Scientific
Research
Publishing

The Water Quality Evaluation in Balihe Lake Based on Principal Component Analysis

Lei Zhang, Meifang Zhong, Yixin Xu, Zhongyu Wang*, Hai Huang

Research Center for Engineering Ecology and Nonlinear Science, North China Electric Power University, Beijing, China
Email: *zhy_wang@ncepu.edu.cn

Abstract

The water pollution situation in Balihe Lake, the biggest tributary of Shaying River Basin in Anhui Province, China, has brought a huge pressure on the improvement of water quality in Huai River. On October 16th, 2017, 11 major pollution indexes were observed at 15 sampling points in Balihe Lake. Based on the data experimentally measured, the water quality in Balihe Lake was analyzed utilizing the Principal Component Analysis (PCA) of SPSS. The result suggested that the major components were oxygenated pollutants, water eutrophication pollutants and ammonia nitrogen, in which oxygenated pollutants played a dominant role. In addition, the upper part of Balihe Lake suffered serious situation and needed a focus on oxygenated pollutants.

Keywords

Balihe Lake, Water Quality, Principal Component Analysis, Environmental Protection

1. Introduction

With the increasingly prominent problem of water environmental pollution, the research on the water quality comprehensive evaluation method becomes particularly important (Sina, 2017; Noori, et al., 2019). In the process of water quality comprehensive evaluation, complex and numerous environmental factors can make the research work heavy and the relevant data analysis difficult, moreover, and even may not find the root cause of water quality deterioration (Mena-Rivera, et al., 2017; Yang, 2010).

At present, there are many commonly used water quality comprehensive evaluation methods, such as comprehensive index method, fuzzy evaluation method, neural network method, etc. (Deng & Li, 2010; Wong & Hu, 2013). Although these methods may also make a good evaluation of water quality status,

it is impossible to determine the main factors affecting water quality (Chang et al., 2011). Principal Component Analysis (PCA) can put forward relevant factors from many variables, determine the main factors affecting water quality, and then get a reasonable explanation (Friedman, Hastie, & Tibshirani, 2010; Olsen, et al., 2012; Zhong, et al., 2018; Sun, et al., 2019). E.g., in Sun et al.'s research, the temporal and spatial patterns of river water quality were analyzed to evaluate the pollution status in a natural river based on PCA method (Sun et al., 2019). Similar studies can also be found for lake ecosystems (Zhong et al., 2018).

In this study, a set of actual sampling data was observed in a freshwater lake and used to evaluate the water quality in the sampling area based on PCA method, in order to get the main factors affecting the water quality in the area, and provide guidance for water environmental governance and improvement.

2. Materials and Method

2.1. Study Area

Balihe Lake is located at the intersection are of Huaihe River and Shaying River in Fuyang City, Anhui Province, P. R. China. As an artificially excavated lake, Balihe lake is originally the largest tributary of Shaying River Basin in Anhui Province. With the geographical coordinates of E116°14'-116°19' and N32°33'-32°36', it belongs to the semi-humid monsoon climate zone in subtropical and warm temperate zones. The total drainage area of Balihe Lake is about 500 km², accounting for about one-eighth of the total area of Anhui Section of Shaying River Basin. Besides, as can be seen in Figure 1, three rivers including Disanhugou River, Liugou River and Wulihugou River flows into the lake.

Water pollution in the Balihe Lake Basin not only seriously affected the economic development of the basin and the stability of the ecosystem, but also affected the ecological environment and water quality of the Shaying River, bringing tremendous pressure to the improvement of water quality in the Huaihe River Basin. Generally speaking, there were two main sources of water pollution in the lake drainage area: first, non-point source pollution along the lake coastline along with rainfall runoff; second, pollutants from these rivers flowing into the lake. Therefore, the analysis of water quality at different locations in Balihe
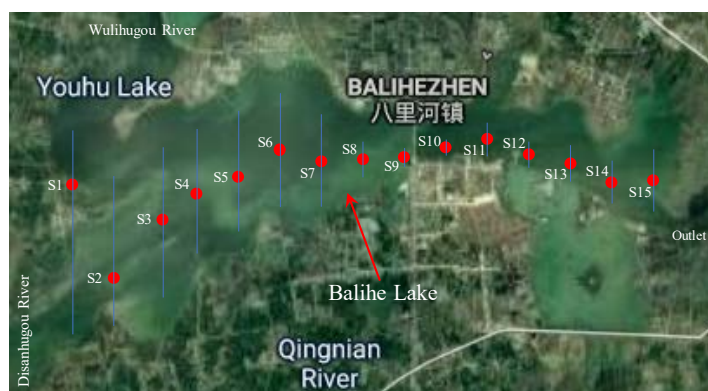

Figure 1. Balihe Lake and the distribution of sampling sites.

Lake is of great significance to the water pollution control work in Huaihe River Basin. In this study, PCA method was applied to the water quality comprehensive evaluation of Balihe Lake. The water pollution status of Balihe Lake was then analyzed comprehensively, and the main pollution factors were identified, which may provide some guidance to the water pollution control of Balihe Lake and Huaihe River Basin.

2.2. Sample Collection and Analysis

In October 2017, a field sampling survey was carried out at 15 sampling sites in Balihe Lake (see Figure 1). Surface water samples were collected because the water depth of the survey was within 10 m. According to the most concerned indicators of water environment monitoring in China, the water quality indicators including dissolved oxygen (DO), total nitrogen (TN), total dissolved nitrogen (TDN), ammonia nitrogen (NH_3-N), nitrate nitrogen (NO_3^--N), nitrous nitrogen (NO_2^--N), total phosphorus (TP), total dissolved phosphorus (TDP), phosphate (PO_4^{3-}), chemical oxygen demand (COD) and chlorophyll a (Chl-a) were measured according to the national standard method (Environmental Protection Administration of People's Republic of China, 2002, 2009).

2.3. PCA Method

Principal component analysis (PCA), also known as principal variable analysis, uses the idea of dimensionality reduction to transform multiple indicators into a few comprehensive indicators under the principle of minimizing the loss of data information (Debels, et al., 2005; Ouyang, 2005). In PCA, the comprehensive index of transformation analysis is usually called principal component. The principal component is a linear combination of the original variables and is not correlated with each other. Therefore, only a few principal components need to be considered to grasp the main contradictions and avoid the problem of collinearity between variables in complex problems, while the main information of the original data is not lost. And as such, the analysis efficiency could be improved significantly. Based on IBM SPSS Statistic 25.0 software, PCA was carried out on 11 water quality indicators of the 15 sampling sites mentioned above.

3. Results and Discussion

3.1. Standardization of the Experimental Data

The original data of these 11 indexes was standardized to eliminate the influence of magnitude and dimension among different data. The standardized data obtained obey the normal distribution with 0 as mean and 1 as standard deviation. Equation (1) is the calculation formula and the results were shown in Table 1.

$$Z_i = \frac{x_i - \frac{1}{m}\sum_{i=1}^{m} x_i}{\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_i - \frac{1}{m}\sum_{i=1}^{m} x_i\right)^2}}, \tag{1}$$

Table 1. Standardized data.

Site	DO	TN	TDN	NH$_3$-N	NO$_3^-$-N	NO$_2^-$-N	TP	TDP	PO$_4^{3-}$	COD	Chl-a
1	−1.238	−1.393	−0.019	−2.018	0.497	−0.952	−0.151	−0.712	−0.544	1.566	−1.196
2	−1.295	−1.286	0.064	−1.604	0.462	−1.164	0.526	−0.18	0.034	1.659	−1.163
3	−1.362	−1.128	0.072	−1.225	0.604	−1.336	0.935	0.375	0.558	1.694	−0.847
4	−1.396	−0.016	1.127	−0.433	1.153	−0.316	2.169	2.254	2.129	1.209	−0.411
5	0.309	0.376	1.048	−0.172	1.339	−1.667	1.8	1.19	1.31	0.226	0.674
6	−0.126	1.81	1.619	0.443	0.994	−0.343	0.636	1.001	0.983	0.051	0.944
7	0.556	1.995	1.604	0.549	1.02	0.386	−0.261	0.493	0.438	−0.038	0.898
8	0.612	1.065	1.234	−0.22	0.923	0.598	−0.584	−0.298	−0.184	0.178	2.741
9	0.331	0.622	−0.525	−0.113	−0.291	1.22	−0.599	−0.015	−0.086	−0.783	−0.473
10	0.432	0.407	−0.608	0.123	−0.22	1.485	−0.764	−0.428	−0.522	−0.88	0.134
11	1.38	−0.443	−0.923	−0.113	−0.486	1.459	−1.315	−1.633	−1.657	−0.892	−0.187
12	1.171	−0.366	−0.938	0.467	−0.805	0.743	−1.11	−1.114	−1.155	−0.847	−0.226
13	0.908	−0.483	−0.962	0.644	−1.053	0.399	−0.764	−0.771	−0.915	−0.765	0.237
14	0.893	−0.359	−0.972	0.75	−1.045	0.439	−0.662	−0.747	−0.773	−0.824	0.372
15	0.219	−0.537	−0.934	0.916	−1.248	−0.104	−0.182	−0.44	−0.544	−0.783	−0.631

where m is the number of sampling sites, x_i is the original index value, and Zi is the standardized value (Yang, 2010; Wu, 2019).

3.2. Maintaining the Integrity of the Specifications

The standardized data are analyzed by PCA method. Table 2 shows that KMO statistic is 0.624 (>0.500), and the significance level of Bartlett's test of sphericity is less than 0.001. It shows that independent variables are interrelated, and the data meet the basic requirements of PCA.

Spearman correlation analysis was used to analyze the correlations between these 11 indicators. And the results of the correlation coefficients were shown in Table 3. The greater the absolute value of the correlation coefficient between two indicators, the stronger the correlation between these two indicators. There is a positive correlation between two different indicators if the correlation coefficient is positive and vice versa. As can be seen in Table 3, there are some strong correlations between some indicators. E.g., 7 indicators have negative correlations with DO, which indicates that these indicators may be oxygen-consuming ones. Chl-a has a positive correlation with DO, which is consistent with the understanding that Chl-a is the main pigment for photosynthesis. Besides, there are strong positive correlations between TP, TDP and PO$_4^{3-}$, which indicates that the water quality information reflected by these indicators does overlap and is suitable for principal component analysis (Singh et al., 2004).

Table 2. KMO and Bartlett's test results.

KMO measure of sampling adequacy		0.624
Bartlett's test of sphericity	Approximate Chi Square	278.493
	Freedom	55
	Significance	0.000

Table 3. Correlation matrix.

Indicator	DO	TN	TDN	NH_3-N	NO_3^--N	NO_2^--N	TP	TDP	PO_4^{3-}	COD	Chl-a
DO	1.000	0.362	−0.232	0.398	−0.241	0.726	−0.683	−0.590	−0.635	−0.754	0.532
TN	0.362	1.000	0.627	0.408	0.374	0.325	0.003	0.349	0.310	−0.320	0.736
TDN	−0.232	0.627	1.000	−0.255	0.898	−0.334	0.574	0.644	0.679	0.504	0.520
NH_3-N	0.398	0.408	−0.255	1.000	−0.602	0.320	−0.237	0.072	−0.019	−0.813	0.253
NO_3^--N	−0.241	0.374	0.898	−0.602	1.000	−0.333	0.564	0.483	0.544	0.670	0.353
NO_2^--N	0.726	0.325	−0.334	0.320	−0.333	1.000	−0.769	−0.531	−0.599	−0.700	0.302
TP	−0.683	0.003	0.574	−0.237	0.564	−0.769	1.000	0.887	0.916	0.631	−0.137
TDP	−0.590	0.349	0.644	0.072	0.483	−0.531	0.887	1.000	0.993	0.383	0.042
PO_4^{3-}	−0.635	0.310	0.679	−0.019	0.544	−0.599	0.916	0.993	1.000	0.470	0.036
COD	−0.754	−0.320	0.504	−0.813	0.670	−0.700	0.631	0.383	0.470	1.000	−0.253
Chl-a	0.532	0.736	0.520	0.253	0.353	0.302	−0.137	0.042	0.036	−0.253	1.000

Generally speaking, there are three principles to determine the number of principal components, they are: 1) the eigenvalue of the principal component λ should be larger than 1; 2) the cumulative variance percentage of the principal components larger than 80% - 85%; 3) the number of principal components should be determined by the mutation of the eigenvalue. Among these principles, the eigenvalue represents the affecting degree of the principal component on the selected indicators, e.g. the explanation of the principal component is not enough if the eigenvalue is less than 1.

According to the explanatory table of total variance (**Table 4**), λ_1, λ_2 and λ_3 corresponding to the 1st, 2nd and 3rd principal components are 5.4466 2.999 and 1.657, respectively. The corresponding percentages of variance are 49.689%, 27.265% and 15.062%. And the cumulative contribution of these three components is up to 92.016%, which matches the first two extraction principles mentioned above. It can be considered that these three principal components included all the information of these 11 environmental indicators. According to the scree plot of PCA eigenvalue curve (**Figure 2**), λ_4 is less than 1 and the curve after this eigenvalue becomes gentler, which means that the explanatory power becomes smaller and the mutation occurs at the eigenvalue λ_4. Above all, the number of principal components is determined at 3.
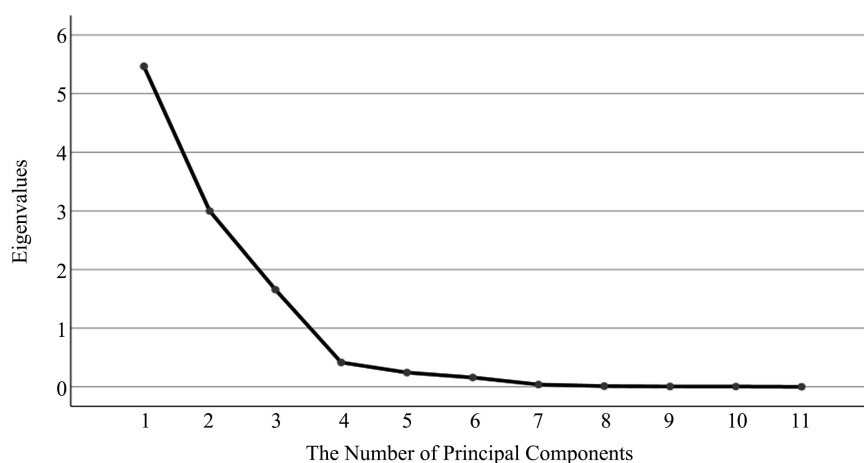
Figure 2. Scree plot of component variance eigenvalues.

Table 4. Total variance explained.

Component	Initial eigenvalue			Extracted square sum of loads		
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage
1	5.466	49.689	49.689	5.466	49.689	49.689
2	2.999	27.265	76.954	2.999	27.265	76.954
3	1.657	15.062	92.016	1.657	15.062	92.016
4	0.416	3.777	95.793			
5	0.243	2.205	97.998			
6	0.159	1.445	99.443			
7	0.039	0.354	99.797			
8	0.012	0.109	99.905			
9	0.005	0.047	99.952			
10	0.005	0.042	99.994			
11	0.001	0.006	100.000			

The factorial load matrix (Table 5) is directly calculated by SPSS, in which, the values are the correlation coefficients between the principal components and the original variables. The absolute value of these coefficients represents the degree of correlation for the relevant relationships, e.g., the greater the absolute value, the stronger the correlation and the closer the relationship. DO, TDN, NO_3^--N, NO_2^--N, TP, TDP, PO_4^{3-} and COD have high loads on the 1st principal component, which indicates that this principal component reflects the information of these eight indicators comprehensively and can be interpreted as the level of oxygen-consuming pollutants in Balihe Lake. Similarly, the 2nd principal component reflects the information of TN and Chl-a comprehensively and can explain the level of water eutrophication. The 3rd principal component mainly reflects the information of NH_3-N.

The factorial load matrix is not the principal component coefficient matrix. By dividing the factor load matrix by the square root of the corresponding principal component eigenvalue, the principal component coefficient matrix (Table 6) can be calculated. By multiplying the obtained component coefficient matrix with the normalized data, the evaluation functions F1, F2, F3 corresponding to each principal component and the comprehensive evaluation function F can be obtained. Based on these evaluation functions, the water quality pollution score of each sampling site can be quantitatively described. The higher the score, the more serious the pollution is. The expressions of each function are as follows:

Table 5. The factorial load matrix.

Indicator	Component		
	1	2	3
DO	**−0.762**	0.489	−0.212
TN	0.062	**0.959**	0.042
TDN	**0.737**	0.597	−0.267
NH$_3$-N	−0.453	0.435	**0.755**
NO$_3^-$-N	**0.743**	0.361	−0.538
NO$_2^-$-N	**−0.771**	0.358	−0.160
TP	**0.924**	−0.015	0.264
TDP	**0.825**	0.288	0.457
PO$_4^{3-}$	**0.879**	0.249	0.386
COD	**0.820**	−0.369	−0.400
Chl-a	−0.074	**0.869**	−0.259

Table 6. The principal component coefficient matrix.

Indicator	Component		
	1	2	3
DO	−0.181	0.171	−0.013
TN	0.032	0.302	−0.105
TDN	0.036	0.249	0.144
NH$_3$-N	0.172	0.013	−0.454
NO$_3^-$-N	−0.051	0.215	0.300
NO$_2^-$-N	−0.168	0.121	−0.027
TP	0.227	−0.024	−0.042
TDP	0.275	0.041	−0.175
PO$_4^{3-}$	0.261	0.041	−0.131
COD	−0.002	−0.038	0.307
Chl-a	−0.085	0.316	0.039

$$F1 = -0.181ZX1 + 0.032ZX2 + 0.036ZX3 + 0.172ZX4 - 0.051ZX5 - 0.168ZX6$$
$$+ 0.227ZX7 + 0.275ZX8 + 0.261ZX9 - 0.002ZX10 - 0.085ZX11 \tag{2}$$

$$F2 = 0.171ZX1 + 0.302ZX2 + 0.249ZX3 + 0.013ZX4 + 0.215ZX5 + 0.121ZX6$$
$$- 0.024ZX7 + 0.041ZX8 + 0.041ZX9 - 0.038ZX10 + 0.316ZX11 \tag{3}$$

$$F3 = -0.013ZX1 - 0.015ZX2 + 0.144ZX3 - 0.454ZX4 + 0.300ZX5 - 0.027ZX6$$
$$- 0.042ZX7 - 0.175ZX8 - 0.131ZX9 + 0.307ZX11 + 0.039ZX11 \tag{4}$$

$$F = \left[\lambda1/(\lambda1 + \lambda2 + \lambda3) \right] F1 + \left[\lambda2/(\lambda1 + \lambda2 + \lambda3) \right] F2$$
$$+ \left[\lambda3/(\lambda1 + \lambda2 + \lambda3) \right] F3 \tag{5}$$
$$= 0.540F1 + 0.296F2 + 0.164F3$$

3.3. Comprehensive Water Quality Evaluation

The scores and ranks of the principal components 1, 2, 3 and the comprehensive principal component were calculated and shown in Table 7. As can be found in the table, 1) according to the ranks for the 1st principal component, the top five sampling sites are 4, 5, 6, 3 and 2 indicating that the oxygen-consuming pollution is relatively serious at three places, and the site 4 is highest one; 2) for the principal component 2, the top five sites are 8, 7, 6, 5 and 10, which means that the eutrophication pollution may be more serious in comparison to the other places; 3) the top five sites for principal component 3 are 1, 2, 3, 8 and 4 where ammonia nitrogen pollution may be serious; 4) the top five places 4, 5, 6, 7, 3 for the comprehensive principal component are similar to that of the 1st principal component. Considering the items 1) and 4) together, it can be concluded that the upper part of the Balihe Lake should be polluted seriously by the oxygen-consuming pollution and should be treated adequately. In addition, the main treatment measures should be oriented to the oxygen-consuming pollutions like living sources, non-point sources, etc. The lower ranks of sites 11, 12, 13, 14 and 15 indicate that the water quality of the lower part of the lake is better and conservative measures may be taken to this area.

4. Conclusion

Based on the field measurements of 11 environmental factors at 15 sampling sites, the water quality in Balihe Lake was evaluated utilizing PCA method. The following conclusions can be drawn.

   1) There are obvious correlations between some of these 11 environmental factors. The 3 extracted principal components accounting for 92.016% of the total variance can well explain the water quality status in Balihe Lake. The 1st, 2nd and 3rd principal component represent the pollution of oxygen consuming pollutants, eutrophication and ammonia nitrogen, correspondingly.

   2) The sampling sites 4, 5, 6, 7 and 3, which have relatively higher PCA scores, are all located in the upper part of the lake. The water quality in these places should be more serious and the main pollutants are oxygen-consuming. Therefore, more attention should be paid to such areas in the future water quality prevention and treatment.

Table 7. Score and ranks of the principal components.

Site	F1 score	F1 rank	F2 score	F2 rank	F3 score	F3 rank	F score	F rank
1	−0.307	8	−1.159	15	1.761	1	−0.220	8
2	0.266	5	−1.107	14	1.412	2	0.048	7
3	0.725	4	−0.921	13	1.128	3	0.304	5
4	1.913	1	0.193	6	0.322	5	1.143	1
5	1.196	2	0.775	4	0.306	6	0.926	2
6	0.818	3	1.469	3	0.036	8	0.882	3
7	0.113	6	1.700	2	0.131	7	0.586	4
8	−0.713	11	1.855	1	0.777	4	0.291	6
9	−0.389	9	0.088	7	−0.378	11	−0.246	9
10	−0.741	12	0.234	5	−0.351	10	−0.389	10
11	−1.699	15	−0.186	8	−0.001	9	−0.973	15
12	−1.100	14	−0.326	10	−0.494	12	−0.771	14
13	−0.760	13	−0.346	11	−0.701	13	−0.628	13
14	−0.687	10	−0.257	9	−0.791	14	−0.577	12
15	−0.101	7	−0.831	12	−1.026	15	−0.469	11

3) The water quality at sites 11, 12, 13, 14 and 15 concentrated in the lower part of the lake as the PCA scores are lower. According to the better water quality in this area, conservative measures may be taken to this area.

Although the results of this study may provide some guidance or inspiration to the water pollution prevention and treatment of Balihe Lake, more research focusing on this topic based on some other methods are necessary in the future.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Chang, K., Gao, J. L., Wu, W. Y., & Yuan, Y. X. (2011). Water Quality Comprehensive Evaluation Method for Large Water Distribution Network Based on Clustering Analysis. *Journal of Hydroinformatics, 13,* 390. https://doi.org/10.2166/hydro.2011.021

Debels, P., Figueroa, R., Urrutia, R., Barra, R., & Niell, X. (2005). Evaluation of Water

Quality in the Chillán River (Central Chile) Using Physicochemical Parameters and a Modified Water Quality Index. *Environmental Monitoring and Assessment, 110,* 301-322. https://doi.org/10.1007/s10661-005-8064-1

Deng, D., & Li, Y. J. (2010). Application of Rough Set and Fuzzy Comprehensive Evaluation Method in Water Quality Assessment. In *International Conference on Computing, Control and Industrial Engineering* (pp. 126-128). Wuhan: IEEE. https://doi.org/10.1109/CCIE.2010.150

Environmental Protection Administration of People's Republic of China (2002). *Surface Water Environmental Quality Standards (GB 3838-2002).* (In Chinese)

Environmental Protection Administration of People's Republic of China (2009). *Monitoring and Analysis Methods of Water and Wastewater* (4th ed.). Beijing: China Environmental Science Press. (In Chinese)

Friedman, J., Hastie, T., & Tibshirani, R. (2010). *The Elements of Statistical Learning* (Vol. 1, No. 10). New York: Springer.

Mena-Rivera, L., Salgado-Silva, V., Benavides-Benavides, C., Coto-Campos, J. M., & Swinscoe, T. H. A. (2017). Spatial and Seasonal Surface Water Quality Assessment in a Tropical Urban Catchment: Burío River, Costa Rica. *Water, 9,* 558. https://doi.org/10.3390/w9080558

Noori, R., Berndtsson, R., Hosseinzadeh, M., Adamowski, J. F., & Abyaneh, M. R. (2019). A Critical Review on the Application of the National Sanitation Foundation Water Quality Index. *Environmental Pollution, 244,* 575-587. https://doi.org/10.1016/j.envpol.2018.10.076

Olsen, R. L., Chappell, R. W., & Loftis, J. C. (2012). Water Quality Sample Collection, Data Treatment and Results Presentation for Principal Components Analysis—Literature Review and Illinois River Watershed Case Study. *Water Research, 46,* 3110-3122. https://doi.org/10.1016/j.watres.2012.03.028

Ouyang, Y. (2005). Evaluation of River Water Quality Monitoring Stations by Principal Component Analysis. *Water Research, 39,* 2621-2635. https://doi.org/10.1016/j.watres.2005.04.024

Sina, Z. (2017). Modification of Expected Conflicts between Drinking Water Quality Index and Irrigation Water Quality Index in Water Quality Ranking of Shared Extraction Wells Using Multi Criteria Decision Making Techniques. *Ecological Indicators, 83,* 368-379. https://doi.org/10.1016/j.ecolind.2017.08.017

Singh, K. P., Malik, A., Mohan, D. et al. (2004). Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality of Gomti River (India)—A Case Study. *Water Research, 38,* 3980-3992. https://doi.org/10.1016/j.watres.2004.06.011

Sun, X. W., Zhang, H. Y., Zhong, M. F., Wang, Z. Y., Liang, X. Q., Huang, T. S., & Huang, H. (2019). Analyses on the Temporal and Spatial Characteristics of Water Quality in a Seagoing River Using Multivariate Statistical Techniques: A Case Study in the Duliujian River, China. *International Journal of Environmental Research and Public Health, 16,* 1020. https://doi.org/10.3390/ijerph16061020

Wong, H., & Hu, B. Q. (2013). Application of Interval Clustering Approach to Water Quality Evaluation. *Journal of Hydrology, 491,* 1-12. https://doi.org/10.1016/j.jhydrol.2013.03.009

Wu, D. F. (2019). Application of Set Pair Model Based on Principal Component Analysis in Water Quality Evaluation. *Water Conservancy Science and Technology and Economy, 25,* 1-7. (In Chinese)

Yang, Y. (2010). Management of Agricultural Pollution in China: Current Status and In-

ternational Experience. In *International Conference on Management and Service Science* (pp. 939-946). Wuhan: IEEE. https://doi.org/10.1109/ICMSS.2010.5576720

Zhong, M. F., Zhang, H. Y., Sun, X. W., Wang, Z. Y., Tian, W., & Huang, H. (2018). Analyzing the Significant Environmental Factors on the Spatial and Temporal Distribution of Water Quality Utilizing Multivariate Statistical Techniques: A Case Study in the Balihe Lake, China. *Environmental Science and Pollution Research, 25,* 29418-29432. https://doi.org/10.1007/s11356-018-2943-9