



The Influence of Large Scale Phosphate Mining on the Water Quality of the Huangbaihe River Basin in China: Dominant Pollutants and Spatial Distributions

Gebrehiwet Legese Reta^{1,2} · Xiaohua Dong^{1,2} · Bob Su^{1,3} · Xiaonong Hu^{1,4} · Huijuan Bo^{1,2} · Hao Wan^{1,2} · Ji Liu^{1,2} · Yinghai Li^{1,2} · Tao Peng^{1,2} · Haibo Ma^{1,2} · Kai Wang⁵ · Shijin Xu⁵

Received: 10 May 2018 / Accepted: 8 March 2019 / Published online: 30 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Large scale phosphate mining in the Huangbaihe River Basin, China has reduced the self-purification capacity of the basin's fresh water. Three years (2014–2016) of monitoring data and chemometric analysis were used to identify the dominant pollutants and define their spatial distribution in the basin. Principal component analysis was applied to determine the contribution of the individual pollutants. Total phosphorus (TP) 53%, water temperature (TEMP) 27%, and total nitrogen (TN) 20% proved to be the dominant problems. A discriminant functions (DF) model was developed to classify the study area into high, moderate, and low pollution zones. The DF coefficients were applied to analyze the correlation between DF and the measured parameters and it was found that TP, TN, and TEMP were positively correlated with the DF, indicating that these parameters were the most important. Finally, the results were compared with the locations of the mining activities, which revealed that TP is higher in the upper sub-basins, Xuanmiaoguan and Tianfumiao, where most of the high pollution zones are located and more than 78% of the areas are affected by the phosphate mines. It is concluded that the phosphate mining is the major source of pollution and TP is the dominant pollutant responsible for the total water quality variation in the river basin. More effective management measures have to be taken to reduce phosphorus runoff into the reservoir watersheds.

Keywords Chemometrics · Discriminant analysis · Principal component analysis

Introduction

As with all mining activities, the extraction and beneficiation of phosphate rock to produce mineral fertilizer has the potential to cause negative environmental impacts (Association

2001). There are two primary concerns regarding the effects of phosphate mining on water resources (Lewelling and Wylie 1993): (1) its hydrological impact due to water usage, landscape, and ecological changes, and (2) its impacts on water quality due to discharges of polluted water, including effluent process water (Chraïti et al. 2016; Jiries et al. 2004), radionuclides (Uyanik et al. 1999), and phosphate runoff (Das 1999). Such discharges can temporarily or permanently change the use of water resources for a particular purpose (Fakayode 2005).

The Huangbaihe River Basin (HRB) and its four reservoirs, Xuanmiaoguan, Tianfumiao, Xibeikou, and Shangji-ahe, are the first-class drinking water source protection zone of Yichang City, China (Qu et al. 2016). There are a total of 28 water sources, providing 775,000 tons of water per day to a population of 2.269 million. Recently, due to large scale phosphate mining in the region, eutrophication has affected the self-purification capacity of the basin's fresh water. Frequent occurrences of algae blooms in reservoirs and along

✉ Xiaohua Dong
xhdong@ctgu.edu.cn

¹ College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China

² Hubei Provincial Collaborative Innovation Center for Water Security, Wuhan 430070, China

³ University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), 7500 AE Enschede, The Netherlands

⁴ Institute of Groundwater and Earth Sciences, Jinan University, Guangzhou 510632, Guangdong, China

⁵ Hydrologic Bureau of Huaihe River Commission, Bengbu 233001, China

the rivers have produced serious drinking water crises, such as in Yichang City in 2013 (Bao et al. 2018).

Long term water quality monitoring programs have proven to be the best approach to understand watershed hydrochemistry and river water quality (Dixon and Chiswell 1996). However, because monitoring usually involves multiple measurements collected at different times and different locations, monitoring programs can generate large volumes of data that cannot be managed by traditional data analysis (Gazzaz et al. 2012). Chemometric (also known as multivariate) statistical techniques have been tested for managing and analyzing large databases composed of multiple variables, and frequently used for water quality parameter reduction (Lopes et al. 2014), water pollution assessment (Boyacioglu 2007), and water quality classification (Kowalkowski et al. 2006). Chemometric techniques can also be used to explore for hidden patterns in a wide range of data to allow us to relate physical information to our chemical knowledge (Brereton 2003). Three chemometric statistical techniques, cluster analysis (CA), discriminant analysis (DA), and principal component analysis (PCA), were used to analyze 3 years (2014–2016) of water quality monitoring data, identify the dominant pollutants, and define the spatial distribution of pollution zones in the river basin.

Materials and Methods

Study Area

The HRB is a tributary of the Yangtze River, China, located between latitudes 31°00′18″N and 31°29′06″N and longitudes 111°03′54″E and 111°27′34″E (Fig. 1). The HRB has four cascaded reservoirs Xuanmiaoguan, Tianfumiao, Xibeikou, and Shangjiahe located upstream, in the upper and lower sections, and downstream of the river. Their respective total storage capacities are: Xuanmiaoguan $40.54 \times 10^6 \text{ m}^3$, Tianfumiao $64.2 \times 10^6 \text{ m}^3$, Xibeikou $210 \times 10^6 \text{ m}^3$, and Shangjiahe $16.46 \times 10^6 \text{ m}^3$.

Phosphate mining covers 28% of the study area, including the four reservoirs (Wang et al. 2016) (Fig. 1). Since 2003, the scale of mining has been rising: 4.74 billion kg in 2004, 6.05 billion kg in 2005, 8.76 billion kg in 2008, and 14.1 billion kg in 2014. The activities include exploitation, exploration, and pre-beneficiation of phosphate ore. Mining effluent and reused process water have been directly discharged into the nearby streams and reservoirs (Bao et al. 2018; Wang et al. 2016).

Water Quality Sampling and Analytical Procedures

Water samples were collected from 26 stations (Fig. 1) once a month between 2014 and 2015. After 2015, some of the stations were equipped with remote, automatic, real-time, 24 h water quality monitoring systems that monitored water temperature (TEMP), pH, dissolved oxygen (DO), electrical conductivity, turbidity, ammonia nitrogen (NH_4^+), permanganate, fluoride (F), total phosphorus (TP), and total nitrogen (TN), using a telemetry terminal and a WEB release system. This allowed the frequency of water quality data collection to be increased to three times a month between 2015 and 2016. The use of such an online system for data collection has been previously validated (Zhu et al. 2010). Water transparency (WT) and chlorophyll-a (Chl.a) were measured manually, when and where possible. Water sample preservation, transportation, and analysis followed the methods stipulated in State Environmental Protection Administration, China (SEPA 2002). The analytical methods are given in Table 1.

Chemometrics Data Analysis

Chemometrics is a subdivision of environmental science that uses environmental measurements as well as multivariate techniques for analysis and interpretation of data (Boyacioglu 2007; Simeonov et al. 2000). This section gives a brief description of the application of the three standard chemometric methods.

Cluster Analysis

CA is an unsupervised classification and is the most widely used technique for classification of cases into unique clusters. CA classifies members based on the similarity between them and gradually links each member until they are joined together to form a complete cluster. It defines the interval between groups based on Euclidean distance functions EDF; (Kannel et al. 2007). The algorithm that exists between analogous groups is referred to as hierarchical agglomerative clustering HAC; (Ward 1963). The end result of HAC can be displayed as a dendrogram. Due to problems of different measurement units, the original input variables for CA were standardized into a zero mean and unit standard deviation before analysis. The EDF applied in the CA has the form of eq. (1) (Kannel et al. 2007)

$$p(x, y) = \sum_{i=1}^n |x_i - y_i|, \quad (1)$$

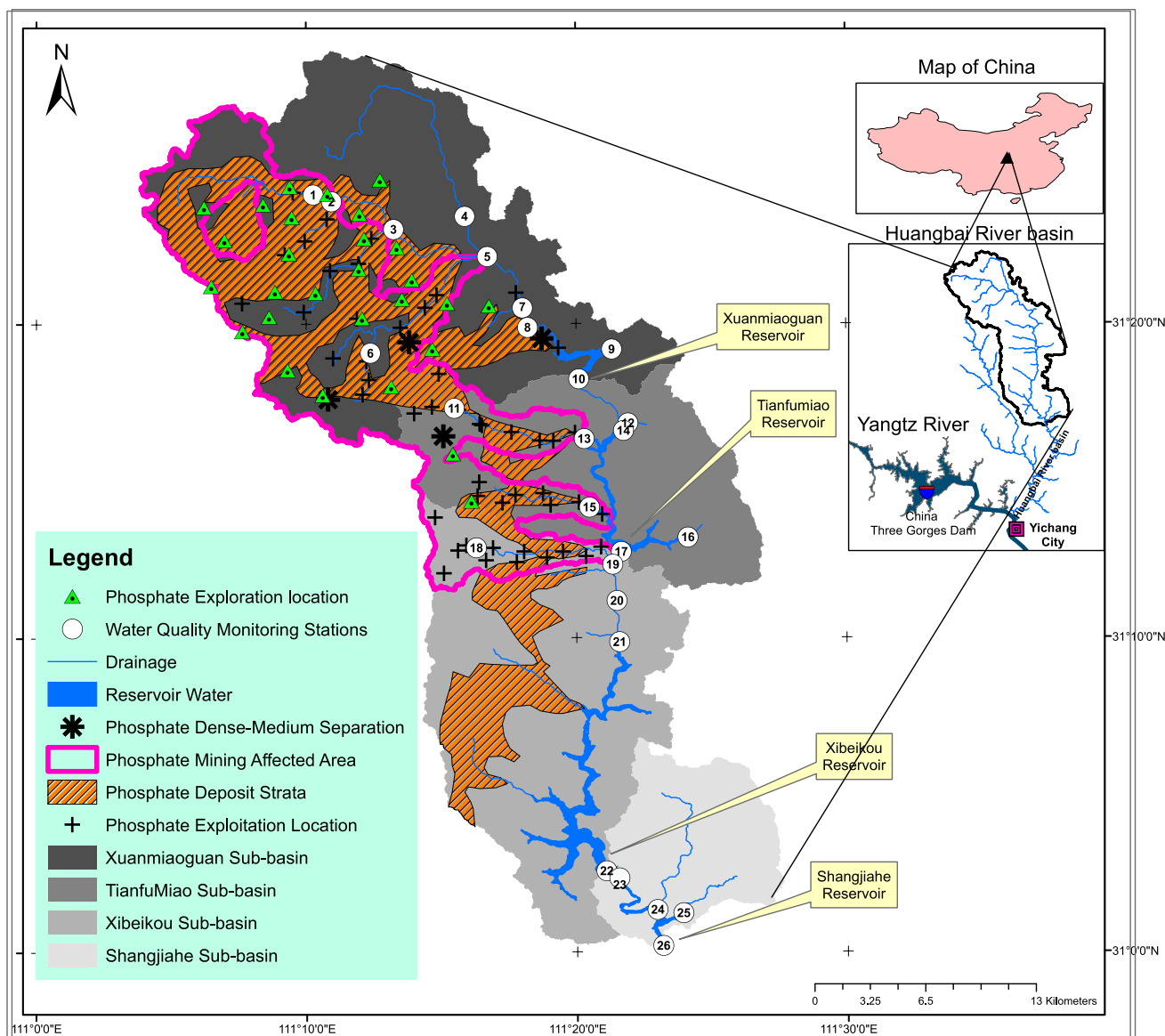


Fig. 1 Location map of the water quality monitoring stations and mining infrastructures in the Huangbai River Basin. Stations 1–10 are located in the Xuanmiaoguan sub-basin, stations 11–17 are located in the Tianfumiao sub-basin, stations 18–22 are located in the

Xibeikou sub-basin, and stations 23–26 are located in the Shangjiahe sub-basin. Monitoring stations 10, 17, 22, and 26 are reservoir water quality monitoring stations located at the outlets of the Xuanmiaoguan, Tianfumiao, Xibeikou, and Shangjiahe reservoirs, respectively

Table 1 Measured parameters and analytical methods

Parameters	Abbreviation	Unit	Analytical method	Detection limit
Water temperature	TEMP	°C	Thermometer	–
pH	pH	mg/L	pH meter	–
Dissolved oxygen	DO	mg/L	Winkler method	0.2
Permanganate index	COD _{mn}	mg/L	Acidic potassium permanganate method	0.5
Ammonium nitrogen	NH ₄ ⁺	mg/L	Auto discrete analyzer	0.025
Total phosphorus	TP	mg/L	Molybdenum blue method	0.01
Total nitrogen	TN	mg/L	Ultraviolet spectrophotometer method	0.05
Fluoride	F	mg/L	Fluoride selective electrode method	0.02
Water transparency	WT	m	Manual, using Secchi disk	–
Chlorophyll-a	Chl.a	mg/L	Spectrophotometer method	0.00011

where $p(x, y)$ is EDF defined by x_i and y_i , and p is the vector space of the parameters.

Discriminant Analysis

DA is a supervised classification technique where known classifications of some observations are used to discriminate them from others in the dataset (Brereton 2003). It forms one or more weighted linear combinations of discriminator variables called discriminant functions DFs; (Tinsley and Brown 2000). There are two main applications of DA for chemometric analysis; (1) classification or discrimination of observation samples between previously identified dependent groups, and (2) determination of the relative contribution of independent parameters to the classification of the dependent groups (Fisher 1936; Tinsley and Brown 2000). The effectiveness of the DF in differentiating between groups is determined by the size of the eigenvalues. The higher the percentage of the eigenvalue, the more effective the DF is in differentiating between groups and vice versa (Aguilera et al. 2001). The Wilks' Lambda test measures the significance of DA, and is an indication of the degree of variance of dependent variables not explained by the DA (Paul et al. 2006). The linear combination for DA, also known as DF, is derived from an equation that takes the form of eq. (2) (Papaioannou et al. 2010):

$$f(G_i) = K_i + \sum_{j=1}^n W_{ij}P_{ij}, \tag{2}$$

where: k_i is a constant; i is the number of groups; n is the number of variables; and w_{ij} is the weight contribution of variable p_{ij} .

Principal Components Analysis

PCA is an advanced data analysis method used to identify the various relationships and interconnections existing among target constituents in multidimensional datasets. It usually serves as a data reduction technique (Lopes et al. 2014) and transforms the original dataset into smaller sets of uncorrelated principal components (PCs) (Gu et al. 2016). The number of significant PCs is ideally equal to the number of significant groups in the dataset (Brereton 2003). In PCA, each principal component (PC) is represented by the scores and factor loadings. The size of the factor loadings of each PC is indicative of the power of the PC to differentiate among the variables. Rotation of the PCs axis generates new sets of factors referred to as varifactors (VFs), which can be used by PCA to extract hidden information that were not identified by PCA (Helena et al. 2000). The general form of PCA has the form of eq. (3) (Singh et al. 2005).

$$Z_{ij} = b_{i1}X_{1j} + b_{i2}X_{2j} + \dots + b_{im}X_{mj}, \tag{3}$$

where: Z is the PC score; a is PC loading; X represents the monitored value of the parameters; i represent PC number; j represents observation number; and m represents the total number of measured parameters.

Results and Discussion

Water Quality Results Evaluation

The descriptive statistics of the 3-year (2014–2016) of experimental parameters from the 26 monitoring stations was separately analyzed for each sub basin (Table 2). For comparison, the recommended environmental quality standard for surface water (China 2002), also known as (GB 3838-2002) is included in each table. The (GB 3838-2002) standards classifies surface water bodies into five classes (Table 2a), analogous to clean to seriously polluted water source. The standards Class I, Class II, and Class III are mainly applicable for drinking water sources, Class IV for industrial use, and Class V for agricultural use.

Considering the toxicity of the pollutants and the location of the station, the numbers of monitored parameters varied among the different stations. WT was measured manually and accessibility was a problem at some of the stations (especially during the deep winter season). Similarly, Chl.a was mainly monitored along the main rivers and inside the reservoir. For this reason, only the 18 stations with common data records were used for chemometric analysis (Fig. 2). Due to problems of no commensurate units, the initial input variables had to be standardized into a zero mean and unit standard deviation before analysis (Kannel et al. 2007). In this study, CA and PCA was applied to the standardized samples and DA was applied to the original, unstandardized samples (Vega et al. 1998). Data analysis was done using SPSS V20 (Norušis 1986).

The basic statistical results showed that there was a higher degree of variation in the nutrient pollutants than in the others. The maximum values for TP, TN, and Chl.a were 286, 72, and 2454 times their minimum values, respectively. The TP measured in the streams (TP_S) also showed greater variation than the TP measured in the reservoir water (TP_R). The maximum TP_S value was 286 times its minimum value, but the maximum values for TP_R was ten times the minimum TP_R . Such large variations in the measured values are likely controlled by the source and location of the pollutant in the basin. Comparison of the mean values of the measured parameters in the HRB with the GB 3838-2002 criteria (Table 2) indicate that the pollution level of DO, pH, F, COD_{mn} and NH_4^+ correlated with *Class I*, TP_S with *Class II*, TP_R with *Class III*, and

Table 2 Descriptive statistics of water quality monitoring samples

	Statistics of measured results				Surface water class limits (GB 3838-2002)					Classification
	Min	Max	Mean	StD	Class I	Class II	Class III	Class IV	Class V	
(a) Xuanmiaoguan sub-basin										
TEMP	2.00	30.00	14.66	6.21	–	–	–	–	–	–
DO	1.92	12.70	9.07	1.48	≥ 7.5	≥ 6.0	≥ 5.0	≥ 3.0	≥ 2.0	Class I
pH	7.20	8.93	8.40	0.26	6.5–8.5	6.5–8.5	6.5–8.5	6.5–8.5	6.0–9.0	Class I
F	0.03	1.09	0.34	0.18	≤ 1.0	≤ 1.0	≤ 1.0	≤ 1.5	≤ 1.5	Class I
COD _{mn}	0.50	12.00	1.35	1.02	≤ 2.0	≤ 4.0	≤ 6.0	≤ 10	≤ 15	Class I
NH ₄ ⁺	0.03	2.69	0.13	0.18	≤ 0.15	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class I
TP _S	0.01	1.37	0.085	0.08	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	Class II
TP _R	0.01	0.10	0.036	0.02	≤ 0.01	≤ 0.025	≤ 0.05	≤ 0.1	≤ 0.2	Class III
TN	0.13	5.98	1.45	0.83	≤ 0.2	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class IV
WT ^a	0.10	4.00	1.65	0.81	–	–	–	–	–	–
Chl.a ^a	0.00011	0.27	0.03	0.05	–	–	–	–	–	–
(b) Tianfumiao sub-basin										
TEMP	3.00	30.00	15.62	6.57	–	–	–	–	–	–
DO	4.00	12.80	9.08	1.43	≥ 7.5	≥ 6.0	≥ 5.0	≥ 3.0	≥ 2.0	Class I
pH	7.60	8.98	8.46	0.19	6.5–8.5	6.5–8.5	6.5–8.5	6.5–8.5	6.0–9.0	Class I
F	0.03	0.97	0.38	0.17	≤ 1.0	≤ 1.0	≤ 1.0	≤ 1.5	≤ 1.5	Class I
COD _{mn}	0.50	6.19	1.40	0.87	≤ 2.0	≤ 4.0	≤ 6.0	≤ 10	≤ 15	Class I
NH ₄ ⁺	0.03	1.89	0.13	0.16	≤ 0.15	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class I
TP _S	0.01	2.86	0.073	0.15	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	Class II
TP _R	0.01	0.1	0.040	0.02	≤ 0.01	≤ 0.025	≤ 0.05	≤ 0.1	≤ 0.2	Class III
TN	0.07	5.07	1.48	0.71	≤ 0.2	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class IV
WT	0.10	4.20	1.78	0.78	–	–	–	–	–	–
Chl.a	0.00011	0.14	0.02	0.02	–	–	–	–	–	–
(c) Xibeikou sub-basin										
TEMP	3.00	30.00	16.04	6.58	–	–	–	–	–	–
DO	6.69	12.80	9.14	1.33	≥ 7.5	≥ 6.0	≥ 5.0	≥ 3.0	≥ 2.0	Class I
pH	7.10	8.98	8.47	0.20	6.5–8.5	6.5–8.5	6.5–8.5	6.5–8.5	6.0–9.0	Class I
F	0.03	0.97	0.40	0.21	≤ 1.0	≤ 1.0	≤ 1.0	≤ 1.5	≤ 1.5	Class I
COD _{mn}	0.50	6.02	1.40	0.82	≤ 2.0	≤ 4.0	≤ 6.0	≤ 10	≤ 15	Class I
NH ₄ ⁺	0.03	0.63	0.11	0.09	≤ 0.15	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class I
TP _S	0.01	0.18	0.053	0.03	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	Class II
TP _R	0.01	0.050	0.030	0.012	≤ 0.01	≤ 0.025	≤ 0.05	≤ 0.1	≤ 0.2	Class III
TN	0.24	12.19	1.52	0.99	≤ 0.2	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class V
WT	0.50	5.20	2.53	0.78	–	–	–	–	–	–
Chl.a	0.00011	0.10	0.01	0.02	–	–	–	–	–	–
(d) Shangjiahe sub-basin										
TEMP	3.00	30.00	16.76	6.79	–	–	–	–	–	–
DO	4.60	12.72	9.06	1.43	≥ 7.5	≥ 6.0	≥ 5.0	≥ 3.0	≥ 2.0	Class I
pH	7.60	8.90	8.35	0.22	6.5–8.5	6.5–8.5	6.5–8.5	6.5–8.5	6.0–9.0	Class I
F	0.02	0.58	0.28	0.11	≤ 1.0	≤ 1.0	≤ 1.0	≤ 1.5	≤ 1.5	Class I
COD _{mn}	0.57	6.00	1.93	1.09	≤ 2.0	≤ 4.0	≤ 6.0	≤ 10	≤ 15	Class I
NH ₄ ⁺	0.03	7.51	0.14	0.53	≤ 0.15	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class I
TP _S	0.01	0.45	0.052	0.04	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	Class II
TP _R	0.01	0.05	0.029	0.01	≤ 0.01	≤ 0.025	≤ 0.05	≤ 0.1	≤ 0.2	Class III
TN	0.22	4.73	1.65	0.77	≤ 0.2	≤ 0.5	≤ 1.0	≤ 1.5	≤ 2.0	Class V
WT	0.50	4.60	2.30	0.83	–	–	–	–	–	–
Chl.a	0.00011	0.10	0.01	0.02	–	–	–	–	–	–

TP_S TP concentrations in streams, TP_R TP concentrations in reservoirs and lakes

^aWT

^bChl.a do not affect classification in GB 3838-2002

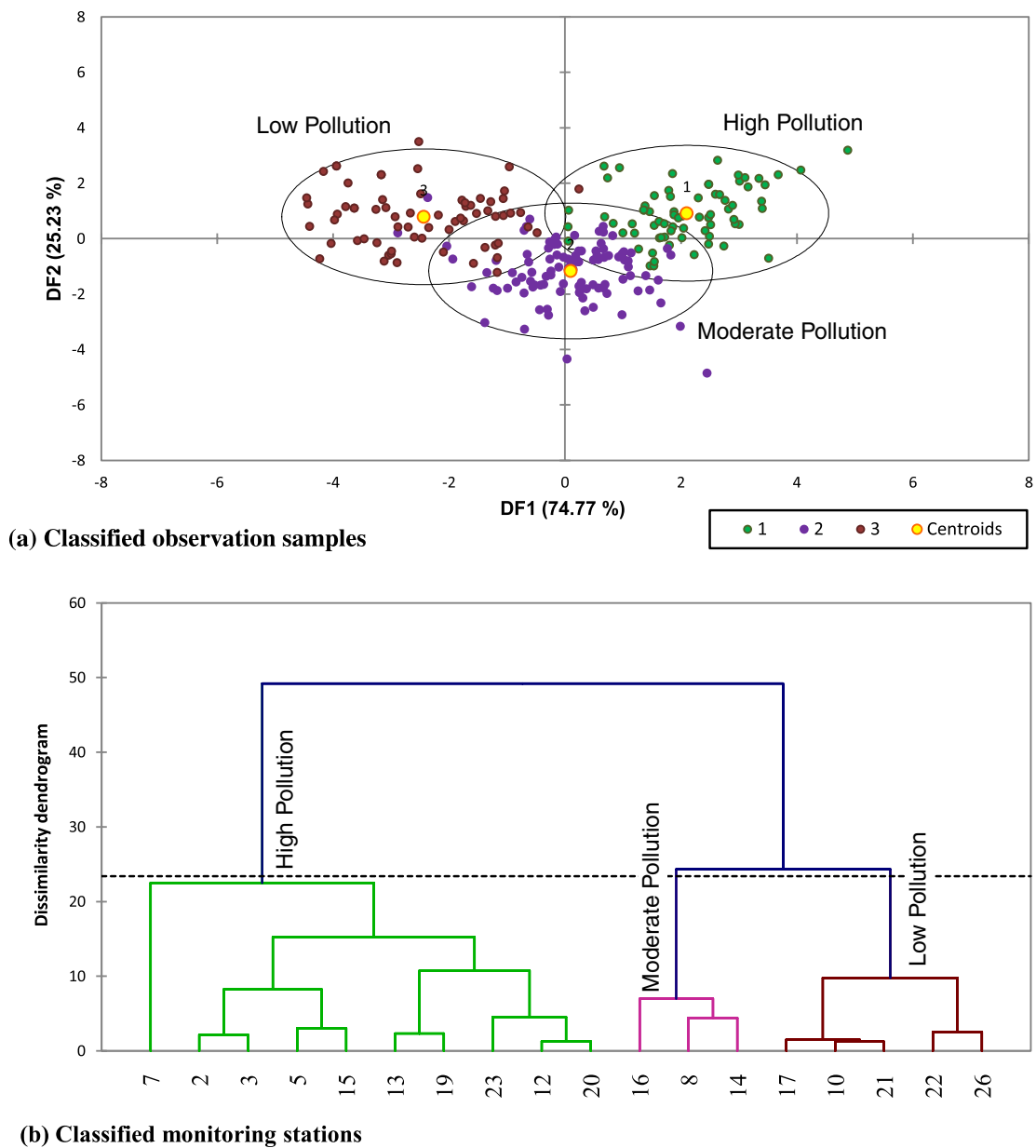


Fig. 2 Classification of pollution zones: **a** classification of observation samples into three pollution zones, and **b** classification of monitoring stations (spatial locations in the river basin) into three pollution zones

TN with *Class IV–Class V*, the worst water quality grade in China’s surface water quality standard (China 2002).

Dominant Pollutants and Contribution

PCA was used to identify the dominant pollutants and to determine their contribution to the total pollution in the river basin. The Kaiser–Meyer–Olkin (KMO) test measure was calculated for each parameter to test the sampling adequacy. KOM values ranges from 0 to 1, where 1 indicates that the

dataset is applicable for PCA, 0 implies the dataset is not applicable for PCA, and a value greater than 0.6 is acceptable (Nalini et al. 2010). The KMO result in this study was 0.66 (Table 3), which may indicate acceptable. A similar test was used by Lopes et al. (2014) when they found a KMO value of 0.6 was acceptable.

According to the Kaiser criterion (Kaiser 1960), only eigenvalues greater than one should be used because subsequent smaller eigenvalues are less able to explain variation among the samples. The best result of the PCA model was

Table 3 Principal component analysis: Kaiser–Meyer–Olkin (KMO) test result

Parameters	TEMP	pH	DO	COD _{mn}	NH ₄ ⁺	TP	TN	F	WT	Chl.a	Average
KMO	0.474	0.471	0.588	0.819	0.733	0.801	0.604	0.447	0.753	0.730	0.663

Table 4 Principal component analysis: total variance explained

PCs	Extracted eigenvalues			Eigenvalues greater than one		
	Eigenvalues	Variability (%)	Cumulative (%)	Eigenvalues	Variability (%)	Cumulative (%)
PC-1	3.166	31.661	31.661	3.166	31.661	31.661
PC-2	1.605	16.054	47.715	1.605	16.054	47.715
PC-3	1.217	12.173	59.888	1.217	12.173	59.888
PC-4	0.966	9.661	69.549			
PC-5	0.861	8.612	78.161			
PC-6	0.780	7.805	85.965			
PC-7	0.548	5.481	91.446			
PC-8	0.420	4.201	95.647			
PC-9	0.315	3.149	98.796			
PC-10	0.120	1.204	100.000			

Table 5 Principal component analysis: principal component matrix

Parameters	PC-1	PC-2	PC-3
TEMP	0.460	−0.797	0.205
pH	−0.343	−0.087	0.597
DO	−0.725	0.594	−0.101
COD _{mn}	0.770	0.237	−0.107
NH ₄ ⁺	0.750	0.252	−0.034
TP	0.783	0.227	0.113
TN	0.265	−0.106	−0.705
F	0.082	0.280	0.439
WT	−0.569	0.002	−0.267
Chl.a	0.380	0.591	0.147

therefore made up of three PCs (Table 4); PCA reduced ten variables into three PCs that explained 60% of the total variance. It can be seen that the first two PCs (PC-1 and PC-2) alone accounted for 48% of the total variations in the river basin.

The values of the factor loadings in each PC indicate the relationship of the PC with that parameter. For the first factor, PC-1 (Table 5), the highest load is for the parameter TP, 0.783, indicating its significance in explaining the variation in PC-1. The second, PC-2, is largely explained by TEMP, 0.797. The third factor, PC-3, is largely explained by TN, 0.705.

The final step is to show how the application of a matrix transformation by the varimax rotation, simplifies the interpretation of the results. The method is accomplished by factor analysis of the PCA results (Horel 1981); the results are called varifactors (VFs) (Table 6). This analysis is very important because it maximizes the gap between the highest

and lowest factor loadings, which can help distinguish the most important pollutants in the water pollution analysis. Three VFs were generated (Table 6), VF-1, VF-2, and VF-3, which explained 31.6%, 16%, and 12% of the total water quality variation, respectively. Successive VFs were very small and their contribution is insignificant (Kaiser 1960); therefore, the percentage score of the three identified contributing VFs can be scaled up to 53%, 27%, and 20%, respectively. The variables with the highest loading in each of the VFs the hallmark variables; (Kothai et al. 2008) were TP (0.793) on VF1, TEMP (−0.943) on VF2, and TN (−0.754) on VF3.

Pollution Zones Discrimination Analysis

Combined application of CA, DA, and water quality variables can be used to discriminate between significant pollution zones. CA classified monitoring stations into three significant pollutions zones (Fig. 2): high pollution (HP), represented by stations 2, 3, 5, 7, 12, 13, 15, 19, 20, and 23, moderate pollution (MP), represented by stations 8, 14, and 16, and low pollution (LP), represented by stations 10, 17, 21, 22, and 26.

The clustering of monitoring stations into three distinct pollution zones indicates that water quality in the river basin varied distinctively and this variation was mainly due to the anthropogenic activities in the river basin. The results (Fig. 2b) can be interpreted with the location of the monitoring stations (Fig. 1), which showed that most of the HP stations are located in the more upstream sub-basins, Xuanmiaoguan and Tianfumiao, and that most of the LP stations are located in the more downstream sub-basins, Xibeikou and Shangjiahe. Figure 2b also shows

Table 6 Principal component analysis: rotated component matrix and the contribution of parameters (%)

Parameters	Rotated component matrix			(% of Contribution		
	VF-1	VF-2	VF-3	VF-1	VF-2	VF-3
TEMP	−0.005	−0.943	−0.012	0.106	29.931	0.449
pH	−0.260	−0.044	0.642	5.910	1.383	23.667
DO	−0.312	0.877	0.146	7.094	27.861	5.371
COD _{mn}	0.759	−0.152	−0.246	17.271	4.817	9.067
NH ₄ ⁺	0.759	−0.147	−0.169	17.264	4.660	6.241
TP	0.793	−0.218	−0.038	18.031	6.927	1.386
TN	0.079	−0.056	−0.754	1.806	1.777	27.795
F	0.270	0.094	0.443	6.148	2.991	16.311
WT	−0.512	0.342	−0.128	11.634	10.850	4.735
Chl.a	0.648	0.277	0.135	14.737	8.802	4.978
(%) of Variance	31.661	16.054	12.173			
Cumulative (%)	31.661	47.715	59.888			
The main pollutants in the river basin				TP	TEMP	TN
Percentage of contribution				53%	27%	20%

Table 7 Discriminate analysis: un-standardized discriminate function coefficients and eigenvalues

Eigenvalue	Un-standardized discriminate function coefficients	
	DF ₁	DF ₂
Eigenvalue	3.006	1.040
(%) of Variance	74.3	25.7
Cumulative (%)	74.3	100
	Parameters	
	TEMP	0.099
	pH	1.861
	DO	−0.513
	COD _{mn}	0.201
	NH ₄ ⁺	2.851
	TP	20.797
	TN	0.321
	F	−4.371
	WT	−0.061
	Chl.a	−6.026
	Intercept	−13.679

that the stream monitoring stations are more polluted than the reservoir monitoring stations (10, 17, 22, and 26; Fig. 1). Generally the findings from the CA analysis indicate that the upstream sub-basins are more mineralized (more polluted) than the downstream sub-basins and that pollution is worse in the tributary streams than in the main river and reservoirs.

DA was used to develop discriminate functions (DFs) that can be used to discriminate the pollution heterogeneity of monitored samples for the monitoring stations. For number dependent groups (K), the number of DFs is k − 1. There are three pollution zones (dependent groups); therefore, the developed DA-model has two DFs (Table 7).

Un-standardized discriminate function coefficients (UDFCs) were used to construct the actual prediction equation, which was used to classify new observations. The eigenvalues (Table 7) were used to give the discriminate criterion. The higher the eigenvalues, the more effective the DF is in differentiating between the groups (Aguilera et al. 2001). 74% of the variance was explained by the first DF (DF₁), which was more effective in differentiating between the groups than DF₂. Therefore, it is statistically valid to develop the DF for the river basin based on DF₁. The DF, developed based on the UDFCs (Table 7) and (Eq. 2) is given in Eq. (4). This is the classical method of classification, used in different studies e.g. Gazzaz et al. (2012).

$$DF_1 = \left[-13.679 + 0.099 \times TEMP + 1.861 \times pH - 0.513 \times DO + 0.201 \times COD_{mn} + 2.851 \times NH_4^+ + 20.797 \times TP + 0.321 \times TN - 4.371 \times F - 0.061 \times WT - 6.026 \times Chla \right] \tag{4}$$

Table 8 Discriminate analysis: test of equality of covariance matrices

Wilks' Lambda test	Value
Lambda	0.119
F (observed value)	43.588
F (critical value)	1.593
DF1	20
DF2	460
p value	<0.0001
Alpha	0.05

The next step was to assess the level of significance of the developed DF. The Wilk's lambda test evaluates the significance of the discriminatory power of the DF. It indicates the amount of the variance of the dependent variable that is not explained by the DA. Wilk's lambda value ranges between 0 and 1, where 0 indicates total discrimination and 1 indicates no discrimination. The Wilks' Lambda test result was 0.119 (Table 8), which is considered significant.

The last step was to use the developed DF to classify the monitored samples for the monitoring stations. Figure 2a shows that the monitored samples can be also classified into three distinct pollution group analogues to HP, MP, and LP. In Fig. 2a, the HP samples display the highest percentage of variance on the DF₁ vs. DF₂ axis and the LP samples display the lowest percentage of variance. The efficiency of the classification can be assessed by the percentage of correctly classified samples (Table 9). In this analysis, the efficiency obtained was high, with 93% of the total correctly classified, and 88% cross-validated (Table 9). The correct assignment

Table 9 Discriminate analysis: original sample classification results and cross-validation result

	Group	Predicted-group-member-ship			Total-sample
		HP	MP	LP	
Original-classification					
Count	HP	70	3	0	73
	MP	8	94	1	103
	LP	0	4	63	67
(%) Cor-rectly classified 93.42%	HP	95.89	4.1	0.0	100
	MP	7.7	91.26	0.97	100
	LP	0	5.9	94.03	100
Cross-validated					
Count	HP	63	8	2	73
	MP	8	91	4	103
	LP	1	5	61	67
(%) Cor-rectly classified 88.48%	HP	86.3	10.9	2.7	100
	MP	7.7	88.35	3.8	100
	LP	1.5	7.5	91.00	100

for each pollution zone was HP 96%, MP 91%, and LP 94%. The cross-validation provides the percent correctly predicted for any samples left unclassified. The correct assignment for cross-validation among the pollution zones was HP 88%, MP 86%, and LP 91%, which indicates that the DF model may be accurate enough for river water quality management in the HRB.

The UDFCs (Table 7) were used to examine the developed DF to determine the relative importance of the monitored parameters in discriminating between the pollution zones (Fig. 2a, b). The parameters with relatively larger weights contribute more to the discriminate power of the DF than parameters with smaller weights. The parameters TP (20.797), NH₄⁺ (2.851), pH (1.861), TN (0.321), COD_{mn} (0.201), and TEMP (0.099), were positively correlated with DF and were important parameters in the discriminating process (Table 7). However, the contribution of TP is greater than the contributions from all of the other selected parameters, indicating that TP was the most important parameter for classifying the observations using DF.

Spatial Correlation of Dominant Pollutants and the Mining Area

The objective of this analysis was to investigate the spatial relationship of dominant pollutants, TP and TN, and the mining activities by comparing concentrations of the pollutants and the proportion of mining area in the mining sub-basins (Figs. 1, Fig. 3a1, a2, b).

More than 260 km² or 28% of the study area is affected by phosphate mining and associated activities. The mining sub-basins are Xuanmiaoguan, Tianfumiao, and Xibeikou. The Shangjiahe sub-basin has no mining activity. The proportion of mining area in the mining sub-basins calculated from (Fig. 1); is Xuanmiaoguan (45%), Tianfumiao (33%), and Xibeikou (11%). The TP_S in each sub-basin (upstream–downstream) were: Xuanmiaoguan 0.085 mg/L, Tianfumiao 0.073 mg/L, Xibeikou 0.053 mg/L, and Shangjiahe 0.052 mg/L (Fig. 3a1). The TP_R in each sub-basin had the same trend: Xuanmiaoguan 0.036 mg/L, Tianfumiao 0.040 mg/L, Xibeikou 0.030 mg/L, and Shangjiahe 0.029 mg/L (Fig. 3a2). The result indicated that TP concentration (TPs and TP_R) of the mining sub-basins varied among sub-basins and decreased sharply from upstream to downstream sub-basin, which corresponded with the areal extent and proximity of the mining area within each sub-basin. The relationship was stronger between TP_S and mining extent than between TP_R and mining extent, which may be because TP_R is affected by dilution due to mixing of the reservoir water with stream water. Even so, TP decreased downstream in both cases, away from the mining sub basins (Fig. 3a1, a2). This result is consistent with the results

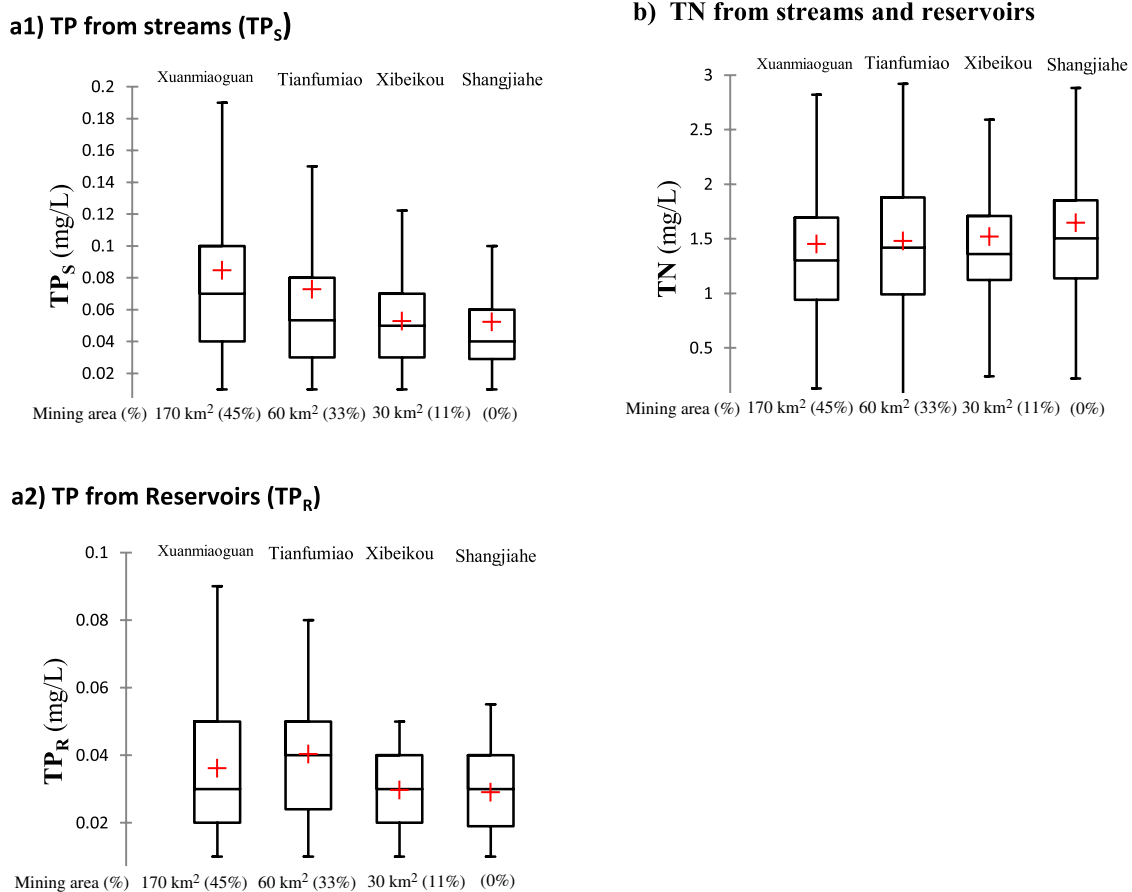


Fig. 3 Box plots that show the spatial correlation of dominant pollutants and proportion of mining area in the Huangbaihe River’s four sub-basins: **a1** total phosphorus from streams and proportion of the mining area, **a2** total phosphorus from reservoirs and proportion of

the mining area, **b** total nitrogen from streams and reservoirs and proportion of the mining area. In the plots, a “+” marks the location of the mean; a line across the box represents the median, and the bottom and top of the box show the locations of the first and third quartiles

from PCA and DA: TP is higher in Xuanmiaoguan and Tianfumiao, where most of the high pollution zones are located, and more than 78% of the areas are affected by the phosphate mines.

Similar analysis of the TN concentrations in the sub-basins (Fig. 3b) and proximity to mining indicated a weak correlation; TN slightly increased downstream, away from the mining areas. The TN concentrations (Fig. 3b) were: Xuanmiaoguan 1.453 mg/L, Tianfumiao 1.483 mg/L, Xibeikou 1.520 mg/L and Shangjiahe 1.647 mg/L.

Comparison of Surface Water Class Limits Among Sub-basins and Between Streams and Reservoirs

The descriptive water quality statistics are presented in Table 2a–d. All of the parameters listed in Table 1, except for WT and Chl.a, are used to assess surface water quality in China (China 2002). China’s surface water quality standard

has separate limits for TP from streams (TP_S) and reservoirs (TP_R). Comparison of surface water class limits among the sub-basins and between the streams and reservoirs (Table 2), indicated that the parameters DO, pH, F, COD_{mn}, and NH₄⁺ classified stream and reservoir water in all of the sub basins into Class I; TPS classified stream water in all sub basins into Class II; TPR classified reservoir water in all sub basins into Class III; TN classified stream and reservoir water in the Xuanmiaoguan and Tianfumiao sub-basins into Class IV and stream and reservoir water in the Xibeikou and Shangjiahe sub-basins into Class V. According to this analysis, the water quality in the upper two sub-basins (Xuanmiaoguan and Tianfumiao) is between Class I and Class IV and water quality in the lower two sub-basins (Xibeikou and Shangjiahe) is between Class I and Class V. Of the seven water quality indicators considered for water quality classification (Table 2), the most important pollutants are TN and TP.

The Impact of Phosphate Mining on Water Quality

To investigate the impact of phosphate mining in each sub-basin, the concentrations of TP transported into the reservoirs, TP_S , were compared to the TP concentrations in the reservoirs, TP_R (Table 10). The TP-ratio computed for each reservoir was: Xuanmiaoguan (2.34), Tianfumiao (1.80), Xibeikou (1.77), and Shangjiahe (1.76) (Table 10). The size of the ratio indicates the relative amount of TP being discharged in each reservoir. Xuanmiaoguan had the highest ratio, indicating that phosphate mining has increased the TP_S in the Xuanmiaoguan sub-basin relative to the other sub-basins. A similar analysis compared the TN concentrations transported into the reservoirs (TN_S) and in the reservoirs (TN_R). The TN-ratio (TN_S/TN_R) computed for each reservoir was: Xuanmiaoguan (1.2), Tianfumiao (0.8), Xibeikou (0.93), and Shangjiahe (0.86) (Table 10). Similar to the TP-ratio, the TN-ratio was higher for the Xuanmiaoguan reservoir than the other three downstream reservoirs, indicating the phosphate mining has increased TN in the Xuanmiaoguan sub-basin relative to the other sub-basins.

The TP-ratio was greater than one ($TP_S/TP_R > 1$) in all of the sub-basins, which indicates that the sub-basin watersheds are more mineralized with TP than the reservoirs. Therefore, the watershed activities may be the main cause of the increased TP in the reservoirs. However, the TN-ratio, especially in three sub-basins (Tianfumiao, Xibeikou, and Shangjiahe) was less than one ($TN_S/TN_R < 1$), indicating that the watershed is less contaminated by TN than the reservoirs. Therefore, the reservoir hydrochemistry may be the main reason for the increased TN in the reservoirs. There are several studies (e.g. Quirós 2003) that show that lakes and reservoirs can increase TN in the water by self denitrification, ammonium emission from sediment, and N_2 fixation.

The general outcome of this analysis is that the Xuanmiaoguan subbasin is more mineralized by nutrient runoff, especially TP, which is released from the phosphate mines. This finding is consistent with the results from previous

work conducted in the same area. For example, Bao et al. (2018) measured the TP in the sediments of the reservoirs and found that TP concentration in the sediments of Xuanmiaoguan reservoir is much higher than that of almost all typical eutrophic lakes or reservoirs in China. However, the TP diffusion fluxes in the sediments of Xuanmiaoguan reservoir are lower than that in Lake Dianchi. Lake Dianchi is one of the most hypertrophic lakes in China (Xiangcan 2003) affected by phosphate mining activities (Yang et al. 2018).

Conclusions

Phosphate mining activities in the area has increased TP in HRB's upstream reservoir watersheds. Water quality data was collected between 2014 and 2016 to provide a synopsis analysis of ten experimentally determined parameters. Factor analysis of the PCA results indicated that TP, TEMP, and TN are the hallmark pollutants, responsible for 53%, 27%, and 20% of the total water pollution variations in the HRB. A DF model was developed to classify observations into high, moderate, and low pollution zones. The DF coefficients were used to analyze the correlation between DF and the measured parameters and it was found that TP, NH_4^+ , pH, TN, COD_{mn} , and TEMP had good positive correlations with DF, indicating that these parameters were the most important for classifying the observations. A spatial comparison of the analytical results and mining activities in the study area was done by comparing the sub-basin average TP_S , TP_R , and TN concentrations and proportion of mining area in the mining sub-basins. The result indicated that TP concentration (TP_S and TP_R) of the mining sub-basins varied among sub-basins and decreased sharply from upstream to downstream sub-basin, which corresponded with the areal extent and proximity of the mining area within each sub-basin. The TP concentrations were higher in the upper sub-basins: Xuanmiaoguan and Tianfumiao, where most of the high pollution zones are located and more than 78% of the areas are affected by the phosphate mines. As a general outcome, this study indicated that these two upstream sub-basins are more mineralized with nutrient runoff, especially TP, which is being released by phosphate mines in these upstream. These results demonstrate the need to reduce phosphorus discharges into the upstream tributaries.

Acknowledgements This research has been supported by the Non-profit Industry Financial Program of Ministry of Water Resources of China (no. 201301066), the National Natural Science Foundation of China (40701024, 41101511, 51409152), and Hubei Provincial Collaborative Innovation Center for Water Security.

Table 10 Comparison of TP and TN concentrations among sub-basins and between streams and reservoirs

Parameter	Xuanmiaoguan	Tianfumiao	Xibeikou	Shangjiahe
TP_S	0.085	0.073	0.053	0.051
TP_R	0.036	0.04	0.03	0.029
TN_S	1.554	1.366	1.475	1.528
TN_R	1.318	1.694	1.579	1.767
$(TP_S/TP_R)^a$	2.343	1.805	1.777	1.767
$(TN_S/TN_R)^b$	1.179	0.806	0.934	0.864

TN_S total nitrogen from streams, TN_R total nitrogen from reservoirs

^aTP-ratio

^bTN-ratio

References

- Aguilera PA, Castro H, Rescia A, Schmitz MF (2001) Methodological development of an index of coastal water quality: application in a tourist area. *Environ Manag* 27:295–301
- Association IFI (2001) Environmental aspects of phosphate and potash mining. UNEP, Paris
- Bao Y, Wang Y, Hu M, Wang Q (2018) Phosphorus fractions and its summer flux from sediments of deep reservoirs located at a phosphate-rock watershed, Central China. *Water Sci Technol* 18:688–697
- Boyacioglu H (2007) Surface water quality assessment by environmental methods. *Environ Monit Assess* 131:371–376
- Brereton RG (2003) Chemometrics: data analysis for the laboratory and chemical plant. Wiley, Chichester
- China E (2002) Environmental quality standards for surface water (GB3838-2002). MEP, Beijing
- Chraïti R, Raddaoui M, Hafiane A (2016) Effluent water quality at phosphate mines in M'Dhillia, Tunisia and its potential environmental effects. *Mine Water Environ* 35:462–468
- Das B (1999) Environmental pollution of Udaisagar lake and impact of phosphate mine, Udaipur, Rajasthan, India. *Environ Geol* 38:244–248
- Dixon W, Chiswell B (1996) Review of aquatic monitoring program design. *Water Res* 30:1935–1948
- Fakayode SO (2005) Impact of industrial effluents on water quality of the receiving Alaro River in Ibadan, Nigeria. *Afr J Environ Assess Manag* 10:1–13
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Eugen* 7:179–188
- Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, Juahir H (2012) Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Mar Pollut Bull* 64:688–698
- Gu Q et al (2016) Assessment of reservoir water quality using multivariate statistical techniques: a case study of Qiandao Lake, China. *Sustain Basel* 8:243
- Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L (2000) Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res* 34:807–816
- Horel JD (1981) A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Mon Weather Rev* 109:2080–2092
- Jiries A, El-Hasan T, Al-Hweiti M, Seiler K-P (2004) Evaluation of the effluent water quality produced at phosphate mines in central Jordan. *Mine Water Environ* 23:133–137
- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Kannel PR, Lee S, Kanel SR, Khan SP (2007) Chemometric application in classification and assessment of monitoring locations of an urban river system. *Anal Chim Acta* 582:390–399
- Kothai P, Saradhi I, Prathibha P, Hopke PK, Pandit G, Puranik V (2008) Source apportionment of coarse and fine particulate matter at Navi Mumbai, India. *Aerosol Air Qual Res* 8:423–436
- Kowalkowski T, Zbytniewski R, Szpejna J, Buszewski B (2006) Application of chemometrics in river water classification. *Water Res* 40:744–752
- Lewelling B, Wylie R (1993) Hydrology and water quality of unmined and reclaimed basins in phosphate-mining areas, west-central Florida. *Water Resour Investig Rep* 93:4002
- Lopes FB, Andrade EM, Meireles A, Becker H, Batista AA (2014) Assessment of the water quality in a large reservoir in semiarid region of Brazil. *Rev Bras Eng Agr Amb* 18:437–445
- Nalini A, Mohamed AF, Chiew EF, Mohamed Z (2010) Supply chain analysis of fresh fruits and vegetables (FFV): prospects of contract farming. *Agric Econ* 56:435–442
- Norušis MJ (1986) SPSS/PC+ for the IBM PC/XT/AT, vol 1. SPSS, Chicago
- Papaioannou A, Dovriki E, Rigas N, Plageras P, Rigas I, Kokkora M, Papastergiou P (2010) Assessment and modelling of groundwater quality data by environmental methods in the context of public health. *Water Resour Manag* 24:3257–3278
- Paul S, Srinivasan R, Sanabria J, Haan PK, Mukhtar S, Neimann K (2006) Groupwise modeling study of bacterially impaired watersheds in Texas: clustering analysis 1. *J Am Water Resour Assoc* 42:1017–1031
- Qu J, Meng X, Ye X, You H (2016) Characteristic variation and original analysis of emergent water source pollution accidents in China between 1985 and 2013. *Environ Sci Pollut Res* 23:19675–19685
- Quirós R (2003) The relationship between nitrate and ammonia concentrations in the pelagic zone of lakes. *Limnetica* 22:37–50
- Simeonov V, Stefanov S, Tsakovski S (2000) Environmental treatment of water quality survey data from Yantra River, Bulgaria. *Microchim Acta* 134:15–21
- Singh KP, Malik A, Sinha S (2005) Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Anal Chim Acta* 538:355–374
- Tinsley HE, Brown SD (2000) Handbook of applied multivariate statistics and mathematical modeling. Academic Press, New York
- Uyanik A, Tinkiliç N, Odabaşoğlu M, Karaca H (1999) Spectrophotometric determination of uranium in waste water of phosphoric acid and fertilizer manufacturing process. *Turk J Chem* 23:275–284
- Vega M, Pardo R, Barrado E, Debán L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res* 32:3581–3592
- Wang K, Lin Z, Zhang R (2016) Impact of phosphate mining and separation of mined materials on the hydrology and water environment of the Huangbai River basin, China. *Sci Total Environ* 543:347–356
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Xiangcan J (2003) Analysis of eutrophication state and trend for lakes in China. *J Limnol* 62:60–66
- Yang W-Q, Xiao H, Li Y, Miao D-R (2018) Vertical distribution and release characteristics of phosphorus forms in the sediments from the river inflow area of Dianchi Lake, China. *Chem Speciat Bioavailab* 30:14–22
- Zhu X, Li D, He D, Wang J, Ma D, Li F (2010) A remote wireless system for water quality online monitoring in intensive fish culture. *Comput Electron Agric* 71:S3–S9