



Water quality assessment and pollution source apportionment using multivariate statistical techniques: a case study of the Laixi River Basin, China

Jie Xiao · Dongdong Gao · Han Zhang ·
Hongle Shi · Qiang Chen · Hongfei Li ·
Xingnian Ren · Qingsong Chen

Received: 19 September 2022 / Accepted: 15 December 2022 / Published online: 10 January 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract Identifying potential sources of pollution in tributaries and determining their contribution rates are critical to the treatment of water pollution in main streams. In this paper, we conducted a multivariate statistical analysis on the water quality data of 12 parameters for 3 years (2018–2020) at six sampling sites in the Laixi River to qualitatively identify potential pollution sources and quantitatively calculate the contribution rates to reveal the tributaries' pollution status. Spatio-temporal cluster analysis (CA) divided 12 months into two parts, corresponding to the lightly polluted season (LPS) and highly polluted season (HPS), and six sampling sites were divided into two regions, corresponding to the lightly polluted region (LPR) and highly polluted region (HPR). Principal component analysis (PCA) was used to determine the potential sources of contamination, identifying four and three potential factors in the LPS and HPS, respectively. The absolute principal component score-multiple linear regression (APCS-MLR) receptor model quantitatively analyzed the

contribution rates of identified pollution sources, and the importance of the different pollution sources in LPS can be ranked as domestic sewage and industrial wastewater and breeding pollution (33.80%) > soil weathering (29.02%) > agricultural activities (20.95%) > natural influence (13.03%). HPS can be classified as agricultural cultivation (41.23%), domestic sewage and industrial wastewater and animal waste (33.19%), and natural variations (21.43%). Four potential sources were identified in LPR ranked as rural domestic sewage (31.01%) > agricultural pollution (26.82%) > industrial effluents and free-range livestock and poultry pollution (25.13%) > natural influence (14.82%). Three identified latent pollution sources in HPR were municipal sewage and industrial effluents (37.96%) > agricultural nonpoint sources and livestock and poultry wastewater (33.55%) > natural sources (25.23%). Using multivariate statistical tools to identify and quantify potential pollution sources, managers may be able to enhance water quality in tributary watersheds and develop future management plans.

J. Xiao · D. Gao (✉) · H. Shi · Q. Chen · Q. Chen
Sichuan Academy of Ecological and Environmental
Science, Chengdu 610041, China
e-mail: hydrogeochemistry@126.com

H. Zhang · X. Ren
Faculty of Geosciences and Environmental Engineering,
Southwest Jiaotong University, Chengdu 610031, China

H. Li
Administrative Committee of Sichuan Tianquan Economic
Development Zone, Ya'an 625000, China

Keywords Surface water quality · Source apportionment · Multivariate statistical techniques · APCS-MLR

Introduction

River water is an important part of water resources and is indispensable in ensuring residents' production and

living water, healthy development of society and economy, and a harmonious balance of the ecological environment (Herojeet et al., 2017; Ma et al., 2021). However, with the diversification and rapid development of industries and the continuous acceleration of urbanization, water quality degradation has become a major worldwide issue, causing environmental problems such as eutrophication of water bodies and excessive heavy metals, which seriously restrict the sustainable development of society (Nong et al., 2020; Piroozfar et al., 2021). Moreover, water quality is not only affected by natural conditions such as climate, geographical environment, and land use types but also by human factors such as industrial pollution, agricultural activities, and domestic sewage (Zhang et al., 2020, 2022). Therefore, understanding the temporal and spatial variations in water quality and accurately identifying the sources of pollution in river water has become a precondition for improving the quality of the water environment.

Because of the complexity of the monitored values of water quality indicators, it is difficult to explain the spatial-temporal variation characteristics of water quality and identify potential pollution sources (Li et al., 2020). Multivariate statistical analysis technology has unique advantages in explaining complicated datasets and is widely applied in water pollution source identification, which mainly includes cluster analysis (CA), factor analysis (FA), and PCA (de Oliveira et al., 2020; Han et al., 2019; Liu et al., 2019; Mir & Gani, 2019; Varol, 2020). In addition, according to the qualitative recognition results of PCA, the receptor model of absolute principal component score-multiple linear regression (APCS-MLR) was established, which has been widely used to quantitatively analyze the contribution rates of latent pollution sources (Cheng et al., 2020; Cho et al., 2022; Yu et al., 2022; Zhang et al., 2022). By combining PCA/FA with the APCS-MLR model, the ability to reveal potential sources of pollution in river water can be improved and provide a more reliable scientific basis for water environment pollution control (Cho et al., 2022).

The Laixi River is located in southwestern China and is an important tributary of the Tuojiang River, which is a first-class tributary of the Yangtze River (Liu et al., 2022). The Yangtze River is an important ecological security barrier central to China's economic activities. With societal development, it is inevitably affected by rapid urbanization, industrial and agricultural development, and high population density,

and the increasing environmental pressure has resulted in severe river water pollution (Cheng et al., 2020; Liu et al., 2022). In recent years, China's government has issued several water pollution control policies (Fu et al., 2020). Based on these policies, in-depth studies have been conducted on the analysis of pollution sources in mainstream rivers, but there are relatively few studies on small watersheds (Chen et al., 2019; Cheng et al., 2020; Haji Gholizadeh et al., 2016; Huang et al., 2010; Kellner et al., 2018; Muangthong & Shrestha, 2015). As the primary pollution source of the mainstream, tributaries have problems, such as comprehensive pollution sources, serious pollutants exceeding the standard, and difficult governance (Yang et al., 2022; Zhang et al., 2020). Therefore, understanding the water quality contamination status of tributaries, combined with source identification techniques for pollution control, is of great significance for mainstream pollution management.

Given the above considerations, this study aimed to combine multivariate statistical methods to better understand potential water pollution sources and realize a quantitative assessment of the apportionment of water pollution sources in the small watershed of the Laixi River. Accordingly, this study (1) explored the spatiotemporal variation characteristics of water quality in the Laixi River Basin; (2) identified the major water quality deterioration parameters; (3) revealed the latent spatiotemporal sources of pollution and quantified their contribution ratios. These results provide a scientific basis for the formulation of effective water quality management and mainstream pollution control policies.

Material and methodology

Study area

The Laixi River, which originates in the Dazu District, Chongqing City, is a major tributary of the lower reaches of the Tuojiang River. The total basin area is 3257 km², and the mainstream length is 238 km. The study area (105°14'57"–105°41'51" E, 28°59'56"–29°20'3" N) was located in Luzhou City, with a river length of 77 km and an area of 814 km² (<http://www.gscloud.cn/>). It belongs to the lower Lacey River, with an average annual flow of 36 m³/s. The Jiuqu and Maxi Rivers are two of the

largest tributaries of the Laixi River in the Luzhou section (Fig. 1). The Jiuqu River has an area of 155 km², a length of 31 km, and an annual average flow of 10 m³/s. The Maxi River has an area of 292 km², a length of 41 km, and an annual average flow of 2.5 m³/s. The climate is classified as subtropical humid. The annual precipitation is extremely uneven, with an average of 1153.7 mm falling primarily between May and September, accounting for approximately 76% of total precipitation. The average annual temperature is 18 °C, with the lowest being -1.6 °C and the highest at 41.3 °C. The study area is dominated by hilly terrain and inclines towards the southwest, with elevations of 218–757.5 m. The land use is primarily cultivated land, accounting for 58.3% of the total, with construction land accounting for less and distributed mainly on both sides of the river.

The Laixi River provides water for the development of industry, agriculture, and the lives of residents in the area through which it flows. However, recently, the Laixi River has been increasingly

polluted, which is not only detrimental to the local development of Luzhou but also has a serious impact on the water environment of the Tuojiang River basin. Water pollution in the Laixi River Basin has a long history, and the water quality has not improved significantly. In 2019, the discharge of chemical oxygen demand (COD_{Cr}), ammonia nitrogen (NH₄⁺-N), and total phosphorus (TP) in the Laixi River reached 5309.1 tons, 675.56 tons, and 330.4 tons, respectively (LEEB, 2021). From 2018 to 2020, concentrations of COD_{Cr}, permanganate index (COD_{Mn}), and TP were the most serious water quality parameters exceeding the standard, with 38.5%, 56.3%, and 19.7% of water quality grades lower than the III of Water Environment Quality Standard (MEPC, 2002). In addition, fluoride ions (F⁻), NH₄⁺-N, dissolved oxygen (DO), and other parameters also exceed the standard to varying degrees. Furthermore, the Laixi River receives various sewage and surface runoff from both sides of the river, and the health of the water ecology has aroused widespread concern.

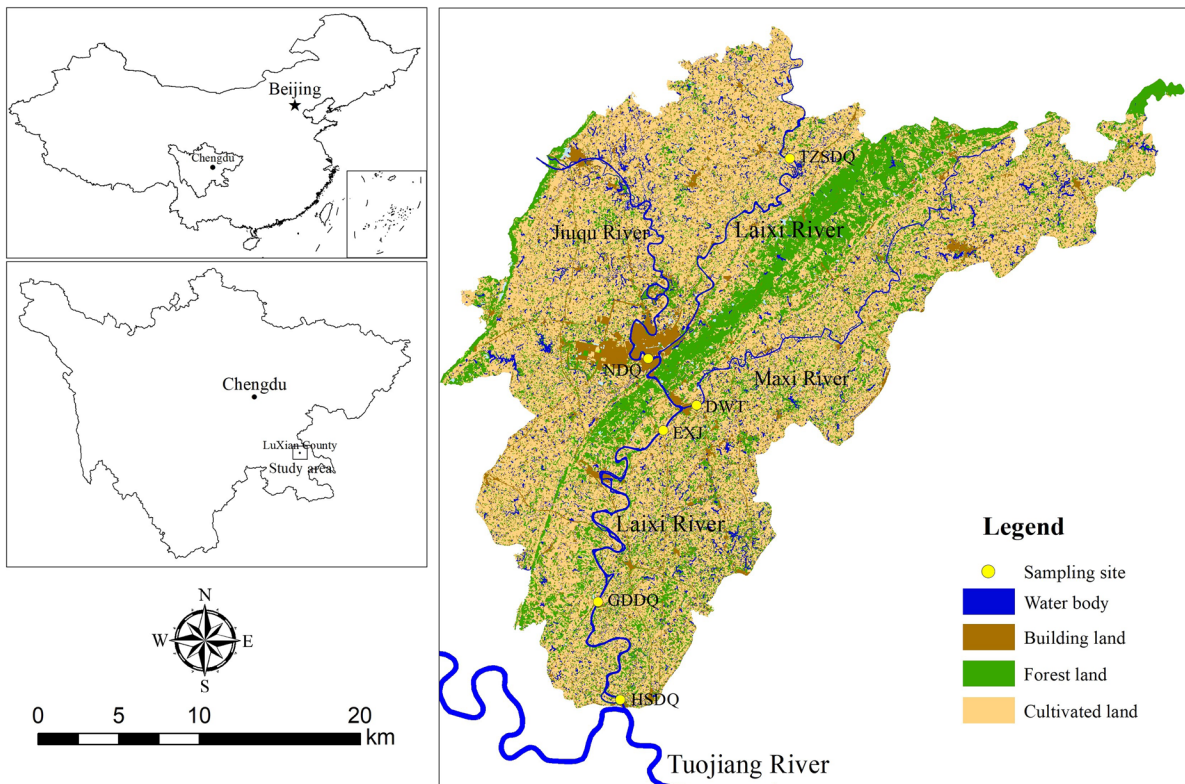


Fig. 1 Map of study area and surface water quality sampling sites in the Laixi River basin, China

Sampling collection and analysis

Monitoring datasets from six sampling sites (Fig. 1), comprising 12 water quality indicators monitored monthly from January 2018 to December 2020, were obtained from the Bureau of Ecology and Environment of Luzhou. Among them, the TZSDQ is at the entrance of the study area; sites EXJ, GDDQ, and HSDQ are distributed downstream; and sites NDQ and DWT are the monitoring points on the tributaries. Although there are a total of 26 monitoring indicators in the monthly water quality evaluation procedure, we selected 12 important parameters because the concentration of monitoring values of some indicators is too low or below the detection limit. The selected water quality parameters included 5-day biochemical oxygen demand (BOD₅), potassium permanganate index (COD_{Mn}), chemical oxygen demand (COD_{Cr}), water temperature (WT), dissolved oxygen (DO), ammonia nitrogen (NH₄⁺-N), hydrogen ion concentration index (pH), electrical conductivity (EC), fluoride (F), anionic surfactant (AS), total nitrogen (TN), and total phosphorus (TP). The descriptive statistics of the river water quality monitoring data sets of each site and month are shown in Table 1, including the mean value and standard deviation. The analytical methods of water quality parameters in this study were carried out in accordance with the instructions in the technical specification requirements for surface water and wastewater monitoring (MEEC, 2002).

To make the data suitable for cluster analysis and principal component analysis, we (a) detected and processed outliers, (b) used statistical methods to supplement missing values with the mean values of the corresponding data groups, and (c) combined abundance, skewness analysis, and the K-S test for normality of the dataset (Katsaounis, 2004). In addition, we standardized the data to eliminate the influence of dimensionality on the mathematical analysis. All data analyses in this study were processed using Excel 2010 and SPSS26.0 software.

Multivariate statistical methods

Cluster analysis

Cluster analysis (CA) is commonly used to categorize complicated datasets and organize items into clusters based on similarities within classes and differences across classes (Herojeet et al., 2017). The

most extensively used clustering approach is hierarchical clustering, typically used in conjunction with PCA to analyze surface water quality (Rezaei et al., 2019). In this study, we use the square of Euclidean distance as a similarity measure to measure the distance between clusters, and Ward's method was used to perform cluster analysis on standardized datasets to minimize the sum of squares of adjacent clusters that could be formed in each step. Clustering results are usually represented by dendrograms, which can more intuitively illustrate the similarity between clusters and help improve the efficiency of the spatiotemporal analysis of water quality (Pinto et al., 2019).

Principal component analysis

PCA is a dimension-reduction technique that is commonly used to reduce original variables into a few unrelated new variables. The unrelated new variables, called principal components, represent most of the original data (Bonansea et al., 2015). Before principal component analysis, it is necessary to determine whether the dataset meets the prerequisite conditions. Before statistical analysis, we can test the correlation between variables through Bartlett's sphericity test, the KMO test, and other methods to determine the applicability of PCA (Kaiser, 1974; Zhang et al., 2010). In general, the KMO test value is greater than 0.5 and the significance of Bartlett's sphericity test is less than 0.05, indicating that PCA is applicable (Li et al., 2020). In this study, the maximum variance method was used for rotation analysis, which could maximize the sum of the squares of loads of each component so that the principal components could explain more variables centrally (Liu et al., 2019). Components with eigenvalues greater than one were defined as the principal components. Generally, the higher the factor load, the greater the influence of the water quality parameter on the corresponding principal component. Factor loadings from 0.3–0.5, 0.5–0.75, and greater than 0.75 were defined as weak, medium, and strong, respectively (Liu et al., 2003). PCA was applied to detect the main influencing variables of each group in the cluster analysis results, and a qualitative analysis of surface water pollution sources in the Laixi River Basin was conducted. The data used for PCA analysis were the measured values of the selected parameters for a total of 36 months over three years.

Table 1 Mean and standard deviation (SD) of physicochemical variables of water quality in the Laixi River Basin

		WT	pH	EC	DO	BOD ₅	COD _{Mn}	COD _{Cr}	NH ₄ ⁺ -N	TN	TP	F ⁻	AS
Season													
Jan	Mean	11.44	7.90	47.43	7.70	2.45	5.99	24.94	0.69	3.97	0.18	0.77	0.06
	SD	0.72	0.22	9.50	1.19	1.16	1.82	6.84	0.71	1.43	0.08	0.22	0.02
Feb	Mean	10.95	7.74	50.46	8.37	2.71	5.94	24.69	0.60	4.14	0.16	0.79	0.06
	SD	1.60	0.16	9.90	1.63	1.18	0.81	5.38	0.50	1.04	0.07	0.27	0.03
Mar	Mean	14.70	8.20	48.67	10.18	3.48	6.53	28.06	0.68	4.93	0.18	0.83	0.10
	SD	2.90	0.55	18.14	3.20	1.12	1.20	9.28	0.73	1.98	0.10	0.24	0.07
Apr	Mean	19.90	7.96	57.43	8.34	3.08	6.59	28.41	0.71	4.16	0.17	0.95	0.07
	SD	1.88	0.36	11.86	2.39	1.20	1.79	10.86	0.82	1.60	0.16	0.41	0.03
May	Mean	24.78	7.63	51.41	7.41	4.12	7.59	31.18	0.75	3.87	0.19	0.83	0.07
	SD	2.51	0.48	10.50	5.69	1.15	1.65	7.51	0.73	1.14	0.08	0.21	0.04
Jun	Mean	24.91	7.54	44.29	6.57	3.26	6.98	24.71	0.35	3.12	0.17	0.70	0.06
	SD	1.22	0.21	17.57	2.14	0.95	1.28	5.03	0.35	0.66	0.06	0.21	0.04
Jul	Mean	26.69	7.60	41.36	6.41	2.96	6.16	21.89	0.35	2.70	0.19	0.61	0.05
	SD	1.85	0.20	9.37	1.67	1.11	0.67	3.75	0.28	0.85	0.05	0.13	0.02
Aug	Mean	29.60	7.96	32.78	6.11	2.48	5.41	17.67	0.17	2.46	0.18	0.53	0.05
	SD	1.02	0.35	7.50	1.03	1.01	0.69	3.25	0.12	0.50	0.03	0.12	0.01
Sept	Mean	29.88	7.73	41.76	5.14	2.80	5.82	19.42	0.18	2.13	0.16	0.59	0.05
	SD	2.59	0.33	7.13	1.51	0.77	1.03	4.66	0.14	0.74	0.03	0.13	0.01
Oct	Mean	21.11	7.69	40.10	7.10	2.28	5.80	21.47	0.32	2.35	0.17	0.67	0.05
	SD	1.93	0.29	8.72	1.56	1.20	0.65	5.43	0.30	0.68	0.04	0.24	0.01
Nov	Mean	18.83	7.77	41.08	6.96	2.46	5.17	20.72	0.20	2.50	0.15	0.63	0.05
	SD	0.66	0.19	8.18	0.98	1.22	0.91	4.75	0.14	0.73	0.04	0.16	0.01
Dec	Mean	14.94	7.77	43.11	7.35	2.22	5.24	19.74	0.22	2.82	0.14	0.70	0.05
	SD	0.72	0.17	13.99	0.88	1.19	0.64	3.59	0.19	0.58	0.03	0.27	0.01
Station													
NDQ	Mean	21.07	7.75	45.32	6.51	3.98	7.53	29.00	0.94	4.06	0.25	1.06	0.07
	SD	6.96	0.36	13.06	3.60	0.50	1.53	9.46	0.90	1.69	0.12	0.28	0.04
DWT	Mean	20.43	7.76	45.48	6.94	3.85	7.72	30.86	0.87	4.49	0.22	0.82	0.08
	SD	6.29	0.29	12.36	1.24	0.48	1.33	8.37	0.90	2.03	0.07	0.16	0.04
HSDQ	Mean	20.41	7.78	45.49	6.68	3.51	6.21	27.80	0.51	2.39	0.15	0.67	0.08
	SD	6.63	0.32	16.06	2.65	0.54	1.50	7.54	0.54	1.09	0.08	0.23	0.04
EXJ	Mean	20.64	7.85	48.06	8.18	3.98	5.27	19.88	0.17	2.94	0.13	0.65	0.04
	SD	6.01	0.53	12.20	2.35	3.11	0.91	4.47	0.14	0.78	0.03	0.14	0.03
GDDQ	Mean	21.32	7.79	39.84	7.33	4.21	6.22	23.94	0.39	3.42	0.17	0.70	0.06
	SD	6.61	0.34	9.74	2.76	1.62	0.91	5.79	0.33	1.22	0.03	0.14	0.03
TZSDQ	Mean	21.08	7.78	37.94	7.30	2.50	5.69	19.75	0.29	3.26	0.16	0.67	0.07
	SD	6.86	0.30	12.30	1.76	1.02	0.99	5.15	0.26	1.26	0.04	0.18	0.02

Units: mg/L, except pH dimensionless, EC ms/m, WT °C

APCS-MLR model

Regression analysis is a widely used quantitative analysis method to analyze the correlation between variables and establish regression equations to simulate

and predict the corresponding variables (Cheng et al., 2020; Ding et al., 2016; Uddin et al., 2022). This study established a multiple linear regression model between the absolute principal component APCS (independent variable) and water quality parameter

concentration (dependent variable) to quantitatively analyze the contribution of each pollution source to the monitoring index. The concentration contribution of each pollution source to all pollutants (C_j) can be expressed as

$$C_j = b_j + \sum_{h=1}^n r_{hj} \times APCS_{hj}$$

where b_j indicates a multiple regression constant for pollutant j , r_{hj} represents the multiple regression coefficients of source h for j , $APCS_{hj}$ is the absolute principal component score of identified pollution source for each sample, $r_{hj} \times APCS_{hj}$ is the contribution of source h to C_j .

In the receptor model analysis, we introduced an absolute zero concentration of the sample. Then, using the score coefficient matrix obtained in the PCA process as the weight value, combined with the absolute zero concentration and standardized concentration of the water quality index, the APCS value of each sample was obtained through mathematical calculations (Wang et al., 2022). For more details on the APCS-MLR model analysis, please refer to Thurston and Spengler (1984).

The linear regression process often has a negative linear regression coefficient and APCS value, which makes the contribution rate tend to have negative inaccuracies in the interpretation of the contribution of pollution sources easily created if the negative numbers are not rectified, which reduces the precision

of the pollution source analysis (Liu et al., 2020). Therefore, this study took the absolute value of the negative contribution rate to optimize our interpretation of the regression results (Haji Gholizadeh et al., 2016; Zhang et al., 2020).

Results and discussion

Spatiotemporal variations of water quality parameters

In order to show the change rule of water quality in the study area more intuitively, considering that the water pollution types of rivers may change in 3 years, we applied the independent sample t -test method to analyze the significant change in water quality in 2018–2020 and judged whether the water quality difference was significant in 3 years. The independent samples t -test (Table 2) showed no significant differences in most of the water quality indicators between the years 2018–2020 ($\text{sig} > 0.05$). The descriptive statistics only found that the $\text{NH}_4^+\text{-N}$ had too high anomalous values, with a variability of 123.5%. Also, the LAS had 12.5% of the monitored samples below the minimum detection. These may have contributed to the low significance of the independent sample t -test (Table 2). In addition, the significance of the tests was greater than 0.05 for most indicators, indicating that the water quality conditions were close between the 3 years. Therefore, we can use the 3-year average to assess the overall water quality status.

Table 2 Results of independent samples t -test

Indicators	T-test value		Sig.		CV	BD
	2018/2019	2019/2020	2018/2019	2019/2020		
WT	0.584	1.007	0.560	0.316	31.52%	0%
pH	2.148	1.443	0.033	0.151	4.70%	0%
EC	0.909	2.429	0.365	0.116	29.39%	0%
DO	0.478	2.200	0.633	0.630	36.78%	0%
BOD ₅	0.662	-0.808	0.509	0.420	43.32%	0%
COD _{Mn}	2.155	-0.943	0.053	0.348	22.48%	0%
COD _{Cr}	1.421	-1.198	0.158	0.233	31.70%	0%
NH ₄ ⁺ -N	2.210	-2.658	0.029	0.009	124.89%	0%
TN	1.485	-3.958	0.140	0.231	43.21%	0%
TP	1.487	-1.549	0.139	0.124	44.14%	0%
F	-0.241	-0.807	0.810	0.421	36.00%	0%
LAS	3.113	-2.290	0.002	0.024	56.23%	12.5%

Bolded: Significance less than 0.05

CV variable coefficient, BD the proportion of concentrations below the detection limit

The descriptive statistical calculation summary of the monthly average concentration and sampling site concentration of 12 water quality indicators is shown in Table 1. In general, the concentration levels of COD_{Cr} and TN were relatively high in both time and space compared to national standards (MEPC, 2002). Table 1 shows that COD_{Cr} and TN were the most serious indicators of pollution each month. The 12-month average COD_{Cr} concentration range is 17.67–31.18 mg/L, most of the months were around the class IV Water Environment Quality Standards (30 mg/L), and the highest mean COD_{Cr} (31.18 mg/L) appeared in May (MEPC, 2002). TN pollution was even more severe; the concentration each month exceeded the class V of water quality standard (2.0 mg/L), with the highest average value of 4.93 mg/L, which appeared in March. In addition, COD_{Mn} was slightly higher than class III standards (6 mg/L) from March to July. Except for COD_{Cr} , COD_{Mn} , and TN, other indicators met the class III standard each month.

As indicated by the sites, the water quality variables TN, COD_{Cr} , COD_{Mn} , BOD_5 , TP, NH_4^+-N , and F differed significantly between the sites. The spatial variation of these parameters may be due to the influence of different levels of urbanization, the intensity of anthropogenic activities, and industrial distribution (Zhang et al., 2022). Indicators representing the concentration of pollutants, such as COD_{Cr} , COD_{Mn} , NH_4^+-N , TN, TP, and F, have the highest concentrations in the NDQ site, showing the most serious pollution. The remaining sites showed small differences in the concentrations of pollution indicators and exhibited a mild to severe pollution status. The average concentration of TN was 2.39–4.06 mg/L in 6 sites, higher than the class V water quality standard (2.0 mg/L). In addition, the concentrations of the other indicators were within the range of class III to class IV water quality standards, with relatively mild pollution.

Temporal cluster analysis

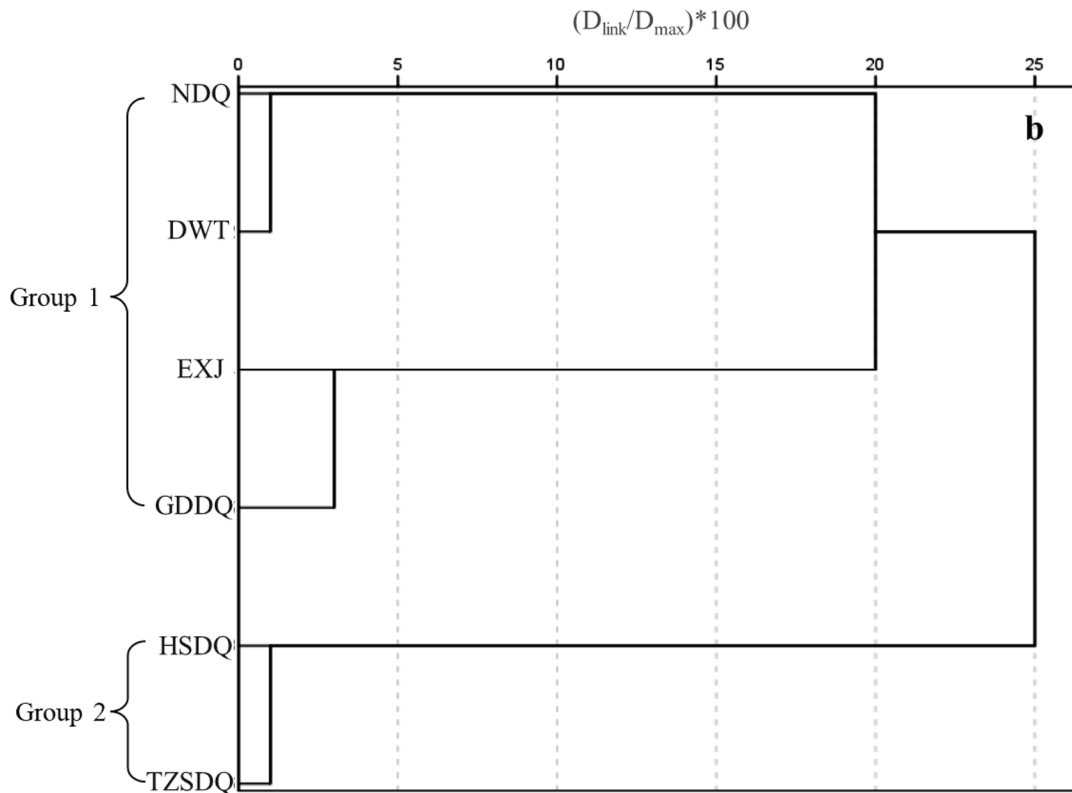
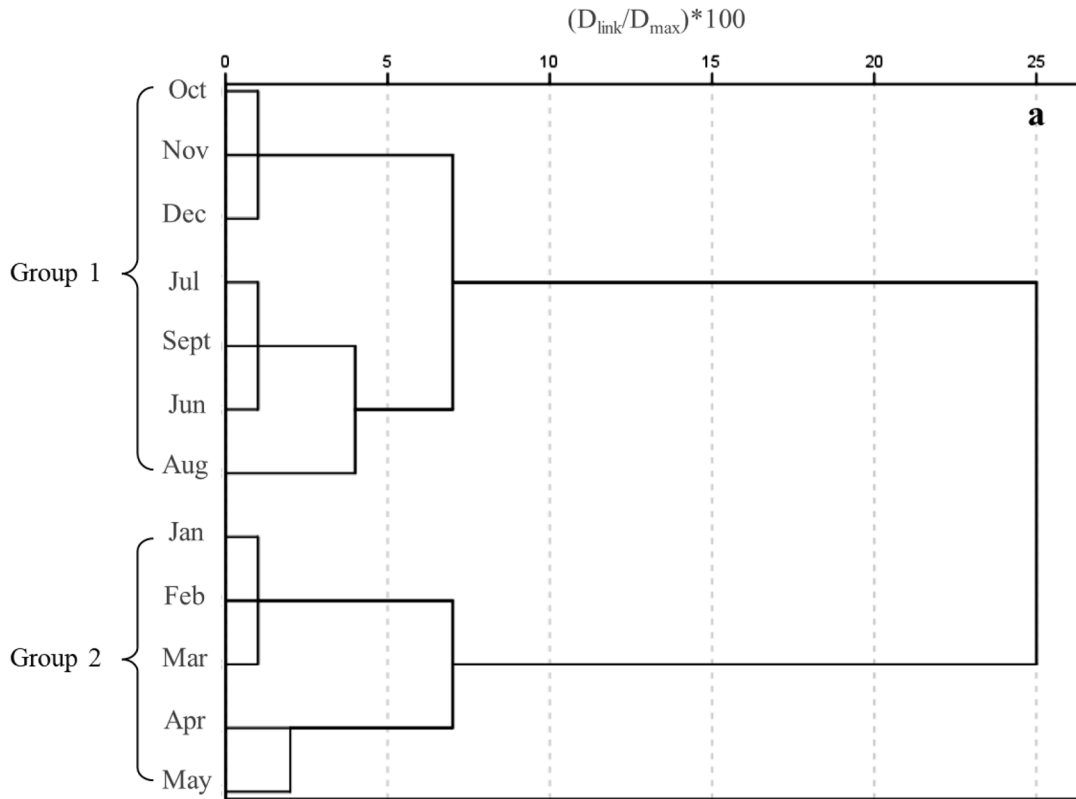
The time-varying characteristics of water quality were discussed in more depth using cluster analysis, and the clustering analysis used monthly averages of 3 years of observed data. According to the variations in the monitoring months, the 12-month average data for 3 years (2018–2020) were classified as two clusters with significant distance differences at $(D_{\text{link}}/D_{\text{max}}) \times 100 < 10$ (Fig. 2(a)). The two groups correspond to two phases with

different levels of pollution in the study area. Group 1 contained seven months from June to December, representing the LPS, accounting for 65–75% of annual rainfall. Group 2 included the remaining months, from January to May, corresponding to the HPS, during which the rainfall was about 25–35% of the annual rainfall. The flow of group 1 was significantly higher than that of group 2.

From the statistical values of parameters in Table 1, it can be seen that COD_{Cr} and TN were the most polluted indicators in the study area. TP is a representative pollution indicator that can indicate various pollution sources. DO is usually regarded as an important parameter for the self-purification ability of rivers. Therefore, four indicators, COD_{Cr} , TN, TP, and DO, were selected for a more in-depth analysis of the temporal trends in river water quality, and the results are shown in Fig. 3. The concentration variation trends of the four selected indicators remained consistent, and the concentration variation range and mean value in HPS were higher than those in LPS. The more serious HPS pollution might be because it corresponds to the dry and planting seasons. Simultaneously, chemical fertilizers and pesticides are used extensively during sowing in spring, flowing into the river through surface runoff to intensify pollution, which is also the main route for phosphorus pollutants and organics to migrate from soil to water systems (Varol, 2020; Verheyen et al., 2015; Zhou & Gao, 2011). DO showed minimum values in HPS, mainly because the DO value was inversely proportional to the concentration of organic pollutants during their degradation (Liu et al., 2020). In addition, the variation in DO concentration complies with the natural law that warmer water can hold less DO (Wang et al., 2013). As shown in Fig. 3, there are some outliers in different seasons, which might be caused by significant differences in the spatial distribution of pollutant concentrations at each sampling site.

Spatial cluster analysis

Spatial CA was used to assess the differences in water quality between different sites across the region. The six monitoring sites were clustered into two statistically significant clusters (groups 1 and 2) at $(D_{\text{link}}/D_{\text{max}}) \times 100 < 25$ (Fig. 2(b)). These groups were identified by judging their water quality, which is largely influenced by land use structure. The sites in each group had similar characteristics and natural



◀**Fig. 2** Cluster analysis dendrogram showing the grouping of sampling months (a) and sampling sites (b) of the study area

background sources. In addition, COD_{Cr} , TN, TP, and DO were selected to investigate the spatial variation differences in water quality between the two clusters, and the corresponding results are shown in Fig. 3.

Group 1 included four sites: NDQ, DWT, EXJ, and GDDQ, which correspond to the highly polluted region (HPR). The four sites are located in the middle of the study area and are mainly surrounded by buildings and cultivated lands (Fig. 1), indicating that the water quality in this area is mainly polluted by domestic sewage, industrial wastewater, and agricultural nonpoint

sources. The remaining sites (HSDQ and TZSDQ) belong to group 2, which corresponds to the LPR. The two sites were situated at the beginning and end of the study area and covered vastly cultivated and forest lands. Besides, HSDQ and TZSDQ were in a lightly polluted state, reflecting the minor impact of agricultural planting and the dilution effect of river water. The variation in the boxplot (Fig. 3) also confirmed this result: the mean value and concentration range of COD_{Cr} , TN, and TP were larger in the HPR, whereas the mean value of DO was smaller and there were more outliers, indicating that group 1 was more seriously polluted. As shown in the boxplot (Fig. 3), outliers indicate extreme values in some months, which may be influenced by periodic rainfall or anthropogenic activities.

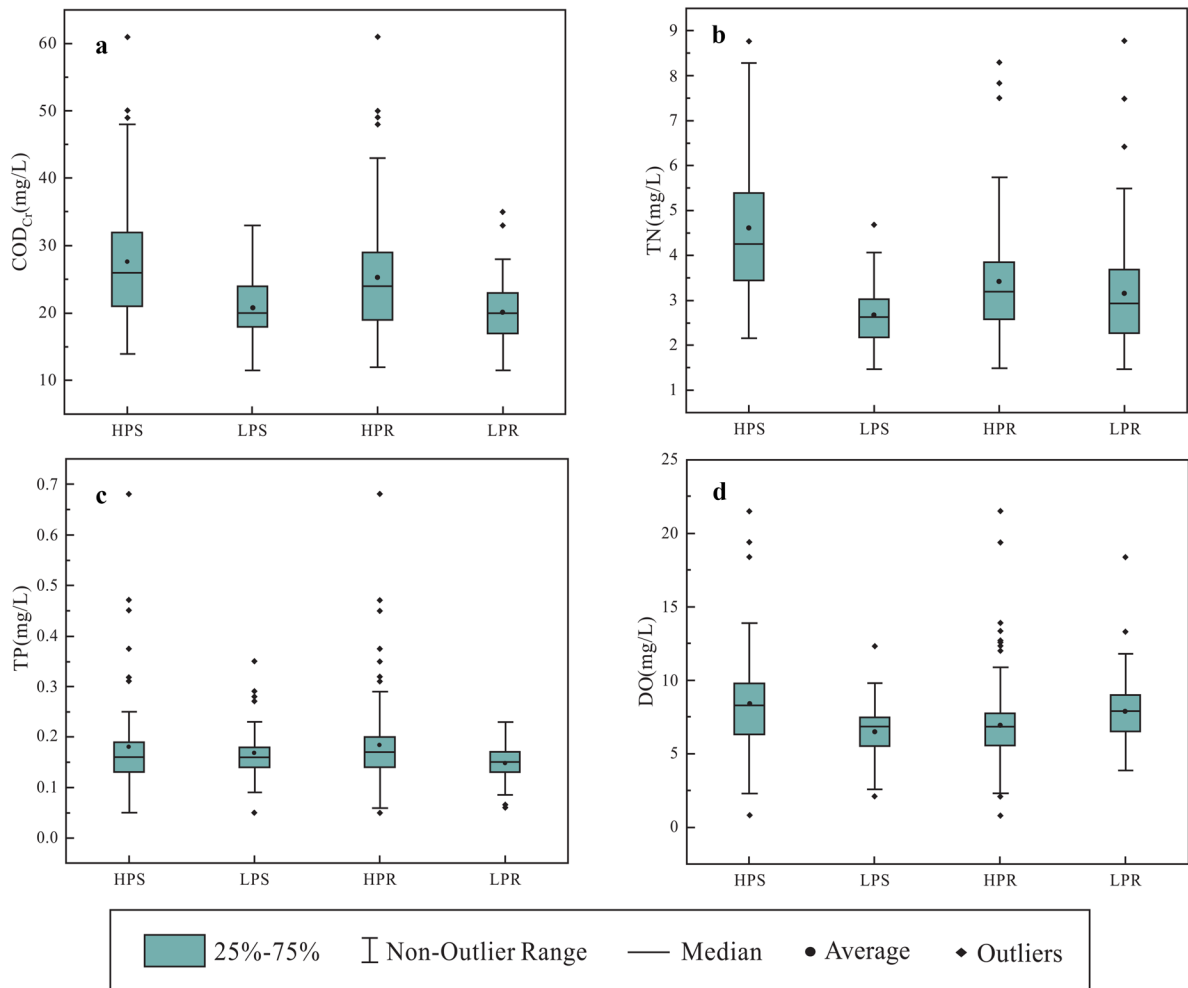


Fig. 3 Temporal and spatial variations of COD_{Cr} , TN, TP, and DO in different groups

Source identification

Source identification in a temporal pattern with PCA

PCA was performed on the temporal datasets to identify potential sources of contamination during the LPS and HPS periods. The KMO values for LPS and HPS were 0.675 and 0.732, respectively, and Bartlett’s sphericity test values were 0.00 (Sig.), indicating that the variables are strongly correlated and the correlation coefficient matrix is not significantly different from the identity matrix, which meets the conditions of PCA analysis (Haji Gholizadeh et al., 2016). Based on the Kaiser rule, four and three principal components with eigenvalues greater than 1 were extracted in LRP and HRP, respectively, totaling 63.87 and 74.65% of the variance in the original monitoring data, respectively. Table 3 shows the PCA results for different regions, including loads of each water quality parameter on each PC, the eigenvalues and variances of extracted PCS, and the total variance variables explained cumulatively.

For LPS, PC1 (31.35% of the total variance) had strong positive loadings on TN and NH₄⁺-N (0.78 and 0.77) and a moderate positive loading on AS (0.61). PC1 is associated with nutrient pollutants (N), which may originate from point-source pollution of sewage treatment plants and factories or nonpoint-source

pollution caused by agricultural cultivation and livestock breeding (Han et al., 2019; Matiatos, 2016; Zheng et al., 2015). In addition, AS has been used extensively in various applications, including domestic and industrial processes (Sasi et al., 2021). Thus, PC1 denotes the point sources of domestic and industrial wastewater and breeding pollution. PC2 had moderate positive loadings on COD_{Mn}, F⁻, COD_{Cr}, EC, and BOD₅ (0.71, 0.67, 0.64, 0.64, and 0.51) and explained 13.13% of the total variance. PC2 reflects the influence of water quality on organic pollutants resulting from anthropogenic activities such as the discharge of domestic sewage and industrial wastewater. Furthermore, this factor has a moderate correlation with F, usually observed in cement plants, mineral smelters, and certain chemical plants (Fu et al., 2020). But in fact, the F⁻ concentrations in all monitored months were below 1 mg/L, indicating either an absence or an extremely low level of contamination. This means that the F⁻ concentration probably originated from the local soil and entered rivers with rainfall runoff (Ma et al., 2020; Meng et al., 2018). Therefore, PC2 can be regarded as a type of mixed pollution influenced by domestic sewage, industrial wastewater, and soil weathering. PC3 and PC4 explained 11.27 and 8.12% of the total variance, respectively, with strong positive loadings on WT (0.90), medium positive loading on pH and TP (0.76 and 0.63), and medium negative

Table 3 Loadings of 12 variables on varimax rotated factors of different seasons in the Laixi River Basin

Parameters	LPS				HPS			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	
TN	0.78	0	-0.18	0.04	AS	0.97	0.1	0.03
NH ₄ ⁺ -N	0.77	0.3	0.06	-0.09	TP	0.92	0.31	-0.07
AS	0.61	0.07	0.14	-0.28	NH ₄ ⁺ -N	0.81	0.22	-0.36
COD _{Mn}	0.33	0.71	0.25	-0.15	EC	0.66	0.14	0.04
F ⁻	0.18	0.67	-0.08	-0.32	COD _{Cr}	0.05	0.87	0.2
COD _{Cr}	0.28	0.64	0.02	-0.28	COD _{Mn}	0.3	0.86	0.23
EC	-0.32	0.64	-0.23	0.33	F ⁻	0.1	0.86	0.36
BOD ₅	0.2	0.51	0.36	-0.49	DO	0.18	0.75	-0.16
WT	-0.02	-0.04	0.9	0.2	BOD ₅	0.23	0.7	0.24
DO	-0.06	-0.04	-0.69	0.31	TN	0.34	0.56	0.02
TP	0.36	0.38	0.63	0.13	pH	0.31	0.18	0.9
pH	-0.01	-0.19	0.02	0.76	WT	-0.01	0.39	0.69
Eigenvalue	3.76	1.58	1.35	1.03	Eigenvalue	4.49	1.29	1.01
Total variance %	31.35	13.13	11.27	8.12	Total variance %	49.89	14.34	10.43
Cumulative %	31.35	44.48	55.75	63.87	Cumulative %	49.89	64.23	74.65

Bolded: maximum PC loading for one parameter

Table 4 Loadings of 12 variables on varimax rotated factors of different regions in the Laixi River Basin

Parameters	LPR				HPR			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	
WT	-0.87	0.31	-0.06	-0.01	NH ₄ ⁺ -N	0.90	-0.07	0.10
EC	0.80	0.17	0.26	-0.03	TP	0.85	-0.15	-0.18
TN	0.76	0.24	0.02	0.28	COD _{Mn}	0.84	0.29	-0.10
BOD ₅	-0.06	0.84	0.12	0.04	COD _{Cr}	0.77	0.32	0.23
COD _{Cr}	0.32	0.76	-0.13	-0.03	BOD ₅	0.63	0.31	-0.07
COD _{Mn}	-0.13	0.71	0.18	0.32	EC	0.17	0.82	0.05
F ⁻	0.51	0.57	0.07	-0.38	TN	-0.25	0.72	0.41
pH	0.02	0.11	0.83	-0.17	F ⁻	0.16	0.64	0.15
DO	0.44	0.03	0.76	-0.14	AS	0.44	0.63	0.19
AS	0.04	0.01	0.71	0.47	pH	-0.03	-0.01	0.73
NH ₄ ⁺ -N	0.11	0.15	0.14	0.75	DO	-0.16	0.32	0.69
TP	0.01	-0.03	-0.36	0.73	WT	-0.18	0.12	-0.59
Eigenvalue	6.04	2.24	1.47	1.12	Eigenvalue	4.79	1.81	1.09
Total variance %	47.37	18.63	12.27	8.83	Total variance %	43.88	18.05	13.11
Cumulative %	47.37	66.00	78.27	87.10	Cumulative %	43.88	61.92	74.03

Bolded: maximum PC loading for one parameter

loading on DO (-0.69). It is generally believed that WT, pH, and DO are mainly affected by temperature changes and natural meteorology, and TP is likely attributed to agricultural activities (Giao et al., 2021; Zhou et al., 2007).

For HPS, the first PC (PC1), accounting for 49.89% of the total variance, had strong positive loadings on AS, TP, and NH₄⁺-N (0.97, 0.92, and 0.81) and moderate positive loading on EC (0.66). Temporally, TP and NH₄⁺-N during HPS were larger than during LPS, likely due to increased agricultural planting activities and surface runoff in spring (Zhang et al., 2020). Therefore, PC1 could be interpreted as agricultural cultivation. PC2 (accounting for 14.34% of the total variance) had strong positive loadings on COD_{Cr}, COD_{Mn}, and F⁻ (0.87, 0.86, and 0.86) and medium positive loadings on DO, BOD₅, and TN (0.75, 0.70, and 0.56). According to the LPS analysis, PC2 mainly represented the sources of domestic sewage, industrial wastewater, and animal waste. PC3 (10.43% of the total variance) had strong positive loading on pH (0.90) and moderate positive loading on WT (0.69), representing the influence of natural factors.

Through PCA, the pollution sources of LPS were identified as domestic sewage, industrial wastewater, and breeding pollution > soil weathering > agricultural activities > natural influence, according to the contribution rate. HPS can be ranked as agricultural

cultivation > domestic sewage, industrial wastewater, and animal waste > natural variations.

Source identification in a spatial pattern with PCA

PCA was also performed on the two spatial groups of the monitoring sites, similar to the temporal groups. The KMO values for LPR and HPR were 0.743 and 0.773, respectively, and Bartlett’s sphericity test values were 0.00 (Sig.). In these two different spatial groups, four and three principal components (PCs) were extracted with eigenvalues > 1, explaining 87.10 and 74.03% of the total variance, respectively. Table 4 shows the PCA results, including the load, eigenvalue, and variance of each PC in the two periods, as well as the cumulative explained variance.

For LPR, PC1 explained 47.37% of the total variance, with strong negative loadings on WT (-0.87) and positive loadings on EC and TN (0.80 and 0.76). The TN in the water body may come from a variety of pollution sources, such as agricultural planting, livestock and poultry breeding, domestic sewage, industrial effluents, etc. (Wang et al., 2013; Zheng et al., 2015). Based on the land use map of the study area, the sites in the LPR were dominated by agricultural and some building land. Owing to the low density of industrial enterprises in LPR, TN seems primarily associated with manure and chemical

fertilizer application and domestic sewage. Thus, PC1 can be considered the effect of agricultural pollution and rural domestic sewage. PC2, explaining 18.63% of the total variance, had strong positive loadings on BOD₅ and COD_{Cr} (0.84 and 0.76) and moderate positive loadings on COD_{Mn} and F⁻ (0.71 and 0.57) (Table 4). This factor might be due to the accumulation of organic pollutants from rural household waste, industrial wastewater, and free-range livestock and poultry pollution (Liu et al., 2020; Najar & Khan, 2012). PC3 (12.27% of the total variance) had strong positive loadings on pH and DO (0.83 and 0.76) and moderate positive loading on AS (0.71). PC3 represents natural influence and domestic sewage (Haji Gholizadeh et al., 2016). PC4 accounted for 8.83% of the total variation and had moderate positive loadings for NH₄⁺-N and TP (0.75 and 0.73). According to the previous analysis, PC4 may represent the pollution caused by N and P fertilizers used in agricultural planting entering rivers through surface scouring.

For HPR, PC1 explained 43.88% of the total variance, with strong positive loadings on NH₄⁺-N, TP, COD_{Mn}, and COD_{Cr} (0.90, 0.85, 0.84, and 0.77) and moderate positive loading on BOD₅ (0.63). According to land use information and the local statistical yearbook (LSB 2018–2020), this area was widely affected by population concentration and urbanization, industrial development, agricultural production, and livestock farming. Hence, combined with the above analysis, PC1 is largely related to municipal sewage with industrial wastewater, agricultural nonpoint sources, and livestock and poultry wastewater (Lap et al., 2021). PC2, accounting for 18.05% of the total variance, had a strong positive loading on EC (0.82) and medium positive loadings on TN, F⁻, and AS (0.72, 0.64, and 0.63). As shown in Table 1, the F⁻ concentration at the NDQ site was higher than 1.0 mg/L, and there were many factories around this site. Thus, considering the previous PCA results, PC2 represents domestic sewage and industrial wastewater. PC3, occupying approximately 13.11% of the total variance, had moderate positive loadings on pH and DO (0.73 and 0.69) and medium negative loading on WT (-0.59). Therefore, this factor can be considered a natural source (Ma et al., 2020).

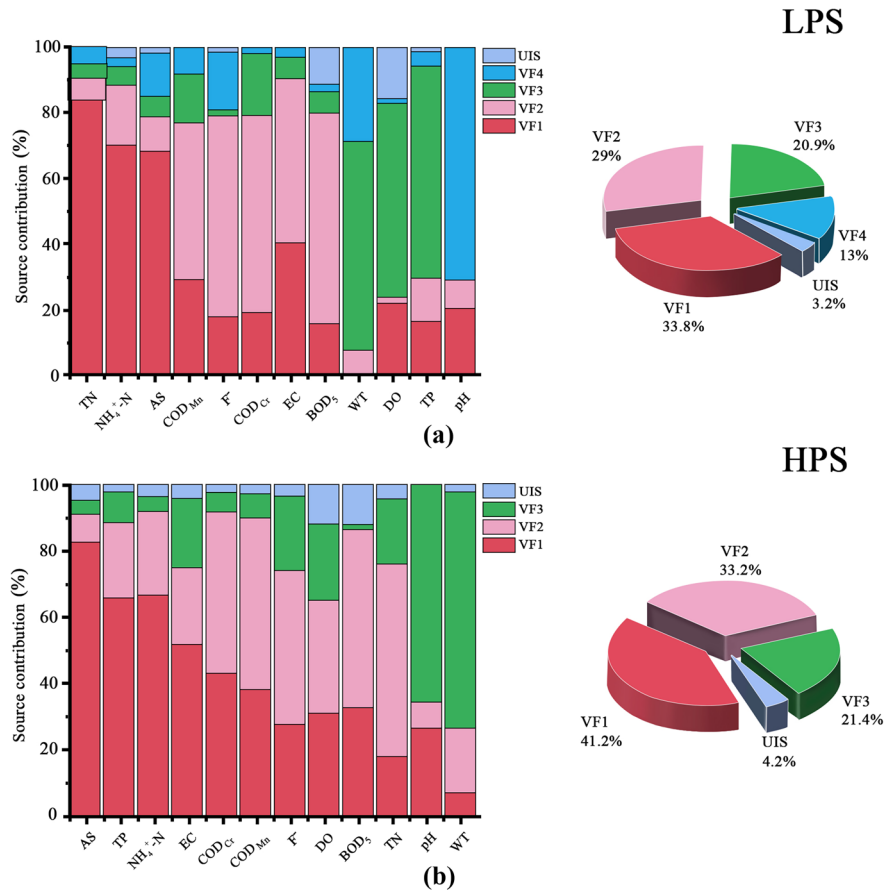
The PCA results showed that there were different amounts and contributions of pollution sources affecting the water quality of LPR and HPR. Pollution sources in the LPR can be ranked as follows: rural domestic sewage > agricultural pollution > industrial effluents and free-range livestock and poultry pollution > natural influence. HPR could be ranked as municipal sewage and industrial effluents > agricultural nonpoint sources and livestock and poultry wastewater > natural sources.

Source apportionment in temporal pattern with APCS-MLR model

On the basis of qualitative identification of pollution sources, the APCS-MLR model was established to quantitatively calculate the contribution rate of each pollution source to LPS and HPS water quality indicators. Figure 4 shows the source apportionment results for the two temporal patterns. In previous studies, a correlation coefficient greater than 0.5 between the observed value and the estimated value indicates a good fit of the model (Haji Gholizadeh et al., 2016; Simeonov et al., 2003). In our work, the modeling results showed that the mean R^2 of 0.62 for LPS and 0.65 for HPS (most parameters were greater than 0.6) reflect the accuracy and applicability of the APCS-MLR model.

For LPS, domestic and industrial wastewater and breeding pollution (PC1) was the first contamination sources, with an average contribution of 33.80%. PC1 includes nutrient indices TN (83.80%), NH₄⁺-N (70.28%), and AS (68.30%). Furthermore, pollution sources come from industrial wastewater and domestic sewage and soil weathering sources (PC2) accounted for 29.02% of total pollution sources, represented as organic indicators COD_{Cr} (60.15%), COD_{Mn} (47.54%), and BOD₅ (64.08%), and F⁻ (61.41%), and EC (49.94%), respectively. The contributions of agricultural activities and natural influences (PC3 and PC4, average contribution of 20.95%) ranged from 0.14 (pH) to 64.53% (TP) for the 12 monitoring parameters. The contributions of WT (63.62%), DO (58.95%), TP (64.53%), and pH (70.81%) mainly come from the pollution sources of PC3 and PC4. In this phase, the unidentified source

Fig. 4 Contributions on the selected water quality variables and average contributions of different pollution sources in LPS (a) and HPS (b) using APCS-MLR model (UIS: unidentified source)



contribution ranges from 0.06 (pH) to 15.78% (DO), it also contributes to each monitoring indicator to varying degrees. This may be because the contaminants come from diverse and complex sources, making it difficult to quantify pollution sources by using the APCS-MLR model (Zhang et al., 2017).

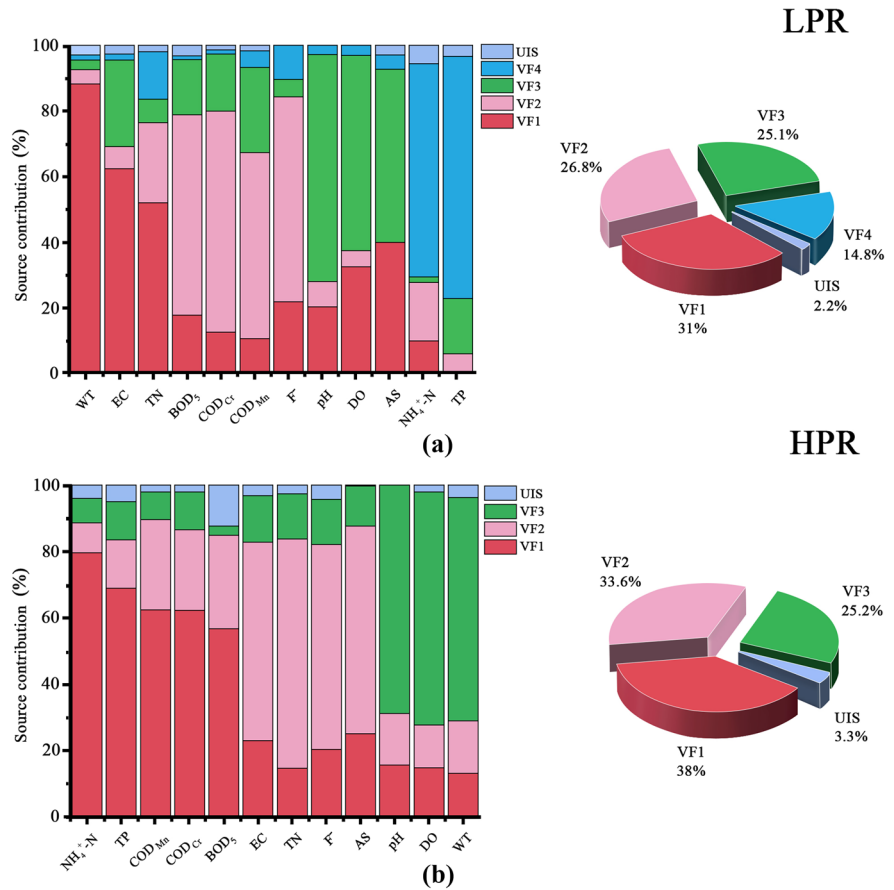
For HPS, most of the parameters were mainly affected by agricultural nonpoint source pollution (PC1, average 41.23%), manifested as nutrient indexes ($\text{NH}_4^+\text{-N}$, 66.68%, and TP, 65.78%), and AS (83.02%) and EC (51.81%) high contribution rates. Domestic sewage, industrial wastewater, and animal waste (PC2) accounted for 33.19% of total pollution sources, represented as organic parameters COD_{Cr} (48.75%), COD_{Mn} (51.93%), BOD_5 (53.77%), nutrient indexes TN (57.91%), and F^- (46.36%), and DO (34.39%). The natural variations contributed 21.43% (PC3), with most responsible for pH (76.47%) and WT (59.93%). Besides, the unidentified sources also caused the river water pollution of HPS, ranging from 0.23 (pH) to 11.92% (BOD_5).

Source apportionment in spatial pattern with APCS-MLR model

The APCS-MLR model was also applied to calculate the contributions of each pollution source to the water quality indicators for LPR and HPR. Similar to the temporal patterns, most of the concentration R^2 values of the 12 selected parameters of LPR and HPR were greater than 0.5, with mean values of 0.70 and 0.65, respectively, shows that the predicted value of the model is consistent with the actual observed value to a high degree, and the final apportionment result is scientific and reliable.

Figure 5 shows the source apportionment results for the two spatial patterns. As shown in Fig. 5 (LPR), the contributions of agricultural pollution and rural domestic sewage were 31.01% of total pollution sources (PC1), mainly represented by WT (88.20%), EC (62.31%), and TN (52.07%). Furthermore, organic pollution from rural household waste,

Fig. 5 Contributions on the selected water quality variables and average contributions of different pollution sources in LPR **a** and HPR **b** using APCS-MLR model (UIS: unidentified source)



industrial effluents, and free-range livestock and poultry pollution (PC2) accounted for 26.82% of total pollution sources, represented by BOD₅ (68.07%), COD_{Cr} (65.95%), COD_{Mn} (56.77%), and F⁻ (56.04%), respectively. Contributions of physicochemical influence and domestic sewage (PC3, average 19.86%) to different water quality indicators ranged from 1.97 (NH₄⁺-N) to 69.66% (pH). PC4 (average 14.82%) represented agricultural sources, and the corresponding contribution rates of NH₄⁺-N and TP were 79.83 and 57.23%. Unidentified sources of pollution contributed to water quality indicators ranging from 0.27 (pH) to 6.52% (NH₄⁺-N). For unidentified sources, the contribution to each monitoring indicator ranges from 0.27 (pH) to 6.52% (NH₄⁺-N). Generally, compared with temporal apportionment, the source contribution of unknown pollution was relatively low in LPR (mean contribution of 2.22%). This indicated that the potential pollution sources in the LPR were accurately identified.

For HPR (Fig. 5), most water quality parameters were significantly affected by municipal sewage with industrial effluents, agricultural sources, and livestock and poultry wastewater (PC1, 37.96%), shown as nutrients index (NH₄⁺-N, 79.78%; TP, 68.98%) and organic pollutants (COD_{Mn}, 62.81%; COD_{Cr}, 61.87%; BOD₅, 56.68%). Domestic sewage and industrial wastewater sources (PC2) explained 33.55% of total pollution sources, represented as EC (69.94%), TN (69.04%), F⁻ (62.21%), and AS (53.22%). The natural sources contributed 25.23% (PC3), with the most responsible for pH (71.73%), DO (67.87%), and WT (67.45%). Unidentified contamination sources contributed to the water quality indicators, ranging from 0.15 (pH) to 12.14% (BOD₅). The average contribution (3.26%) of unidentified pollution sources in HPR was roughly similar to that of LPR, indicating that potential pollution sources in HPR were basically identified completely.

Conclusions

In this study, the spatial and temporal distribution patterns of monitoring parameters in Laixi River were discussed by using multivariate statistical techniques, and the contribution of potential pollution sources in different spatial and temporal categories to selected monitoring indicators was clarified. The CA results showed that the 12 months were divided into two clusters, consistent with the LPS and HPS. The spatial clustering results showed that the six monitoring sites in the study area were divided into two groups with different pollution statuses: the LPR and the HPR. The number of pollution sources under different spatial and temporal conditions was determined from the PCA results. Finally, the relative contribution of the sources was quantified using the APCS-MLR model.

For LPS, domestic and industrial wastewater and breeding pollution (PC1), with a contribution rate of 33.80%, and for HPS, source pollution from agricultural activities (PC1), with a contribution rate of 41.23%, were the main pollution sources in river water quality. These were followed by industrial effluents, domestic sewage, and soil weathering (PC2) with a 29.02% contribution and agricultural activities and natural influence (PC3 and PC4) with a 20.95% contribution to LPS, and domestic sewage, industrial wastewater, and animal waste (PC2) with 33.19% contribution and natural variations (PC3) with 21.43% contribution to HPS. The four identified latent sources of contamination in LPR were rural domestic sewage > agricultural pollution > industrial effluents and free-range livestock and poultry pollution > natural influence, with average contributions of 31.01%, 26.82%, 25.13%, and 14.82%, respectively. While in HPR, the three identified latent pollution sources were municipal sewage and industrial effluents > agricultural nonpoint sources and livestock and poultry wastewater > natural sources, with average contributions of 37.96%, 33.55%, and 25.23%, respectively.

The results of this paper illustrate that multivariate statistical analysis methods can serve as excellent exploratory tools for analyzing and interpreting complex water quality datasets and identifying and assigning pollution sources. In addition, this evaluation can help managers and decision-makers gain an in-depth understanding of the main pollution sources

of the study area and provide a reference for formulating more reasonable and reliable pollution control strategies in tributary watersheds.

Acknowledgements We are grateful to the Luzhou Ecological Environment Bureau (LEEB) for providing the water quality data. The authors are grateful to the editors and the anonymous reviewers for their constructive comments and suggestions.

Author contribution Jie Xiao and Dongdong Gao wrote the main manuscript text and made contributions to the methodology. Han Zhang did the formal and data analysis. Hongle Shi, Qiang Chen, and Qingsong Chen did the investigation. Hongfei Li and Xingnian Ren prepared figures and tables.

Funding This study is supported by the Sichuan Science and Technology Program (grant no. 2021JDZH0030) and the National Natural Science Foundation of China (grant nos. 51979237 and 52170104).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

- Bonansea, M., Ledesma, C., Rodriguez, C., & Pinotti, L. (2015). Water quality assessment using multivariate statistical techniques in Río Tercero Reservoir, Argentina. *Hydrology Research*, 46(3), 377–388. <https://doi.org/10.2166/nh.2014.174>
- Chen, R., Teng, Y., Chen, H., Hu, B., & Yue, W. (2019). Groundwater pollution and risk assessment based on source apportionment in a typical cold agricultural region in Northeastern China. *Science of the Total Environment*, 696(19), 133972. <https://doi.org/10.1016/j.scitotenv.2019.133972>
- Cheng, G., Wang, M., Chen, Y., & Gao, W. (2020). Source apportionment of water pollutants in the upstream of Yangtze River using APCS-MLR. *Environmental Geochemistry and Health*, 42(11), 3795–3810. <https://doi.org/10.1007/s10653-020-00641-z>
- Cho, Y., Choi, H., & Lee, M. G. (2022). Sources using multivariate statistical techniques and APCS-MLR model to assess surface water quality. *water*, 14, 793–812.
- de Oliveira, J. F., Fia, R., Nunes, B. S. B., Siniscalchi, L. A. B., de Matos, M. P., & Fia, F. R. L. (2020). Nitrogen and phosphorus removal associated with changes in organic loads from biological reactors monitored by multivariate criteria. *Water, Air, and Soil Pollution*, 231(10), 511. <https://doi.org/10.1007/s11270-020-04858-7>

- Ding, J., Jiang, Y., Liu, Q., Hou, Z., Liao, J., Fu, L., & Peng, Q. (2016). Influences of the land use pattern on water quality in low-order streams of the Dongjiang River Basin, China: A multi-scale analysis. *Science of the Total Environment*, 551, 205–216. <https://doi.org/10.1016/j.scitotenv.2016.01.162>
- Fu, D., Wu, X., Chen, Y., & Yi, Z. (2020). Spatial variation and source apportionment of surface water pollution in the Tuo River, China, using multivariate statistical techniques. *Environmental Monitoring and Assessment*, 192(12), 745. <https://doi.org/10.1007/s10661-020-08706-3>
- Giao, N. T., Anh, P. K., & Nhien, H. T. H. (2021). Spatiotemporal analysis of surface water quality in Dong. *water*, 13(3), 336. <https://doi.org/10.3390/w13030336>
- Haji Gholizadeh, M., Melesse, A. M., & Reddi, L. (2016). Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. *Science of the Total Environment*, 566–567, 1552–1567. <https://doi.org/10.1016/j.scitotenv.2016.06.046>
- Han, Q., Tong, R., Sun, W., Zhao, Y., & Yu, J. (2019). Anthropogenic influences on the water quality of the Baiyangdian Lake in North China over the last decade. *Science of the Total Environment*, 134929. <https://doi.org/10.1016/j.scitotenv.2019.134929>
- Herojeet, R., Rishi, M. S., Lata, R., & Dolma, K. (2017). Quality characterization and pollution source identification of surface water using multivariate statistical techniques, Nalagarh Valley, Himachal Pradesh, India. *Applied Water Science*, 7(5), 2137–2156. <https://doi.org/10.1007/s13201-017-0600-y>
- Huang, F., Wang, X., Lou, L., Zhou, Z., & Wu, J. (2010). Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water Research*, 44(5), 1562–1572. <https://doi.org/10.1016/j.watres.2009.11.003>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Katsaounis, T. I. (2004). Analyzing multivariate data. *Technometrics*, 46(2), 254–255. <https://doi.org/10.1198/tech.2004.s798>
- Kellner, E., Hubbart, J., Stephan, K., Morrissey, E., Freedman, Z., Kutta, E., & Kelly, C. (2018). Characterization of sub-watershed-scale stream chemistry regimes in an Appalachian mixed-land-use watershed. *Environmental Monitoring and Assessment*, 190, 586. <https://doi.org/10.1007/s10661-018-6968-9>
- Lap, B. Q., Nam, N. H., Anh, B. T. K., Linh, T. T. T., Quang, L. X., Toan, V. D., et al. (2021). Monitoring water quality in Lien Son irrigation system of Vietnam and identification of potential pollution sources by using multivariate analysis. *Water, Air, and Soil Pollution*, 232(5), 187. <https://doi.org/10.1007/s11270-021-05137-9>
- LEEB. (2021). Research report on optimization and improvement of ecological zoning control in Luzhou. Luzhou Ecological Environment Bureau. (In Chinese).
- LSB. (2018–2020). Luzhou statistical yearbook. Luzhou Statistics Bureau. (In Chinese).
- Li, Q., Zhang, H., Guo, S., Fu, K., Liao, L., Xu, Y., & Cheng, S. (2020). Groundwater pollution source apportionment using principal component analysis in a multiple land-use area in southwestern China. *Environmental Science and Pollution Research*, 27(9), 9000–9011. <https://doi.org/10.1007/s11356-019-06126-6>
- Liu, C. W., Lin, K. H., & Kuo, Y. M. (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. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- Liu, D., Li, X., Zhang, Y., Lu, Z., Bai, L., Qiao, Q., & Liu, J. (2022). Spatial-temporal distribution of phosphorus fractions and their relationship in water-sediment phases in the tuojiang river, China. *Water (switzerland)*, 14(1), 27. <https://doi.org/10.3390/w14010027>
- Liu, L., Dong, Y., Kong, M., Zhou, J., Zhao, H., Tang, Z., et al. (2020). Insights into the long-term pollution trends and sources contributions in Lake Taihu, China using multivariate statistical analyses models. *Chemosphere*, 242. <https://doi.org/10.1016/j.chemosphere.2019.125272>
- Liu, L., Tang, Z., Kong, M., Chen, X., Zhou, C., Huang, K., & Wang, Z. (2019). Tracing the potential pollution sources of the coastal water in Hong Kong with statistical models combining APCS-MLR. *Journal of Environmental Management*, 245(May), 143–150. <https://doi.org/10.1016/j.jenvman.2019.05.066>
- Ma, W., Meng, L., Wei, F., Opp, C., & Yang, D. (2021). Spatiotemporal variations of agricultural water footprint and socioeconomic matching evaluation from the perspective of ecological function zone. *Agricultural Water Management*, 249(January), 106803. <https://doi.org/10.1016/j.agwat.2021.106803>
- Ma, X., Wang, L., Yang, H., Li, N., & Gong, C. (2020). Spatiotemporal analysis of water quality using multivariate statistical techniques and the water quality identification index for the Qinhuai River Basin, East China. *Water (switzerland)*, 12(10), 2764. <https://doi.org/10.3390/w12102764>
- Matiatos, I. (2016). Nitrate source identification in groundwater of multiple land-use areas by combining isotopes and multivariate statistical analysis: A case study of Asopos basin (Central Greece). *Science of the Total Environment*, 541, 802–814. <https://doi.org/10.1016/j.scitotenv.2015.09.134>
- Meng, L., Zuo, R., Wang, J. S., Yang, J., Teng, Y. G., Shi, R. T., & Zhai, Y. Z. (2018). Apportionment and evolution of pollution sources in a typical riverside groundwater resource area using PCA-APCS-MLR model. *Journal of Contaminant Hydrology*, 218(April), 70–83. <https://doi.org/10.1016/j.jconhyd.2018.10.005>
- MEEC. (2002). Technical specifications requirements for monitoring of surface water and waste water (HJT 91-2002). Ministry of Ecology and Environment of the People's Republic of China, Beijing (In Chinese).
- MEPC. (2002). Environmental quality standards for surface water, GB 3838–2002. Ministry of Environmental Protection of China, Beijing. (In Chinese).
- Mir, R. A., & Gani, K. M. (2019). Water quality evaluation of the upper stretch of the river Jhelum using multivariate statistical techniques. *Arabian Journal of Geosciences*, 12(14), 445. <https://doi.org/10.1007/s12517-019-4578-7>
- Muangthong, S., & Shrestha, S. (2015). Assessment of surface water quality using multivariate statistical techniques: Case study of the Nampong River and Songkhram River, Thailand. *Environmental Monitoring and Assessment*, 187(9), 548. <https://doi.org/10.1007/s10661-015-4774-1>
- Najar, I. A., & Khan, A. B. (2012). Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis. *Environmental*

- Earth Sciences*, 66(8), 2367–2378. <https://doi.org/10.1007/s12665-011-1458-1>
- Nong, X., Shao, D., Zhong, H., Liang, J. (2020). Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Research*, 178. <https://doi.org/10.1016/j.watres.2020.115781>
- Pinto, C. C., Calazans, G. M., & Oliveira, S. C. (2019). Assessment of spatial variations in the surface water quality of the Velhas River Basin, Brazil, using multivariate statistical analysis and nonparametric statistics. *Environmental Monitoring and Assessment*, 191(3), 1–13. <https://doi.org/10.1007/s10661-019-7281-y>
- Piroozfar, P., Alipour, S., Modabberi, S., & Cohen, D. (2021). Using multivariate statistical analysis in assessment of surface water quality and identification of heavy metal pollution sources in Sarough watershed, NW of Iran. *Environmental Monitoring and Assessment*, 193(9), 1–20. <https://doi.org/10.1007/s10661-021-09363-w>
- Rezaei, A., Hassani, H., Hassani, S., Jabbari, N., Fard Mousavi, S. B., & Rezaei, S. (2019). Evaluation of groundwater quality and heavy metal pollution indices in Bazman basin, southeastern Iran. *Groundwater for Sustainable Development*, 9(July), 100245. <https://doi.org/10.1016/j.gsd.2019.100245>
- Sasi, S., Rayaroth, M. P., Aravindakumar, C. T., & Aravind, U. K. (2021). Alcohol ethoxysulfates (AES) in environmental matrices. *Environmental Science and Pollution Research*, 28(26), 34167–34186. <https://doi.org/10.1007/s11356-021-14003-4>
- Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsas, D., Anthemidis, A., et al. (2003). Assessment of the surface water quality in Northern Greece. *Water Research*, 37(17), 4119–4124. [https://doi.org/10.1016/S0043-1354\(03\)00398-1](https://doi.org/10.1016/S0043-1354(03)00398-1)
- Thurston, & Spengler. (1984). Assessment of source contributions to inhalable particulate pollution in metropolitan. *Atmospheric Environment*, 19(1), 9–25.
- Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2022). A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. *Water Research*, 219, 118532. <https://doi.org/10.1016/j.watres.2022.118532>
- Varol, M. (2020). Spatio-temporal changes in surface water quality and sediment phosphorus content of a large reservoir in Turkey. *Environmental Pollution*, 259, 113860. <https://doi.org/10.1016/j.envpol.2019.113860>
- Verheyen, D., Van Gaalen, N., Ronchi, B., Batelaan, O., Struyf, E., Govers, G., et al. (2015). Dissolved phosphorus transport from soil to surface water in catchments with different land use. *Ambio*, 44(2), 228–240. <https://doi.org/10.1007/s13280-014-0617-5>
- Wang, J., Yang, J., & Chen, T. (2022). Source appointment of potentially toxic elements (PTEs) at an abandoned realgar mine: Combination of multivariate statistical analysis and three common receptor models. *Chemosphere*, 307, 135923. <https://doi.org/10.1016/j.chemosphere.2022.135923>
- Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., et al. (2013). Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-Environment Research*, 7(1), 30–40. <https://doi.org/10.1016/j.jher.2012.10.003>
- Yang, C., Zeng, Z., Zhang, H., Gao, D., Wang, Y., He, G., Liu, Y., Wang, Y., & Du, X. (2022). Distribution of sediment microbial communities and their relationship with surrounding environmental factors in a typical rural river, Southwest China. *Environmental Science and Pollution Research*, 29, 84206–84225. <https://doi.org/10.1007/s11356-022-21627-7>
- Yu, L., Zheng, T., Yuan, R., & Zheng, X. (2022). APCS-MLR model: A convenient and fast method for quantitative identification of nitrate pollution sources in groundwater. *Journal of Environmental Management*, 314(April). <https://doi.org/10.1016/j.jenvman.2022.115101>
- Zhang, H., Li, H., Gao, D., & Yu, H. (2022). Source identification of surface water pollution using multivariate statistics combined with physicochemical and socioeconomic parameters. *Science of the Total Environment*, 806, 151274. <https://doi.org/10.1016/j.scitotenv.2021.151274>
- Zhang, H., Li, H., Yu, H., & Cheng, S. (2020). Water quality assessment and pollution source apportionment using multi-statistic and APCS-MLR modeling techniques in Min River Basin, China. *Environmental Science and Pollution Research*, 27(33), 41987–42000. <https://doi.org/10.1007/s11356-020-10219-y>
- Zhang, Q., Wang, H., Wang, Y., Yang, M., & Zhu, L. (2017). Groundwater quality assessment and pollution source apportionment in an intensely exploited region of northern China. *Environmental Science and Pollution Research*, 24(20), 16639–16650. <https://doi.org/10.1007/s11356-017-9114-2>
- Zhang, Z., Tao, F., Du, J., Shi, P., Yu, D., Meng, Y., & Sun, Y. (2010). Surface water quality and its control in a river with intensive human impacts—a case study of the Xiangjiang River, China. *Journal of Environmental Management*, 91(12), 2483–2490. <https://doi.org/10.1016/j.jenvman.2010.07.002>
- Zheng, L. Y., Yu, H. B., & Wang, Q. S. (2015). Assessment of temporal and spatial variations in surface water quality using multivariate statistical techniques: A case study of Nenjiang River basin, China. *Journal of Central South University*, 22(10), 3770–3780. <https://doi.org/10.1007/s11771-015-2921-z>
- Zhou, F., Huang, G. H., Guo, H. C., Zhang, W., & Hao, Z. (2007). Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong. *Water Research*, 41(15), 3429–3439. <https://doi.org/10.1016/j.watres.2007.04.022>
- Zhou, H., & Gao, C. (2011). Assessing the risk of phosphorus loss and identifying critical source areas in the Chaohu Lake watershed, China. *Environmental Management*, 48(5), 1033–1043. <https://doi.org/10.1007/s00267-011-9743-z>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.