A multivariate analysis of the spatial variations of water quality during high-flow period in the Chaobai River (Beijing, China) restored by reclaimed water

Rui Zhao, Hongmei Bu, Xianfang Song and Yinghua Zhang

ABSTRACT

Reclaimed water has demonstrated its broad applications in social construction to alleviate the contradiction of water shortage in Beijing, China. Using multivariate statistical analysis, the current study investigated the spatial variations of water quality in the Chaobai River restored by reclaimed water during the high-flow period. Hierarchical cluster analysis (CA) classified the 11 sampling sites into four clusters, namely most polluted, highly polluted, moderately polluted, and lowly polluted sections. The Kruskal-Wallis test showed that pH, TDS, EC, Ca\(^{2+}\), Mg\(^{2+}\), Cl\(^{-}\), SO\(_4^{2-}\), NO\(_3^{-}\), and TN had significant spatial differences among four clusters (\(p < 0.05\)). Mean value of total nitrogen (TN) in the most polluted site exceeded the guideline (15 mg/L) of the Water Quality Standard for Scenic Environment Use, reaching 22.3 mg/L. Principal component analysis (PCA) extracted three principal components (PCs) accounting for 81.5% of the total variance in the data set of water quality. The three PCs reflected the chemical characteristics of reclaimed water, mineral pollution, and nutrient pollution, respectively. With the ordination biplot of sampling sites defined by the first and second PCs, PCA provided a classification of sampling sites based on the similarity of pollution sources, which supported the results of CA. The results revealed that water quality of the Chaobai River restored by reclaimed water was affected by untreated domestic and agricultural sewage with nitrogen and minerals being the main pollutants along the river basin. This study showed rivers restored by reclaimed water had significant spatial variations of water quality, demonstrating effectiveness of multivariate statistical methods on water quality analysis.

Key words | cluster analysis, principal component analysis, reclaimed water, the Chaobai River, water quality

HIGHLIGHTS

- Water quality of the river restored by reclaimed water showed significant spatial variations.
- Nitrogen pollution was severe compared to standard values.
- CA and PCA indicated similar spatial patterns of water quality variations.
- Besides reclaimed water, untreated domestic and agricultural sewage were also pollution sources.
- The effectiveness of multivariate statistical methods on water quality analysis was demonstrated.
INTRODUCTION

Reclaimed water use is viewed as one part of the solution to China’s water shortage. Reclaimed water has demonstrated its broad applications in miscellaneous sectors including agricultural irrigation (Wang et al. 2017), industrial production (Chu et al. 2004), groundwater replenishment (Jin et al. 2010), domestic use (Hurlimann & McKay 2007), and urban landscaping around the world (Chen et al. 2017). As the urban water shortage grows in China, more and more cities have developed municipal wastewater reclamation systems, such as Beijing. Beijing has been integrating wastewater reclamation into its municipal water resources allocation system since 2003, and its annual volume of reclaimed water use has risen rapidly from 210 in 2003 to 1050 million tons in 2017 (Beijing Water Authority 2017). Accordingly, reclaimed water has ranked second to groundwater in the total water supplies in Beijing. Reclaimed water in Beijing was mostly applied in scenic environment use (i.e. reclaimed water is mainly used to maintain the landscape environment), accounting for 92.3% of total reclaimed water utilization (Beijing Water Authority 2017). Prominent development has been achieved in this field with reclaimed water used to replenish scenic rivers and lakes including the Chaobai River, the Longtan Lake, the Dragon-shaped Water System, and so on (Li & Zhang 2011).

However, water quality of reclaimed water has been a major concern since various pollutants remain in wastewater even after treatments (Zhao et al. 2015). Organic matter, nitrogen, and phosphorus are believed to be the main pollutants restricting the quality of reclaimed water, hindering the large-scale consumption of reclaimed water for scenic environment use in China (Chen et al. 2017). Considering the high background values of nutrients in reclaimed water and their close relationship with algal growth and bacterial growth (Li et al. 2011), water quality monitoring in water bodies receiving reclaimed water is primary to guarantee a satisfactory water environment. Noticeably, denitrification, which reduces nitrate nitrogen (NO₃⁻-N) in the river water and ultimately produces nitrogen (N₂) or nitrous oxide (N₂O), is active in reclaimed water (He et al. 2016), playing a significant role in the removal of nitrogen. Additionally, photosynthesis and respiration of phytoplankton strongly influence the water chemistry, causing the changes of many parameters such as pH, DO, and TDS (Yang et al. 2016). Therefore, characteristics of reclaimed water quality show significant variations along river courses (Wang et al. 2014), making water quality assessment and management more difficult. However, limited studies have thoroughly discussed the spatial patterns of water quality and pollution sources.

We used a multivariate statistical approach to evaluate the water quality in the Chaobai River. Hierarchical cluster analysis (CA) is widely applied to detect the distance and relatedness among variables and subsequently group multiple variables into clusters (Razmkhah et al. 2010). Principal component analysis (PCA) or factor analysis (FA) can reduce data dimensionality by extracting the most significant components and rendering a simpler representation of the data (Wallace et al. 2016). In Jajrood River (Iran), Out-Meygoon was identified as the most polluted station by both CA and PCA, with CA presenting an overview of the problem and PCA giving supporting explanations (Razmkhah et al. 2010). According to a study on the spatial and temporal variations of organic pollutants in rivers, CA grouped sampling sites distributing three key rivers in Romania into three clusters and PCA extracted 87.3% information from data set variability, therefore the combination of CA and PCA provided a simplified and optimized monitoring program that could reduce costs and retain important information (Feher et al. 2016). These studies showed the combined application of CA and PCA is capable of detecting the spatial and temporal differences in water quality and identifying the contribution of pollution sources in rivers (Razmkhah et al. 2010; Chounlamany et al. 2017). However, among all the studies on assessment of reclaimed water quality, the combination of CA and PCA is still scarce (Wang et al. 2014).

In this study, the Shunyi reaches of the Chaobai River, a scenic river restored by reclaimed water to maintain the landscape function, was chosen for water quality evaluation. Moreover, the northern reach of the Chaobai River, which is only in flow during high-flow period, was included for the first time. There is no lining constructed at the bottom of
the Chaobai River, therefore the Chaobai River is quite typical in representing the conditions of natural rivers receiving reclaimed water. The main objectives of this study were to (1) analyze the spatial variations of water quality characteristics, (2) identify the major pollution factors and sources influencing the water quality in the Chaobai River, and (3) verify the reliability of multivariate statistics in finding water quality patterns in the study area. This research will provide useful information on water quality characteristics in the river restored by reclaimed water using multivariate statistical methods.

MATERIALS AND METHODS

Study area

The study area (40°03′-40°11′N, 116°38′-116°45′E, Spherical Mercator) is located in Shunyi district, Beijing, where the Chaobai River is restored by reclaimed water (Figure 1). The study area is within a temperate semi-arid and semi-humid monsoon climate zone, characterized by cold, windy and dry winters, and hot and humid summers. The annual average precipitation from 1959 to 2017 was 644.8 mm (National Meteorological Science Data Center http://data.cma.cn). Most rainfall occurs in summer from July to September, and rainfall in July accounts for 32% of annual precipitation.

Meteorological droughts occurred 128 times during 1960–2011 in the Chaobai River Basin, most of which lasted one or two months (Xu et al. 2019). Therefore, the river reaches between Xiangyang Gate Dam and Henan Rubber Dam (Figure 1) in the Chaobai River had dried up since 1999. The ‘Water Diversion Project from Wenyu River to Chaobai River’ was initiated in October 2007. Reclaimed water is discharged from the outlet in the Jian River, and then replenishes the southeastern reach and the northern reach of the Chaobai River for landscaping and ecological restoration (Figure 1). By the end of 2018, the actual water volume transferred by the water diversion project had accumulated to 278 million m³. Larger amounts of reclaimed water were diverted into the river reaches from July to September in 2017, thus a higher river water supply was observed during this period (Figure 2). Since no lining was constructed at the bottom of the river courses, the physicochemical properties of the river water can be actively affected by biochemical reactions in sediments.

Water sampling and monitored parameters

To investigate the spatial distribution of water quality, 11 sampling sites were selected in the study area (Figure 1). According to direction of water flow, the Jian River was
upstream, while the southeastern reach and northern reach of the Chaobai River were downstream (Figure 1). Sampling sites S1-3 were in the Jian River, S4-7 in the southeastern reach of the Chaobai River, and S8-11 in the northeastern reach. The Jian River and the southeastern reach of the Chaobai River were perennial river reaches with continuous supply of reclaimed water, while the northeastern reach was only in flow during the high-flow period (July, August, and September). Therefore, the water sampling was conducted in September 2017 when the higher water supply was observed (Figure 2).

Sixteen variables relating to water quality were measured, namely water temperature (Temp), dissolved oxygen (DO), pH, total dissolved solid (TDS), electrical conductivity (EC), calcium (Ca²⁺), potassium (K⁺), magnesium (Mg²⁺), sodium (Na⁺), fluoride (F⁻), chloride (Cl⁻), sulfate (SO₄²⁻), ammonium nitrogen (NH₄-N), nitrate nitrogen (NO₃-N), nitrate nitrogen (NH₄-N), total nitrogen (TN), and total phosphorus (TP). The sampling, preservation, and transportation of the water samples were performed according to the standard methods for the examination of water and wastewater (NEPB 2002). Variables including Temp, EC, DO, TDS, and pH were directly measured in situ using a multiparameter water quality monitoring instrument (YSI Incorporated, USA). River water samples were collected in polyethylene plastic bottles which had been pre-rinsed three times with distilled water in the field. Before laboratory analysis, water samples were kept in the refrigerator at 4 °C (NEPB 2002). Anions of F⁻, Cl⁻, and SO₄²⁻ were determined using ion chromatography (Dionex ICS-3000, USA). Cations of Ca²⁺, K⁺, Mg²⁺, and Na⁺ were determined on the inductively coupled plasma optical emission spectrometer (Perkin Elmer, USA). Water samples for NO₃-N, NH₄-N, and TN analysis were all acidified to pH < 2 by sulfuric acid in situ (NEPB 2002). The concentrations of NH₄-N and NO₃-N were determined using the continuous flow analyzer (SEAL Analytical, UK). TN was determined by alkaline potassium persulfate digestion UV spectrophotometric method on the automatic chemical analyzer (Bran + Luebbe AutoAnalyzer 3, Germany). TP was analyzed by digestion and a colorimetric method on the spectrophotometer (Alliance Smart Chem 200, France). Detection limits were 0.025, 0.02, 0.05, and 0.01 mg/L for NH₄-N, NO₃-N, TN, and TP, respectively.

### Statistical analysis

The Shapiro-Wilk method was used to verify if the original data was normally distributed, and the Bartlett method was applied to test homogeneity of variance of the data set. Since the data set failed to meet the standard normal distribution and homogeneity of variance, the nonparametric Kruskal-Wallis test was applied to estimate the spatial differences of water quality variables. Correlations remain reliable even when non-normality and heteroskedasticity are not extreme (Zar 2010). The Kendall correlation coefficient was used to analyze the relationships between variables, harmonizing data analyses as all nonparametric methods.

Hierarchical cluster analysis (CA) and principal component analysis (PCA) were used to investigate the spatial patterns of water quality. To avoid misclassification and eliminate the influence of different units of measurement, the raw data was firstly standardized through z-score transformation (the mean and variance were respectively set to 0 and 1) before CA and PCA were performed on the data set (Singh et al. 2004). CA was used to group sampling sites into clusters to achieve maximum similarities within a cluster and maximum dissimilarities among different clusters (Razmkhah et al. 2010). Ward’s method was applied to assess the similarities of water quality in diverse sampling sites based on squared Euclidean distances. Then, Silhouette Width value, an index measuring how much an object associates with its cluster (Rousseeuw 1987), was calculated to determine the optimal number of clusters for the data set.

![Figure 2](image-url) | Monthly water supply of reclaimed water in Shunyi (Beijing) reaches of the Chaobai River in 2017.
and higher Silhouette Width value means closer associations, therefore a better classification scheme. PCA was performed on the correlation matrix to identify the most effective variables by reducing data dimensionality and extracting the contribution of each principal component (PC). Eigenvalues served as a measure of how much PCs explained the original data. Eigenvectors were selected following Kaiser-Guttman criterion, which firstly computed the mean of all eigenvalues and then interpreted only the axes with eigenvalues larger than the mean (Borcard et al. 2011). Correlations of PCs and original variables were expressed as loadings, and individual transformed observations of sampling sites were given by scores (Alberto et al. 2001).

All data analysis was performed in the R software environment.

RESULTS AND DISCUSSION

Cluster analysis and spatial similarities

Cluster analysis was used to detect the similarities of sampling sites based on chemical compositions of river water in spatial scale (Bu et al. 2016). Sampling sites were grouped into four clusters according to the dendrogram rendered by CA (Figure 3). The Kruskal-Wallis test showed that values of seven variables, namely pH, TDS, EC, Ca$^{2+}$, Mg$^{2+}$, Cl$^{-}$, SO$_4^{2-}$, NO$_3$-N, and TN, were significantly different among the four clusters ($p < 0.05$; Figure 4).

Cluster 1 included sampling sites S1 and S2 located in the Jian River (Figure 3). The mean values of EC, TDS, Ca$^{2+}$, Mg$^{2+}$, Cl$^{-}$, SO$_4^{2-}$, NO$_3$-N, and TN in cluster 1 were all highest among the four clusters ($p < 0.05$; Figure 4), reaching 952.5 ± 129.5 (mean ± S.D.) μS/cm, 461 ± 63, 63.9 ± 4.1, 23.2 ± 0.2, 104.6 ± 11.1, 114.5 ± 20.9, 7.7 ± 3.7, and 22.3 ± 1.4 mg/L, respectively. Thus, sampling sites in cluster 1 were most polluted in the study area. As reclaimed water was for scenic environment use in the study area, its water quality was supposed to meet the standards of Water Quality Standard for Scenic Environment Use in China (GB/T18921-2002) (AQSIQ 2003), which stipulates that TN concentration should be lower than 15 mg/L. However, this threshold value for TN has been argued to be not strict enough to guarantee ecological safety of rivers and lakes receiving reclaimed water (Zhou et al. 2011). The mean concentration of TN in cluster 1 (22.3 mg/L) exceeded the threshold (15 mg/L) by 7.3 mg/L. Therefore, nitrogen pollution was extremely serious in the Jian River. Considering that sampling sites S1 and S2 were near the outlet of reclaimed water (Figure 1), the two sites were assumed to be mainly affected by reclaimed water directly discharged from the wastewater treatment plant. Reclaimed water quality varies greatly in different studies depending on their wastewater reclamation processes including coagulation-sedimentation-filtration, micro-flocculation reclamation processes, ultra-filtration membrane process, membrane bioreactor (MBR) technology, and so on (Zhao et al. 2015). Though MBR technology, which combines biological treatment with membrane separation process, was used for reclaimed water supplying the Chaobai River, salts and nitrogen remained and caused poor water quality in the Jian River (Yang et al. 2016). Therefore, the removal efficiency of pollutants (especially nutrient load) by wastewater treatment plants is technologically limited (Marti et al. 2004).

Cluster 2 contained four sampling sites (Figure 3). Of these sampling sites, S8-10 were in the northern reach of the Chaobai River while S3 was at the confluence of the Jian River and the Chaobai River. Therefore, cluster 2 represented the condition of the northern reach of the Chaobai River. Compared with cluster 1, these four sampling sites in cluster 2 were moderately polluted considering they have significantly lower mean values of EC.
(667.5 ± 49.3 μS/cm), TDS (311.8 ± 25.4 mg/L), Ca²⁺ (49.8 ± 7.5 mg/L), NO₃⁻ (1.3 ± 0.6 mg/L), and TN (12.3 ± 4.1 mg/L) than these in cluster 1 (p < 0.05). Moreover, though the concentrations of NH₄⁺ were not significantly different among the four clusters (p > 0.05), the highest mean concentration of NH₄⁺ (0.12 ± 0.05 mg/L) was observed in cluster 2. Since the amount of reclaimed water supplied to the Chaobai River fluctuated over a year (Figure 2), the northern reach of the river was only in flow during the high-flow period (July, August, and September). Therefore, previous studies on the Chaobai River have excluded it, making the water quality of the northern reach unclear (Wang et al. 2014; He et al. 2016; Yang et al. 2016). We found that the northern reach of the Chaobai River, as a seasonal river reach, showed distinct water chemistry from the other river reaches. During the low-flow period, the northern reach was almost dry and overgrown with plants. Subsequently, when reclaimed water recharged into the northern reach during the high-flow period, the living environment of microorganisms in river water changed with the extra input of organic matter (e.g., carbon and nitrogen) from litter of plants and sediments (Chen et al. 2017).

Sampling sites S4-S6 in cluster 3 in the southeastern reach of the Chaobai River (Figure 3) were less polluted with lowest mean values of TDS (242.6 ± 19.3 mg/L), EC

Figure 4 | Variations of the 16 water quality variables in different clusters in Shunyi (Beijing) reaches of the Chaobai River, China (Letters of a, b, and c denote significant different mean values in spatial scale according to the Kruskal-Wallis tests (p < 0.05), separately).
(529.7 ± 36.2 μS/cm), K⁺ (10.9 ± 0.2 mg/L), Mg²⁺ (15.9 ± 0.6 mg/L), Na⁺ (58.3 ± 2.1 mg/L), Cl⁻ (51.7 ± 2.3 mg/L), SO₄²⁻ (71.3 ± 5.1 mg/L), NO₃⁻-N (0.29 ± 0.02 mg/L), and NH₄⁺ (0.06 ± 0.01 mg/L). The southeastern reach was characterized by an oxygen-enriched environment (with DO concentrations of 6.4–11.6 mg/L), and high alkalinity of river water (with pH values of 7.5–10.2) (Figure 4). With sufficient DO in river water, bacterial aerobic denitrification is active with average N₂ fluxes about 25 mmol N₂-m⁻²-d⁻¹ at the water-air interface, playing a crucial role in nitrogen removal in the Chaobai River (He et al. 2016). Moreover, high pH values cause precipitation and co-precipitation of minerals which contribute to the decrease of cation and anion concentrations (Yang et al. 2016). Thus, biochemical reactions are the natural reasons for the improvement of water quality in cluster 3. In addition, hydrophytes including Phragmites australis, Potamogeton crispus, Hydrocharis dubia, and Iris tectorum are planted along the southeastern reach of the Chaobai River. These aquatic plants are widely applied for remediation of surface water and wastewater since they can assimilate nutrients and other pollutants in water (Xu et al. 2018; Zhao et al. 2020). Therefore, better water quality in cluster 3 not only demonstrated self-purification ability of river water, but also proved effectiveness of ecological restoration measures using aquatic plants.

Sampling sites S7 and S11 were in cluster 4, and they were in the southeastern reach and the northern reach of the Chaobai River, respectively (Figure 3). Although they were both far away from the reclaimed water outlet, the two sites were highly polluted. The mean values of water quality variables including pH, EC, TDS, K⁺, Mg²⁺, Na⁺, F⁻, Cl⁻, and TP were highest or second highest among the four clusters, reaching 9.5 ± 0.1, 715.5 ± 5.5 μS/cm, 325 ± 3, 16.5 ± 0.7, 23.3 ± 0.2, 111.1 ± 3.3, 0.82 ± 0.12, 104.6 ± 9.6, and 0.09 ± 0.04 mg/L, respectively. Since reclaimed water volume was limited and the two sites were not close to the reclaimed water outlet, reclaimed water was supposed to have less influence on chemical characteristics of river water at S7 and S11 (Martí et al. 2004). Moreover, sampling site S7 was in front of the Xiangyang Gate Dam and S11 was near the Suzhuang Dam (Figure 1), where residential and agricultural uses were close to the two dams. Thus, the water quality deterioration at S7 and S11 may be caused by domestic sewage and agricultural surface runoff containing potash fertilizer, phosphate fertilizer, fluoride and other minerals (Rothwell et al. 2010). It was estimated that around 1 million tons of untreated sewage, apart from reclaimed water, is discharged into the Chaobai River every year (Shen et al. 2018). Therefore, input of exogenous pollution instead of reclaimed water was the main factor affecting water quality at S7 and S11.

Principal component analysis and source identification

Significant correlations existed between many variables (p < 0.01 or p < 0.05) (Table 1), which indicated that the variance of each variable could be explained by others (Singh et al. 2004). PCA is most effective to reduce the dimensionality of the data set when close correlations are observed among variables (Primpas et al. 2010). Therefore, based on the correlation matrix of variables, PCA extracted three most representative PCs accounting for 81.5% of the total variance in the data set of river water quality. Eigenvalues markedly decreased after the first one (Table 2). Loadings, which are defined as the projections of variables on the axes of PCs, represented the correlation coefficients between three PCs and water quality parameters (Table 2). Absolute loading values of 0.3–0.5, 0.5–0.75, and >0.75 are defined as weak, moderate, and strong loadings, respectively.

PC1, which accounted for 46.9% of the total variance, was positively correlated with TDS, EC, Ca²⁺, Mg²⁺, Cl⁻, SO₄²⁻, and NO₃⁻-N, but negatively correlated with pH. Since most variables were significantly correlated with PC1, PC1 generally reflected the overall chemical characteristics of river water supplemented with reclaimed water. Noticeably, concentrations of these variables showing strong or moderate loadings in PC1 all changed significantly in spatial scale (Figure 4). According to previous studies in the same study area, denitrification and precipitation of minerals are active in the Shunyi reaches of the Chaobai River (He et al. 2016; Yang et al. 2016), and the two chemical reactions are closely related to the changes of pH values. Moreover, biomass of phytoplankton is high in reclaimed water with average algal density of 2991.18 × 10⁴ cells per liter (Li & Zhang 2011). Respiration and photosynthesis of phytoplankton influenced the equilibrium of carbonate (CO₃²⁻) and bicarbonate (HCO₃⁻), which also affected pH. 

Corrected Proof
Table 1 | Correlation matrix of 16 variables based on Kendall correlation analysis in Shunyi (Beijing) reaches of the Chaobai River, China

<table>
<thead>
<tr>
<th></th>
<th>Temp</th>
<th>EC</th>
<th>DO</th>
<th>TDS</th>
<th>pH</th>
<th>Ca²⁺</th>
<th>K⁺</th>
<th>Mg²⁺</th>
<th>Na⁺</th>
<th>F⁻</th>
<th>Cl⁻</th>
<th>SO₄²⁻</th>
<th>NO₃⁻N</th>
<th>NH₄⁻N</th>
<th>TN</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>0.183</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO</td>
<td>0.093</td>
<td>0.110</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDS</td>
<td>-0.183</td>
<td>0.999*</td>
<td>0.11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>0.183</td>
<td>-0.594*</td>
<td>-0.11</td>
<td>-0.564*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>-0.220</td>
<td>0.309</td>
<td>-0.147</td>
<td>0.309</td>
<td>-0.610*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K⁺</td>
<td>0.037</td>
<td>0.709**</td>
<td>0.257</td>
<td>0.709</td>
<td>-0.345</td>
<td>0.164</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>-0.037</td>
<td>0.709**</td>
<td>0.330</td>
<td>0.709</td>
<td>-0.345</td>
<td>0.091</td>
<td>0.927**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na⁺</td>
<td>-0.110</td>
<td>0.345</td>
<td>0.404</td>
<td>0.345</td>
<td>-0.055</td>
<td>-0.127</td>
<td>0.564</td>
<td>0.636*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F⁻</td>
<td>0.343</td>
<td>-0.208</td>
<td>0.438</td>
<td>-0.208</td>
<td>0.397</td>
<td>-0.585</td>
<td>0.019</td>
<td>0.094</td>
<td>0.549*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl⁻</td>
<td>-0.220</td>
<td>0.600*</td>
<td>0.367</td>
<td>0.603*</td>
<td>-0.455</td>
<td>0.127</td>
<td>0.745**</td>
<td>0.818**</td>
<td>0.600*</td>
<td>0.057</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO₄²⁻</td>
<td>-0.183</td>
<td>0.418</td>
<td>0.183</td>
<td>0.418</td>
<td>-0.782**</td>
<td>0.382</td>
<td>0.345</td>
<td>0.453</td>
<td>0.127</td>
<td>-0.170</td>
<td>0.627**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO₃⁻N</td>
<td>-0.093</td>
<td>0.514</td>
<td>-0.093</td>
<td>0.514</td>
<td>-0.587*</td>
<td>0.734**</td>
<td>0.294</td>
<td>0.220</td>
<td>-0.073</td>
<td>-0.572*</td>
<td>0.183</td>
<td>0.367</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH₄⁻N</td>
<td>-0.250</td>
<td>0.343</td>
<td>0.173</td>
<td>0.343</td>
<td>-0.496</td>
<td>0.420</td>
<td>0.038</td>
<td>0.038</td>
<td>-0.114</td>
<td>-0.119</td>
<td>0.114</td>
<td>0.458</td>
<td>0.443</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>-0.220</td>
<td>0.164</td>
<td>-0.440</td>
<td>0.164</td>
<td>-0.309</td>
<td>0.491</td>
<td>-0.127</td>
<td>-0.127</td>
<td>-0.127</td>
<td>-0.623*</td>
<td>-0.091</td>
<td>0.091</td>
<td>0.367</td>
<td>-0.114</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>-0.020</td>
<td>-0.117</td>
<td>-0.295</td>
<td>-0.117</td>
<td>0.428</td>
<td>-0.389</td>
<td>-0.156</td>
<td>-0.156</td>
<td>0.039</td>
<td>0.081</td>
<td>-0.195</td>
<td>-0.545</td>
<td>-0.373</td>
<td>-0.510</td>
<td>0.117</td>
<td>1</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).
values. Therefore, negative correlations existing between pH and other variables in PC1 (EC, TDS, Ca$^{2+}$, Mg$^{2+}$, SO$_4^{2-}$, and NO$_3$-N) were significant ($p < 0.05$ or $p < 0.01$; Table 1), showing that pH was indicative of variation of water chemistry (Jain 2002). PC2, with 22.4% of the total variance, had moderate and positive correlations with DO, K$^+$, and Na$^+$, and strong and positive correlation with F$^-$. Therefore, PC2 mainly represented mineral pollution. Concentrations of variables showing strong loadings in PC2 were relatively stable in river water (Figure 4). PC3, explaining 12.2% of the total variance, was positively correlated with TN and TP, but negatively correlated with NH$_4$-N. Hence, PC3 pointed to nutrient pollution in river water.

The ordination biplot of sampling sites, which was defined by PC1 (chemical characteristics of reclaimed water) and PC2 (mineral pollution), was shown (Figure 5). Distances among sampling sites were approximations of their Euclidean distances in multidimensional space. With the ordination biplot of sampling sites, PCA provided a classification of sampling sites based on the similarity of pollution sources, which supported the results of CA (Razmkhah et al. 2010). Higher site scores of sampling sites on PC suggested stronger influence of PC (Chounlamany et al. 2017). Sampling sites in cluster 1 (S1 and S2) and cluster 4 (S7 and S11) had the highest scores in PC1 and PC2 respectively, therefore the two groups of sampling sites were more seriously polluted. However, these sampling sites were mainly affected by different pollution sources since they had obvious spatial heterogeneity in the biplot defined by PCA (Figure 5; Razmkhah et al. 2010). As to sampling sites in cluster 2 (S3 and S8-S10), their scores in PC1 were lower than cluster 1, and their scores in PC2 were below cluster 4, therefore sampling sites in cluster 2 were poorly affected by both reclaimed water and other mineral pollution. Moreover, S4-S6 in cluster 3 had negative scores in both PC1 and PC2, indicating that the main pollution sources in the river have the weakest influence on these sampling sites. According to site scores and eigenvalues, total site scores were calculated by PCA and served as a new pollution index (Figure 6). Higher values of total site scores implied a higher degree of pollution conditions (Bu et al. 2016). It was obvious that sampling sites in the same cluster had similar values of total site score (Figure 6), indicating similar water quality. Moreover, total site scores defined by PCA also identified cluster 1 as the most polluted and cluster 3 as the least

### Table 2 | Principal component loadings of the 16 variables based on PCA in Shunyi (Beijing) reaches of the Chaobai River, China

<table>
<thead>
<tr>
<th>Variables</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>−0.304</td>
<td>0.302</td>
<td>0.132</td>
</tr>
<tr>
<td>DO</td>
<td>0.052</td>
<td>0.624</td>
<td>0.324</td>
</tr>
<tr>
<td>pH</td>
<td>−0.794</td>
<td>0.255</td>
<td>−0.230</td>
</tr>
<tr>
<td>TDS</td>
<td>0.823</td>
<td>−0.063</td>
<td>−0.221</td>
</tr>
<tr>
<td>EC</td>
<td>0.821</td>
<td>−0.023</td>
<td>−0.213</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td>0.710</td>
<td>−0.433</td>
<td>0.252</td>
</tr>
<tr>
<td>K$^+$</td>
<td>0.483</td>
<td>0.607</td>
<td>−0.235</td>
</tr>
<tr>
<td>Mg$^{2+}$</td>
<td>0.728</td>
<td>0.439</td>
<td>0.034</td>
</tr>
<tr>
<td>Na$^+$</td>
<td>0.452</td>
<td>0.702</td>
<td>−0.134</td>
</tr>
<tr>
<td>F$^-$</td>
<td>−0.123</td>
<td>0.843</td>
<td>0.060</td>
</tr>
<tr>
<td>Cl$^{-}$</td>
<td>0.684</td>
<td>0.461</td>
<td>−0.081</td>
</tr>
<tr>
<td>SO$_4^{2-}$</td>
<td>0.756</td>
<td>0.105</td>
<td>0.030</td>
</tr>
<tr>
<td>NO$_3$-N</td>
<td>0.727</td>
<td>−0.323</td>
<td>−0.234</td>
</tr>
<tr>
<td>NH$_4$-N</td>
<td>0.273</td>
<td>−0.076</td>
<td>−0.603</td>
</tr>
<tr>
<td>TN</td>
<td>0.313</td>
<td>−0.286</td>
<td>0.627</td>
</tr>
<tr>
<td>TP</td>
<td>−0.440</td>
<td>0.074</td>
<td>0.549</td>
</tr>
<tr>
<td>Eigenvalues</td>
<td>7.96</td>
<td>3.81</td>
<td>2.08</td>
</tr>
<tr>
<td>Proportion explained (%)</td>
<td>46.9</td>
<td>22.4</td>
<td>12.2</td>
</tr>
<tr>
<td>Cumulative proportion (%)</td>
<td>46.9</td>
<td>69.3</td>
<td>81.5</td>
</tr>
</tbody>
</table>

Bold values represent strong and moderate loadings (>0.50 or <−0.50).
polluted, the same results as CA. Water quality at S1 was identified as the worst with S1 reaching the highest in total site score. However, as the distances between sampling sites and the reclaimed water outlet increased, river water quality changed gradually. Total site scores provided additional evidence to the inference by CA that the hydrochemistry of river water was influenced by multiple factors.

Anthropogenic activities impose enormous changes on rivers, causing water quality deterioration and public health impairment (Razmkhah et al. 2010; Bu et al. 2016). Similar to this study, domestic and agricultural wastewaters are identified as an important contribution to river water pollution in many river basins (Jain 2002; Feher et al. 2016; Chounlamany et al. 2017). For example, Chounlamany et al. (2017) found that the use of fertilizers during the dry season and at the end of the rainy season accounts for NO$_3$ loading in a segment of Marikina River. Therefore, common measures, such as integrated real-time control technology used in river reaches with high incidence of water pollution (Meng et al. 2017), are applicable to manage river water quality in the study area. In addition, the Chaobai River is characterized by using reclaimed water as its major water source, which is quite different from ordinary surface waters. According to the above analysis, reclaimed water from the treatment plant contained high background values of nutrients and minerals, causing poor water quality on the upper reach of the Chaobai River. Moreover, a study carried out in the Feng-qing Lake, which is a landscape lake supplemented with reclaimed water, also demonstrated the close relationship between the supply of reclaimed water and the growth of phytoplankton (Zhao et al. 2015). Accordingly, in addition to monitoring and controlling water pollution along the river, it is also critical to improve water quality standards for reclaimed water and encourage the development of wastewater treatment technologies in the long run.

CONCLUSIONS

In this study, the spatial characteristics of water quality were evaluated in the Chaobai River using multivariate statistical methods. Eleven sampling sites were grouped into four clusters by CA based on pollution levels. Seven monitored water quality parameters, namely pH, TDS, EC, Ca$^{2+}$, Mg$^{2+}$, Cl$^-$, SO$_4^{2-}$, NO$_3$-N, and TN, showed significant spatial differences among four clusters ($p < 0.05$). Mean value of TN concentrations in the most polluted cluster was 22.3 mg/L, exceeding the threshold of the GB/T18921-2002 Standard. Therefore, nitrogen was the main pollutant in the study area. The water chemistry in the northern reach of the Chaobai River, which was excluded in previous studies due to only being in flow during the high-flow period, was distinct from the other river reaches, caused by the intermittent supply of reclaimed water. PCA reduced the data dimensionality by extracting three principal components to explain 81.5% of the total variance of the data set. Three principal components were related to the chemical characteristics of reclaimed water, mineral pollution, and nutrient pollution, respectively. Besides reclaimed water supplied from the wastewater treatment plant, untreated domestic and agricultural sewage were also the main pollution sources. Total site scores defined by PCA provided supporting information for the results of CA, showing sampling sites in the Jian River were most polluted. Therefore, the Jian River is a key area for water quality monitoring in the future.

Generally, this study distinguished pollution levels of different river reaches and identified the main pollutants in river reaches receiving reclaimed water, providing useful information for water quality management. Furthermore, this study proved that CA and PCA led to similar results, with CA presenting an overview of the spatial patterns of water quality and PCA providing supporting information.
details. However, the sampling period of this study was short, and results may be affected by occasional circumstances. Thus, long-term monitoring of the water quality in the Chaobai River is needed to ensure safe water and sound ecological status of the river after receiving reclaimed water.

ACKNOWLEDGEMENTS

This research was supported by the Beijing Natural Science Foundation (Grant No. 8172044) and the National Natural Science Foundation of China (Grant No. 41730749). The authors express sincere gratitude to Anran Liao, Zijuan Chen and Leixin Sang for their assistance during the fieldwork and laboratory analysis. The authors also thank two anonymous reviewers for their valuable comments.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES


Razmkhah, H., Abrishamchi, A. & Torkian, A. 2010 Evaluation of spatial and temporal variation in water quality by pattern


First received 4 December 2020; accepted in revised form 17 March 2021. Available online 25 March 2021