water

MDPI

Article

Use of Multivariate Statistical Techniques to Study Spatial Variability and Sources Apportionment of Pollution in Rivers Flowing into the Laizhou Bay in Dongying District

Jing Su [1,†], Yanhua Qiu [1,*,†], Yuling Lu [2], Xiaosong Yang [1] and Songyan Li [1]

[1] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China; sujing-418@163.com (J.S.); yangxiaosong15@mails.ucas.ac.cn (X.Y.); lisongyan17@mails.ucas.edu.cn (S.L.)
[2] Shandong Zhiteng Environmental Testing Co. Ltd., Dongying 257000, China; 3324343448@qq.com
* Correspondence: qiu_yan_hua@163.com
† These authors contributed equally to this work.

Abstract: Spatial variability and source apportionment of river pollution flowing into the Bohai Sea are of great significance to the pollution liability and development of control strategies to reduce the terrestrial discharge of pollution in the ocean. In this study, ten water quality variables from 14 monitoring sites in rivers flowing into Laizhou Bay were obtained to investigate the spatial variation and pollution sources in Dongying District from 2018–2019. The survey area was divided into a low pollution (LP) zone and a high pollution (HP) zone by cluster analysis based on ten indicators. Principle component analysis/factor analysis with a geographic information system was performed to identify the four main pollution sources in the survey area. Compared with the positive matrix factorization model, the absolute principal component score-multiple linear regression (APCS-MLR) model was more appropriate for the source apportionment of pollution in the surface water of Dongying District. The point source pollution of domestic sewage (23.6%) was the most crucial pollution source of water in the LP zone, followed by non-point pollution from agricultural activity (16.4%). The contribution rate in the HP zone analyzed by the APCS-MLR model followed a decreasing order: point source pollution from domestic sewage (28.5%) > non-point pollution source of overland runoff (14.8%) > point source pollution of hybrid wastewater (12.4%) > point source pollution from industries sewage (10.6%). Therefore, the spatial distribution and sources of pollution in the investigated area should be considered while developing control measures to reduce the discharge of pollution to Laizhou Bay.

Keywords: multivariate statistical techniques; spatial variation; source apportionment; APCS-MLR; PMF; Laizhou Bay

check for updates

Citation: Su, J.; Qiu, Y.; Lu, Y.; Yang, X.; Li, S. Use of Multivariate Statistical Techniques to Study Spatial Variability and Sources Apportionment of Pollution in Rivers Flowing into the Laizhou Bay in Dongying District. *Water* **2021**, *13*, 772. https://doi.org/10.3390/w13060772

Academic Editor: Bahram Gharabaghi

Received: 26 January 2021
Accepted: 8 March 2021
Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Bohai Sea, a semi-enclosed sea, is surrounded by land on three sides with poor water exchange conditions. As one of the Bohai Sea bays, Laizhou Bay is of ecological and economic importance because of its biological diversity and economic and environmental value. Numerous water pollution events have been reported in Laizhou Bay [1–3]. Spatiotemporal variation of water quality is available in large regions [4], but small-scale water quality variation has rarely been investigated. Little information is available in the literature on spatial variability and source apportionment of pollution in rivers flowing into Laizhou Bay. The current method to determine pollution liability is through the inspection of sewage outlets in China. Because of this, it is impossible to identify the source of pollution if the underground sewage outlets are relatively concealed. This has hindered the responsibility for pollution and the development of control measures to reduce land-based pollution discharged into Laizhou Bay.

The coastal areas of Laizhou Bay include Dongying City, Weifang City, and Yantai in Shandong Province. Among them, Dongying City, located in the northern part of Laizhou Bay, is well developed in both industry and agriculture. The surface water quality of rivers in Dongying City is a critical factor that specifies the ecological and environmental changes in Laizhou Bay. In 2008, approximately 2100 tons of chemical oxygen demand (COD), 393 tons of biochemical oxygen demand (BOD), 368 tons of ammonia nitrogen (NH_3–N), 85 tons of total phosphorus (TP), and 48 tons of total nitrogen (TN) were monitored in the Bohai Sea due to land-based emissions from Dongying City [5]. Moreover, the enterprise is distributed on both sides of the rivers in the city, including the salt chemical industry, petroleum exploitation, petrochemical industry, machine manufacturing, and metal processing [6]. Owing to this, the rivers flowing into Laizhou Bay in Dongying City inevitably contain pollutants [7,8]. Therefore, it is of great significance to carry out pollution accountability and formulate pollution control measures to explore the characteristics of spatial variability and pollution sources in the rivers in Dongying City.

The continued strengthening of water pollution prevention and control has improved the ability to monitor the quality of the water environment in coastal areas. Reliable water quality estimates are available for the continuous monitoring of river hydro-chemical and biological properties [9,10]. However, owing to a large number of existing databases, it has become challenging to interpret and draw meaningful conclusions. Fortunately, multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), and factor analysis (FA) can be used to simplify large and complex water quality datasets and better explain both temporal and spatial variations [11–15]. Receptor-based models, such as the absolute principal component score-multiple linear regression (APCS-MLR) and positive matrix factorization (PMF), have also been used to quantitatively analyze the contributions of pollution sources in rivers [16–18]. Combining the above methods would be expected to comprehensively address spatial variability and pollution sources in rivers.

In the present study, the ZhiMai River and GuangLi River flowing into Laizhou Bay in Dongying District were selected to investigate the spatial variability and pollution sources in surface water quality at a small scale using multivariate statistical techniques combined with a geographic information system (GIS). The aims were to (1) analyze the spatial variation of water quality in rivers, (2) identify potential factors influencing the water quality of rivers, and (3) reveal the potential pollution sources and quantify their contribution to the pollution of surface water in rivers. Insights from the present study could provide scientific guidance for pollution liability and the development of control measures by local conditions to reduce the discharge of pollution into Laizhou Bay.

2. Materials and Methods

2.1. Study Area

The Dongying District is located in the downtown area of Dongying city, Shangdong Province, and lies between latitudes 37°14′13″ N and 37°31′57″ N, and between longitudes 118°12′42″ E and 118°59′52″ E (Figure 1). Dongying District is 67.5 km in width from east to west and 26.5 km in length from north to south, with a total area of 1155.62 km^2. Land use types include cropland, forestland, grassland, wetland, water bodies (rivers, canals, ditches), anthropogenic areas, and barelands (Figure 2). Nine rivers are situated in the Dongying District, including a natural river (Yellow River) and eight drainage rivers or drainage canals. The GuangLi and ZhiMai rivers originate from the Yellow River and flow into the Bohai Sea via the Laizhou Bay. The GuangLi River has several tributaries, including the Laoguangpu, Liugan, and Wugan rivers. The ZhiMai River consists of the Xinguangpu and Wujia rivers. The Dongying River flows into the Yihong River and then into the Guangli River (Figure 1).

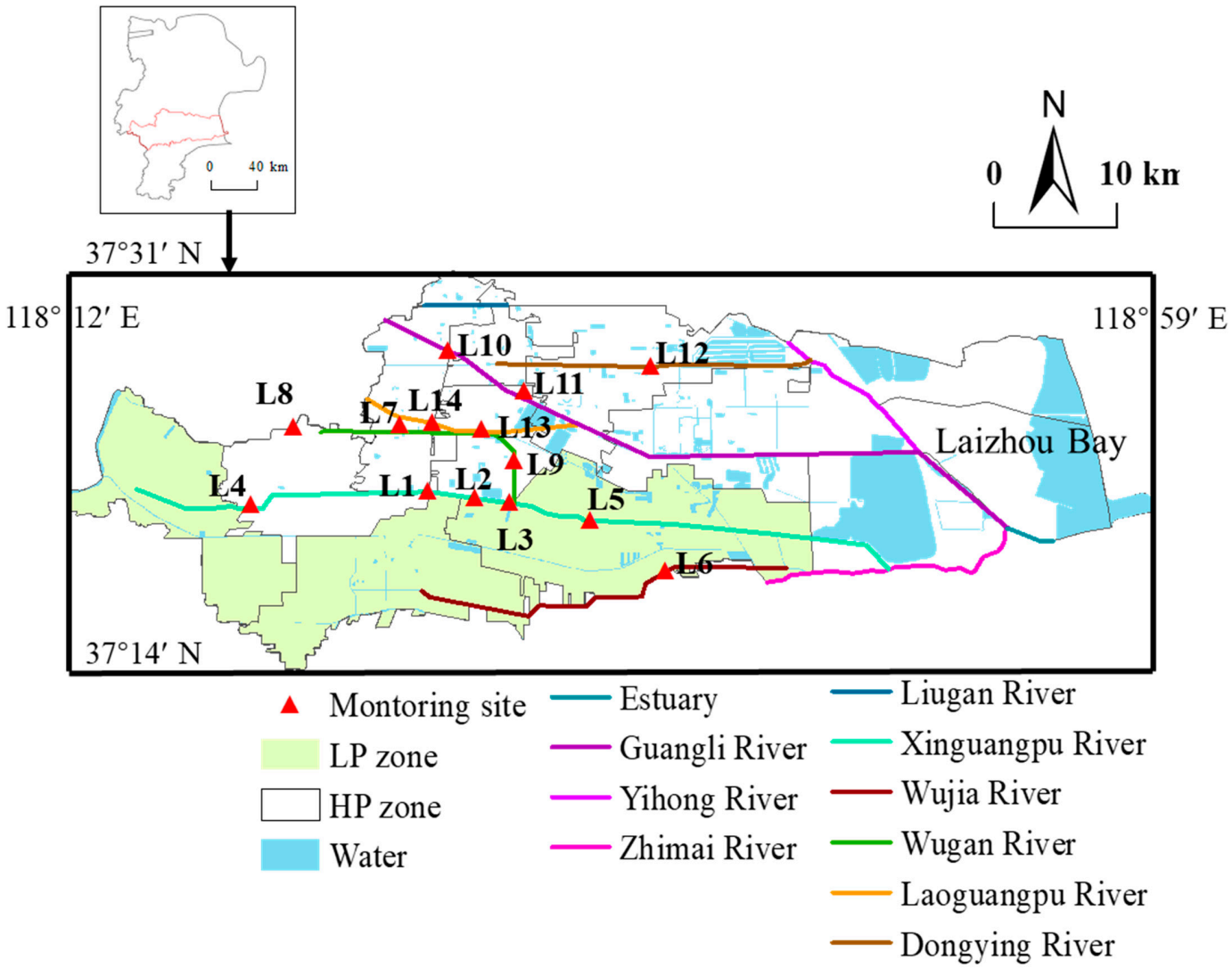


Figure 1. Map of the rivers and monitoring sites for water quality variables in Dongying District.

Dongying District has a warm temperate continental monsoon climate. The average annual temperature of Dongying District was 14.7 °C. The lowest temperature often occurs in January, and the hottest months are usually July and August. The average annual precipitation is 472.8 mm, and the average annual evaporation is 1900.8 mm.

2.2. Sample Collection and Analysis

Water samples were collected from 14 monitoring sites along the tributaries of the GuangLi and ZhiMai rivers from May 2018 to July 2019 (Figure 1 and Table S1). The meteorological conditions at the time of sampling are presented in Figure S1. Twenty-four samples were collected every 15 days (twice a month) from each site on the river axis. A bucket was used to collect 1 L of water samples from 10 to 20 cm at each sampling point. The samples were then placed in a polyethylene bottle and brought back to the laboratory for freezing. The selected chemical variables that could affect the quality of surface water included pH, dissolved oxygen (DO), BOD, COD, NH_3-N, TN, TP, permanganate index (PI), petroleum hydrocarbon (PHC), and the number of fecal coliforms (*F.coli*). These variables were determined based on GB3838-2002 [19] and are described briefly as follows. The pH value was measured using the glass electrode method, while DO was determined using the electrochemical probe method, and BOD was determined by dilution and inoculation. The COD was measured using the dichromate method, while PI was determined using

the potassium permanganate method. Nessler's reagent spectrophotometry was used to determine the NH_3-N, TN content was determined by UV spectrophotometry using alkaline potassium persulfate digestion, and TP was determined using ammonium molybdate spectrophotometry. The PHC was determined using infrared spectrophotometry, while *F.coli* was determined using the multi-tube fermentation method. It should be noted that water temperature and turbidity did not need to be determined in this study. This is because water temperature results from the effect of natural variations on water quality, including the season, and turbidity is related to other water quality indicators monitored in this study, including organic pollutants, inorganic pollutants, and microorganisms. The seasonal variation was not obvious in this study because the pollution of the city was mainly affected by the discharge of wastewater but not by rainfall, thus, these factors were excluded.

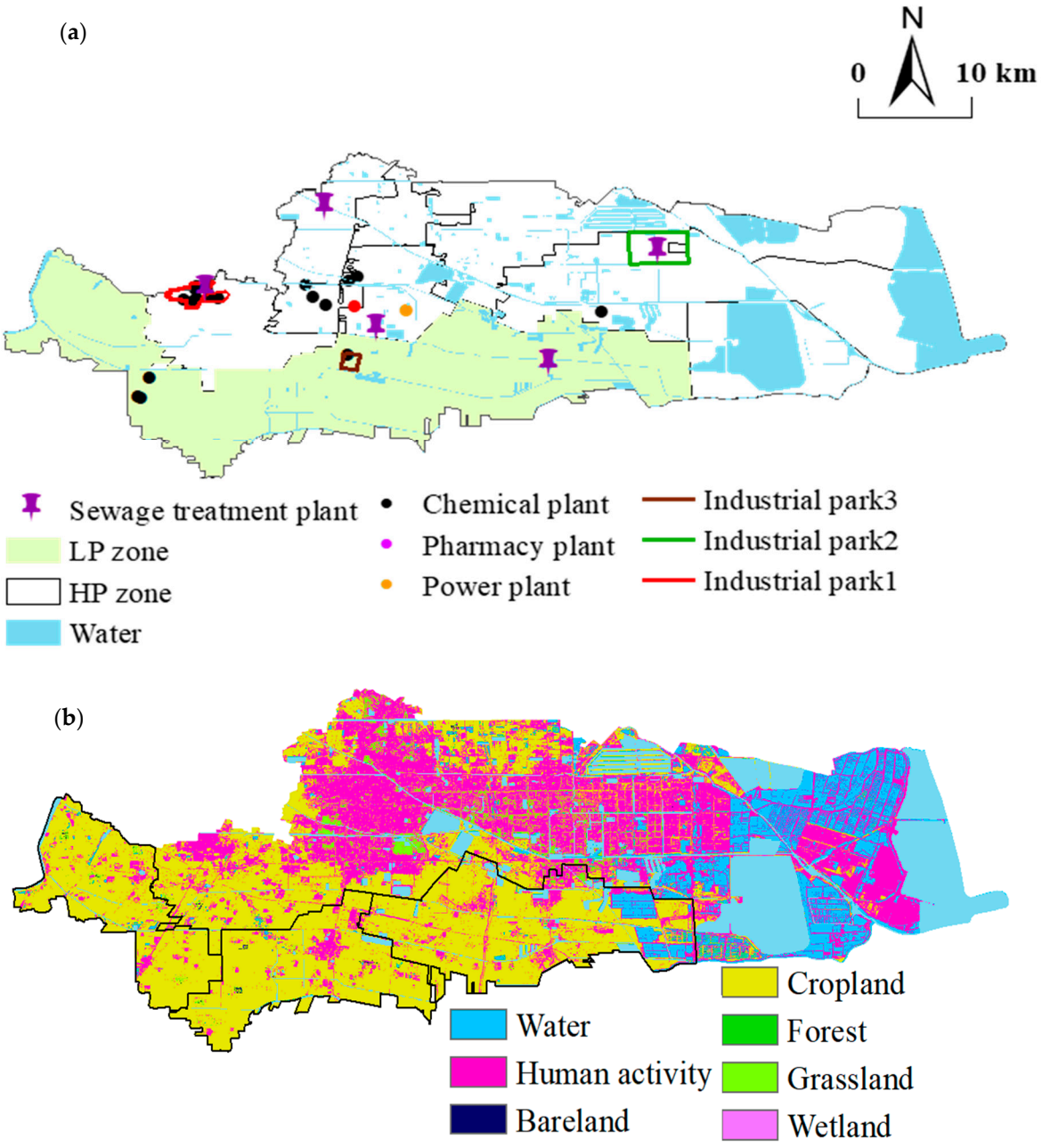


Figure 2. Maps of industry and sewage treatment plants (**a**), and land use types (**b**) in Dongying District.

The GIS data include land use types and the distribution of industrial parks, plants, and sewage. Landsat Thematic Mapper images with 30 m resolution for 2015 were used to map land use in the study area (Figure 2b). Land use data were obtained from Tsinghua University (http://data.ess.tsinghua.edu.cn/fromglc2015_v1.html, accessed on 29 May 2019). The distribution of industrial parks, plants, and sewage in 2019 was obtained by the Environment Protection Ministry (Figure 2a and Table S2). The animal husbandry Ministry of Dongying District provided information on livestock and poultry in 2019 (Table S3).

2.3. Data Treatment and Multivariate Statistical Techniques

2.3.1. Pollution Distribution Based on the Cluster Analysis (CA)

Cluster analysis was used to identify different spatial distribution characteristics based on distance or similar variables [20]. As one of the most common CA methods, hierarchical cluster analysis was adopted to reduce the dimensionality of the original data dramatically in this study [21]. The Ward method was chosen to perform clusters with minimum intra-cluster variance, and squared Euclidean distances were applied to act as similarity indicators.

2.3.2. Determination of the Critical Factors of Water Quality Changes Based on Principal Component Analysis (PCA)/Factor Analysis (FA)

Principal component analysis and FA were conducted to analyze the key factors of water quality changes in each zone divided by CA and identify the primary pollution sources. The principal component analysis is a multivariate statistical method that transforms correlated original variables into new and linearly uncorrelated variables to reduce the dimensions of the original dataset [22]. Factor analysis further reduces the contribution of the less significant variables to simplify the data structure by rotating the axis defined by the PCA. From this, the new variables called varifactors (VF) are constructed to explain unobservable, hypothetical, and latent sources that lead to surface water pollution. An eigenvalue greater than one is regarded as a component.

2.3.3. Estimating the Contribution of Pollution Sources Based on the APCS-MLR Model

As a receptor model, the APCS-MLR model can quantitatively describe the contribution of pollution sources to the water quality indicators, which has been widely used in the analysis of water environmental pollution sources in recent years [16,18], and is expressed as follows in Equation (1):

$$C_{ik} = r_{i0} + \sum_{j=1}^{p} r_{ij} \times APCS_{jk} \quad (1)$$

where i is the number of water quality variables in the analysis, k is the number of observations, j is the number of the components influencing the water quality variable, C_{ik} is the concentration of the ith parameter during observation k, r_{i0} is the constant obtained by the multiple linear regression (MLR) component of the analysis of the water quality variable i, r_{ij} is the regression coefficient of pollution source j to variable i, and $APCS_{jk}$ is the absolute component score for the jth component, which was obtained by Equations (2)–(6) [23].

$$(Z_k)_i = \frac{C_{ik} - \overline{C_i}}{\sigma_i} \quad (2)$$

$$(Z_0)_i = \frac{0 - \overline{C_i}}{\sigma_i} \quad (3)$$

$$(A_z)_{jk} = \sum_{j=1}^{p} S_{ij} \times (Z_k)_i \quad (4)$$

$$(A_0)_j = \sum_{j=1}^{p} S_{ij} \times (Z_0)_i \quad (5)$$

$$APCS_{jk} = (A_z)_{jk} - (A_0)_{jk} \quad (6)$$

where $(Z_k)_i$ is the standardized value of the *i*th variable on the *k*th observation, C_i and σ_i are the average concentration and standard deviation of the *i*th variable, respectively; S_{ij} is the score coefficient of the *j*th component for the *i*th variable, $(Z_0)_i$ is the standardized value of absolute zero concentration, $(A_z)_{jk}$ is the component score of the *j*th component, $(A_0)_i$ is the component score of absolute zero concentration, and $APCS_{jk}$ is the absolute component score for the *j*th component.

A negative value may be obtained in the APCS-MLR model, which results in the loss of pollution source analysis accuracy. According to Gholizadeh et al. [10], a measure of converting negative to positive values was used to calculate the contribution rate of pollution sources to water quality variables.

2.3.4. Estimating Contribution of Pollution Sources Based on the PMF Model

Similar to the APCS-MLR model, the PMF model, first proposed by Paatero et al. [24] as a receptor model, is another multivariate statistical analysis tool and has been widely used for the identification of pollution sources. Based on a weighted least-squares fit, the PMF model decomposes the original data matrix into a factor score matrix, factor loading matrix, and residual matrix, which was expressed as follows in Equation (7):

$$X_{ij} = E_{ij} + \sum_{k=1}^{p} \left(G_{ik} F_{kj} \right) \quad (7)$$

where X_{ij} is the original data matrix containing the concentration of the *j*th parameter in the *i*th sample, G_{ik} is the matrix with the contributions of *k*th source pollution to the *i*th sample, F_{kj} is the factor loading matrix for with the *k*th factor of the *j*th parameter, and E_{ij} is the residual matrix for the *j*th parameter in the *i*th sample.

The optimal matrices G and F are obtained by minimizing the objective function Q based on the weighted least-squares method for limiting and iterative calculations (Equation (8)):

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{e_{ij}}{u_{ij}} \right)^2 \quad (8)$$

where u_{ij} is the uncertainty of the *j*th variable in the *i*th sample. In addition, an additional uncertainty of 20%, as suggested, was considered in this study to represent the uncertainty associated with field sampling operations, such as sample collection, transport, and storage [25,26]. The number of factors is an important parameter for obtaining optimal results in the PMF model process. In this study, the number of factors was determined by PCA/FA results to compare between the APCS-MLR and PMF models [10]. Further detailed information can be found in the EPA PMF 5.0 User Guide.

2.4. Statistical Analysis

Excel 2016 and SPSS 22 were used to process and analyze the data. The non-parameter test, Kruskal–Walls ANOVA, was used to compare different zones divided by CA. SPSS 22 was used to perform CA, PCA, and ACPS-MLR, while EPA PMF 5.0 was used to handle the PMF model. The resulting data are shown as a mean ± standard deviation. ArcGIS 10.6 was used to handle the map of land use and industrial distribution.

3. Results

3.1. Characteristics of the Water Pollution

Table 1 describes the annual mean of ten variables at the 14 monitoring sites (twenty-four samples in each site). The pH ranges from 7.00 to 9.34, while the concentration of DO varied between 3.07 and 9.01 mg/L, and the concentration of BOD ranged from 2.3 to 10 mg/L. The COD concentrations ranged from 13 to 51 mg/L, while the concentration of NH_3-N varied between 0.038 and 1.98 mg/L and the concentrations of TN and TP ranged from 13–51 and 0.013–0.9 mg/L, respectively. The concentration of PI ranged from 1.88 to 14.9 mg/L, and the concentration of PHC varied between 0.005 and 0.49 mg/L. These results indicate that the surface water in Dongying District was polluted to some degree despite the fact that pH, PI, and PHC in all samples satisfied grade IV according to the GB3838-2002 standard [19]. The concentrations of BOD, COD, NH_3-N, TP, and TN exceeded grade IV of the national surface water GB 3838-2002 in some samples [19]. Overall, the pollution caused by organic pollutants appeared to be the most serious. Furthermore, the amount of *F.coli* in some samples was much higher than the 20,000 MPN/L threshold of grade IV [19].

Table 1. Statistical description of water quality indicators in Dongying District during 2018–2019.

Site	pH	DO	BOD	COD	NH_3-N	TP	TN	PI	PHC	*F.coli*
		mg/L								MPN/L
L1	7.72 ± 0.29	6.05 ± 0.81	7.49 ± 1.48	34.00 ± 5.22	0.87 ± 0.38	0.22 ± 0.10	1.43 ± 0.27	5.97 ± 3.26	0.048 ± 0.074	3246 ± 3065
L2	7.77 ± 0.29	5.51 ± 0.98	7.37 ± 1.60	32.79 ± 3.75	0.78 ± 0.33	0.25 ± 0.09	1.43 ± 0.23	6.07 ± 3.08	0.037 ± 0.058	5513 ± 4602
L3	7.87 ± 0.27	5.58 ± 1.19	7.37 ± 1.44	32.79 ± 5.75	0.81 ± 0.43	0.22 ± 0.09	1.43 ± 0.33	6.04 ± 3.38	0.032 ± 0.05	4679 ± 4066
L4	7.84 ± 0.33	6.09 ± 1.19	7.65 ± 1.56	32.25 ± 1.56	0.86 ± 0.52	0.25 ± 0.10	1.42 ± 0.31	5.7 ± 3.37	0.058 ± 0.096	3330 ± 1885
L5	7.77 ± 0.28	6.25 ± 0.87	6.72 ± 1.69	31.92 ± 6.93	0.81 ± 0.42	0.25 ± 0.10	1.46 ± 0.24	5.41 ± 2.79	0.058 ± 0.093	3163 ± 2293
L6	7.73 ± 0.31	5.37 ± 0.95	7.14 ± 1.33	32.05 ± 5.82	1.07 ± 0.41	0.27 ± 0.10	1.37 ± 0.19	6.97 ± 3.05	0.045 ± 0.076	2851 ± 1402
L7	7.71 ± 0.18	5.32 ± 1.31	7.33 ± 1.31	32.13 ± 5.81	1.09 ± 0.48	0.30 ± 0.19	1.52 ± 0.28	6.78 ± 2.88	0.071 ± 0.196	3815 ± 2959
L8	7.76 ± 0.28	5.25 ± 1.25	7.30 ± 1.26	32.75 ± 4.90	1.03 ± 0.42	0.24 ± 0.11	1.42 ± 0.28	7.19 ± 3.32	0.049 ± 0.077	5311 ± 3868
L9	7.80 ± 0.21	5.70 ± 1.29	7.42 ± 1.27	33.46 ± 5.41	1.08 ± 0.55	0.27 ± 0.09	1.39 ± 0.28	6.31 ± 3.20	0.049 ± 0.081	5314 ± 4491
L10	7.69 ± 0.46	5.14 ± 0.98	8.16 ± 1.05	34.71 ± 3.71	1.08 ± 0.37	0.27 ± 0.09	1.57 ± 0.31	7.16 ± 3.05	0.071 ± 0.121	3343 ± 2473
L11	7.70 ± 0.41	6.05 ± 0.83	7.73 ± 1.27	33.70 ± 3.46	1.31 ± 0.42	0.25 ± 0.12	1.40 ± 0.25	7.09 ± 2.99	0.043 ± 0.061	4391 ± 2922
L12	7.88 ± 0.61	5.57 ± 1.14	7.38 ± 1.10	33.42 ± 2.65	1.14 ± 0.50	0.21 ± 0.09	1.32 ± 0.31	6.53 ± 2.58	0.037 ± 0.058	3752 ± 3493
L13	7.90 ± 0.38	5.47 ± 1.22	7.57 ± 1.17	33.29 ± 4.37	0.96 ± 0.45	0.29 ± 0.10	1.44 ± 0.30	7.05 ± 3.15	0.056 ± 0.085	4210 ± 3752
L14	7.95 ± 0.35	5.48 ± 1.30	7.78 ± 1.13	33.29 ± 5.00	1.00 ± 0.46	0.25 ± 0.10	1.43 ± 0.23	7.22 ± 3.01	0.068 ± 0.118	5377 ± 3812

Note: All data is mean ± standard deviation for the annual value of the bi-monthly sampling events.

3.2. Characteristics of the Spatial Distribution

The average concentration of the sample site was used to perform CA to gain insights into the distribution characteristics and trace the pollution sources. Figure S2 depicts the dendrogram rendered by the CA, All 14 river sampling sites were divided into two statistically significant clusters: cluster 1, consisting of L1, L2, L3, L4, and L5; and cluster 2, consisting of L6, L7, L8, L9, L10, L11, L12, L13, and L14. The spatial variations of clusters 1 and 2 are displayed in Figure 3. There was a significant difference in water pollution at the monitoring sites between both clusters. Compared to cluster 1, the concentrations of BOD, COD, NH_3-N, PPI, and *F.coli* were significantly higher in cluster 2. However, the DO of cluster 1 was significantly superior to that of cluster 2. Thus, clusters 1 and 2 corresponded to relatively low pollution and very high pollution, respectively, and could be defined accordingly as the low pollution (LP) and high pollution (HP) zones. The LP zone is distributed in the southwestern Dongying District, while the HP zone is distributed in the northeastern Dongying District (Figure 1). In addition, Figure S3 displays the maps of spatial variation of the annual mean of ten water quality indicators before standardization in Dongying District. The results shown in Figure S3 are consistent with those of the CA. The wet and dry season variations were not obvious in this study because the pollution of the city was mainly affected by human activities and not by rainfall.

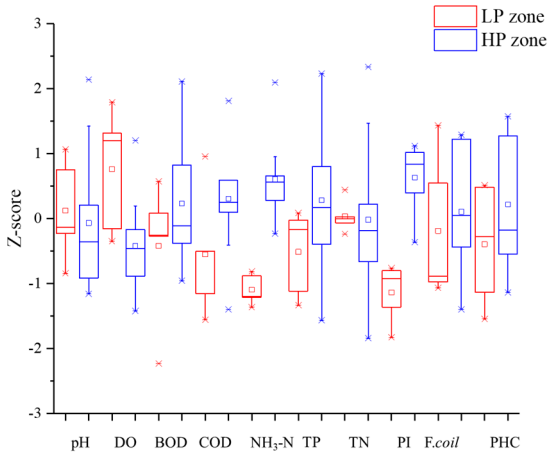


Figure 3. Spatial variation of water indictors in Dongying District.

3.3. Number of Pollution Sources Identified Based on the PCA Analysis

For PCA/FA, the Kaiser–Meyer–Olkin (KMO) test was used to verify the sampling adequacy. When KMO value was significantly higher than 0.5 [27], PCA/FA could be used for further analysis. In the two pollution zones, the KMO values were 0.577 (LP zone) and 0.673 (HP zone), with p values close to zero, respectively. Thus, PCA/FA could be used significantly to reduce the initial database dimensionality and was valid for source identification.

Based on the PCA/FC results (Table 2), four factors in the LP zone could explain 63.5% of the total variance. Factor 1 (VF1) contributed approximately 23.6% of the total variance and showed a strongly positive load on PI and PHC, described as follows. The Dongying District is a typical petrochemical city. Figure 2a indicates that the industrial park3 and some petrochemical industries are located in the LP zone, and wastewater from these industries can be discharged into adjacent rivers. Furthermore, PI is an indicator of organic/inorganic matter pollution. Therefore, VF1 represents point source pollution from the discharge of sewage in the petrochemical industry. Factor 2 (VF2, 15.4% of the total variance) showed a moderately positive load on NH_3-N and TN. As shown in Figure 2b, cropland was predominantly located in the LP zone, thus nitrogen fertilizer is likely the major source of N in rivers [28–30]. According to the Statistical Yearbook in 2019, approximately 2600 tons of nitrogen fertilizers was applied to croplands in Dongying District. Therefore, VF2 could be explained by the non-point source pollution of agricultural activity due to fertilizers. Factor 3 (VF3, 14.1% of the total variance) showed a moderately positive load on TP and *F.coli*. According to the information provided by the Animal Husbandry Ministry of Dongying District in 2019 (Table S3), livestock farms were mainly located in the LP zone, thus livestock manure could be the dominant source of bacteria in adjacent rivers. Further, there were no sewage or manure treatment facilities in rural areas, resulting in the direct discharge of pollutants. Moreover, fecal pollution is associated with soil erosion and surface runoff, and non-point sources tend to contribute significantly to the fecal pollution of surface waters [31]. Consequently, VF3 represents the fecal pollution sources with points and non-points. Factor 4 (VF4, 10.3% of the total variance) exhibited a moderately positive load on BOD which is a comprehensive indicator of organic pollutants [18]. As shown in Figure 2b, the rural area was distributed in the LP zone, thus organic matter from untreated domestic wastewater could be discharged into adjacent rivers. From this, VF4 represents the point source of pollution of the domestic sewage discharge.

Table 2. Loadings, eigenvalue, percentage of variance, cumulative percentage variance of ten selected variables on varimax rotated factors in low pollution (LP) and high pollution (HP) zones.

Variable	LP Zone				HP Zone			
	VF1	VF2	VF3	VF4	VF1	VF2	VF3	VF4
pH	−0.241	−0.575	0.223	0.362	−0.536	0.209	0.513	−0.109
DO	−0.449	−0.361	0.049	0.525	−0.601	0.397	0.167	0.079
BOD	0.278	0.150	0.072	0.775	0.109	−0.029	0.858	0.074
COD	0.318	−0.144	−0.718	0.047	0.065	−0.289	0.015	−0.695
NH_3-N	0.139	0.776	0.081	0.002	0.002	−0.363	0.084	0.647
TP	0.031	−0.278	0.744	0.259	−0.232	0.678	−0.090	−0.155
TN	−0.250	0.583	−0.009	0.092	0.149	0.708	0.121	0.174
PI	0.836	0.009	−0.117	0.036	0.812	−0.045	0.100	−0.041
F. Coli	0.349	0.164	0.551	−0.479	0.623	0.088	−0.147	0.326
PHC	0.830	−0.003	−0.011	0.034	0.763	0.062	0.216	−0.205
Eigenvalue	2.364	1.542	1.409	1.033	2.503	1.327	1.125	1.074
Variance	23.642	15.417	14.094	10.327	25.028	13.270	11.252	10.741
Cumulative variance	23.642	39.059	53.153	63.480	25.028	38.298	49.550	60.291

In the HP zone, four significant factors accounted for 60.3% of the total variance in the dataset. Among these factors, VF1 (25.0% of the total variance) showed a strongly positive load on PI and PHC, moderate load on *F.coli*, and negative load on pH and DO, which represented the point source pollution of hybrid sewage discharged into the river from the outlet (Table S2). VF2 in the HP zone (13.3% of the total variance) showed a moderately positive load on TP and TN, representing the influence of non-point source pollution of soil erosion due to urbanization. Accordingly, many researchers have found that urbanization activities increase N and P in surface water [32–34]; furthermore, soil erosion in this area has been severe owing to rapid urban development over the past decade [34,35]. Similar to VF4 in the LP zone, VF3 (11.7% of the total variance) showed a strongly positive load on BOD, which might be attributed to the point source discharge of municipal wastewater. VF4 (10.7% of the total variance) had a moderately positive load on NH_3-N and negative loading on COD. The COD originated primarily from industrial wastewater. Industrial park 1 (chemical industries) and industrial park 2 (manufacturing) were mainly distributed along the riverbank of Dongying District (Figure 2b), producing a large amount of acid-containing and organic wastewater. The untreated sewage from these plants exhibited high COD. Furthermore, it is known that a large amount of NH_3-N can arise from the decomposition of organic pollutants [10,17,36]. Thus, VF4 represents the effect of point source pollution from this industrial sewage.

3.4. Contributions of Pollution Sources Based on the APCA-MLR Model

The APCA-MLR model was used to quantify the contribution of each pollution source to water quality. According to the results shown in Table S4, the determination coefficient (R^2) of MLR was greater than 0.5, indicating a good consistency between the predicted and measured values by line fitting, and that the source apportionment results were reliable [10]. The ratio of the average predicted and measured values was close to 1, which showed that the APCS-MLR receptor model had good applicability for source apportionment. Figures 4 and 5 show the contribution of pollution sources for the ten studied variables and the average contribution of pollution sources based on the APCS-MLR model in the LP and HP zones.

In the LP zone, most contributions were predominately influenced by point source pollution of domestic sewage (S4, 23.6%), shown as DO (56.5%) and BOD (66.6%). In addition, non-point source pollution from agricultural activity (S2) accounted for 16.4% of the total pollution sources as NH_3-N (38.1%) and TN (26.9%). The average contribution of the pollution source of fecal matter (S3) and point source pollution of petrochemical sewage discharge (S1) accounted for 7.8% and 3.4%, respectively, which were significantly lower than those of S2 and S4. Therefore, the contributions of pollution sources in the LP zone should be ranked in the decreasing order as follows: point source pollution of

domestic sewage > non-point source pollution of agricultural activity> pollution source of fecal matter > point source pollution of petrochemical sewage.

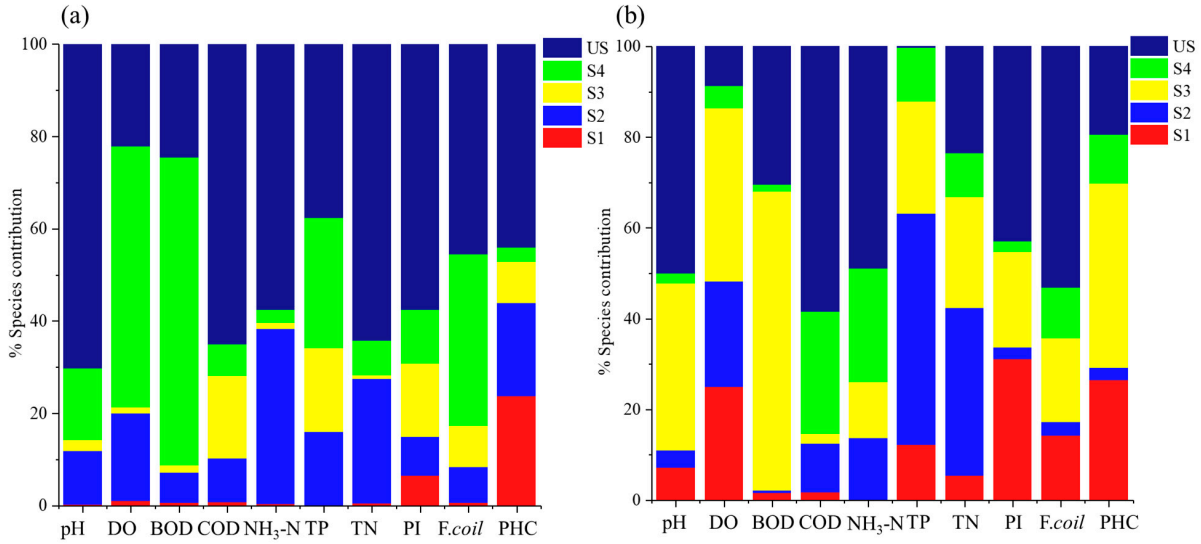


Figure 4. Contributions of pollution sources for ten studied variables estimated based on the absolute principal component score-multiple linear regression (APCS-MLR) model in the LP (**a**) and HP (**b**) zones. The S1, S2, S3, S4, and US in the LP zone were identified as point source pollution from petrochemical sewage, non-point source pollution from agricultural activity, pollution source of fecal matter, point source pollution of domestic sewage, and one undefined source, respectively. The S1, S2, S3, S4, and US in the HP zone were identified as point source pollution of hybrid wastewater discharged, non-point source of pollution of the overland runoff, point source pollution of domestic sewage, point source pollution of industrial sewage, and one undefined source, respectively.

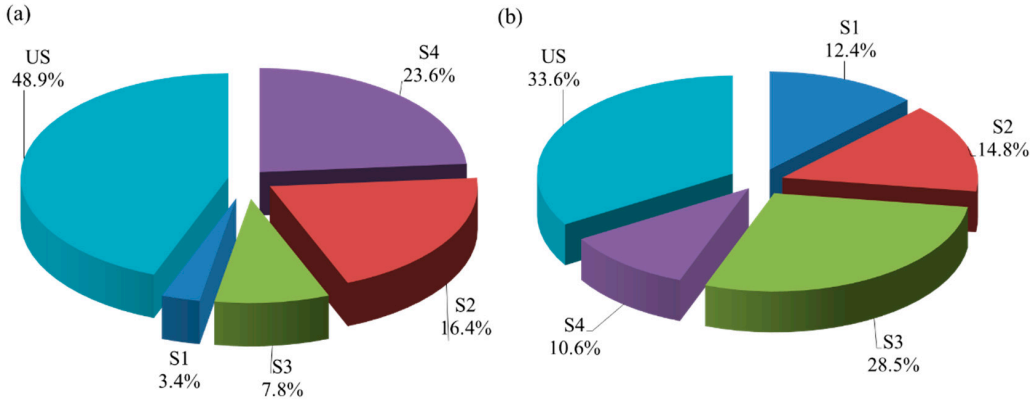


Figure 5. Average contributions of pollution sources to the water quality based on the APCS-MLR model in the LP (**a**) and HP (**b**) zones. The S1, S2, S3, S4, and US in the LP zone were identified as point source pollution of petrochemical sewage, non-point sources pollution from agricultural activity, pollution source of fecal matter, point source pollution of domestic sewage, and one undefined source, respectively. The S1, S2, S3, S4, and US in the HP zone were identified as point source pollution of hybrid wastewater discharged, non-point source of pollution of overland runoff, point source pollution of domestic sewage, point source pollution of industrial sewage, and one undefined source, respectively.

In the HP zone, the point source discharges of municipal domestic wastewater (S3) contributed the most at 28.5%, shown as BOD (65.8%). Non-point source pollution from overland runoff (S2) accounted for 14.8% of the total pollution sources, represented as TP (50.9%). The point source pollution of hybrid wastewater discharged to the river by outlets (S1) accounted for 12.4% of the total pollution sources, represented as PI (31.1%) and DO (24.9%). The average contribution of point source pollution from industrial sewage (S4) accounted for 10.6%, shown as COD (58.5%) and NH_3-N (48.9%). S4 was significantly lower

than those of S2 and S3. This could contribute to the strict monitoring and management of sewage plants. The HP zone's pollution sources might follow the decreasing order as follows: point source pollution from domestic sewage > non-point source of pollution of the overland runoff > point source pollution from hybrid wastewater discharged into the river by outlets > point source pollution from industrial sewage.

The APCS-MLR model results showed that point source pollution was the most critical source of surface water pollution in Dongying District, followed by non-point pollution from agriculture and overland runoff. In general, the LP zone was mainly affected by point source pollution from untreated domestic sewage, followed by non-point source pollution from agriculture. Therefore, it is essential to strengthen the infrastructure construction of sewage treatment facilities in rural areas. Additionally, the use of fertilizers and pesticides on croplands needs to be controlled. In the HP zone, industrial and domestic wastewater may be one of the most critical pollutant sources. This study has found that water quality is influenced by non-point source pollution of overland runoff caused by urbanization. Therefore, the monitoring of sewage treatment plants and factory discharge outlets should be strengthened, and substandard wastewater discharge should be prohibited in urban areas. In the progress of urbanization, the green area should be increased to reduce soil loss.

3.5. Contribution of Pollution Sources Based on the PMF Model

The "Factor Figureprints" of water quality variables are depicted in Figure 6 to indicate the contribution of pollution sources based on the PMF model. In the LP zone, the contributions of source 1 (S1) to COD and PI were 47.5% and 45.9%, respectively. Therefore, S1 was identified as the source of industrial sewage pollution. Source 2 (S2) was mainly characterized by a PHC of 88.8%, which could be considered a point source of wastewater discharged from petrochemical enterprises. Source 3 (S3) was dominated by *F.coli* (100%), NH_3-N (69.1%), and TN (65.0%), which could be considered to represent the non-point source pollution of overland runoff due to agricultural activity. The contribution of source 4 (S4) to TP, DO, BOD, and pH was 72.0%, 56.0%, 53.9%, and 47.5%, respectively. Thus, S4 represents the point source pollution of the domestic sewage.

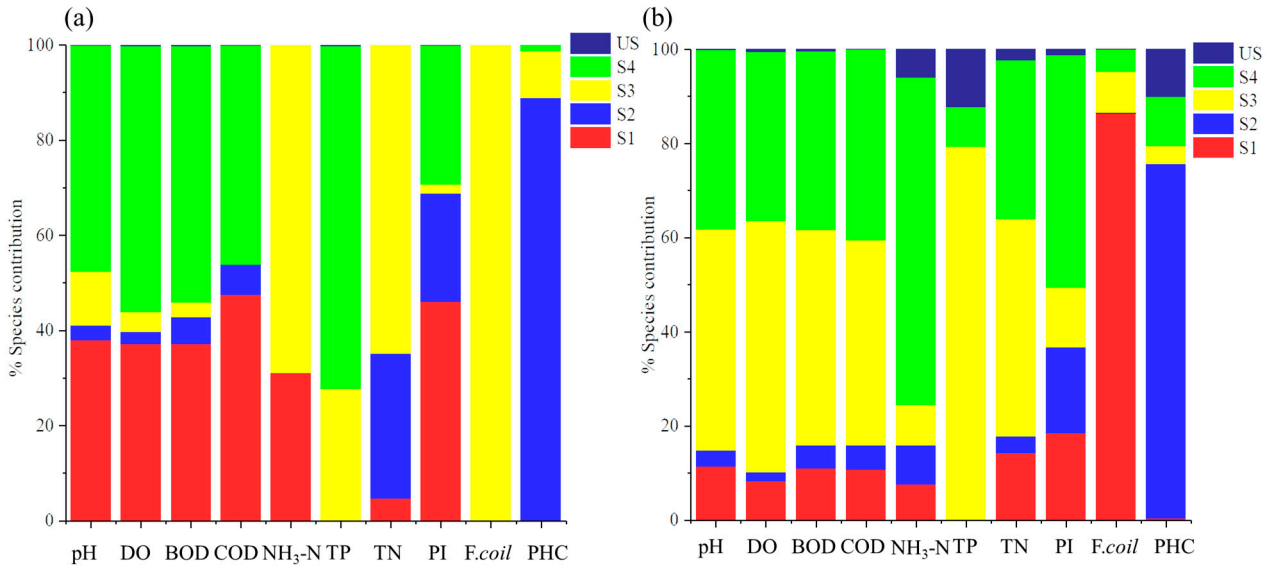


Figure 6. Contributions of pollution sources for ten studied variables based on the positive matrix factorization (PMF) model in the LP (**a**) and HP (**b**) zones. The S1, S2, S3, S4, and US in LP zone were identified as point source pollution of industrial sewage, point source of petrochemical enterprises wastewater, non-point source pollution of overland runoff, point source pollution of domestic sewage, and one undefined source, respectively. The S1, S2, S3, S4, and US in the HP zone were identified as non-point source pollution of overland runoff, point source of petrochemical enterprises wastewater, point source pollution of domestic sewage, point source pollution of industrial sewage, and one undefined source, respectively.

In the HP zone, S1 was dominated by *F.coli* (86.4%), which is considered to represent the non-point source pollution of overland runoff. The contribution of S2 to the PHC was 75.3%. Therefore, S2 represents the point source pollution of wastewater discharged from petrochemical enterprises. S3 was mainly characterized by a TP of 79.2%, DO of 53.4%, pH of 47.0%, TN of 46.2%, and BOD of 45.7% in the HP zone. Therefore, S3 could be considered as a point source of pollution in domestic sewage. The contributions of S4 to NH_3-N and PI were 69.6% and 49.5%, respectively, thus, S4 represents the point source pollution of industrial wastewater.

The suggested average contribution of pollution sources is shown in Figure 7, illustrating that the point source pollution of industrial sewage, point source of wastewater discharged from petrochemical enterprises, non-point source pollution of overland runoff, point source pollution of domestic wastewater, and unidentified sources for different variables in the LP zone were 22.5%, 17.4%, 31.2%, 28.7%, and 0.1%, respectively. The suggested average contributions of the non-point source pollution of overland runoff, point source pollution of wastewater discharged from petrochemical enterprises, point source pollution of domestic sewage, point source pollution of industrial sewage, and unidentified sources for different variables in the HP zone were 16.8%, 12.1%, 34.9%, 32.9%, and 3.4%, respectively.

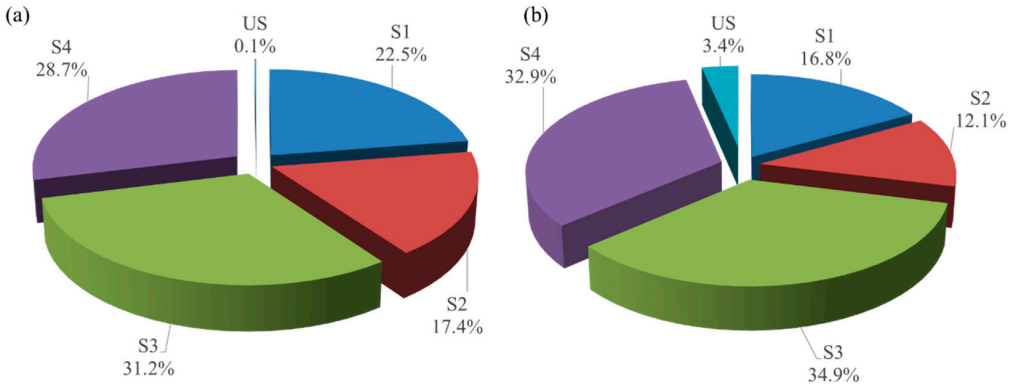


Figure 7. Average contributions of pollution sources to the water quality based on the PMF model in the LP (**a**) and HP (**b**) zones. The S1, S2, S3, S4, and US in (**a**) were identified as point source pollution of industrial sewage, point source of petrochemical enterprises wastewater, non-point source pollution of overland runoff, point source pollution of domestic sewage, and one undefined source, respectively. The S1, S2, S3, S4, and US in (**b**) were identified as non-point source pollution of overland runoff, point source of petrochemical enterprises wastewater, point source pollution of domestic sewage, point source pollution of industrial sewage, and one undefined source, respectively.

Similar to the APCS-MLR model, the PMF model results showed that point source pollution was the most critical pollution source of surface water in Dongying District, followed by the non-point pollution of overland runoff.

3.6. Comparison of Pollution Source Contribution Based on the APCS-MLR and PMF Models

The contribution of pollution sources based on the APCS-MLR and PMF models differed significantly. Therefore, R^2 and the ratio of the predicted and observed values were used to compare the application and accuracy between the APCS-MLR and PMF models (Table S4). According to Liu et al. and Simeonov et al. [14,37], the source apportionment results would be relatively good when the ratio of predicted and observed values was close to 1.0, and R^2 was greater than 0.5. In this study, the ratio of predicted and observed values in the APCS-MLR model varied between 0.7 to 1.43, and the average R^2 was 0.60 (Table S4), whereas the ratio of predicted and observed values in the PMF model ranged between 1.00 and 1.08, and the average of R^2 was 0.42 (Table S4). In addition, the scatter plots between the predicted and observed values from the results of the two PMF and

APCS-MLR models are displayed in Figures S4 and S5 for the LP and HP zone, respectively. From these figures, the APCS-MLR model had better goodness of fit for most of the water quality variables based on the fitting equation and R^2. On the other hand, the procedure and dataset involved in the APCS-MLR model were relatively simple compared to the PMF model [10,38], although there was a lower specificity for the APCS-MLR model to use the concentration data without uncertainty values. Therefore, the accuracy of the APCS-MLR model was superior to that of the PMF model in terms of quantitative analysis in this study.

4. Conclusions

In this study, multivariate statistical methods, including CA and PCA/FA, were used to survey the spatial variation in river water quality. The APCS-MLR and PMF models identified potential factors influencing the water quality of the river, revealed the potential pollution sources, and quantified their contribution to the pollution of surface water. According to the degree of pollution, the study area was classified into LP and HP zones by CA. The contribution obtained by the APCS-MLR model was more reliable in identifying the pollution source in Dongying District, by comparison, the R^2 and the ratio of predicted and observed values between the APCS-MLR and PMF models. The APCS-MLR model with PCA/FA was used to identify four critical pollution sources in the LP and HP zones. In the LP zone, the contributions of pollution sources at the LP zone should be ranked as point source pollution of domestic sewage > non-point source pollution of agricultural activity > pollution source of fecal matter > point source pollution of petrochemical sewage. The contribution in the HP zone followed the decreasing order as follows: point source pollution from domestic sewage > non-point pollution source of overland runoff > point source pollution of hybrid wastewater > point source pollution from industrial sewage. These findings suggest that the rivers flowing into Laizhou Bay in Dongying District are primarily polluted by the point source pollution of domestic sewage and non-point source pollution of overland off. In considering the limitations of this study, the monitored indicators consisted of only ten water quality variables, and the sampling time period was only one year. Therefore, further work needs to be conducted using more monitored indicators and a longer sampling time period to better identify the undefined sources and understand the spatial variability.

Supplementary Materials: The following are available online at https://www.mdpi.com/2073-4441/13/6/772/s1, Figure S1. Meteorological conditions at the time of sampling from May 2018 to July 2019. (a): average temperature, (b): humidity, (c): weed speed, (d): visibility; Figure S2. Dendrogram from hierarchical cluster analysis of sampling sites based on ten water quality variables in Dongying District; Figure S3. Maps of the spatial variation showing the annual mean of ten water quality indicators before standardization in Dongying District; Figure S4. Scatter plots comparing predicted and observed values from the results of the APCS-MLR and PMF models in the LP zone. (a): pH, (b): DO, (c): BOD, (d): COD, (e): NH3-N, (f): TP, (g): TN, (h): PI, (i): F.coli, and (j): PHC; Figure S5. Scatter plots comparing predicted and observed values from the results of the APCS-MLR and PMF models in the HP zone. (a): pH, (b): DO, (c): BOD, (d): COD, (e): NH3-N, (f): TP, (g): TN, (h): PI, (i): F.coli, and (j): PHC; Table S1. Longitude and latitude of monitoring sites; Table S2. River outlets in the LP and HP zones; Table S3. Number of livestock and poultry farms and types of livestock in the LP and HP zones; Table S4. Mean source contributions to different variables by the APCS-MLR and PMF models in the LP and HP zones.

Author Contributions: J.S. and Y.Q. performed the experiments and prepared the draft; Y.L. performed the determination of samples; X.Y. and S.L. conceptualized the study and analyzed the data. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are deeply indebted to the National Key Research and Development Project (No. 2016YFD0800100-04), and Hu Zhengyi for his help in preparing this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, B.; Yi, Q.; Deng, X. Analysis and Evaluation of Water Quality in the Laizhou Bay and Its Adjacent Areas in Summer, 2017. *Coast. Eng.* **2018**, *37*, 44–52.
2. Xia, B.; Zhang, X.; Cui, Y.; Chen, B.; Chen, J.; Guo, F.; Zhang, Y.; Huang, C. Evaluation of the physicochemical environment and nutrition status in Laizhou Bay and adjacent waters in summer. *Prog. Fish. Sci.* **2009**, *30*, 103–111.
3. Zhuang, W.; Gao, X. Acid-volatile sulfide and simultaneously extracted metals in surface sediments of the southwestern coastal Laizhou Bay, Bohai Sea: Concentrations, spatial distributions and the indication of heavy metal pollution status. *Mar. Pollut. Bull.* **2013**, *76*, 128–138. [CrossRef] [PubMed]
4. Song, D.; Gao, Z.; Zhang, H.; Xu, F.; Zheng, X.; Ai, J.; Hu, X.; Huang, G.; Zhang, H. GIS-based health assessment of the marine ecosystem in Laizhou Bay, China. *Mar. Pollut. Bull.* **2017**, *125*, 242–249. [CrossRef] [PubMed]
5. Qu, Y.H. The Investigation and Evaluation of Land-based Pollutant Poured into the Sea in Dongying City. Ph.D. Thesis, Ocean University of China, Qingdao, China, 2009.
6. Liu, Q.; Wang, F.; Meng, F.; Jiang, L.; Li, G.; Zhou, R. Assessment of metal contamination in estuarine surface sediments from Dongying City, China: Use of a modified ecological risk index. *Mar. Pollut. Bull.* **2018**, *126*, 293–303. [CrossRef]
7. Ludwig, W.; Dumont, E.; Meybeck, M.; Heussner, S. River discharges of water and nutrients to the Mediterranean and Black Sea: Major drivers for ecosystem changes during past and future decades? *Prog. Oceanogr.* **2009**, *80*, 199–217. [CrossRef]
8. Ministry of Ecology and Environment of the People's of Republic of China. China Marine Ecological Environment Bulletin in 2018. Available online: http://www.mee.gov.cn/hjzl/shj/jagb/ (accessed on 29 May 2019).
9. Singh, K.P.; Malik, A.; Sinha, S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—A case study. *Anal. Chim. Acta* **2005**, *538*, 355–374. [CrossRef]
10. Haji Gholizadeh, M.; Melesse, A.M.; Reddi, L. Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. *Sci. Total Environ.* **2016**, *566–567*, 1552–1567. [CrossRef]
11. Liu, P.; Hoth, N.; Drebenstedt, C.; Sun, Y.; Xu, Z. Hydro-geochemical paths of multi-layer groundwater system in coal mining regions—Using multivariate statistics and geochemical modeling approaches. *Sci. Total Environ.* **2017**, *601–602*, 1–14. [CrossRef] [PubMed]
12. Dugga, P.; Pervez, S.; Tripathi, M.; Siddiqui, M.N. Spatiotemporal variability and source apportionment of the ionic components of groundwater of a mineral-rich tribal belt in Bastar, India. *Groundw. Sustain. Dev.* **2020**, *10*, 100356. [CrossRef]
13. Zanotti, C.; Rotiroti, M.; Fumagalli, L.; Stefania, G.A.; Canonaco, F.; Stefenelli, G.; Prévôt, A.S.H.; Leoni, B.; Bonomi, T. Groundwater and surface water quality characterization through positive matrix factorization combined with GIS approach. *Water Res.* **2019**, *159*, 122–134. [CrossRef] [PubMed]
14. Liu, L.; Dong, Y.; Kong, M.; Zhou, J.; Zhao, H.; Tang, Z.; Zhang, M.; Wang, Z. Insights into the long-term pollution trends and sources contributions in Lake Taihu, China using multi-statistic analyses models. *Chemosphere* **2020**, *242*, 125272. [CrossRef]
15. Yang, Y.H.; Zhou, F.; Guo, H.C.; Sheng, H.; Liu, H.; Dao, X.; He, C.J. Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environ. Monit. Assess.* **2010**, *170*, 407–416. [CrossRef]
16. Zhang, H.; Li, H.; Yu, H.; Cheng, S. Water quality assessment and pollution source apportionment using multi-statistic and APCS-MLR modeling techniques in Min River Basin, China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 41987–42000. [CrossRef]
17. Cheng, G.; Wang, M.; Chen, Y.; Gao, W. Source apportionment of water pollutants in the upstream of Yangtze River using APCS–MLR. *Environ. Geochem. Health* **2020**, *42*, 3795–3810. [CrossRef] [PubMed]
18. Zhao, H.Y.; Zhao, R.; Sun, Y.Y.; Zheng, M.X.; Su, Q.; Fu, X.M.; Ding, H.Y. Watershed pollution source analysis based on receptor model and control unit division: Taking Zhangjiakou section of Yongding River as an example. *J. Environ. Eng. Technol.* **2020**, *10*, 758–768.
19. Environment Protection Ministry of China. *Environmental Quality Standards for Surface Water*; Environment Protection Ministry of China: Beijing, China, 2002.
20. Plotkin, J.B.; Chave, J.; Ashton, P.S. Cluster Analysis of Spatial Patterns in Malaysian Tree Species. *Am. Nat.* **2002**, *160*, 629–644. [CrossRef]
21. Shrestha, S.; Kazama, F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environ. Model. Softw.* **2007**, *22*, 464–475. [CrossRef]
22. Helena, B. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* **2000**, *34*, 807–816. [CrossRef]
23. Thurston, G.D.; Spengler, J.D. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmos. Environ.* **1985**, *19*, 9–25. [CrossRef]
24. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126. [CrossRef]

25. Soonthornnonda, P.; Christensen, E.R. Source apportionment of pollutants and flows of combined sewer wastewater. *Water Res.* **2008**, *42*, 1989–1998. [CrossRef] [PubMed]
26. Zhang, H.; Cheng, S.; Li, H.; Fu, K.; Xu, Y. Groundwater pollution source identification and apportionment using PMF and PCA-APCA-MLR receptor models in a typical mixed land-use area in Southwestern China. *Sci. Total Environ.* **2020**, *741*, 140383. [CrossRef]
27. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36. [CrossRef]
28. Mishra, A.; Kar, S.; Raghuwanshi, N.S. Modeling Nonpoint Source Pollutant Losses from a Small Watershed Using HSPF Model. *J. Environ. Eng.* **2009**, *135*, 92–100. [CrossRef]
29. Wiens, J.H. Agricultural runoff and water pollution. *Can. Water Resour. J.* **1980**, *5*, 78–89. [CrossRef]
30. Srinivas, R.; Singh, A.P.; Dhadse, K.; Garg, C. An evidence based integrated watershed modelling system to assess the impact of non-point source pollution in the riverine ecosystem. *J. Clean. Prod.* **2020**, *246*, 118963. [CrossRef]
31. Reitz, A.; Hemric, E.; Hall, K.K. Evaluation of a multivariate analysis modeling approach identifying sources and patterns of nonpoint fecal pollution in a mixed use watershed. *J. Environ. Manag.* **2021**, *277*, 111413. [CrossRef]
32. Johnson, L.; Richards, C.; Host, G.; Arthur, J. Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshw. Biol.* **1997**, *37*, 193–208. [CrossRef]
33. Boyer, E.W.; Goodale, C.L.; Jaworski, N.A.; Howarth, R.W. Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern U.S.A. *Biogeochemistry* **2002**, *57*, 137–169. [CrossRef]
34. Zhou, T.; Wu, J.; Peng, S. Assessing the effects of landscape pattern on river water quality at multiple scales: A case study of the Dongjiang River watershed, China. *Ecol. Indic.* **2012**, *23*, 166–175. [CrossRef]
35. Ho, K.C.; Hui, K.C.C. Chemical contamination of the East River (Dongjiang) and its implication on sustainable development in the Pearl River Delta. *Environ. Int.* **2001**, *26*, 303–308. [CrossRef]
36. Mikkelsen, R. Ammonia emissions from agricultural operations: Fertilizer. *Better Crops* **2009**, *93*, 9–11.
37. Simeonov, V.; Stratis, J.A.; Samara, C.; Zachariadis, G.; Voutsa, D.; Anthemidis, A.; Sofoniou, M.; Kouimtzis, T. Assessment of the surface water quality in Northern Greece. *Water Res.* **2003**, *37*, 4119–4124. [CrossRef]
38. Wu, J.; Margenot, A.J.; Wei, X.; Fan, M.; Zhang, H.; Best, J.L.; Wu, P.; Chen, F.; Gao, C. Source apportionment of soil heavy metals in fluvial islands, Anhui section of the lower Yangtze River: Comparison of APCS–MLR and PMF. *J. Soils Sediments* **2020**, *20*, 3380–3393. [CrossRef]