Assessment of water quality and source apportionment in a typical urban river in China using multivariate statistical methods

Jingshui Huang, Ruyi Xie, Hailong Yin and Qi Zhou

ABSTRACT

Water quality in urban rivers is a product of the interactions of human activities and natural processes. To explore water quality characteristics and to assess the impacts of natural and anthropogenic processes on urban river systems, we used multivariate statistical techniques to analyse water quality of a typical urban river in eastern China. Cluster analysis grouped the sites into four clusters which were affected by wastewater treatment plant effluent, untreated domestic sewage, tributaries and shipping, respectively. Cluster analysis provided scientific basis for optimizing the monitoring scheme. Three latent factors obtained from principal component analysis/factor analysis were interpreted as wastewater treatment plant effluent, untreated domestic sewage and surface runoff. Absolute principal component analysis indicated that most of the total dissolved phosphorus, nitrite, total dissolved nitrogen, and total nitrogen, Na, K and Cl resulted from the wastewater treatment plant effluent, most of the ammonia, dissolved organic carbon, sulfate and Mg resulted from the surface runoff. The pollution control measures for nitrogen and phosphorus were proposed based on the source apportionment results. The present study showed that the multivariate statistical methods are effective to identify the main pollution sources, quantify their relative contributions and provide useful water management suggesitions in urban rivers.

Key words | multivariate statistical techniques, source apportionment, urban river, water quality

Jingshui Huang
Ruyi Xie (corresponding author)
Hailong Yin
Qi Zhou
College of Environmental Science and Engineering,
Tongji University,
Shanghai 200092
China
E-mail: *cugxry@163.com*

INTRODUCTION

The surface water quality in a region is largely determined by natural processes and anthropogenic influences, such as urbanization, industrial and agricultural activities and the exploitation of water resources. Estimates have indicated that 60% of the world will reside in urban areas by 2030, which will put great pressure on urban rivers (United Nations 2016). The water quality in urban rivers has deteriorated for several reasons. First, incomplete sewage interception results in the discharge of high concentrations of pollutants directly into river systems. Second, various types of polluted runoff mixed with untreated domestic sewage are discharged into urban rivers during storm events from combined sewer overflow (CSO). Third, because the demands for water are increasing, more and more cities are intercepting water upstream in drinking water reservoirs, which results in higher pollution concentrations in the urban sections of rivers and smaller and lower flow velocity. Finally, the polluted sediments in the river bed serve as an internal source of pollution, which further degrades the river water quality.

Therefore, regular water quality monitoring is often carried out by relevant agencies to obtain better information regarding urban river water quality. However, the very large and complex data matrix that includes a large number of water quality parameters measured at various sites and during different seasons is often difficult to interpret, making it difficult to draw meaningful conclusions from the data (Singh *et al.* 2005). To understand the large

doi: 10.2166/ws.2018.002

amount of information hidden in the matrix, the data could be analysed using multivariate statistical techniques, such as cluster analysis (CA), principal component analysis/factor analysis (PCA/FA) and multiple linear regression on absolute principal component scores (APCS-MLR).

Multivariate statistical methods have been used to study large rivers with various types of land use in their basins (e.g. rural, agricultural, urban and industrial land use) to assess their water quality and understand their temporal and spatial variations (Shrestha *et al.* 2008; Baborowski *et al.* 2012; Chen *et al.* 2013). The studies indicate the usefulness of multivariate statistical techniques for analysing and interpreting complex datasets for river water quality assessments, the identification of pollution sources and understanding temporal/spatial variations, which can be used for effective river water quality management (Shrestha & Kazama 2007). However, most previous studies have focused on rivers with large basins and different land use types. Current knowledge on the effectiveness of these methods for identification and allocation of the pollution sources in urban watersheds are still limited.

Therefore, the focus of this study is to apply multivariate statistical methods to obtain detailed temporal and spatial water quality characteristics and to identify pollution sources and quantify their contributions to river water quality in an urban river, Nanfei River of Hefei City, China. The suggestions to restore urban river water quality in fast developing areas in China were further presented.

METHODS

Study area and sampling sites

With economic development, Hefei City has been expanding rapidly. Over the past ten years, the population of Hefei City increased by 70% from 2005 to 2015 (reaching 7.8 million), and the gross domestic product increased by 560% from 2005 to 2015 (reaching ¥566 billion). As shown in Figure 1, the urban land had fast expansion in the early 21st century, which is a typical characteristic of the urbanization progress in China. The speed of urban expansion has placed great pressure on the Nanfei River, which is used for drinking, industrial and agricultural water and receives large amounts of wastewater from the city. Therefore, the Nanfei River is considered as a typical urban river in China in this study.

The Nanfei River is approximately 70 km long and has a drainage area of approximately 1,446 km^2. The length of urban section of the Nanfei River is approximately 17 km from the Dongpu Reservoir to a rubber dam (Figure 2). The subcatchment area of this reach is 108.84 km^2 and more than 70% of the subcatchment area consists of urban land. The average flow in the studied reach is 7.75 $m^3 s^{-1}$. As shown in Figure 2, there are two tributaries in the urban section: Sili River and Banqiao River. Because of the existence of Dongpu Reservoir, which is used for drinking water supply, Nanfei River receives no significant inflow from upstream. The Wangtang Wastewater Treatment Plant

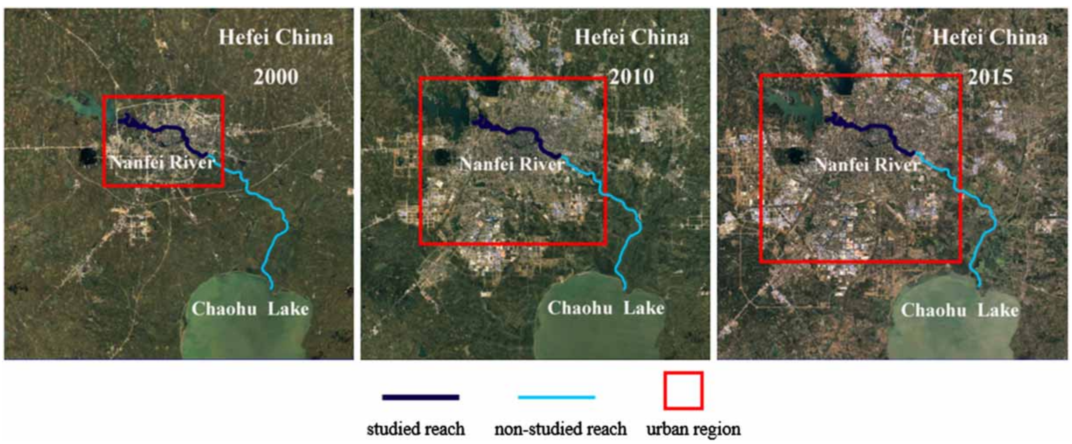

Figure 1 | The urbanization progress of Hefei City from 2000 to 2015. Picture redrawn from Google and NASA China.

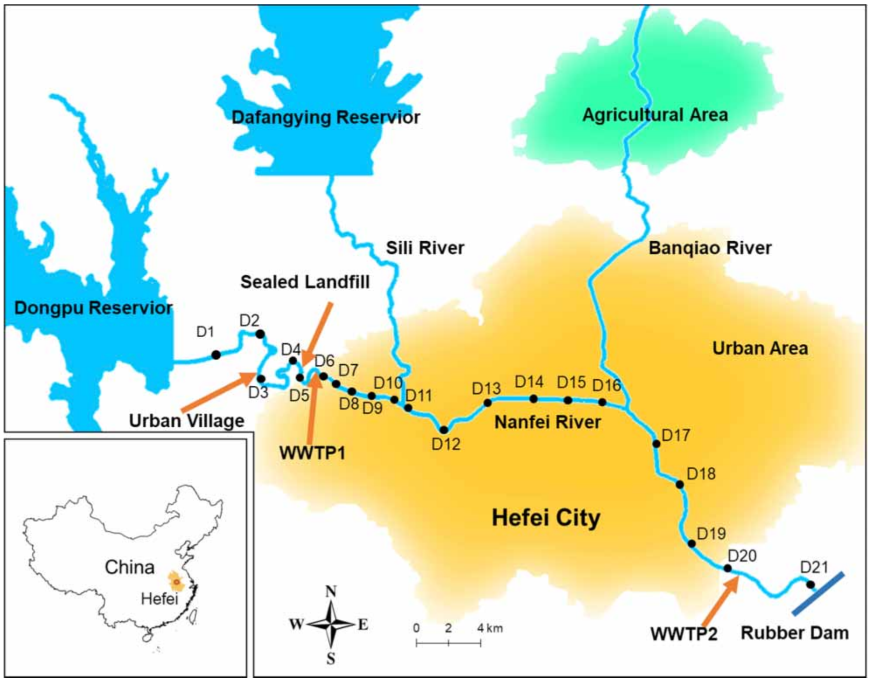

Figure 2 | Study area and monitoring sites. Black dots show sampling sites D1-D21 along the river. The yellow and green areas depict the urban and rural areas, respectively. Notably, reservoirs are located upstream of the Nanfei River and the Sili River. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/ws.2018.002.

(WWTP1) is located upstream of the urban reach and has an annual average discharge of approximately 2.21 $m^3 s^{-1}$. The Wangxiaoying Wastewater Treatment Plant (WWTP2) is located downstream of the urban section and nearly one kilometre upstream of the rubber dam and has an average discharge of 3.66 $m^3 s^{-1}$. The WWTP effluent accounts for 75% and 53% of the streamflow during the dry and wet seasons, respectively. An urban village is located approximately 2 km downstream from Dongpu Reservoir and directly discharges untreated domestic sewage into Nanfei River through drains. In addition, a landfill that has been sealed for many years is located 4 km downstream of Dongpu Reservoir. The 2 km farthest downstream of the studied reach is used for shipping.

During the dry season (from October to April), 17 monitoring sites (D1 to D16 and D21) were sampled in October 2015. During the wet season (from May to September), water samples were collected in July 2016 at 19 sampling sites along the river (D1-D6, D8 and D10-D21). The distance between the sampling sites was approximately 500 to 1,000 m. The intensive sampling network was designed to determine the spatial water quality variations of the river. Sampling, preservation, the transport of the water samples to the laboratory and sample storage were conducted in accordance with standard methods (Environmental Protection Agency of China 2002b).

Monitored parameters and analytical methods

Five field parameters were temperature (T), dissolved oxygen (DO), pH, turbidity and electrical conductivity (EC) were measured on-site using an HACH HQ40D Portable Meter and a HACH 2100Q Portable Turbid Meter. The 11 laboratory analysis parameters included the total dissolved phosphorus (TDP), nitrate (NO_3^-), ammonia (NH_4), total dissolved nitrogen (TDN), total nitrogen (TN), dissolved organic carbon (DOC), sodium (Na), potassium (K), magnesium (Mg), sulfate (SO_4^{2-}) and chloride (Cl) contents were determined by using the national standard methods (Environmental Protection Agency of China 2002b). A summary of the basic statistical results is presented in Table 1, and a complete dataset including 16 variables of 36 samples is presented in Supplementary Material (Table S1, available with the online version of this paper). Compared with environmental quality standards for surface water of China (Environmental Protection Agency of China 2002a), the

Table 1 | Statistical description of water quality parameters

Parameters	Unit	Mean	S.D.[a]	Minimum	Maximum
T	°C	26.74	2.86	23.2	31.7
pH	–	7.84	0.55	7.27	9.96
DO	$mg\ L^{-1}$	5.95	3.86	0.88	13.60
EC	$uS\ cm^{-1}$	485.3	112.95	226	636
Turbidity	NTU	14.88	6.78	3.92	29.30
TDP	$mg\ P\ L^{-1}$	0.70	0.34	0.07	1.44
NO_3^-	$mg\ N\ L^{-1}$	2.92	3.13	0.13	8.48
NH_4	$mg\ N\ L^{-1}$	5.31	4.10	0.13	11.15
TDN	$mg\ N\ L^{-1}$	8.85	3.36	0.45	13.85
TN	$mg\ N\ L^{-1}$	9.91	3.37	0.78	14.11
DOC	$mg\ L^{-1}$	5.24	1.45	3.33	11.72
Na	$mg\ L^{-1}$	25.56	7.09	7.65	32.53
K	$mg\ L^{-1}$	11.33	6.51	2.77	41.78
Mg	$mg\ L^{-1}$	8.65	1.41	6.56	10.92
SO_4^{2-}	$mg\ L^{-1}$	40.97	15.62	18.72	64.50
Cl	$mg\ L^{-1}$	47.80	20.72	10.96	115.85

[a]S.D means standard deviation.

water quality of Nanfei River was worse than the 5th class (worst) water quality, which stipulates standard values of 2 $mg\ L^{-1}$ for NH_4 and 0.4 $mg\ L^{-1}$ for TP.

Data treatment and multivariate statistical methods

Multivariate statistical methods play an important role in reducing the dimensionality of data and obtaining information that will be useful for water quality assessments and surface water management (Simeonov *et al.* 2003). The analysis tools used in this study included Excel 2016 and SPSS (Statistical Product and Service Solutions) 20.0. Spearman's rank correlation coefficient (R) was used to calculate the correlation between the variables. The correlations between the variables and seasons were also computed by using Spearman's rank correlation.

CA is a type of multivariate technique used to assemble objects based on the characteristics that they possess (Shrestha & Kazama 2007). In CA, cases (or variables) are classified based on their characteristics. Cases in the same cluster are similar and cases in different clusters are very different. In this study, CA was applied to all of the monitoring sites to recognize the spatial distribution patterns in different seasons. In large amount of data, high overlap and correlations between variables result in many obstacles when applying statistical methods. PCA/FA is a solution that drastically reduces the dimensionality of data without loss of much of the original information. PCA has the following four features: (a) usually, the number of factors is much smaller than the number of variables; (b) factors can preserve most of the information obtained from the variables; (c) the linear relationship between the factors are not significant; and (d) factors are prone to be named and explained. The new groups of variables, known as varifactors (VFs), are extracted through Varimax (variance maximizing) by rotating the axis defined by PCA, which is generally used to reallocate the variance and present better interpretations of the factors (Singh *et al.* 2005).

APCS-MLR was first applied to air pollution problems and has gradually become used in water pollution source apportionment (Simeonov *et al.* 2003; Nazeer *et al.* 2016). After determining the possible sources by using PCA, source contributions were estimated by using a multiple regression of the APCSs of the sample mass concentrations. PCA assumes that the total concentration of each constituent is made up of the sum of elemental concentrations from each identified anthropogenic or natural source component. A detailed description of this modelling method can be found elsewhere (Guo *et al.* 2004; Singh *et al.* 2005). A brief introduction to APCS will be explained in Supplementary Material (available online).

RESULTS AND DISCUSSION

Temporal variations of water quality

When computing Spearman's rank correlation, each season was regarded as a numerical value in the data file (Dry = 1 and Wet = 2) and was correlated with all other measured parameters. The computation results showed that water temperature had the highest correlation coefficient (Spearman's R = 0.860, $p < 0.001$) for the seasons (Table S2, available online). In addition, eight other parameters were significantly correlated with the seasons ($p < 0.001$): SO_4^{2-} (R = 0.859), Mg (R = 0.859), NH_4 (R = 0.749), DOC (R = 0.694), TDP (R = 0.653), EC (R = 0.634), TN (R = 0.616) and NO_3^- ($R = -0.524$). To some extent, these parameters can be considered as the major indicators of temporal variations in Nanfei River (Figure 3). The water temperature was clearly significantly different between different seasons. As a very common type of nitrogen fertilizer, ammonium sulfate is widely used to facilitate plant growth. Surface runoff, especially agricultural runoff, carries large amounts of ammonia and sulfate from the fertilizer applied to fields during the wet season. The soil in the Nanfei River Basin is acidic yellow brown soil, which has an intense leaching capacity and easily loses Mg; hence, large amounts of Mg are washed away with agricultural runoff in the wet season (Kalff 2011). The high correlation relationship between SO_4^{2-} and Mg proves that they are likely generated from the same pollution source. The higher concentrations of SO_4^{2-} and Mg during the wet season than dry season could result from soil erosion, which results in the transport of these chemicals and soil in runoff (Penha et al. 2016). Because intensive rainfall events occur during the wet season, large amounts of wet-weather overflow that contains high concentrations of NH_4, DOC and TDP are directly discharged into the river with sewer overflow from pumping stations without treatment. The significant positive correlation of TN (rather than TDN) with the season illustrated that several nitrogen particles were discharged into the river by both urban and agricultural runoff. For EC, more ions were carried into the river with the runoff. On the other hand, EC has a positive relationship with temperature (Kalff 2011). These factors lead to increasement in EC during the wet season. However, regarding the variables that were negatively correlated with the season, the dilution effect was dominant for NO_3^- because the NO_3^- concentration in the runoff was relatively low during the wet season.

Spatial variation of water quality

To study spatial variations and optimize the monitoring scheme, CA was used to detect groups with similarities among sampling sites. Figure 4 demonstrates the dendrograms of the cluster results from the dry and wet seasons, respectively.

During both the dry and wet seasons, the urban river can be clustered into four groups along its length. The background water quality of the urban river can be determined from the Dongpu Reservoir to the urban village. From the urban village to WWTP1, the river is mainly polluted by untreated domestic sewage from the urban village. Notably, the sealed landfill seriously impacts the water quality during the wet season and consequently should be monitored and carefully examined during wet periods. The water quality of the section from WWTP1 downstream to D18 is mainly influenced by the WWTP effluent. The section from D19 to the rubber dam is affected by the vertical diffusion between the water column and the sediment due to navigation. The spatial variations of the water quality in the Nanfei River depend on the location and the type of the pollution sources and the navigation situation, etc. Therefore, it is suggested that only four representative sampling sites in the dry season and five representative sampling sites in the wet season are needed to effectively obtain the water quality characteristics of the entire Nanfei River.

Data structure determination and source identification

PCA/FA was applied to the normalized dataset to compare the compositional patterns among the analysed water samples and to identify the factors that influenced each one. All the sampling sites were regarded as cases in the dry and wet seasons, and only 11 pollutant variables were considered (excluding T, DO, turbidity, EC and pH). The Kaiser–Meyer–Olkin test value was $0.702 > 0.5$, indicating that the dataset was suitable for PCA/FA. Three factors with eigenvalues >1 were sorted, which explained 88% of the variation in the dataset. The factor loadings were classified as 'strong,' 'moderate' and 'weak,' corresponding to the absolute loading values of 0.75, 0.75–0.50 and 0.50–0.30, respectively (Liu et al. 2003). The loadings of the variables on the significant varifactors (VFs) are shown in Table 2. VF1 explained 51% of the total variation in the dataset and showed strong loadings of Na, Cl, TDN and K and moderate loadings of TN and NO_3^-. Here, VF1 refers to the point source of the WWTP effluent based on the following evidence. First, the concentrations of Na, Cl and K dramatically increased where the WWTP1 effluent was discharged into the river. Second, the concentrations of TN and TDN were significantly different between the reaches upstream and downstream of the WWTP1. Lastly, high NO_3^- concentrations and low NH_4 concentrations distinguish the WWTP effluent from the untreated domestic

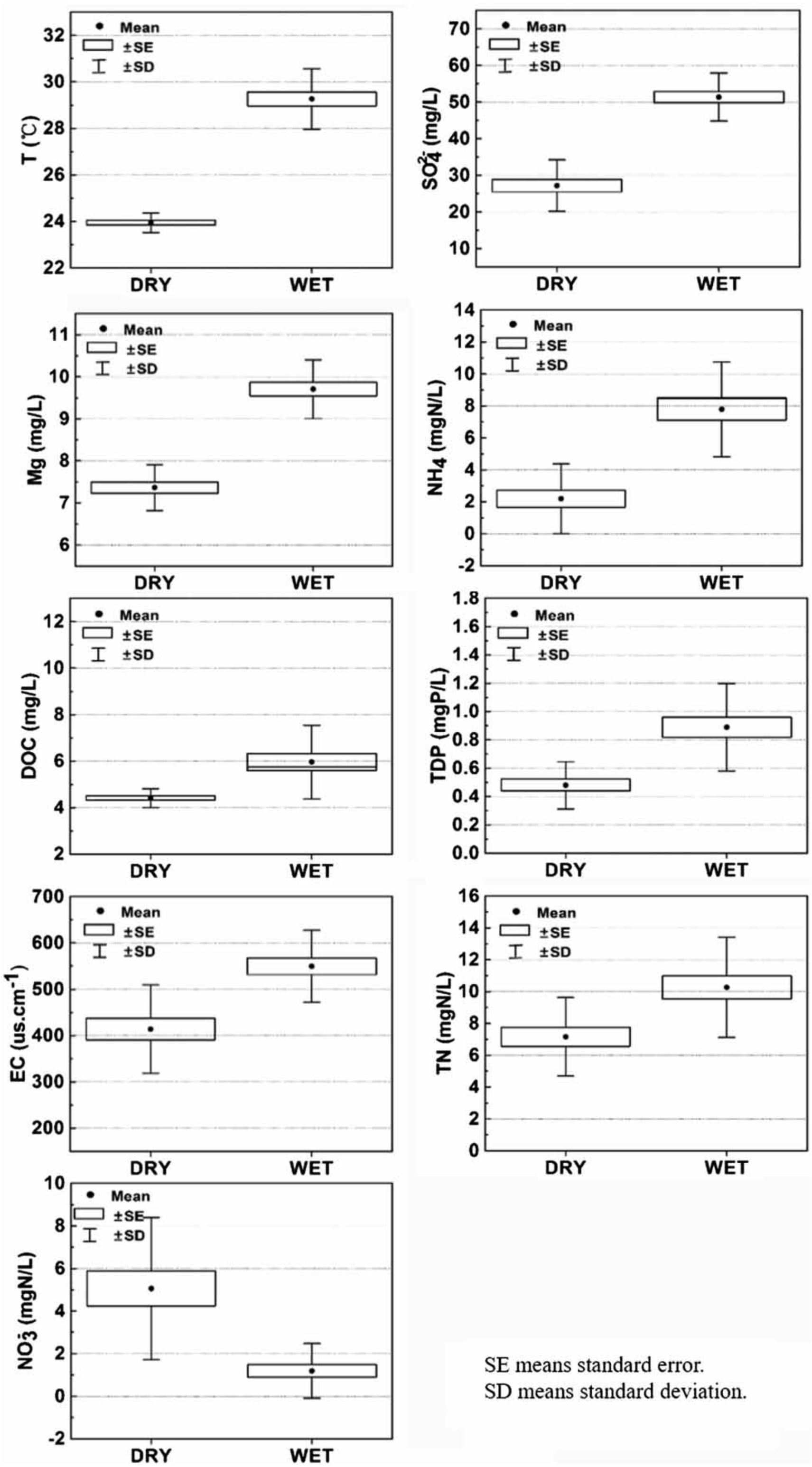

Figure 3 | The temporal variations of the highly seasonally associated surface water quality parameters in the Nanfei River.

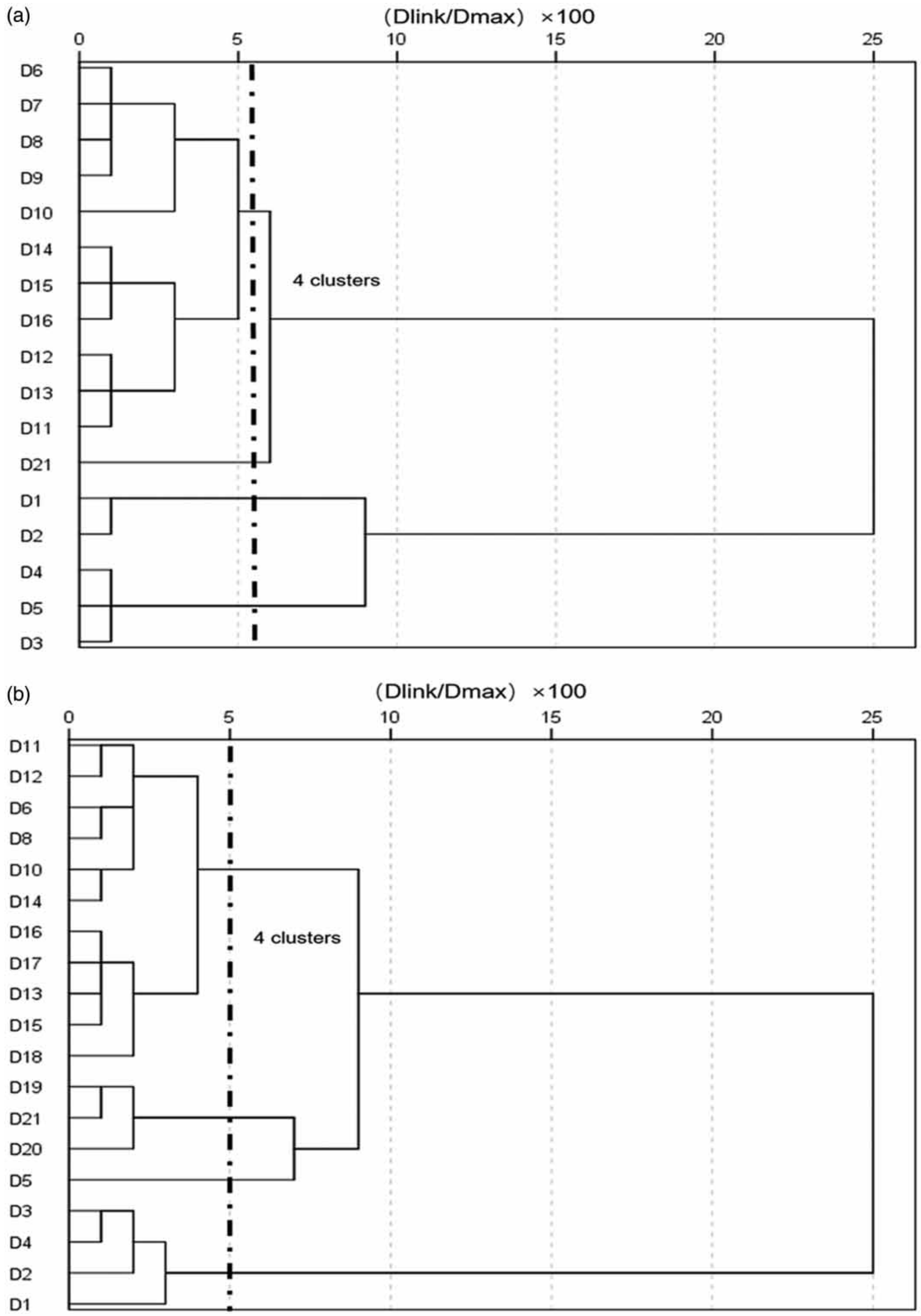

Figure 4 | Dendrogram showing the sampling site clusters in the Nanfei River: (a) dry season; (b) wet season.

wastewater. In addition, VF1 only had moderate NO_3^- loading and weak NH_4 loading, which supports this inference. Next, VF2 accounted for 26% of the total variation and showed strong positive loadings of NH_4, TDP and DOC, moderate loadings of TN and TDN. NH_4, TDP and DOC are believed to be typical pollutants in untreated domestic sewage, since they can easily be removed during the treatment processes at WWTPs, these constituents are present at much lower concentrations in the WWTP effluent. Therefore, VF2 was considered to represent untreated domestic

Table 2 | Loadings of the experimental variables on the significant varifactors (VFs)

Parameter	VF1	VF2	VF3
Na	**0.964**[a]	−0.051	0.011
Cl	**0.898**	0.278	0.160
TDN	**0.802**	0.492	−0.194
K	**0.762**	0.187	0.120
TN	0.685	0.662	0.091
NH_4	0.177	**0.896**	0.300
TDP	0.443	**0.843**	0.188
DOC	0.140	**0.824**	0.233
NO_3^-	0.536	−0.610	−0.459
SO_4^{2-}	0.041	0.201	**0.945**
Mg	0.106	0.279	**0.934**
Eigenvalue	5.649	2.900	1.148
% Total variance	51.350	26.361	10.435
Cumulative	51.350	77.711	88.146

[a]Bold numbers indicate the strong influence loadings for VFs.

wastewater. The urban village serves as a point source of untreated domestic wastewater during the dry and wet seasons. Nevertheless, the untreated domestic sewage that mixes with the storm water discharged by CSO during rainfall contributes to the pollutant loads at VF2, especially during the wet season. Finally, VF3 explained 10% of the total variation in the data and showed strong SO_4^{2-} and Mg loadings, which is in accordance with the Spearman's rank correlation analysis. It is found that the samples' factor scores (Figure 5) on VF3 were all negative in dry season, while Factor Score 3 of the samples in wet season were almost all positive, which means that VF3 is a highly season-related source. During the wet season, large amounts of SO_4^{2-} and Mg can be washed off from the fertilized soil and brought into the river in the surface runoff (Nurdogan *et al.* 1998). Therefore, VF3 can be considered as receiving pollution from surface runoff.

The factor scores of samples on VFs are shown in Figure 5. It is helpful to understand the temporal and spatial distributions in the Nanfei River. A higher score means more significant influence of that component. Combined with the cluster results, Figure 5 provides information about the different features among different clusters, which helps to understand why samples are grouped or not grouped into a cluster.

Source apportionment

In this study, S1–S3 were used to represent the source types identified by PCA/FA, which included WWTP effluent, untreated domestic sewage, surface runoff, respectively. In the regression equation, the intercept of the regression equation represents unknown sources. If the PCA is successful, the constant should tend toward zero (Thurston & Spengler 1985). Different from mass balance, the current methods 'concentration balance' allow the percentage values less than zero and greater than 100%. In APCS-MLR, source contribution estimates are not constrained to be nonnegative (Guo *et al.* 2004; Singh *et al.* 2008). The contributions of different sources could be negative because of dilution effects. The results of the receptor model through the APCS-MLR for source apportionment are presented in Table 3 as annual averages.

All of the R^2 values were greater than 0.8, which corresponds to the fraction of the variance of the measured concentrations that can be attributed to the variance in the calculated concentrations (Singh *et al.* 2008). The multiple regression indicated good agreement between the calculated and measured values.

The WWTP effluent (S1) contributed 19–231% of the concentrations of the constituents in the river. The WWTP effluent was the main source for TDP, NO_3^-, TDN, TN, Na, K and Cl. Compared with the very high NO_3^- concentrations in S1, S2 and S3 contributed largely to the flow but had much lower NO_3^- concentrations. Therefore, it is not surprising that S2 and S3 resulted in negative NO_3^- contributions due to dilution effect.

Although none of the parameters were primarily affected by S2, S2 still accounted for 27% of the NH_4 and 16% of the TDP. After the implementation of the sewage interception project, Nanfei River received less pollution from untreated domestic wastewater during the dry season. However, during the wet season, the untreated wastewater discharged by CSO during rainfall events accounted for a large part of S2. However, the impacts of CSO have not been sufficiently considered by environmental authorities. These results are consistent with the results of a previous study in which direct loading estimates were used to identify sources of ammonia and TDP pollution in Nanfei River (Huang *et al.* 2016).

S3 accounted for 10–99% of the total pollutants' concentrations (excluding NO_3^-). S3 contributed the most part of

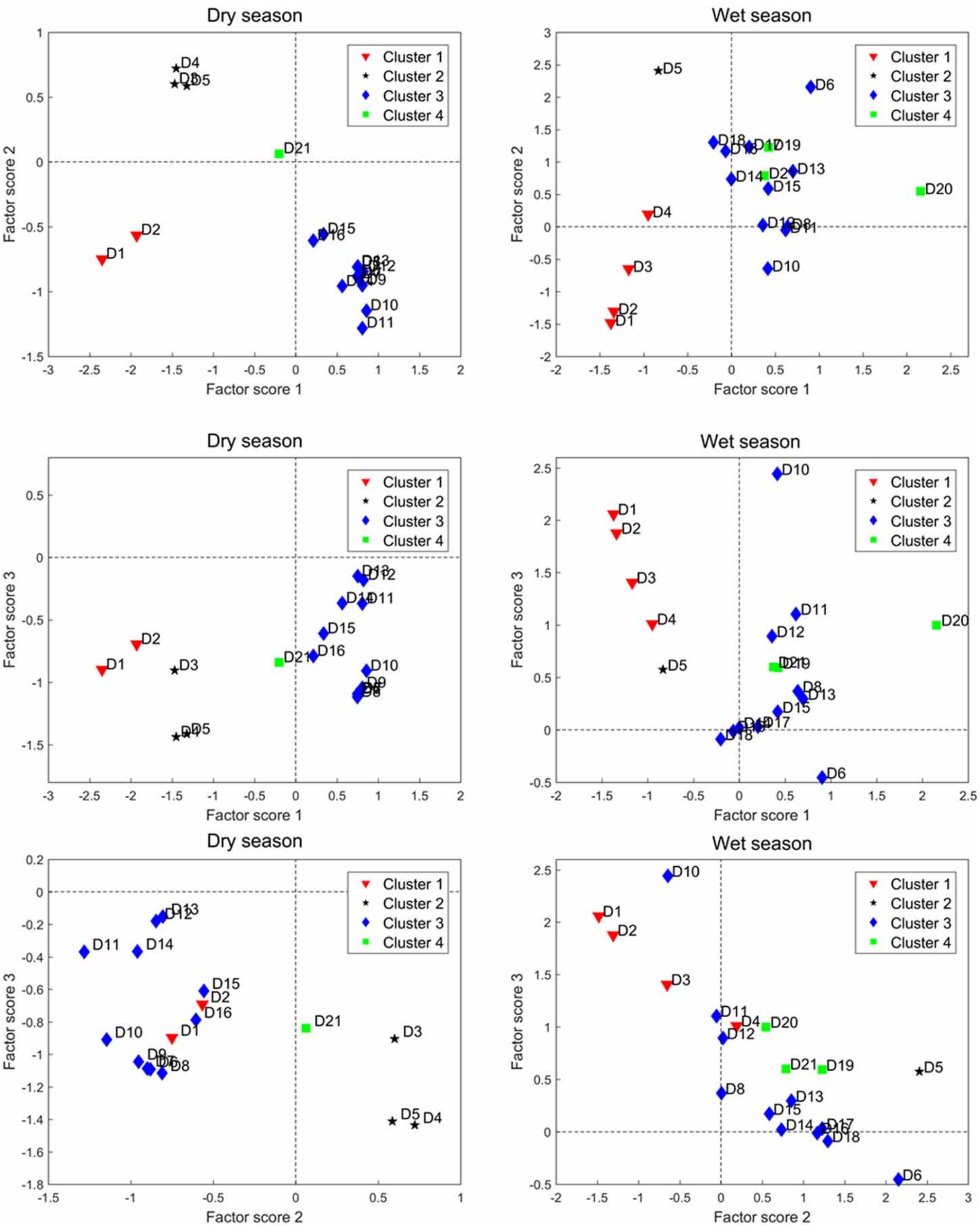

Figure 5 | Factor scores of samples on three principal components.

Table 3 | Source contribution to the different water quality parameters

	Source type						
Parameter	S1 (%)	S2 (%)	S3 (%)	Calculated mean(C)	Measured mean(M)	Ratio[b]	R^2
TDP	**60.38**[a]	16.25	24.26	0.71	0.70	1.007	0.99
NO_3^-	**230.87**	−25.30	−110.74	2.77	2.92	0.949	0.89
NH_4	–	27.45	**73.50**	5.37	5.31	1.010	0.96
TDN	**92.61**	7.09	–	8.81	8.85	0.995	0.99
TN	**79.47**	10.00	10.69	9.82	9.91	0.991	0.99
DOC	33.11	9.43	**55.37**	5.13	5.24	0.979	0.97
Na	**89.36**	**–**	10.02	25.62	25.56	1.002	0.99
K	**103.40**	–	–	11.66	11.27	1.034	0.89
Mg	18.91	2.04	**77.72**	8.53	8.65	0.986	0.99
SO_4^{2-}	–	2.65	**99.83**	41.67	40.97	1.017	0.98
Cl	**97.89**	4.40	–	49.10	48.00	1.023	0.98

[a]Bold numbers indicate the largest source for each variable.
[b]Ratio is between calculated mean and measured mean.

NH_4, implying that pollution from non-point sources cannot be ignored and should be controlled to reduce the NH_4 concentrations in the river.

Environmental management suggestions

For a long time, Nanfei River has been the most polluted tributary of Chaohu Lake in terms of TP and NH_4 (Yang et al. 2013). Based on the source apportionment, the WWTP effluent contributes the most of the TDP, and the surface runoff contributes the most to the NH_4. Considering that the WWTPs in the study area have already implemented the strictest discharge standard for TP (0.3 mg L^{-1}) of China, it would be extremely costly and difficult to further improve the P removal efficiency. Therefore, the residual 40% of phosphorous generated from untreated domestic sewage and surface runoff should be controlled. Despite the difficulties associated with non-point source control, it is important to control non-point source pollution to reduce NH_4 concentrations and TP concentration in Nanfei River Basin. Therefore, changing fertilizer application methods and improving the utilization efficiency of fertilizer are effective measures that can be used to reduce nutrient loading from non-point source. As for untreated domestic sewage, more strict and thorough sewage interception measures should be implemented, especially for direct sewage discharge from the urban village and illegal discharges along the river.

According to the above results, WWTP effluents contribute most of TN in Nanfei River; therefore, improving the nitrogen removal efficiency of WWTPs would effectively reduce the TN loading from Nanfei River to Chaohu Lake. Currently, the WWTPs implement the discharge standards of 5 mg L^{-1} for NH_4 and 15 mg L^{-1} for TN, which are higher than the strictest standards of China (1.5 mg L^{-1} for NH_4 and 10 mg L^{-1} for TN). It indicates that there is still room for improvement of N reduction in the WWTPs. The N removal efficiency could be improved through adding carbon sources to strengthen the denitrification processes in WWTPs.

CONCLUSION

The sample sites were grouped into four sections through CA, which is useful for understanding the spatial variations of water quality and can be applied to optimize the future spatial monitoring network. WWTP1 is a critical demarcation point for river water quality because it is a significant pollution source in Nanfei River and it alters the amount and type of pollutants in the receiving water.

PCA identified three latent pollution sources, and FA made them easy to interpret. According to the PCA/FA results, WWTP effluent, untreated domestic sewage (including CSO) and surface runoff are important pollution sources of Nanfei River.

The APCS-MLR provided the source apportionment result of Nanfei River. TDP, NO_3^-, TDN, TN, Na, K and Cl were primarily generated from WWTP effluent. Although the contributions of untreated domestic sewage are small, it does contribute some to the nutrient load. Meanwhile, NH_4, DOC, Mg and SO_4^{2-} are mainly generated from surface runoff. According to the source apportionment results, specific control schemes were proposed for reducing the nitrogen and phosphorous in a target-oriented and effective way.

The multivariate statistical methods are effective to identify the main pollution sources and quantify their relative contributions in urban rivers.

ACKNOWLEDGEMENTS

This study was supported by China's Major Science and Technology Project on Water Bodies Pollution Control and Treatment (Grant No. 2013ZX07304-002). The authors would like to sincerely thank the Drainage Administration Office of Hefei City for their help with sample collection in the Nanfei River.

REFERENCES

Baborowski, M., Simeonov, V. & Einax, J. W. 2012 Assessment of water quality in the Elbe River at flood water conditions based on cluster analysis, principal components analysis, and source apportionment. *Clean-Soil, Air, Water* **40** (4), 373–380.

Chen, D., Lu, J., Huang, H., Liu, M., Gong, D. & Chen, J. 2013 Stream nitrogen sources apportionment and pollution control scheme development in an agricultural watershed in eastern China. *Environment Management* **52** (2), 450–466.

Environmental Protection Agency of China 2002a *Environmental Quality Standards for Surface Water*. Environmental Protection Agency of China, Beijing, China (in Chinese).

Environmental Protection Agency of China 2002b *Methods of Monitoring and Analysis for Water and Wastewater*, 4th edn. Environmental Protection Agency of China, Beijing, China (in Chinese).

Guo, H., Wang, T. & Louie, P. K. 2004 Source apportionment of ambient non-methane hydrocarbons in Hong Kong: application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model. *Environmental Pollution* **129** (3), 489–498.

Huang, J., Yin, H., Jomma, S., Rode, M. & Zhou, Q. 2016 *Identification of Pollution Sources in A Rapidly Developing Urban River Catchment in China*. Paper presented at the European Geosciences Union, Vienna, Austria, April 17–22.

Kalff, J., Gu, B., Liu, Z. & Li, K. 2011 Limnology inland water ecosystems. In: *Water Forum*, 1st edn. Higher Education Press, Beijing, China (in Chinese).

Liu, C., Lin, K. & Kuo, Y. 2003 Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment* **313** (1–3), 77–89.

Nazeer, S., Ali, Z. & Malik, R. N. 2016 Water quality assessment of River Soan (Pakistan) and source apportionment of pollution sources through receptor modeling. *Archives of Environmental Contamination and Toxicology* **71** (1), 97–112.

Nurdogan, Y., Goldman, E. & Dawes, S. S. 1998 Optimizing chemical treatment of a power plant water softener. *Water Science and Technology* **38** (4), 347–354.

Penha, A. M., Chambel, A., Murteira, M. & Morais, M. 2016 Influence of different land uses on groundwater quality in southern Portugal. *Environmental Earth Sciences* **75** (7), 622.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling & Software* **22** (4), 464–475.

Shrestha, S., Kazama, F. & Nakamura, T. 2008 Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. *Journal of Hydroinformatics* **10** (1), 43–56.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. 2003 Assessment of the surface water quality in Northern Greece. *Water Research* **37** (17), 4119–4124.

Singh, K. P., Malik, A. & Sinha, S. 2005 Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques – a case study. *Analytica Chimica Acta* **538** (1–2), 355–374.

Singh, K. P., Malik, A., Kumar, R., Saxena, P. & Sinha, S. 2008 Receptor modeling for source apportionment of polycyclic aromatic hydrocarbons in urban atmosphere. *Environmental Monitoring and Assessment* **136** (1–3), 183–196.

Thurston, G. D. & Spengler, J. D. 1985 A quantitative assessment of source contributions to inalable particulate matter pollution in metropolitan Boston. *Atmosphere Environment* **19** (1), 9–25.

United Nations, Department of Economic and Social Affairs, Population Division 2016 *The World's Cities in 2016 – Data Booklet* (ST/ESA/SER.A/392).

Yang, L., Lei, K., Meng, W., Fu, G. & Yan, W. 2013 Temporal and spatial changes in nutrients and chlorophyll-α in a shallow lake, Lake Chaohu, China: an 11-year investigation. *Journal of Environmental Sciences* **25** (6), 1117–1123.

First received 11 May 2017; accepted in revised form 20 December 2017. Available online 3 January 2018