Proceedings of the 7th World Congress on Civil, Structural, and Environmental Engineering (CSEE'22)
~~Lisbon, Portugal~~ Virtual Conference – April 10 – 12, 2022
Paper No. ICEPTP 166
DOI: 10.11159/iceptp22.166

Determination of Primary Surface Water Pollution Indicators by Multivariate Statistical Techniques in an Industrialized Basin.

Halil N. Oruc[1], Mehmet Salim Oncel[1,2], Guleda Onkal Engin[3], Fatih Gulgen[4], Mehmet Karpuzcu[5], Sinem Vural[2], Batuhan Kılıç[4], Meltem Çelen[1]
[1]Gebze Technical University Institute of Earth and Marine Sciences, 41400 Gebze, Kocaeli Turkey
horuc@gtu.edu.tr; mkocal@gtu.edu.tr
[2]Gebze Technical University Environmental Engineering Department, 41400, Gebze/Kocaeli
[3]Yildiz Technical University, Civil Engineering Faculty, Environmental Engineering Department, 34220 Davutpasa, Esenler, Istanbul, Turkey
[4]Yildiz Technical University Geomatic Engineering Department, 34220, Esenler/İstanbul
[5]Hasan Kalyoncu University, Civil Engineering Faculty, Şahinbey/Gaziantep

Abstract **-** This study aimed to investigate the point and diffuse pollution sources of high total phosphorus (TP) pollution detected in the dry and wet seasons of the industrialized and urbanized Saz-Cayirova Basin through field observation and multivariate statistical techniques. In this context, nineteen water quality parameters were analyzed in surface water samples collected monthly between June 2020 and July 2021 from nineteen different sites along the Saz-Cayirova stream. Firstly, two reference sites representing a better surface water status were determined and assessed the water quality on the stream tributaries affecting the wastewater treatment plant in the two Organized Industrial Zones (OIZ). Secondly, hierarchical cluster analysis (HCA) and principal component analysis (PCA) were performed to evaluate the complex water quality dataset and reveal the latent sources of TP pollution. The results showed that the tributaries in the pressure of OIZ discharges were highly enriched in COD, TOC, NO_3-N, NH_3-N, and TP concentrations compared to the reference level. Compared with reference sites, the concentrations of the four heavy metals examined were at a plausible level. Besides, the pollution sources of TP were industrial processes wastewater such as dyeing, washing, phosphating, domestic wastewater from OIZ, urban diffuse waters with organic character, and continuous discharges of undefined sources. The seasonal variation of TP values is observed relatively low, indicating that the stream network is greatly affected by point source pollution. Our observation and analysis imply that the treatment technologies adopted by the OIZ wastewaters plant is sufficient to treat heavy metals. However, measures need to be formulated to remove TP and organic pollution from the secondary production process.
Keywords**:** Phosphorus, Surface water quality, Multivariate statistical techniques, Point source pollution

1. Introduction

The physicochemical characteristics of surface waters are a key indicator for identifying and minimizing anthropogenic pressures on the stream. The most important pollution type of surface water in developed countries has been identified as organic and nutrient pollution [1]. Phosphorus is a nutrient derivative, usually in the form of pyrophosphate, metaphosphate, polyphosphate, and organically bound phosphate in spring waters and wastewaters [2]. Approximately 80% of the phosphorus pollution sources directly affect the water quality originate from industrial and domestic discharges [3]. Although the characteristics of domestic discharges generally have low variation, the total phosphorus concentration released from industries to the receiving environment varies with different processes. If a large amount of phosphorus and its derivatives are found in the aquatic environment, it can decrease dissolved oxygen in the water body due to the increase of minerals and organic nutrients, namely eutrophication. Besides, the high phosphorus load from streams to the marine environment under high runoff conditions may lead to dissolved organic matter derived from phytoplankton and algae [4].

The latest example of the negative impact of nutrients on the marine environment is the mucilage formation observed in the Marmara Sea, an inland sea in Turkey, at the end of 2020. With the effect of temperatures, it covered the sea surface and bottom intensively in the spring and summer of 2021 [5]. The primary motivation for this formation is the uncontrolled growth of microorganisms in the marine environment due to over-nutrition [6], [7]. Previous research proved that these nutrients' primary sources are the surface current from the Black Sea [8]. Furthermore, point/non-point source pollution of the small and large-scale basins around the Marmara Sea is also cited as a secondary source of mucilage formation [9], [10].

Saz-Cayirova Basin is dominated by industrial and residential areas (73% of the entire basin), agricultural activities are not carried out, residential discharges are collected in a separate channel, and infrastructure systems are developed compared to other basins flowing into the Marmara Sea. However, field and laboratory analysis shows that the total phosphorus concentrations in the entire basin are relatively high. Thus, the stream having high energy in the region was detected to provides an intense phosphorus load to the Marmara Sea. Due to these basin characteristics, this research aimed to investigate the point and diffuse pollution sources of total phosphorus throughout the basin on temporal and spatial scale using multivariate statistical techniques. In this context, a water quality data set consisting of nineteen parameters and nineteen monitoring sites was created. Firstly, two reference points were determined on the tributaries to calculate the effect of industrial discharges, which create the most significant pressure on the stream network. Subsequently, the pollution level of the sites directly affected by the industrial discharges was analyzed by comparing them to these reference points. Secondly, nineteen sites were grouped spatially through hierarchical clustering analysis. Depending on this grouping, dominant pollution sources were interpreted using principal component analysis.

2. Material and Method

2.1. Study Site

The Saz-Cayirova basin is located in the Kocaeli, which is placed in the northwest of Turkey and has a coast on the Marmara Sea. Saz-Cayirova is a small-scale industrialized and urbanized basin that includes Saz and Cayirova streams (Figure 1). The sub-basin areas of Saz and Cayirova are 20 km^2 and 30 km^2, orderly. The total length of the two-stream tributaries is 10 km, and their width varies between 2 and 20 m [11]. Climatically, this basin is the transition zone between the Mediterranean and the Black Sea climates. The upstream parts of the Saz Stream are primarily under pressure from point discharges originating from Gebze (GOIZ) and Automotive Supply Organized Industrial Zones (OSOIZ). In the downstream parts of the Saz Stream, small/large-scale factories discharge their treated wastewater into the stream. Some of the Cayirova tributaries have been rehabilitated and taken underground by local authorities. Even so, field observations by the research team showed that Cayirova tributaries carry industrial wastewater with different production processes. The local administrators in the region noted that the loads coming from the urban or rural areas of the basin are collected in a separate channel and are not discharged into the Saz Stream.

2.2. Water Sampling and Analysis

The water quality (WQ) monitoring campaign was carried out monthly from June 2020 to July 2021, with a focus on the Saz stream tributaries owing to the inclusion of the Cayirova tributaries in the channel. The campaign includes nineteen monitoring sites (GOIZ and OSOIZ reference points + 2 Cayirova stream outlets + 2 Saz-Cayirova main channel + 13 Saz tributaries). In order to maximize the visibility of these points in the map, the points, namely C1, C2, S6, GOIZ-REF, OSOIZ-REF, and S13, were visualized by a move to them approximately 55 meters (Figure 1). OSOIZ-REF and GOIZ-REF represent rural areas not affected by industrial discharges, and anthropogenic pressures are minimal throughout the basin.

A total of fourteen sampling campaigns were carried out throughout the study. Thirteen measurements were accomplished separately at sites S12 and S13. Twelve measurements were made asunder in S10 and S11. Nine measurements were performed separately at S5 and S6 sites. Eleven and two measurements were performed for GOIZ-REF and OSOIZ-REF, respectively. Fourteen measurements were conducted at each remaining site. Additionally, the tributaries' pollution status under the influence of OSOIZ and GOIZ were visualized with the OSOIZ-REF and GOIZ-REF sites. This comparison was accomplished to represent the hydrological wet (October-April) and dry (May-September) seasons.

Heavy metals (Iron-Fe, Zinc-Zn, Copper-Cu, Lead-Pb), alkaline earth metals (Sodium-Na, Magnesium-Mg, Calcium-Ca), nutrients (Total phosphorus-TP, Ammonia nitrogen-NH_3-N, Nitrate nitrogen - NO_3-N), in-situ (pH, dissolved oxygen-DO, electrical conductivity-EC, oxidation-reduction potential-ORP), organic carbon (Chemical oxygen demand-COD, biological oxygen demand-BOD, total organic carbon-TOC), Total Suspended Solids (TSS),

Alkalinity, a total of the 19 WQ parameters were measured in this research. All WQ parameters were expressed as mg/L, except pH, EC (μS/cm), ORP (mV).
Water quality parameters were analyzed according to Standard Methods Analyzed for the Examination of Water and Wastewater [12]. The sampling, preservation, transportation, and analysis of the water samples were performed according to standard methods [13]. The analytical data quality was ensured through careful standardization, procedural blank measurements, and duplicate samples.

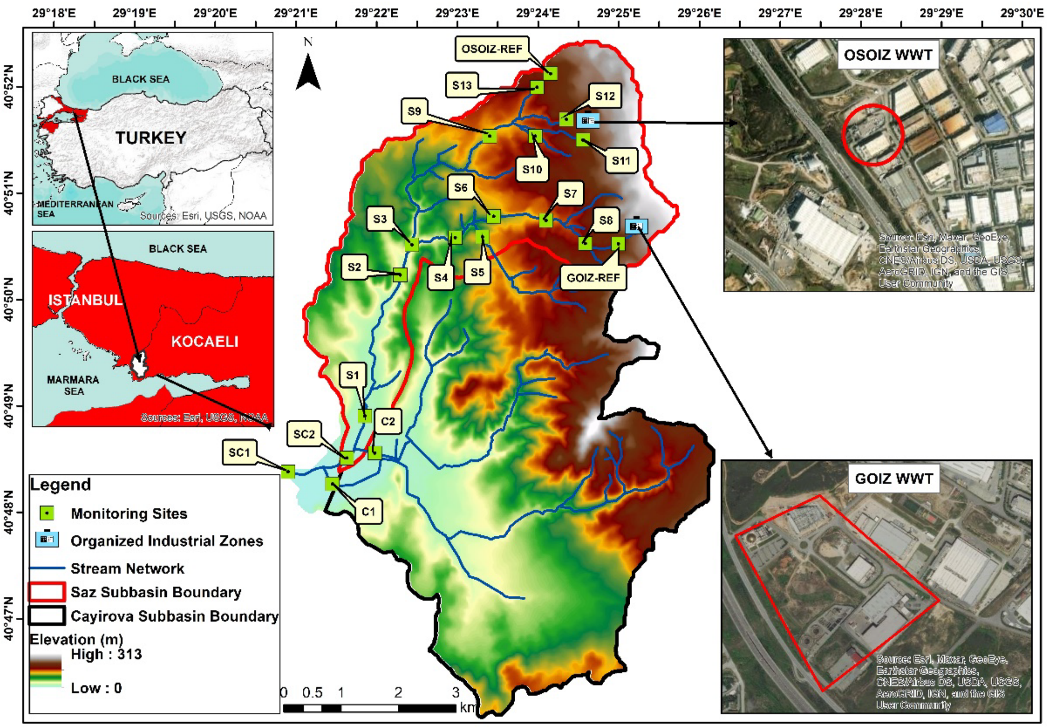

Figure 1. Location map and monitoring sites of Saz-Cayirova basin.

2.3. Exploratory Data Analysis and Multi-Variate Statistical Techniques

The applicability and reliability of the PCA depend on the size of the dataset. Tran et al. [14] suggested that a variable in a multivariate dataset should be at least size of 30. This number of data may not be available from a single monitoring site. The application of cluster analysis techniques is recommended in reducing both time and cost, provided they are validated by field observations [15]. In this study, 17 monitoring stations (excluding reference points) were divided into five groups by field observations and hierarchical clustering analysis technique. The dataset in each group was divided into wet and dry seasons and subjected to PCA.

The consistency of each WQ parameter to the normal distribution was tested with Kolmogorov-Smirnov (K/S) test. K/S results demonstrated that all parameters were distributed log-normally at the 95% confidence interval. In fact, the environmental data matrix usually follows a log-normal distribution [16]. In this context, Kaiser-Meyer-Olkin (KMO) and Barlett's Sphericity tests were applied to the datasets before PCA. A KMO value close to 1 indicates that the sum of the partial correlations is less than the sum of the correlations [17]. The KMO value varies between 0.68 and 0.82 among the grouped datasets. Barlett's Sphericity test results showed significant relationships between variables at the 0.05 significance level. The two test results showed that the WQ dataset examined was suitable for PCA. Before HCA and PCA, the dataset was standardized with the z-scale transformation technique to prevent misclassification due to different units and the data dimensionless [18], [19].

The dendrogram (tree diagram) is the most common approach that visualizes intuitive similarity relationships among variables according to a pre-determined selection criterion. HCA is an unsupervised pattern recognition technique in which clusters with higher elements are created in a mathematical order starting from the most similar pair of parameters [20], [21]. A is widely used to statistically determine the similarities/differences of WQ parameters at different sites [22], [23]. In this study, HCA was performed on standardized values by the Ward2 method using Euclidean distances, which represent a measure of difference.

PCA is the basis of machine learning techniques and an orthogonal linear algebra that transforms related variables into unrelated variables called principal components (PCs) based upon Pearson's correlation matrix [24]. The main purpose of the PCA method is to explain the relationship with fewer variables while preserving the relationship between the original data as much as possible. PCs are specified in descending order from whichever best describes the variance of the dataset (PC1, PC2, PC3 ... PCn). The eigenvalue of PCA is an indicator of the importance of PCs on the original data set. Conversely, PCA outputs are difficult to interpret using eigenvalues in environmental datasets. The VARIMAX rotation technique is preferred in many studies to interpret PCA outputs [25], [26]. Further details for the working principles of VARIMAX can be found in the study conducted by Kaiser [27].

3. Results and Discussion

Table 1 and Table 2 demonstrate the annual mean values and spatial distributions of WQ parameters measured at 19 stations (17 monitoring + 2 references) across the basin. The highest pH, DO, EC, NH_3-N, NO_3-N, TP, TOC, COD, BOD, TSS parameters were determined at C1 and C2 sites, representing Cayirova stream outlets. At the same time, the lowest ORP values were also observed at these stations. A noticeable decrease was observed in these values as of April 2021 (Figure not shown). This situation can be explained by the measures taken to reduce the discharges after the mucilage event in the Marmara Sea. In the tributaries of the Saz Stream, except for TP, the variation scale of these parameters is relatively low. TP was detected approximately two times more in S6, S7, and S8 sites, representing GOIZ affluent than other sites. This shows the direct effect of GOIZ discharges on TP concentrations. Heavy metals and alkaline earth metals did not have a high variation throughout the basin.

Table 1. Annual mean values with standard deviations of WQ parameters in nineteen monitoring sites.

Monitoring sites	pH -	ORP (mV)	TOC (mg/L)	Alkalinity (mg/L)	NH_3-N (mg/L)	TSS (mg/L)	Na (mg/L)	Ca (mg/L)	Mg (mg/L)
SC1	8.9 ± 1.2	105 ± 43	14.9 ± 7.5	300 ± 104	9.3 ± 8.2	0.06 ± 0.09	184 ± 74	83 ± 48	15.4 ± 8.7
SC2	8.7 ± 0.8	108 ± 41	11.4 ± 5.6	278 ± 83	7.1 ± 5.1	0.04 ± 0.04	147 ± 41	82 ± 34	14.4 ± 4.7
C1	9.1 ± 1.2	62 ± 62	18.3 ± 4.5	282 ± 94	14.9 ± 8.2	0.12 ± 0.19	199 ± 111	84 ± 48	12.4 ± 6.5
C2	8.5 ± 1.2	57 ± 53	37.4 ± 19.3	355 ± 110	24.2 ± 12.7	0.13 ± 0.12	119 ± 24	73 ± 30	15.2 ± 5.7
S1	8.8 ± 0.7	125 ± 40	8.9 ± 2.1	272 ± 77	4.8 ± 2.9	0.01 ± 0.006	162 ± 59	89 ± 39	14.8 ± 4.5
S2	8.3 ± 0.5	141 ± 49	8.2 ± 1.8	306 ± 66	5.4 ± 2.9	0.02 ± 0.03	157 ± 43	91 ± 33	14.7 ± 3.8
S3	8.2 ± 0.5	150 ± 44	6.2 ± 1.1	267 ± 46	4.6 ± 1.4	0.02 ± 0.03	112 ± 19	97 ± 36	15.2 ± 4.4
S4	8.4 ± 0.3	127 ± 47	10.6 ± 2.5	324 ± 100	5.3 ± 4.1	0.06 ± 0.18	195 ± 61	86 ± 32	14.5 ± 4.3
S5	8.1 ± 0.3	149 ± 51	5.2 ± 1.4	282 ± 59	3.4 ± 1.1	0.03 ± 0.03	53 ± 10	74 ± 32	17.9 ± 7.7
S6	8.0 ± 1.5	118 ± 52	11.1 ± 2.4	333 ± 79	4.3 ± 2.1	0.09 ± 0.23	201 ± 66	66 ± 23	11.9 ± 3.6
S7	8.5 ± 0.5	119 ± 45	12.8 ± 3.9	353 ± 101	5.4 ± 2.4	0.05 ± 0.1	239 ± 63	88 ± 35	12.7 ± 4
S8	8.4 ± 0.3	117 ± 40	13.1 ± 3.3	362 ± 103	5.6 ± 3.5	0.02 ± 0.04	269 ± 57	92 ± 39	12.4 ± 4.3
S9	7.9 ± 0.6	139 ± 52	5.8 ± 1.6	238 ± 54	4.5 ± 2	0.05 ± 0.07	109 ± 28	101 ± 37	14.8 ± 5.3
S10	7.7 ± 0.3	129 ± 81	5.2 ± 1.7	308 ± 70	4.9 ± 6.5	0.03 ± 0.05	83 ± 33	95 ± 41	18.8 ± 7.1
S11	8.2 ± 0.4	139 ± 53	14.2 ± 23	202 ± 57	3.5 ± 1.4	0.05 ± 0.07	82 ± 32	89 ± 69	13.6 ± 7.2
S12	7.8 ± 0.3	143 ± 57	11.4 ± 21.4	248 ± 58	3.3 ± 1.1	0.03 ± 0.03	118 ± 38	99 ± 37	14.4 ± 5.7
S13	7.6 ± 0.3	105 ± 71	3.7 ± 1.4	289 ± 47	3.1 ± 0.7	0.01 ± 0.01	65 ± 19	83 ± 27	16.9 ± 4.8
GOIZ-REF	7.8 ± 0.5	154 ± 61	3.9 ± 1.9	245 ± 35	2.9 ± 0.9	0.12 ± 0.19	116 ± 61	90 ± 27	29.3 ± 10.1
OSOIZ-REF	7.9 ± 0.2	85 ± 35	4.2 ± 0.6	188 ± 21	2.8 ± 0.8	0.03 ± 0.02	167 ± 87	73 ± 39	14.4 ± 6.8

Table 2. Annual mean values of water quality parameters (mean + SD) in nineteen monitoring sites.

Monitoring sites	DO (mg/L)	EC (uS/cm)	COD (mg/L)	BOD (mg/L)	NO_3-N (mg/L)	TP (mg/L)	Fe (mg/L)	Zn (mg/L)	Cu (mg/L)	Pb (mg/L)
SC1	5.7 ± 2.3	1250 ± 434	51.4 ± 45.1	13 ± 22	3.1 ± 0.9	1.4 ± 1.03	0.31 ± 0.24	0.08 ± 0.08	0.05 ± 0.05	0.015 ± 0.015
SC2	6.5 ± 2.7	1072 ± 257	48.9 ± 36.3	11 ± 16	3.1 ± 0.8	0.9 ± 0.7	0.3 ± 0.16	0.07 ± 0.04	0.05 ± 0.04	0.017 ± 0.016
C1	2.5 ± 1.7	1285 ± 515	81.3 ± 28.5	18 ± 24	2.6 ± 0.7	1.1 ± 1.0	0.4 ± 0.64	0.06 ± 0.06	0.17 ± 0.23	0.016 ± 0.015
C2	6.4 ± 2.3	1018 ± 243	209 ± 111	111 ± 81	4.3 ± 1.6	3.6 ± 2.2	0.3 ± 0.15	0.07 ± 006	0.07 ± 0.07	0.023 ± 0.020
S1	7.8 ± 2.1	1057 ± 251	35.4 ± 14.0	3.4 ± 8.5	3.2 ± 0.8	0.6 ± 0.4	0.3 ± 0.14	0.07 ± 0.04	0.04 ± 0.03	0.012 ± 0.013
S2	8.6 ± 2.2	1139 ± 257	26.8 ± 8.6	1.3 ± 3.1	3.3 ± 0.8	0.7 ± 0.4	0.4 ± 0.2	0.07 ± 0.03	0.09 ± 0.12	0.008 ± 0.009
S3	8 ± 0.4	968 ± 201	23.8 ± 8.9	0.9 ± 2.7	3.8 ± 1.43	0.5 ± 0.2	0.3 ± 0.24	0.04 ± 0.02	0.05 ± 0.04	0.013 ± 0.02
S4	9.2 ± 1.8	1258 ± 321	36.4 ± 9.2	1.6 ± 5.1	3 ± 0.6	0.9 ± 0.7	0.5 ± 0.3	0.09 ± 0.02	0.04 ± 0.04	0.01 ± 0.01
S5	8.5 ± 1.4	666 ± 177	41.2 ± 31.1	0.6 ± 1.6	2.6 ± 0.8	0.1 ± 0.1	0.6 ± 0.2	0.02 ± 0.02	0.05 ± 0.04	0.013 ± 0.010
S6	9.9 ± 1.2	1320 ± 380	39.2 ± 9.2	0.000	2.9 ± 0.7	1.1 ± 0.9	0.7 ± 0.4	0.14 ± 0.04	0.06 ± 0.05	0.015 ± 0.017
S7	8.4 ± 0.7	1496 ± 348	49.5 ± 22	5.9 ± 15.4	2.9 ± 0.8	1.4 ± 1.1	0.6 ± 0.3	0.1 ± 0.03	0.05 ± 0.04	0.012 ± 0.02
S8	8.9 ± 1.2	1618 ± 358	44.9 ± 14.2	0.9 ± 1.4	3 ± 0.9	1.2 ± 1.6	0.6 ± 0.4	0.14 ± 0.07	0.05 ± 0.04	0.060 ± 0.160
S9	8.9 ± 1.7	1001 ± 190	21.5 ± 9.3	0.4 ± 1.04	5.3 ± 1.8	0.4 ± 0.2	0.33 ± 0.2	0.04 ± 0.02	0.06 ± 0.05	0.011 ± 0.02
S10	6.7 ± 2.5	850 ± 213	22.2 ± 23.9	2.2 ± 7.2	1.4 ± 0.7	0.2 ± 0.7	0.4 ± 0.3	0.03 ± 0.03	0.04 ± 0.05	0.011 ± 0.011
S11	8.3 ± 1.7	691 ± 219	102 ± 126	38 ± 81	2.3 ± 1.2	0.3 ± 0.4	0.2 ± 0.14	0.07 ± 0.06	0.04 ± 0.05	0.028 ± 0.039
S12	8.4 ± 1.2	961 ± 233	43.9 ± 89.8	16 ± 53	5.7 ± 3.4	0.8 ± 0.8	0.4 ± 0.6	0.07 ± 0.04	0.04 ± 0.04	0.017 ± 0.020
S13	8.3 ± 1.4	746 ± 274	16.7 ± 10.1	1.3 ± 3.5	1.6 ± 0.6	0.1 ± 0.1	1.36 ± 0.65	0.04 ± 0.03	0.04 ± 0.04	0.006 ± 0.006
GOIZ-REF	8.1 ± 1.3	1217 ± 531	13.9 ± 6.3	0.000	1.9 ± 0.8	0.1 ± 0.1	0.1 ± 0.08	0.02 ± 0.02	0.03 ± 0.02	0.017 ± 0.022
OSOIZ-REF	8.6 ± 1.1	984 ± 595	9.6 ± 6.4	0.000	1.6 ± 0.5	0.000	0.11 ± 0.02	0.03 ± 0.02	0.06 ± 0.04	0.023 ± 0.006

A correctly positioned reference sites are critical for mathematically express the impact of industrial pressures on the basin. Given the field observations and the structure of the WQ dataset, based on the data from the OSOIZ-REF and GOIZ-REF sites, the variations of 9 WQ parameters were examined for the wet season (May to September) and dry season (October to April) (Figure 2). The pollution status of stations S4, S6, S7, and S8 were evaluated according to GOIZ-REF, while S3, S9, S11, and S12 were compared to OSOIZ-REF. Meanwhile, COD, TOC, NH_3-N, NO_3-N, TP, Fe, Zn variables were above the reference values for all sites. Cu and Pb concentrations were measured below the reference values in many sites, especially in the dry season. The abandoned metallic sulphide mine site, where Pb, Zn, and Cu are extracted, is located in the NW and SE parts of the Pelitli Settlement of Gebze. One hundred ten thousand tons of geological reserves have been determined in the mine site [28]. Thus, the relatively high values measured at OSOIZ-REF and GOIZ-REF sites can be associated with Pb, Zn, Cu in the soil content in the basin. This unique characteristic was taken into account in the pollution source analysis section.

Moreover, the close relationship between these three elements and soil-derived elements can be considered as expected at the Saz-Cayirova basin scale. Otherwise, the concentrations of NH_3-N and NO_3-N measured at the S9 were higher than other sites affected by OSOIZ. Greenhouse activities are carried out near the S9 site, and nitrogen fertilizers are commonly use around this site. The relatively high nitrogen concentration measured during the dry season of S9 supports this inference.

Organic carbon derivatives (COD, TOC) were detected above the reference values in 8 sites. A relatively low difference was observed between the dry and wet season (except for TP) values of the sites affected by GOIZ discharges. Another side, higher concentrations were measured in the wet season at the sites affected by OSOIZ discharges. The COD concentration measured at S11 and S12 during the wet season is 100 times and 60 times that of the reference site. TOC values at these sites were measured 6 and 5 times higher, respectively. TOC and COD may accumulate in the sediment over this stream network, since these concentrations showed a descending order from the OSOIZ effluent to the leading network (S11, S12, S9, S3). In consideration of nutrient concentrations, the most dramatic difference was seen in TP relative to reference values. Among sites representative of OSOIZ discharges, the highest TP concentration was measured at site S12. As well, TP values did not show a high difference between the two seasons. The highest TP concentrations were monitored in the dry season at sites representing the GOIZ discharges. Fe and Zn were measured higher than those of OSOIZ in the networks representing GOIZ output. In addition, these metals' wet and dry season values were measured close to each other at sites S4, S6, S7, S8.

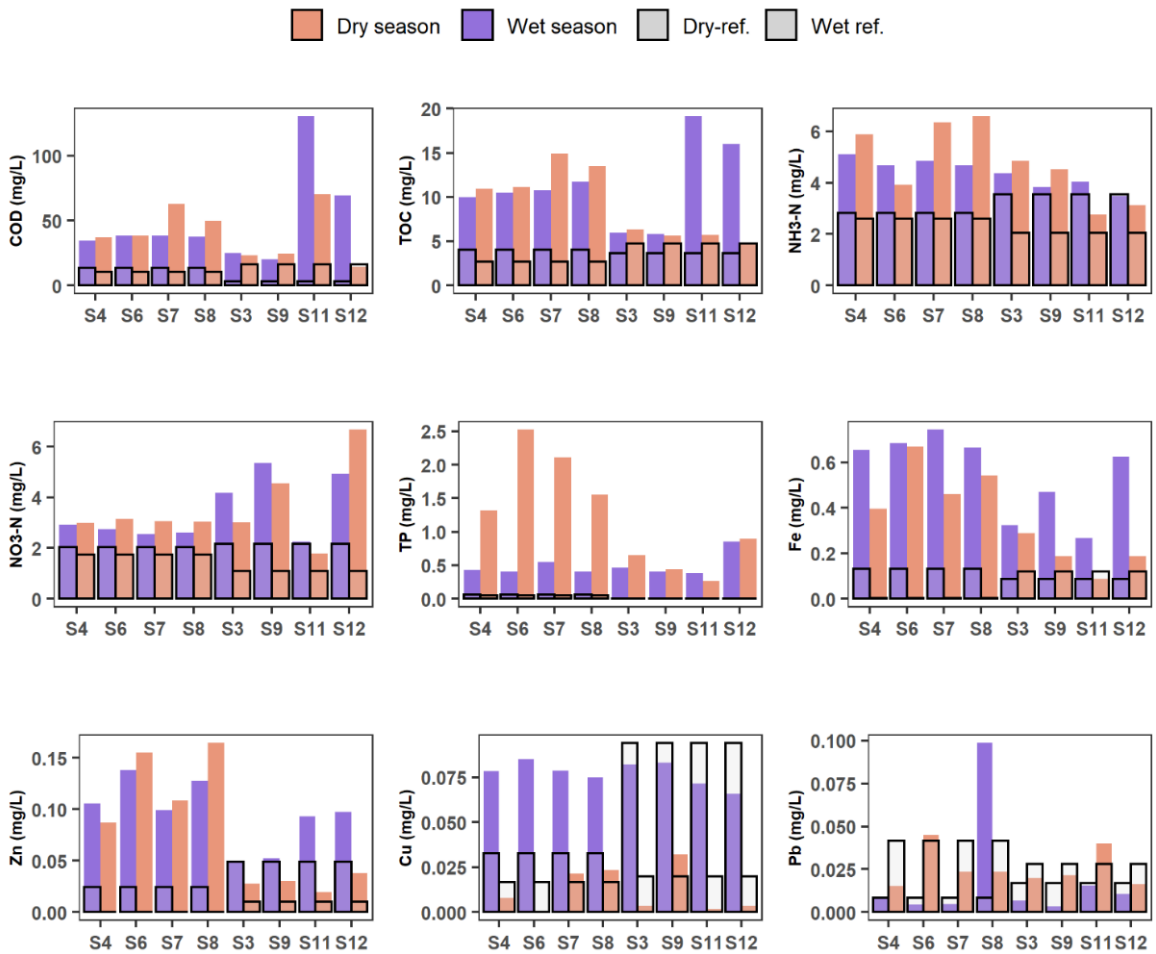

Figure 2. Status of organic carbon, nutrients and heavy metals measured at sites on the OSOIZ and GOIZ stream networks compared to reference sites.

3.2. Hierarchical Clustering Analysis (HCA)

HCA was applied to survey spatial similarities in terms of WQ and determine the number of groups examined in PCA (Figure 3). Statistically significant, five groups were created. Gr-1 represents the sites (S3, S9, S11, S12) under pressure from OSOIZ and various sizes of residential areas (especially around S3). According to the dendrogram, S3 and S9 sites are the most similar sites, followed by S12 and S11. S11 and S12 represent two different networks that of OSOIZ discharges. That is why the pollution level of the S3 and S9 may be more affected by network pollution represented by the S12.

Gr-2 represents sites (S4, S6, S7, S8) under pressure from GOIZ and individual factories of various sizes. Among the sites clustered in this group, S4 to S7 and S6 to S8 show the closest characters. Considering the S6 and S7 on the map, the high similarity between the two sites is to be expected. The relatively low similarity of the two sites within the same group can be explained by the point or non-point leakage discharges located between the two points. Gr-3 is dominated by sites on the leading network (SC1, SC2, S1, S2). While SC2 and S1 present the closest relationship within the group, they are followed by S2 and SC1, respectively. This high similarity is expected as there are approximately 100 meters between sites SC2 and S1. The fact that SC1 is the most different site in the group represents the irregular load of the Cayirova river network at this time.

Gr-4 represents sites (S5, S10, S13) that are not under pressure from GOIZ and OSOIZ and are generally influenced by urban diffuse pollution. In the dendrogram, S5 and S10 showed more similar features than S13. In the dendrogram, S5 and S10 show more similarity to each other than S13. While S13 represents the rural area close to OSOIZ but not affected by discharges, S10 and S13 represent areas with more urban character. Gr-5 was dominated by two sites (C1, C2) where the Cayirova stream networks outlets. Due to the mixed pollution load affecting the Cayirova stream networks, the similarity index (y-axis on the dendrogram) is not surprised to low between the two sites. Overall, the HCA results are consistent with field observations.

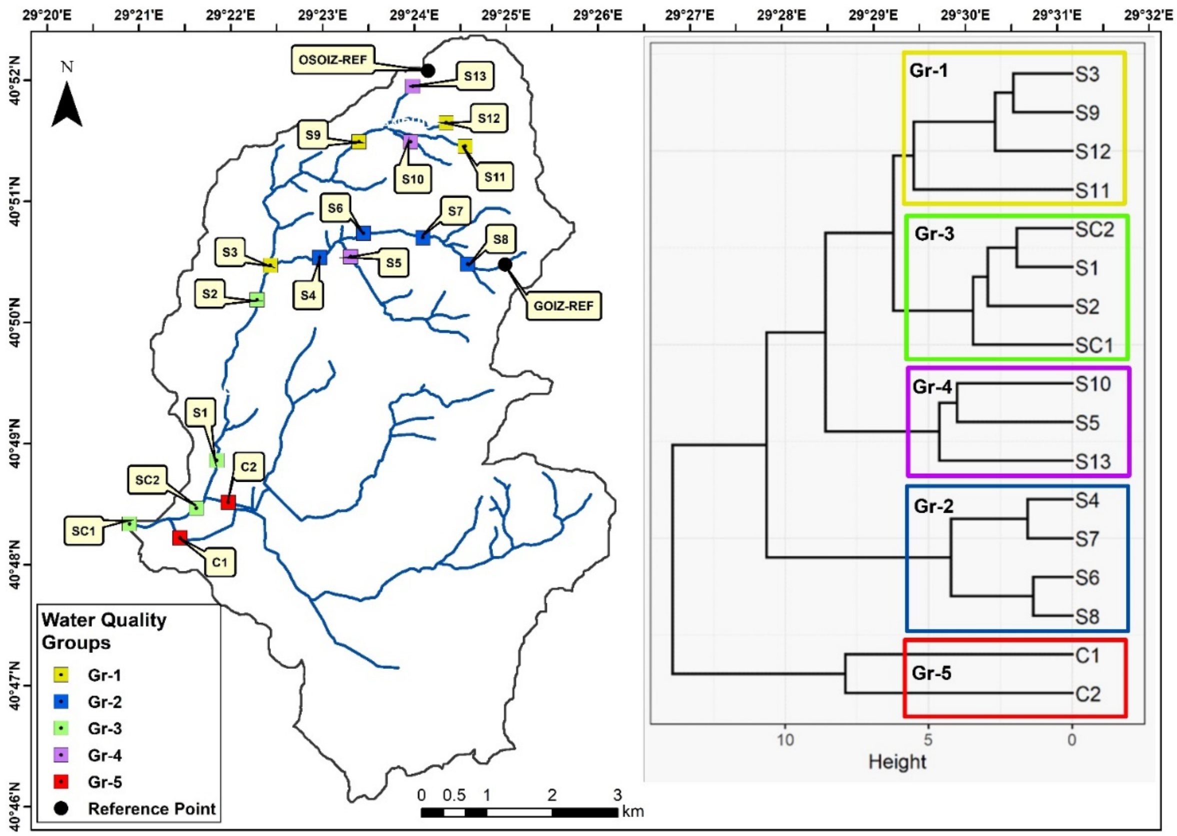

Figure 3. Dendrogram showing clustering of seventeen sites according to WQ characteristics of Saz-Cayirova stream.

3.3 Principal Component Analysis (PCA)

PCA was run for the first four groups (Gr-1, Gr-2, Gr-3, Gr-4) considering the wet and dry seasons. Since some of the Cayirova tributaries were included in the canal due to rehabilitation works, as stated before, monitoring studies could not be carried out in these tributaries. In this regard, it is unreliable to conduct a pollution source assessment by looking at the only two sites representing the Cayirova stream outlet.

PCA was applied based on 456 data matrices (24 sampling x 19 variables) and 532 data matrices (28 sampling x 19 variables) for the dry and wet seasons of Gr-1 and Gr-2, respectively. For both seasons of Gr-3, 532 data matrices (28 sampling x 19 variables) were considered. PCA was set based on 266 data matrices (14 sampling x 19 variables) and 380 data matrices (20 sampling x 19 variables) for the dry and wet season of Gr-4. Liu et al. [29] recommended classifying the PC load as vital (> ±75), moderate (±0.50-0.75), and weak (±0.30-0.50). In this study, PCA results were visualized (Figure 4 and Figure 5) with this classification in mind. The number of PCs was determined by considering components with eigenvalues greater than 1 [17]. Consequently, three statistically significant PCs were identified for each Gr-1, Gr-2, and Gr-4 in dry and wet seasons. Four PCs were evaluated in each of Gr-3's two seasons.

In the wet season of the Gr-1, PC1 (51.6% of the total variance) has a robust positive load on COD, BOD, TOC and strong negative load on DO. Also, it has a moderate positive load on NH_3-N, Na and negative loadings on pH. Oruc et al. [11] pointed out that the TOC content in the soil structure of the Saz-Cayirova basin is relatively low (0.2-3.2%). In this respect, organic carbon derivatives should be attributed to point source pollution rather than natural processes. Moreover, the contribution of BOD and NH_3-N to this PC also proves the impact of domestic wastewater, which comes from the facilities used by the approximately 20,000 employees working at OSOIZ. Thus, PC1 can be attributed to the "Mixture of industrial and domestic wastewater".

PC2 accounts for 24.5% of the total variance, with TP, Fe, and Zn as strong positive loadings. Also, pH contributed a moderate negative load on this PC. Automotive parts manufacturing firms are predominantly in the OSOIZ, and phosphate coating, a phase of steel coating surface treatments, is commonly applied in these processes. Zinc phosphate, iron phosphate, and manganese phosphate are commonly used in these coating types and can be applied in two ways: dipping or spraying. A rust precipitator line is also used in these processes [30]. The high negative contribution of ORP to this PC (-0.8) supports the characterization of wastewater produced by this process. In this status, PC2 most probably represents "Wastewater from metal industry processes". Cu, Pb, Na, Ca, Mg, TSS, Alkalinity, EC, DO, and TP was classified as strongly or weakly in the PC3 (9.6% of total variance). The correlation of Pb and Cu with alkaline earth metals is expected due to the geologic reserve in the basin. For the wet season representing the study period (2020-2021), 605 mm of precipitation was recorded, approximately 75 mm more than the previous two years. This situation depends on natural processes, which accelerate the weathering of ions in the soil and their diffusion into the aquatic environment. Straightforwardly, PC3 stands for "Soil weathering".

In the dry season of Gr-1, PC1 (46.2% of total variance) has a strong positive load on EC, TSS in addition to the variables in PC1's wet season. The inclusion of these WQ variables in this component indicates the degradability of water-soluble organic carbon derivatives under relatively low flow conditions. Considering that industrial activities are not affected by seasonal changes, it can be said that the most important source of pollution in two seasons is the "Mixture of OSOIZ's industrial and domestic wastewater". PC2 explains 27.2% of the total variance, which is strongly represented by the Fe, Zn, and TP variables derived from metal industrial processes, and moderately explained by NO_3-N, which is known to have the characteristics of domestic wastewater and animal manures. Residential discharges are collected in a separate channel and treated in the Gebze urban advanced wastewater treatment plant at the basin outlet. Hence, NO_3-N concentrations can originate from poultry farming around the S3 site clustered in Gr-1. Therefore, PC2 most likely represents a "Mixture of industrial metal processes and poultry farming activities". PC3 (16.3% of total variance) contributed strongly or moderately to parameters Cu, Pb, and earth alkaline metals. During the dry season of the study period, the recorded precipitation (265 mm) increased by approximately 55 mm compared to the previous two years. The Saz-Cayirova Basin is mainly composed of low-mineralized alluvial soil. So, alkaline earth metal concentrations in surface waters due to weathering can be regarded as in the wet season. Therefore, as in the wet season, PC3 can be explained by "Soil weathering".

In the wet season of Gr-2, PC1 (52.2% of the total variance) is dominated by heavy metals (Fe, Zn, Cu, Pb), reminiscent of GOIZ and the industries on the stream network. The strong clustering of the TOC parameter in PC1 supports the wastewaters of industrial processes. There are many sectors such as metal, plastic, textile, chemical products manufacturing in GOIZ. Heavy metal-weighted wastewater may occur from the processes of each sector mentioned. PC1 is defined as "Mixed industrial wastewater, that of mainly metal". PC2, which accounts for 23.9% of the total variance, was strongly associated with COD, NO_3-N, TP, Ca, Mg, and moderately ORP, Alkalinity. Ca and Mg are known as ions representing the hardness of water, and washing processes are carried out to remove them from the water. Large amounts of phosphorus and nitrogen are released from wastewater from the dyeing process [31]. PC2 is considered to represent "Industrial dyeing and washing processes". PC3 explains Na concentrations highly, and TP, TSS, Mg, Ca contribute moderately to this PC. This PC is most likely attributable to anthropogenic sources. However, to explain PC3, more detailed information about the production processes in GOIZ is required. For this reason, PC3 has been defined as an "Undefined pollution source".

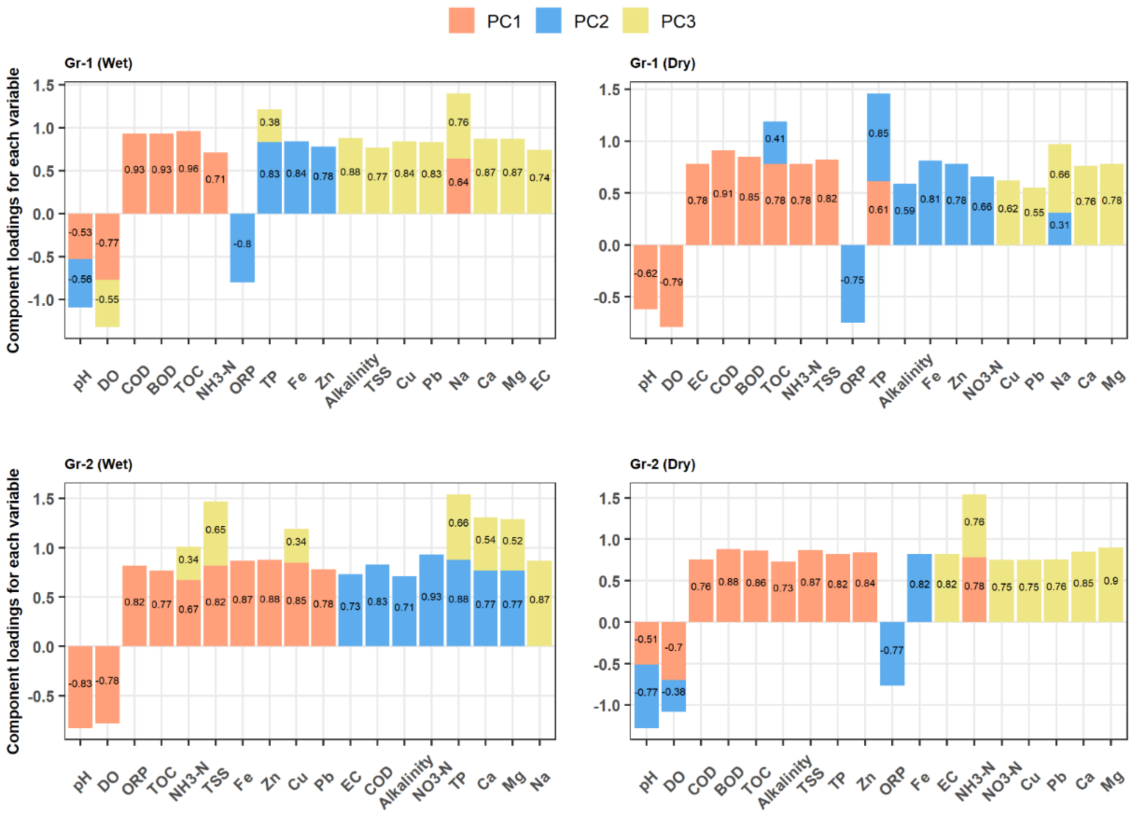

Figure 4. Varimax rotated loadings of significant principal components (eigenvalue>1) for the wet and dry seasons of Gr-1 and Gr-2.

In the dry season of Gr-2, PC1, explaining 54.8% of the total variance, is thought to express "Mixed industrial wastewaters". Because this component is powerfully explained by organic carbon derivatives, TP, TSS, and Zn. PC1 represents wastewater originating not only from metal industries but also from many industrial processes in GOIZ. PC2 (22.9% of total variance) has a strong positive load on Fe and a robust negative load on pH and ORP. The Fe value did not show remarkable changes in the dry and wet seasons at the sites representing GOIZ effluent and was the highest at the S6 (see Figure 2). This issue may be due to the Fe source from GOIZ and the various iron, steel, and recycling industries operating on this network. Therefore, PC2 is most likely attributable to "Mixed industrial wastewater, that of mainly metal". PC3 is strongly characterized by EC, NH_3-N, NO_3-N, Cu, Pb, Ca, Mg. As in the wet season of Gr-2, PC3 is defined as an "Undefined pollution source" during the dry season.

Gr-3 represents the main channel of the Saz Stream, which is the combined load from tributaries affected by OSOIZ and GOIZ. In the wet season of Gr-3, PC1, which explains the highest variance (41.8%), has a strong negative loading to pH and DO. In contrast, it has a solid favorable loading to EC, TSS, Fe, Zn, Mg, along with organic carbon and nutrients derivatives. In particular, the contribution of NO_3-N and NH_3-N to this PC and the high contribution of BOD refers to domestic discharges derived from OIZs. As mentioned above, the clustering of Fe, TP, Zn variables together is linked to the industrial processes in the basin scale. These ions are more accessible to weathering during the wet season, and under the influence of precipitation and relatively low pH, they can contribute to the PC1 together. Thus, PC1 stands for "Industrial and domestic discharges, mainly that of industrials". PC2 (22.9% of total variance) has a high loading on Cu and Ca, while it has a moderate loading on alkalinity and Na. Soil weathering observed in the northern region, which is under the pressure of OSOIZ discharges, is likely to have a higher effect on the main channel with increased weathering processes. Expressing PC2 with "Soil weathering" would not be wrong for the Saz-Cayirova basin. PC3 was highly positively correlated with ORP while moderately loaded by Fe, Cu, and Ca. PC3 denotes "Mixed industrial wastewater, that of mainly metal". PC4, accounts for 10.1% of the total variance, correlated only with Zn and Pb. Many vehicles using unleaded gasoline, such as trucks and transport vehicles, pass on the highway near the Saz Stream main channel. Hence, PC4 most probably refers to "Vehicle emissions".

In the dry season of Gr-3, PC1 (46.3% of total variance) has a strongly positive loading with EC, NH_3-N, TP, Fe, Cu, Ca, and pH has a strong negative loading with DO. PC1 refers to "Industrial and domestic discharges, mainly that of industrials", as in the wet season of Gr-3. Organic carbon derivatives, TSS, and Fe strongly contributed to PC2 (19.2% of total variance), and ORP has a strong negative load on this PC. Considering that organic carbon derivatives are the dominant pollutants here, PC2 can be associated with "Organic carbon derived OIZ discharges". PC3 (13.6% of total variance) has a strong positive loading with NO_3-N. Also, TP contributed moderately to this PC. In this status, PC3 can be attributed to "Urban diffuse pollutions". PC4 (9.3% of total variance) has a substantial load on Zn and Pb, as in the wet season. This situation is related to "Vehicle emissions". In addition, the contribution of the same elements to PC4 in two seasons supports vehicle emissions.

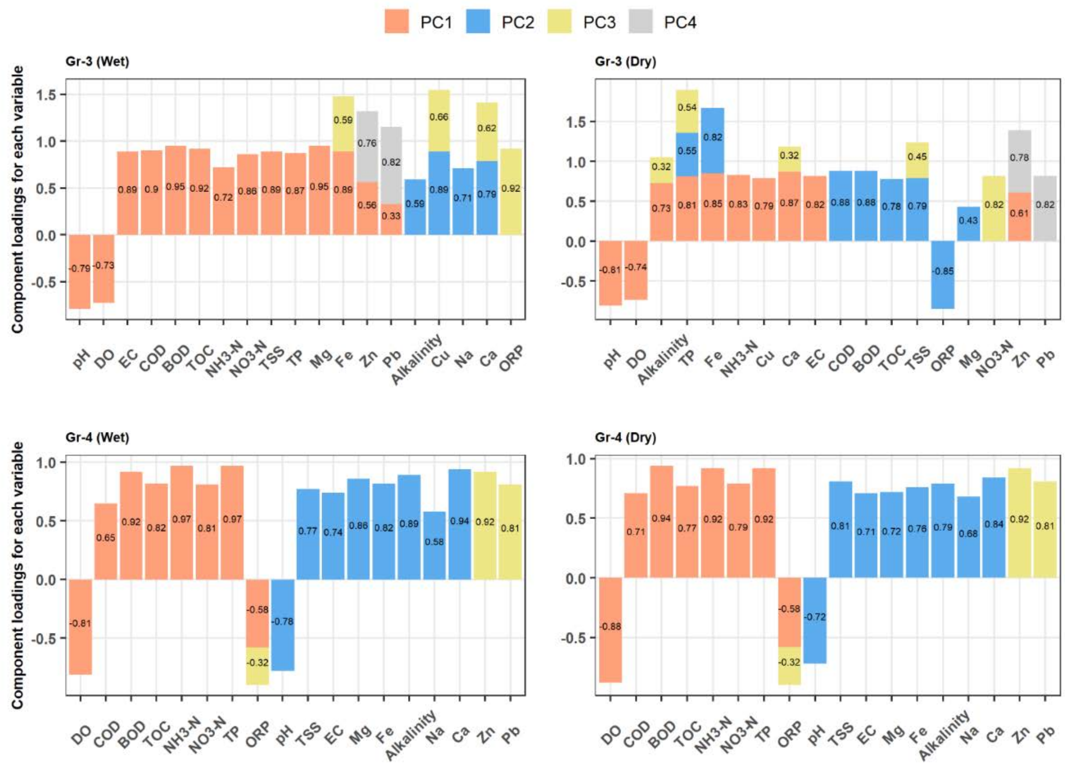

Figure 5. Varimax rotated loadings of significant principal components (eigenvalue >1) for the wet and dry seasons of Gr-3 and Gr-4.

In both seasons of Gr-4, the same elements clustered together with the same PCs. This may indicate that the pollution in this group is loaded from a continuous source. PC1 (47.3% of total variance) has a robust positive load on BOD, TOC, NH_3-N ve NO_3-N. Also, this PC is negatively characterized by DO and ORP. Due to the rapid chemical reactions of nitrogen derivatives in the aquatic environment, this is actual proof that the source of pollution is continuous. PC1 can be recognized as "Urban waters with organic character". EC, Alkalinity, TSS, Fe, Ca, and Mg positively contribute to PC2 (25.5% of total variance). pH has a high negative contribution to this PC. In particular, small and medium-sized industries are placed around the S10, and residential areas are around the S5. Therefore, PC2 can be attributed to "Continuous discharges of undefined source/s". PC3 has a high positive correlation with Pb and Zn. There is a highway with heavy vehicle traffic near the S10 and S13 sites. Clustering of only these two elements together in both seasons brings to mind "Vehicle emissions".

4. Conclusion

This study provides insights into the neglected total phosphorus pollution of the Saz-Cayirova stream, which is just one one of the industrialized and urbanized basins that transport high nutrient loads to the Marmara Sea. In the current research, research, a total of 19 WQ variables, including heavy metals, organic carbon derivatives, in-situ measurement parameters, parameters, and nutrients, were checked at 19 monitoring sites in the entire basin to determine the critical sources of total phosphorus pollution. Compared with reference sites, the concentrations of the four heavy metals examined were at a reasonable level. In addition to the four heavy metals studied, more specific heavy metals, such as antimony (Sb), manganese (Mn), and nickel (Ni), can also be measured, allowing more accurate assessments. Organic carbon derivatives stand out as WQ parameters of serious concern throughout the basin. Despite innovative technologies for heavy metal removal in OIZ wastewater treatment plants, measures for TP and organic pollutants originating from secondary production processes need to be developed urgently.

The main sources of the total phosphorus pollution in the whole basin can be listed as industrial processes wastewater such as dyeing, washing, phosphating, domestic wastewater from industries, urban diffuse waters with organic character, and continuous discharges of undefined sources. Our observations showed that TP pollution sources are more affected by GOIZ's discharges than the OSOIZ's discharges. Moreover, the relatively low seasonal variation in parameter concentrations indicates that point sources are more dominant than urban diffuse sources. Based on the results from field observations and statistical techniques, the following improvement strategies to minimize TP concentrations can be suggested; (1) Incorporate unpolluted tributaries into the stream network, (2) Add technologies to reduce the organic pollution in OIZ wastewater treatment plants, (3) Achieve a comprehensive approach to protect natural resources across the basin, (4) Conduct nutrient reduction simulations with modeling techniques using best management practices for each identified sub-groups.

Acknowledgment

This study has been financially supported by Scientific and Technological Research Council of Turkey (TUBITAK) Project number 119Y032 and Scientific Research Projects of Gebze Technical University (No: BAP 2020-A-101-11). Additionally, this study represents the preliminary evaluation of the TUBİTAK Project (No: 121G066).

References

[1] EEA, "European waters; Assessment of status and pressures," European Environment Agency, Copenhagen, 2018.
[2] C. S. Reynolds and P. S. Davies, "Sources and bioavailability of phosphorus fractions in freshwaters: a British perspective," *Biol Rev Camb Philos Soc,* pp. 27-64, 2001.
[3] E. Commision, "Evaluation of the Urban Waste Water Treatment Directive," European Comission, Brussels, 2019.
[4] L. Ngatia, J. M. Grace, D. Moriasi and R. Taylor, "Nitrogen and Phosphorus Eutrophication in Marine Ecosystems," in *Monitoring of Marine Pollution,* 2019.
[5] B. Toklu-Alicli, S. Polat and N. Balkis-Ozdelice, "Temporal variations in the abundance of picoplanktonic Synechococcus (Cyanobacteria) during a mucilage event in the Gulfs of Bandırma and Erdek," *Estuarine, Coastal and Shelf Science,* vol. 233, 2020.
[6] UNEP, "Assessment of the state of microbial pollution in the Mediterranean Sea," UNEP/MAP, Athens, 1995.
[7] M. S. Myklestad, "Release of extracellular products by phytoplankton with special emphasis on polysaccharides," *Science of Total Environment,* vol. 165, no. 3, pp. 155-164, 1995.
[8] M. Strokal and C. Kroeze, "Nitrogen and phosphorus inputs to the Black Sea in 1970–2050," *Regional Environmental Change,* vol. 13, pp. 179-192, 2013.
[9] Z. Uysal, "A net-plankton study in the Bosporus junction of the Sea of Marmara," *Turkish Journal of Botany,* vol. 20, pp. 321-327, 1996.
[10] O. Baştürk, S. Tugrul, A. Yilmaz and C. Saydam, "Health of the Turkish Straits, Chemical and Environmental Aspects of the Sea of Marmara," METU-Institute of Marine Sciences, 1990.

[11] H. Oruc, M. Çelen, F. Gülgen, M. S. Öncel, S. Vural ve B. Kılıç, “Sensitivity of the SWAT Model to Soil Data Parameterization; Case Study in Saz - Çayırova Stream, Turkey,” *5th EurAsia Waste Management Symposium*, İstanbul, 2020.
[12] APHA, "Standard Methods For Examination of Water and WasteWater," American Public Health Association, Washington, DC, USA, 1998.
[13] APHA, "Standard methods for the examination of water and waste water," American Public Health Association., Washington, DC, 1995.
[14] V. B. Tran, H. Ishidaira, T. Nakamura, T. N. Do ve K. Nishida, "Estimation of Nitrogen Load with Multi-pollution Sources Using the SWAT Model: a Case Study in the Cau River Basin in Northern Vietnam," *Journal of Water and Environment Technology,* pp. 106-119, 2017.
[15] M. H. Gholizadeh, A. M. Melesse and L. Reddi, "Analysis of spatiotemporal trends of water quality parameters using cluster analysis in South Florida," *World Environmental and Water Resources Congress 2016. American Society of Civil Engineers, Reston, VA,* pp. 519-528, 2016.
[16] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Third Edition ed., Amsterdam: Elsevier, 2011.
[17] H. F. Kaiser, "An index of factorial simplicity," *Psychometrika,* pp. 31-36, 1974.
[18] K. P. Singh, A. Malik, D. Mohan and S. Sinha, "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study," *Water Research,* pp. 3980-3992, 2004.
[19] C. W. Liu, K. H. Lin and Y. M. Kuo, "Application of factor analysis in the assessment of ground- water quality in a blackfoot disease area in Taiwan," *The Science of the Total Environment,* pp. 77-89, 2003.
[20] J. A. Ludwig and J. F. Reynolds, "Statistical Ecology: A Primer on Methods and Computing," in *Wiley-Interscience Publication,* New York, 1988.
[21] M. Vega, "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis," *Water Research,* pp. 3581-3592, 1998.
[22] L. Belkhiri, A. Boudoukha, L. Mouni and T. Baouz, "Multivariate statistical characterization of groundwater quality in Ain Azel plain, Algeria," *African Journal of Environmental Science and Technology,* pp. 526-534, 2010.
[23] A. Tiri and N. Lahbari, "Assessment of the quality of water by hierarchical clusterand variance analyses of the Koudiat Medouar Watershed,East Algeria," *Appl. Water Sci.,* 2014.
[24] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology,* pp. 417-441, 1933.
[25] Y. Yang, T. Li, T. Zhang and Q. Yu, "Time dimension analysis; Comparison of Nanjing local driving cycles in 2009 and 2017," *Sustainable Cities and Society,* pp. 1112-1121, 2020.
[26] K. L. Zhao, W. J. Fu, Q. Z. Qiu, Z. Q. Ye, Y. F. Li, H. Tunney and C. Y. Dou, "Spatial patterns of potentially hazardous metals in paddy soils in a typical electrical waste dismantling area and their pollution characteristics," *Geoderma,* pp. 453-462, 2019.
[27] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psycohometrika,* pp. 187-200, 1958.
[28] M.T.A, “Türkiye Doğal Kaynaklar ve Ekonomi Bülteni,” Maden Tetkik ve Arama Genel Müdürlüğü, Ankara, 2010.
[29] C.-W. Liu, K.-H. Lin and Y.-M. Kuo, "Application of Factor Analysis in the Assessment of Groundwater Quality in a Backfoot Disease Area in Taiwan," *Science of The Total Environment,* vol. 313, no. 3, pp. 77-89, 2003.
[30] J. K. Horne and D. C. Schneider, "Spatial variance of mobile aquatic organisms: capelin and cod in Newfoundland coastal waters," *Philosophical Transactions of the Royal Society* , pp. 633-641, 1997.
[31] E. G. Srinath and S. C. Pillai, "Phosphorus in Sewage, Polluted Waters, Sludges, and Effluents," *The Quarterly Review of Biology,* vol. 41, no. 4, pp. 384-407, 1966.