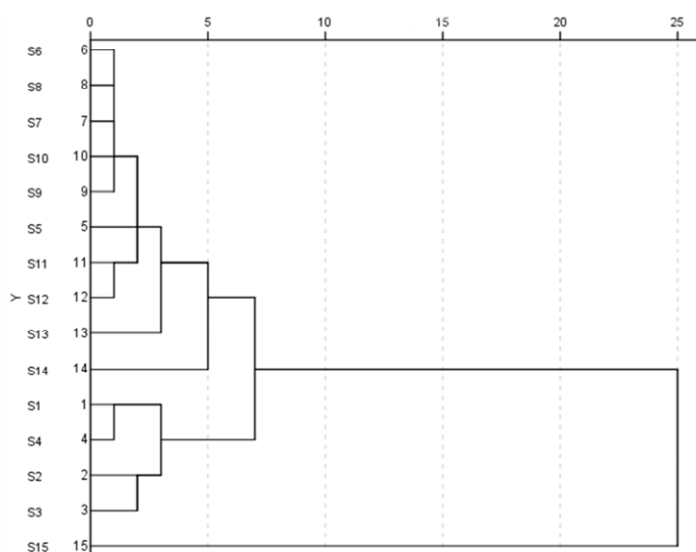


Fig. 2 Dendrogram based on hierarchical clustering for 15 sampling sections.

**Spatial variation in river water quality.** CA was used to analyze the spatial variation among 15 sampling sections, which was obtained by within-group linkage using Euclidean distance. As the Fig. 2 showed, the water quality in the investigated river was grouped into 3 clusters according to the internal correlation among sampling sites.

The cluster 1, including section 5 to section 14, was in high level polluted by water contamination. Section 5 to section 14 were located on the middle-down stream of this river, where had lots of chemical, pharmaceutical, petroleum chemical, iron-steel and electroplating factories.

These industrial effluents discharged into this river and probably became significant point contamination sources. In addition, livestock activities were attributed to non-point source for this section pollution. The cluster 2 was formed by sections 1, 2, 3 and 4, which was perceived less polluted by water contamination compared with Cluster 1. These sections were located on upstream, where near the urban population living quarters receives domestic waste water. The cluster 3 only had section 15 located on the estuary of Duliujian River. This section was mainly influenced by natural condition, especially the sea water intrusion.

## Conclusion

In this study, 5 water quality parameters from 45 sampling sites were investigated and evaluated using multivariate statistical techniques. Spearman's rank correlation result shows  $\text{NH}_4^+\text{-N}$  was significant negative with EC and  $\text{NO}_3^-\text{-N}$  had a strong correlation with WT during investigated period. PCA was used to identify the main factor of water contamination in the Duliujian River. PCA results showed water contamination was mainly polluted by sea water intrusion, industrial effluents (PC1) and livestock operations, domestic wastewater (PC2). CA results showed spatial similarity group. Cluster 1 suggested middle-downstream was probably more polluted by industrial effluents. Cluster 2 suggested up stream was polluted domestic wastewater in a moderate level. Cluster 3 showed the last section was probably polluted by natural condition. In a summary, environmental managers should pay more attention to urban wastewater and industrial effluents, and carry out improvement in livestock operations. The information from the assessment of water quality suggests that multivariate statistical is a good method to analyze the spatial changes in water quality.

## Acknowledgments

This work was financially supported by the National Special Water Programs (No. 2015ZX07203-011).

## Reference

- [1] A. Barakat, M. E. Baghdadi and J. Rais: International Soil & Water Conservation Research, Vol. 4 (2016), p. 284-292.
- [2] J. D. Pérez-Gutiérrez, J. O. Paz and M. L. M. Tagert. Agricultural Water Management, Vol. 187 (2017), p. 131-139.
- [3] Y. Wang, P. Wang and Y. Bai: A Journal of Hydro-environment Research, Vol. 7 (2013), p. 30-40.
- [4] A. Kurunc, S. Ersahin and N. K. Sonmez: Science of the Total Environment, Vol. 554-555 (2016), p. 53.
- [5] C. W. Liu, K. H. Lin and Y. M. Kuo: Science of the Total Environment, Vol. 313 (2003), p. 77-89.
- [6] T. Y. Ling, N. Gerunsin and C. L. Soo: Journal of Chemistry, 2017, Vol. 2017 (2017), p. 1-16.

- [7] Ministry of Environmental Protection of the People's Republic of China: *Determination methods for examination of water and wastewater (4th edition)*. (China Environmental Science Press, China 2002 )
- [8] A. Acquavita, I. F. Aleffi and C. Benci: *Regional Studies in Marine Science*, Vol. 2 (2015), p. 132-144.
- [9] P. S. Harikumar, A. Aravind, S. Vasudevan: *Journal of Water Resource & Protection*, Vol. 9 (2017), p. 83-97.
- [10] M. Kim, Y. Kim and H. Kim: *Desalination & Water Treatment*, 57 (2016), p. 12508-12517.