Water Resour Manage (2010) 24:959–978
DOI 10.1007/s11269-009-9481-7

Water Quality Assessment Using Multivariate Statistical Methods—A Case Study: Melen River System (Turkey)

Rabia Koklu · Bulent Sengorur · Bayram Topal

Received: 4 December 2008 / Accepted: 3 July 2009 /

Abstract This study is focused on water quality of Melen River (Turkey) and evaluation of 26 physical and chemical pollution data obtained five monitoring stations during the period 1995–2006. It presents the application of multivariate statistical methods to the data set, namely, principal component and factor analysis (PCA/FA), multiple regression analysis (MRA) and discriminant analysis (DA). The PCA/FA was employed to evaluate the high–low flow periods correlations of water quality parameters, while the principal factor analysis technique was used to extract the parameters that are most important in assessing high–low flow periods variations of river water quality. Latent factors were identified as responsible for data structure explaining 72–97% of the total variance of the each data sets. PCA/FA was supported with multiple regression analysis to determine the most important parameter in each factor. It examines the relation between a single dependent variable and a set of independent variables to best represent the relation in the each factor. Obtained important parameters provided us to determine the major pollution sources in Melen River Basin. So factors are conditionally named soil structure and erosion, domestic, municipal and industrial effluents, agricultural activities (fertilizer, irrigation water and livestock wastes), atmospheric deposition and seasonal effects factors. DA applied the data set to obtain the parameters responsible for temporal and spatial variations. Assessment of high–low flow period changes in surface water quality is an important aspect for evaluating temporal and spatial variations of

---

R. Koklu (✉) · B. Sengorur
Environmental Engineering Department, Sakarya University,
Esentepe Campus, 54187 Sakarya, Turkey
e-mail: rkoklu@sakarya.edu.tr

B. Topal
Industrial Engineering Department, Sakarya University,
Esentepe Campus, 54187 Sakarya, Turkey

river pollution. The aim of this study is illustration the usefulness of multivariate statistical analysis for evaluation of complex data sets, in Melen River water quality assessment identification of factors and pollution sources, for effective water quality management determination the spatial and temporal variations in water quality.

Keywords Surface water quality · Multivariate statistical analysis · Parameter reduction · Melen River

1 Introduction

The surface water quality is a matter of serious concern today. Rivers due to their role in carrying off the municipal and industrial wastewater and run-off from agricultural land in their vast drainage basins are among the most vulnerable water bodies to pollution (Singh et al. 2005). Flow in rivers is a function of many factors including precipitation, surface runoff, interflow, groundwater flow and pumped inflow and outflow. Seasonal variations of these factors have a strong effect on flow rates and hence on the concentration of pollutants in the river water (Vega et al. 1998). It is imperative to prevent and control the rivers pollution and to have reliable information on the quality of water for effective management (Singh et al. 2005). For effective pollution control and water resource management, it is required to interpreting a large number of water quality data. Results in a huge and complex data matrix comprised of a large number of physico-chemical parameters, which are often difficult to interpret and draw meaningful conclusions (Dixon and Chiswell 1996).

Factor analysis, which includes principal component analysis (PCA) is a very powerful technique applied to reduce the dimensionality of a data set consisting of a large number of inter-related variables, while remaining as much as possible the variability present in data set. This reduction is achieved by transforming the data set into a new set of variables, the principal components (PCs), which are orthogonal (non-correlated) and are arranged in decreasing order of importance (Panda et al. 2006).

Principal component analysis provides information on the most meaningful parameters, which describe whole data set rendering data reduction with minimum loss of original information (Singh et al. 2004). PCA has allowed the identification of a reduced number of latent factors with pollution sources such as spatial (pollution from anthropogenic origin) and temporal (seasonal and climatic) sources of variation affecting quality and hydrochemistry of river water have been differentiated and assigned to polluting sources (Shrestha and Kazama 2007; Simeonov et al. 2003; Kowalkowski et al. 2006; Pekey et al. 2004; Vega et al. 1998). At the same time PCA has allowed the explaining of related parameters by only one factor (Boyacıoğlu and Boyacıoğlu 2006; Kannel et al. 2007; Kottı et al. 2005; Kowalkowski et al. 2006; Sengörür and İsa 2001; Singh et al. 2004) and exposing of the important factor responsible for seasonal changes in river water quality (Ouyang 2005; Ouyang et al. 2006).

Multiple Regression examines the relation between a single dependent variable and a set of independent variables to best represent the relation in the population.

It was used to investigate relationships between water quality parameters (physico-chemical and biological) and landscape characteristics (Amiri and Nakane 2009; Sliva and Williams 2001; Wang 2001; Singh et al. 2005) and to investigate between water quality and pathogen indicators (Mallin et al. 2000; Crowther et al. 2001).

Discriminant Analysis is a statistical method which obtains to discriminate variables between two or more naturally occurring groups. DA was used for identification of water quality variables responsible for spatial and temporal variations in river water quality (Shrestha and Kazama 2007; Singh et al. 2004, 2005). Kowalkowski et al. (2006) used DA to classify and, thus, to confirm the groups found by means of the cluster analysis.

In the present study, a large data matrix, obtained during an 11-year (1995–2006) monitoring program, is subjected to PCA/FA, MRA and DA techniques. The aims of this study are to demonstrate the application of the data reduction techniques (PCA/FA) to evaluate the importance of various water quality parameters, regression techniques (MRA) to determine the most meaningful parameters responsible for water quality and discriminant techniques (DA) identification of water quality variables responsible for spatial and temporal variations in river water quality.

2 Materials and Methods

2.1 Study Area

The Melen River is in the north of the Turkey which has got a complex river system that supports a variety of uses, including irrigation systems in agricultural lands, drinking water and several different industries wastewater. The return flow from all these users is directly discharged into the river. The Big Melen River receives some tributary streams, Aksu Stream, Ugur Stream, the Small Melen River and Asar Stream. Aksu Stream, Ugur Stream and Asar Stream contribute the Small Melen River and then, the Small Melen River converges with Big Melen River a derivation canal before discharging into Efteni Lake. River System calls the Big Melen River after Efteni Lake. The Melen River Basin covers about 2,317 km^2. It flows into the Black Sea with 48.2 m^3/s average flow according to measurement results between 1995–2006. Düzce City (Turkey) is the biggest urban settlement in the Melen River Basin. Important pollution sources in basin are domestic, industrial wastewater and agricultural run-off. The Melen River is also affected by non-point sources of pollution including fertilizers from farm effluents, livestock wastes and septic tanks effluents or point sources of pollution including domestic, some industries wastewater discharges and treatment plant effluents. Asar Stream drainage basin has intensive industrial activities and daily residence plants effluents. The Small Melen River receives sewage wastewater discharges from Düzce City after wastewater is treated, livestock wastes and run-off from dumps. The other towns in the basin have no sewage treatment plant. Basin has density agricultural activity and also intensive forest cover (at high regions). Therefore, mineral concentration comes from the alluvion soil structure (gravel, sand, clay, silt) (Düzce Environment State Report 2007).

The Melen River System Drainage Basin and Sample Points are shown in Fig. 1. The data for five sample points were obtained from monitoring programs of Government Water Association in Ankara, Turkey.

2.2 Measurement Stations and Parameters

River water data were obtained at high and low flow periods from at three different sites on river and at two different sites on its tributaries, namely Small Melen Pasakonagi (1KMP), Big Melen Pakmaya (2BMP), Big Melen Aydogan (3BMA) and Aksu Stream (4AC), Ugur Stream (5US) (Fig. 1) once 2–3 month during a period of 11 years (1995–2006).

The data were obtained from by analyzing of Government Water Association monitoring program. pH, temperatures (T, °C), electrical conductivity (EC, mohm/cm), Suspended Solids (SS, mg L^{-1}), turbidity (turb, NTU), flow (Q, m^3 sn^{-1}), total alkalinity (M-Alk, mg L^{-1}), total hardness (T-Hard, mg L^{-1}), total dissolved solids (TDS, mg L^{-1}), dissolved oxygen (DO, mg L^{-1}), 5-days biochemical oxygen demand (BOD_5, mg L^{-1}), chemical oxygen demand (COD, mg L^{-1}), ammonium nitrogen (NH_4–N, mg L^{-1}), nitrate nitrogen (NO_3–N, mg L^{-1}), nitrite nitrogen (NO_2–N, mg L^{-1}), chloride (Cl^-, mg L^{-1}), sulphate (SO_4^{2-}, mgL^{-1}), phosphate (PO_4^{3-}, mg L^{-1}), sodium (Na^+, mg L^{-1}), potassium (K^+, mg L^{-1}), calcium (Ca^{2+},

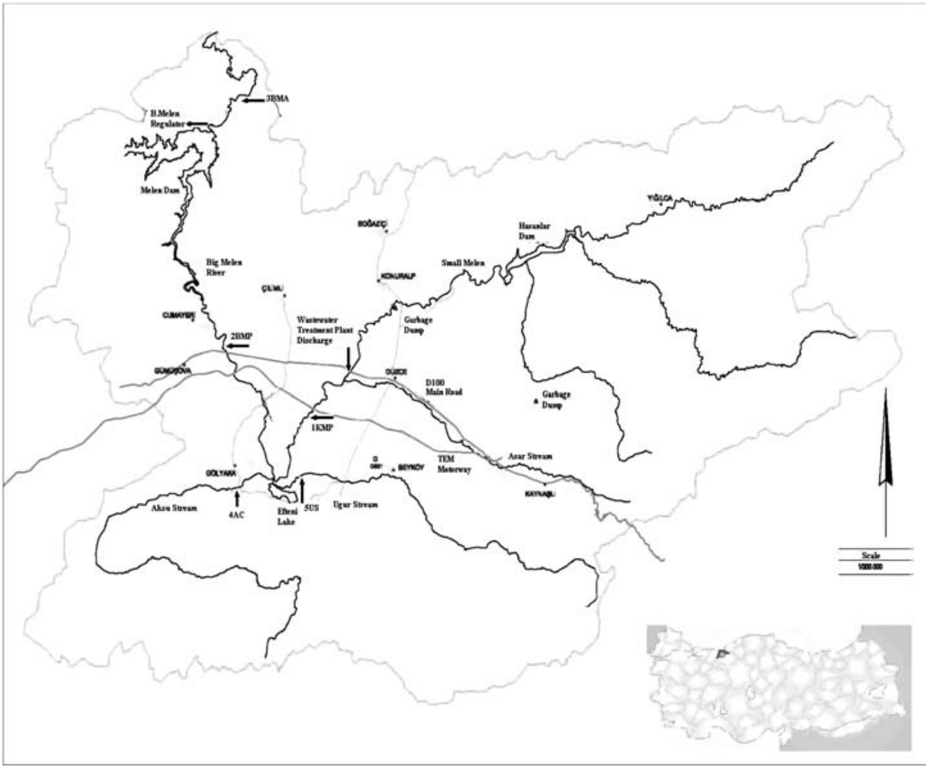

Fig. 1 Big Melen River system drainage basin and sample points

mg L^{-1}), magnesium (Mg^{2+}, mg L^{-1}), boron (B^{3+}, mg L^{-1}), ferrous (Fe^{2+}, mg L^{-1}), manganese (Mn^{2+}, mg L^{-1}) and fecal streptococcus (F-Strip, MPN/100 mL) and total coliform (T-Coli, MPN/100 mL), Escherichia coliform (E-Coli, MPN/100 mL) were analyzed by laboratory of Government Water Association using standard methods (APHA-AWWA-WPCF 1999). The basic statistics of the 11 years data set on river water quality is summarized Tables 1 and 2.

2.3 Data Analysis and Statistical Methods

In this study, the water quality parameters were grouped in two different periods (wet and dry season) by investigating the months of rainy–dry and high–low flow. They were analyzed and their mean values and standard deviations are summarized in Tables 1 and 2. Before the applying of the multivariate statistical techniques, experimental data were normalized within the range 0.1 to 0.9 in order to avoid misclassification due to wide differences in data dimensionality. tends to minimize the influence of difference of variance of variables and eliminates the influence of different units of measurement and renders the data dimensionless.

$$x_i = 0.8\frac{(x - x_{min})}{(x_{max} - x_{min})} + 0.1 \tag{1}$$

where x_i is the normalized value of a certain parameter, x is the measured value for this parameter, x_{min} and x_{max} are the minimum and maximum values in the database for this parameter, respectively (Dogan et al. 2009).

The water quality data were subjected to different multivariate statistical techniques to explore the parameters which were responsible of the water quality variation in different season. PCA/FA was applied to the normalized data of each station for two periods. Varimax rotating method was used for applying PCA/FA. Corresponding f_{Hij} (f: Factor; H: High flow period; i: Station number; j: Factor number) and f_{Lij} (f: Factor; L: Low flow period; i: Station number; j: Factor number), variable loadings and explained variance are determined. Liu et al. (2003) classified the factor loadings as ‘strong’, ‘moderate’ and ‘weak, corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively. FA/PCA results were exposed to MRA. Multiple regression analysis examines the relation between a single dependent variable and a set of independent variables to best represent the relation in the each factor. MRA models were used to determine the most meaningful parameters responsible for the water quality. DA was used for identification of water quality variables responsible for spatial and temporal variations in river water quality. Discriminant functions related to the parameters of temporal and spatial variations were obtained. All the mathematical and statistical computations were made using Excel 2003 and SPSS 13.0.

2.3.1 Factor Analysis/Principal Component Analysis (FA/PCA)

Principal Component Analysis is a data analysis method focused on a particular collection of variables. Consider the form of the first principal component. The score for individual i on component, c_{i1}, uses weights $w_{11},\ldots\ \ldots,w_{p1}$ in the linear combination

$$c_{i1} = y_{i1}w_{11} + y_{i2}w_{22} + \cdots + y_{ip}w_{p1} \tag{2}$$

Table 1 Stations on Melen River

Parameter	Small Melen Pasakonagi (1KMP)				Big Melen Pakmaya (2BMP)				Big Melen Aydogan (3BMA)			
	High flow		Low flow		High flow		Low flow		High flow		Low flow	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
B	0.0475	0.11587	0.0795	0.07742	0.0581	0.09633	0.0677	0.08217	0.0602	0.08889	0.1020	0.11536
BOD_5	4.08	1.87873	6.7167	4.40176	2.6635	0.90927	3.1157	1.04185	3.8703	1.49243	3.3807	1.39066
Ca	51.37	6.17434	63.7368	10.21307	50.5462	6.86686	58.3333	8.70433	50.8951	8.62621	60.6394	7.64316
Cl					5.6277	1.47281	6.6364	2.73616	7.5225	3.42923	11.1296	5.39073
COD	25.20	18.39929	28.0040	14.2745	21.09654	21.39284	19.14857	18.3558	25.5949	15.5193	20.9885	9.48354
DO	10.79	1.40016	7.4410	1.44745	10.9442	1.76924	7.8690	1.08177	11.1451	1.29681	9.2265	1.28497
EC	320.25	40.15583	426.6000	75.87413	324.6923	45.51287	383.8636	65.73127	336.4098	65.8777	422.746	66.4634
E-Coli	40,285.71	36,687.72844	45,150.000	55,508.88	12,665.38	16,774.62	13,850.00	30,932.31	11,843.33	10,519.64	3,464.153	4,327.901
Fe	5.78	9.57239	5.8230	12.46912	3.4515	2.79640	2.4776	2.32617	6.0580	11.39549	1.4176	1.83772
F-Strip	11,767.85	12,975.56260	13,690.526	19,078.83	4,264.00	5,383.68	1,842.63	4,111.59	4,677.50	5,667.76	856.230	1,662.86
K	1.57	1.60401	1.9370	1.05099	1.3428	1.20105	1.6486	0.70016	3.1562	2.90994	5.8651	4.02289
M-Al	138.50	17.25946	184.6500	31.07804	144.2308	19.3464	172.954	28.23663	143.7131	25.31098	181.9085	25.10340
Mg	5.17	1.60529	8.4925	2.21323	6.7142	1.89166	8.7076	2.48830	6.5233	2.27620	9.0390	2.79976
Mn	0.28	0.29327	0.2702	0.14969	0.2063	0.12753	0.3573	0.56382	0.5261	0.96813	0.3129	0.66239
Na	6.65	1.30346	10.7670	3.40073	6.5460	1.39539	9.6071	2.77759	8.6145	4.20530	13.7623	5.80187
NH_4–N	0.48	0.23733	1.2021	0.91370	0.2569	0.14375	0.7012	0.47846	0.4163	0.25618	0.4490	0.36091
NO_2–N	0.010	0.00592	0.0544	0.03359	0.0117	0.01004	0.0667	0.04700	0.0538	0.05113	0.1180	0.07961
NO_3–N	0.61	0.18921	0.6985	0.24917	0.6242	0.18433	0.6510	0.24269	0.9066	0.53159	1.6158	0.78287
o-PO_4	0.23	0.10784	0.5585	0.29224	0.1881	0.12129	0.3624	0.22770	0.2615	0.23602	0.5263	0.30728
pH	7.70	0.29986	7.6135	0.24616	7.6546	0.29037	7.5841	0.22620	7.9205	1.49599	7.9483	0.84060
pV					1.9000	0.56745	2.0024	0.51651	2.5151	0.80339	2.9642	1.15375
Q	28.44	25.07708	5.8193	2.01603	51.9456	28.87547	18.8956	14.57005	80.5977	48.07875	19.9859	12.773
SO_4	17.29	4.08728	24.5650	6.61850	15.8885	3.12581	18.0238	3.87684	19.7025	16.86799	20.9620	4.75411
SS	171.62	167.55881	110.6842	123.77311	163.3654	152.986	92.7619	103.427	217.6557	285.9269	47.0571	52.2169
T	10.39	4.43337	20.9500	2.28208	9.3846	4.34582	19.6818	3.79707	10.1148	4.25088	19.1127	5.09776
T-Coli	180,642.85	285,869.72	127,968.42	165,747.56	70,960.00	178,760.42	43,460.00	62,146.73	49,066.66	68,566.83	18,006.34	27,376.3
TDS	195.42	26.37961	255.4500	43.92275	197.8846	28.99148	233.4545	39.83983	206.5574	43.30070	236.478	82.4956
TH	149.76	18.90056	196.2000	33.17434	154.0000	20.07934	181.6429	30.22960	154.1967	27.03644	187.6761	23.65241

Table 2 Stations on tributaries

Parameters	Aksu stream (4AC)				Ugur stream (5US)			
	High flow		Low flow		High flow		Low flow	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
B	0.0452	0.09633	0.0586	0.09234	0.0448	0.10282	0.0857	0.07922
BOD_5	1.7326	0.88114	1.1568	0.95032	4.6122	13.52943	2.2048	3.04577
Ca	34.6741	7.78087	41.4091	7.05852	51.2000	8.14220	49.3429	6.04720
Cl	3.7785	5.36870	4.9286	8.44243	2.9641	1.24395	3.7219	2.04861
COD	18.4581	25.121	13.558	8.44752	15.8837	13.69328	19.5762	18.5334
DO	11.4030	1.30123	9.6041	0.92624	11.4356	1.32321	9.0771	1.16962
EC	225.00	38.751	269.545	34.21045	311.78	46.96507	338.6190	29.7833
E-Coli	705.81	1,326.23	4,411.66	17,338.73	603.185	987.129	2,405.50	5,143.916
Fe	5.5163	14.632	5.0800	17.56965	4.8519	9.01978	4.4119	5.71896
F-Strip	270.03	422.031	414.142	870.951	326.629	680.5081	1,469.80	3,756.597
K	1.1162	1.57691	0.8068	0.6198	1.2962	1.41083	0.9933	0.5029
M-Al	104.37	19.12388	129.295	20.087	149.129	22.2116	158.238	14.1815
Mg	5.6696	2.12625	7.6345	1.72440	6.5467	1.66446	9.5800	2.57759
Mn	0.1191	0.13788	0.1461	0.30092	0.1561	0.18425	0.2696	0.43115
Na	4.4985	1.28228	4.7900	1.2158	5.6138	1.29423	8.1024	2.34447
NH_4–N	0.1351	0.09074	0.1562	0.12358	0.1111	0.07045	0.1464	0.21028
NO_2–N	0.0057	0.00568	0.0046	0.00410	0.0030	0.00294	0.0028	0.00236
NO_3–N	0.4793	0.17038	0.3986	0.12225	0.3052	0.13569	0.2871	0.20035
o-PO_4	0.1176	0.15327	0.0732	0.08191	0.0907	0.11629	0.0376	0.03360
pH	7.7796	0.43413	8.0591	0.46600	7.9948	0.39073	8.0414	0.33959
pV	1.4900	0.95015	1.2768	0.68762	1.5541	0.88982	1.3905	0.70740
Q	7.6010	6.91109	2.8314	1.98392	6.4019	5.61770	1.7422	1.65328
SO_4	10.6593	2.80977	12.4682	4.60749	13.1333	2.51304	17.0905	5.24890
SS	114.59	149.982	54.1905	79.6201	254.667	566.6619	244.10	400.2680
T	9.3148	4.05526	20.0682	3.04058	9.2963	4.34843	22.0476	3.61215
T-Coli	2,384.62	3,696.26	4,644.04	10,537.40	2,532.07	4,476.029	53,965.00	222,774.28
TDS	144.33	21.92820	167.636	18.68675	186.703	27.77356	203.28	19.68284
TH	110.00	20.83590	134.977	21.95855	154.981	23.28935	162.78	17.23265

This linear combination is chosen so that the sum of squares of c_1 is as large as possible subject to the condition that ${w^2}_{11} + \ldots \ldots \ldots + {w^2}_{p1} = 1$. The second principal component is another linear combination of y_j

$$c_{i2} = y_{i1}w_{12} + y_{i2}w_{22} + \cdots + y_{ip}w_{p2} \tag{3}$$

where the variance c_2 is the maximal, subject to the conditions that corr(c_1, c_2) = 0 and that ${w^2}_{12} + \ldots \ldots \ldots + {w^2}_{p2} = 1$. The criterion of summarizing the information in p variables by a few components is valuable as a means of reducing the number of variables needed in an analysis (Tinsley and Brown 2000).

FA follows PCA. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules, and constructing new variables, also called varifactors (VF). PCA of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation (raw) generating VFs (Brumelis et al. 2000; Singh et al. 2004, 2005).

The FA can be expressed as:

$$y_{ji} = f_{j1}z_{i1} + f_{j2}z_{i2} + \cdots + f_{jm}z_{im} + e_{ij} \tag{4}$$

where y is the measured variable, f is the factor loading, z is the factor score, e the residual term accounting for errors or other source of variation, i the sample number and m the total number of factors (Tinsley and Brown 2000).

2.3.2 Multiple Regression Analysis (MRA)

Multiple regression analysis is a statistical tool for understanding the relationship between two or more variables. Multiple regression examines the relation between a single dependent variable and a set of independent variables to best represent the relation in the population. The technique is used for both predictive and explanatory purposes within experimental or nonexperimental designs (Tinsley and Brown 2000). When there are an arbitrary number of explanatory variables, the linear regression model takes the following form:

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \ldots \ldots \ldots . + \beta_m.X_m + e_{ij} \tag{5}$$

where Y represents the dependent variable and $X_1 \ldots X_m$ represent the different independent variables, $\beta_0, \ldots\ldots\ldots, \beta_m$ represent the regression coefficient and e represents the random error (Freund and Wilson 1998). The error term e represents the collective unobservable influence of any omitted variables. In a linear regression each of the terms being added involves unknown parameters, which are estimated by "fitting" the equation to the data using least-squares.

2.3.3 Discriminant Analysis (DA)

Discriminant Analysis is a statistical method which obtains to discriminate variables between two or more naturally occurring groups. It calculates mathematical weights for scores on each discriminator variable that reflect the degree to which scores on that variable differ among the groups being discriminated. It forms one or more weighted linear combinations of discriminator variables called discriminant functions. Each discriminant function has the general form:

$$D = a + b_1.X_1 + b_2.X_2 + \ldots\ \ldots\ \ldots. + b_P.X_P \tag{6}$$

where D is the discriminant score (z score), a is the Y-intercept of the regression line, b is the discriminant function coefficient, X is the discriminator variable raw score, and p is the number of discriminator variable (Tinsley and Brown 2000).

3 Results and Discussions

3.1 High Flow Period

PCA of the 26 parameters constituted seven PCs explaining about 85.05% of the total variance at the 1KMP Station water quality data set for high flow period as seen in Table 3. f_{H11} has strong positive loadings on Ca^{2+}, M-Al, TH, EC, TDS; negative loadings on Q and moderate negative loadings on Fe^{2+}, COD and positive loadings on E-Coli. It was formed 25.84% of total variance. Flow must be measured in water quality monitoring program. f_{H12} accounting for 12.95% of total variance has strong positive loadings on K^+, F-Strip, o-PO_4^{3+} and moderate loadings on B^{3+}, SO_4^{2-} and weak loadings on Na according to PCA results. f_{H13} accounting for 10.05% of total variance has strong positive loadings on NO_2–N, NO_3–N and moderate loadings on BOD_5; weak loadings on T-Coli, NH_4–N according to PCA results. Whereas, fourth, fifth, sixth and seventh factors accounted for the total variance of 10.80%; 8.761%; 8.936% and 7.692%, were correlated with DO and T; pH and SS; Mg^{2+} and Mn^{2+} respectively.

When FA/PCA and MRA were investigated together, TH 98% or TDS and COD, E-Coli could explain the f_{H11}. This factor represents soil structure and subsequent run-off which can be interpreted as a mineral component of the river and anthropogenic facilities. This clustering of variables points to a common origin for these minerals likely from dissolution of limestone and gypsum soils (Vega et al. 1998). It may be noted that gypsum is widely used as soil modifier in the river catchments (Singh et al. 2004). MRA was applied the data, 66% of parameters in f_{H12} were classified correctly by using o-PO_4^{3-}. We say that o-PO_4^{3-}, SO_4^{2-} and F-Strip could explain the f_{H12} which can be interpreted as agricultural activities such as livestock waste and atmospheric deposition in basin. 55% of parameters in f_{H13} were classified correctly by using NO_2–N and NH_4–N which can be interpreted as organic contamination mainly from domestic wastewater and run-off from dump. NO_2–N is usually associated with active biological process influenced by organic pollution. DO 61% and T explain the f_{H14} as a seasonal factor; pH and SS explain the f_{H15} which describes industrial activities; Mg^{2+} and Mn^{2+} explain the f_{H16} and f_{H17} factor.

PCA of the 28 parameters constituted five PCs explaining about 85.2% of the total variance at the 2BMP Station water quality data set for high flow period as seen in Table 3. f_{H21} explained 30.81% of total variance and was strong positively contributed

Table 3 High flow period FA/PCA results

High flow																						
1KMP Station varimax rotated results								2BMP Station varimax rotated results						3BMA Station varimax rotated results								
Variable	Factors							Variable	Factors					Variable	Factors							
	f_{H11}	f_{H12}	f_{H13}	f_{H14}	f_{H15}	f_{H16}	f_{H17}		f_{H21}	f_{H22}	f_{H23}	f_{H24}	f_{H25}		f_{H31}	f_{H32}	f_{H33}	f_{H34}	f_{H35}	f_{H36}	f_{H37}	f_{H38}
Ca	0.934	−0.031	0.127	−0.207	−0.117	−0.065	−0.114	TH	0.968	0.023	−0.199	−0.012	0.026	Cl	0.925	0.012	0.105	0.099	−0.044	0.123	0.072	0.019
M-Al	0.920	0.013	0.123	−0.039	−0.136	0.245	−0.147	EC	0.959	0.166	−0.112	−0.062	0.098	EC	0.905	−0.359	−0.045	0.068	−0.003	−0.085	0.026	0.088
TH	0.918	0.051	0.143	−0.148	−0.108	0.257	−0.103	TDS	0.948	0.227	0.018	−0.014	−0.035	Na	0.901	−0.074	0.081	−0.047	−0.128	0.065	−0.010	0.174
EC	0.915	0.199	0.219	−0.108	−0.137	0.122	0.074	M-Al	0.936	0.033	−0.278	0.024	0.057	TDS	0.900	−0.279	0.020	0.108	−0.058	−0.132	0.024	0.083
Q	−0.866	0.006	−0.151	−0.060	0.219	−0.077	0.175	Ca	0.875	0.014	−0.277	−0.105	0.319	M-Al	0.881	−0.407	−0.107	−0.013	0.013	−0.117	0.065	0.020
TDS	0.799	0.271	0.338	−0.043	−0.075	0.008	0.087	Na	0.826	0.277	0.079	−0.072	0.184	TH	0.847	−0.433	−0.119	0.108	−0.013	−0.101	0.079	0.028
Fe	−0.611	0.376	0.279	−0.110	0.495	−0.112	−0.069	SO_4	0.778	0.257	0.377	−0.292	−0.195	NO_3–N	0.842	0.240	0.011	−0.081	0.003	−0.088	−0.033	0.335
E-Coli	0.552	0.460	0.173	0.302	−0.182	−0.416	−0.183	COD	−0.659	0.019	0.577	−0.077	0.221	Ca	0.761	−0.471	−0.158	0.091	0.084	−0.024	0.012	0.040
COD	−0.523	0.453	0.233	−0.059	0.501	−0.306	−0.187	Cl	0.584	0.077	−0.044	−0.473	0.188	NO_2–N	0.738	−0.194	0.193	−0.262	−0.116	−0.024	−0.082	−0.195
K	−0.111	0.859	−0.050	−0.028	0.101	0.098	0.195	T-Coli	0.151	0.961	−0.092	0.053	0.021	Mg	0.701	−0.193	0.000	0.089	−0.225	−0.214	0.198	0.000
F-Strip	0.151	0.819	−0.078	−0.216	−0.138	−0.032	−0.138	E-Coli	0.209	0.950	−0.040	0.134	0.075	SO_4	0.553	−0.330	0.301	0.163	−0.118	0.045	0.239	0.133
o-PO_4	0.272	0.770	0.268	0.030	0.054	0.080	0.091	SS	−0.544	0.798	−0.023	−0.049	0.187	o-PO_4	0.544	0.122	0.270	−0.334	−0.098	0.011	−0.058	0.248
B	−0.260	0.555	−0.131	0.211	0.281	0.168	0.173	BOD_5	0.372	0.790	0.204	0.181	0.182	K	0.538	−0.018	0.414	−0.132	−0.051	0.211	0.336	0.003
SO_4	0.253	0.535	0.054	−0.023	−0.102	0.531	0.327	Mn	0.366	0.756	0.062	0.195	−0.257	SS	−0.283	0.910	0.196	0.006	0.033	−0.057	−0.042	−0.023
Na	0.330	0.479	0.203	0.205	−0.129	.409	0.343	F-Strip	0.126	0.712	0.632	0.163	−0.069	Fe	−0.228	0.893	0.128	0.018	0.001	−0.032	−0.016	0.002
NO_2–N	0.165	0.052	0.858	0.173	−0.010	0.337	−0.137	o-PO_4	0.025	0.114	0.962	0.140	−0.006	Turb	−0.286	0.885	0.063	−0.008	−0.015	−0.038	0.038	−0.006
NO_3–N	0.294	−0.232	0.742	−0.078	0.149	−0.215	0.134	B	−0.033	−0.040	0.949	0.177	−0.031	Q	−0.501	0.565	−0.115	0.110	−0.181	0.494	−0.051	−0.063
BOD_5	0.239	0.475	0.574	−0.176	−0.041	0.068	0.373	K	−0.178	0.144	0.804	−0.375	−0.192	pH	0.003	−0.383	−0.025	0.043	0.319	0.030	−0.027	0.173
T-Coli	0.345	0.193	0.443	0.407	−0.209	0.090	−0.265	pV	−0.321	−0.198	0.765	−0.015	0.285	BOD5	0.187	−0.110	0.895	0.135	0.062	0.096	0.145	0.135
NH_4–N	0.408	0.225	0.435	0.043	−0.210	−0.369	0.291	Fe	−0.565	0.492	0.591	−0.048	−0.085	pV	0.145	0.255	0.777	−0.149	0.129	−0.104	0.071	−0.193
T	−0.056	−0.062	−0.080	0.942	0.064	0.068	0.069	NO2–N	0.340	0.164	0.167	0.831	0.191	COD	−0.018	0.479	0.528	0.091	0.209	−0.093	0.161	0.066
DO	0.227	0.055	−0.132	−0.888	−0.190	0.040	0.083	T	−0.175	0.085	−0.114	0.796	−0.188	Mn	−0.219	0.265	0.517	0.138	0.007	0.017	−0.090	0.076
pH	0.321	0.156	−0.004	−0.194	−0.812	0.004	−0.100	DO	0.256	−0.324	−0.003	−0.762	0.061	T	0.045	−0.032	−0.140	−0.930	0.070	0.006	0.019	0.007
SS	−0.146	0.289	−0.111	0.200	0.639	0.009	0.512	pH	0.537	0.036	−0.053	−0.601	0.063	DO	0.155	−0.008	−0.002	0.891	0.154	0.035	0.015	−0.090
Mg	0.395	0.201	0.108	0.064	−0.033	0.826	−0.023	NO_3–N	0.341	0.137	0.106	0.003	0.852	F-Strip	−0.071	0.043	0.054	0.111	0.872	0.079	−0.003	−0.081
Mn	−0.317	0.159	0.068	−0.076	0.178	0.054	0.848	Mg	0.492	0.028	0.139	0.225	−0.717	E-Coli	−0.194	−0.060	0.156	−0.044	0.854	0.056	0.033	0.099
								NH_4–N	0.334	0.297	0.242	0.469	0.529	T-Coli	−0.090	−0.141	0.022	0.014	0.183	0.919	0.098	−0.053
								Q	−0.300	0.277	0.139	0.176	−0.356	B	0.106	0.034	0.149	0.006	0.027	0.072	0.973	0.004
														NH_4–N	0.405	−0.130	0.068	−0.140	0.080	−0.085	0.019	0.842
Variance	25.84	12.95	10.05	10.80	8.76	8.936	7.692	Variance	30.81	17.9	16.80	11.48	8.10	Variance	28.95	12.18	10.58	8.87	7.42	5.76	6.56	5.11
%	25.84	38.79	48.85	59.66	68.42	77.35	85.05	%	30.81	48.79	65.60	77.08	85.19	%	28.95	41.14	51.72	60.59	68.02	73.78	80.35	85.46

by TH, EC, TDS, M-Al, Ca^{2+}, Na^{+}, SO_4^{2-}; moderate positively contributed by Cl^{-} and negatively contributed by COD. f_{H22} explained 17.9% of total variance and was strong positively contributed by T-Coli, E-Coli, SS, BOD_5, Mn^{2+}; moderate positively contributed by F-Strip. f_{H23} explained 16.8% of total variance and strong correlated with o-PO_4^{3-}, B^{3+}, K^{+}, pV; moderate correlated with Fe^{2+}. f_{H24} formed 11.48% of total variance and strong positively correlated with NO_2–N, T; negatively correlated with DO and moderate negatively correlated with pH. f_{H25} accounted for the total variance of 8.1% and was strong positively contributed by NO_3–N; moderate negatively contributed by Mg^{2+}; moderate positively contributed by NH_4–N and weak negatively contributed by Q.

FA/PCA and MRA were investigated together and Ca 95.6% and COD, Na^{+}, TDS could explain the f_{H21}. This factor represents soil structure which can be interpreted as a mineral component of the river and urban run-off such as industrial, commercial, residential. MRA was applied the data, 86,7% of parameters in f_{H22} were classified correctly by using T-Coli and SS. f_{H22} represents the bacteriological and anthropogenic pollution such as municipal wastewater and agricultural activities such as livestock wastes. 55.5% of parameters in f_{H23} were classified correctly by using o-PO_4^{3-} which can be interpreted as diffuse agricultural activities and point sewage treatment work. Phosphorus absorbs strongly onto soil particles, most of the diffuse load enters the river during run-off events in autumn and winter, under conditions of high river flow (Mainstone and Parr 2002). $Fe^{2+}$and pV can be added the measurement program. 73.6% of parameters in f_{H24} were classified correctly by using DO and pH. This factor represents the industrial discharges. f_{H25} was explained by NO_3–N, Mg^{2+}, NH_4–N, Q according to MRA and this factor represents agricultural activities and atmospheric deposition. Nitrate is more associated with the use of organic and inorganic fertilizers (Maillard and Santos 2008; Vega et al. 1998).

PCA of the 29 parameters constituted eight PCs explaining about 85.46% of the total variance at the 3BMA Station water quality data set for high flow period as seen in Table 3. f_{H31} has strong positive loadings on Cl^{-}, EC, Na^{+}, TDS, M-Al, TH, NO_3–N, Ca^{2+} and moderate loadings on NO_2–N, Mg^{2+}, SO_4^{2+}, o-PO_4^{3-}, K^{+}. It was formed 25.95% of total variance. f_{H32} accounting for 12.18% of total variance has strong positive loadings on SS, Fe^{2+}, Turb; moderate positive loadings on Q and weak negative loadings on pH according to PCA results. f_{H33} accounting for 10,58% of total variance has strong positive loadings on $BOD_{5,}$ pV; moderate positive loadings on COD, Mn^{2+}. Whereas, fourth, fifth, sixth, seventh and eight factors accounted for the total variance of 8.87%; 7.42%; 5.76%; 6.56% and 5.11%, were strong correlated with DO and T; F-Strip, E-Coli; T-Coli; B^{3+}; NH_4–N respectively.

FA/PCA and MRA were investigated together and Ca^{2+} 99% and SO_4^{2-}, NO_2–N, o-PO_4^{3-} could explain the f_{H31}. This factor represents soil structure and agricultural activities such as fertilizer. MRA was applied the data, 92% of parameters in f_{H32} were classified correctly by using Turb. and Q. This factor explains the erosion from upland areas during rainfall events (Shrestha and Kazama 2007). f_{H33} were classified correctly by using BOD_5, pV, COD, Mn^{2+} which can be interpreted as industrial discharge to the river. 56% of parameters in f_{H34} were classified correctly by using DO and T as a seasonal factor. f_{H35} were classified correctly by using F-Strip and E-Coli which sources are domestic wastewater. f_{H36}, f_{H37} and f_{H38} were explained by T-Coli, B, NH_4–N respectively according to MRA.

3.2 Low Flow Period

PCA of the 26 parameters constituted six PCs explaining about 88.96% of the total variance at the 1KMP Station water quality data set for low flow period as seen in Table 4. f_{L11} has strong positive loadings on TH, EC, M-Al, SO_4^{2-}, Ca^{2+}, Na^+, TDS, o-PO_4^{3-}, NH_4–N, Mg^{2+} and strong negative loadings on DO; moderate positive loadings on NO_2–N and weak positive loadings on K. It was formed 38.7% of total variance. f_{L12} accounting for 20.28% of total variance has strong positive loadings on F-Strip, E-Coli, T-Coli, COD, SS, Fe^{2+} according to PCA results. Whereas, third, fourth, fifth and sixth factors accounted for the total variance of 9.68%; 7.99%; 7.13% and 5.16%, were correlated with NO_3–N and Q; B^{3+} and T; BOD_5 and Mn^{2+}; pH respectively.

When FA/PCA and MRA were investigated together, TS 99% and DO, K, NO_2–N, SO_4^{2-} could explain the f1. This factor represents soil structure which can be interpreted as a mineral component of the river. MRA was applied the data and 94% of parameters in f_{L12} were classified correctly by using SS and COD, F-Strip. We say that f_{L12} can be represents bacteriological and anthropogenic pollution such as municipal wastewater discharge and agricultural activities such as livestock wastes. NO_3–N and Q constituted f_{L13} which can be interpreted as agricultural activities such as fertilizer. B^{3+}, T explains the f_{L14} as soil erosion and seasonal factor; BOD_5 and Mn^{2+} explain the f_{L15} which describes the domestic wastewater factor; pH explains the f_{L16}.

PCA of the 28 parameters constituted five PCs explaining about 97.73% of the total variance at the 2BMP Station water quality data set for low flow period as seen in Table 4. f_{L21} explained 40.35% of total variance and was strong positively contributed by TH, M-Al, EC, SS, Mg^{2+}, Ca^{2+}, TDS, Na^+, NO_2–N, o-PO_4^{3-}, Cl^-, T and moderate positively contributed by NH_4–N. f_{L22} explained 17.29% of total variance and was strong positively contributed by COD, Mn^{2+}, F-Strip, Fe^{2+}. f_{L23} explained 10.99% of total variance and strong positively correlated with B, K and moderate positively correlated by NO_3–N. f_{L24} formed 9.78% of total variance and positive correlated with BOD_5, pV; negative correlated with $DO_,$. f_{L25} accounted for the total variance of 9.55% and was contributed by pH, SO_4^{2-}, Q, E-Coli. f_{L26} was represented by T-Coli with forming 9.13% of total variance.

FA/PCA and LRA were investigated together and Ca 99% and SS, T, NH_4–N, o-PO_4^{3-} could explain the f_{L21}. This factor represents soil structure which can be interpreted as a mineral component of the river. MRA was applied the data and 93% of parameters in f_{L22} were classified correctly by using COD and Fe. This factor could explain anthropogenic pollution such as industrial wastewater discharge. B^{3+}, K^+ and NO_3–N constituted f_{L23} which can be interpreted as agricultural activities such as irrigation water. BOD_5, pV and DO explain the f_{L24} as domestic wastewater; pH, SO_4^{2-}, Q and E-Coli explain the f_{L25} which describes the anthropogenic factor such as municipal wastewater discharge; T-Coli explains the f_{L26}.

PCA of the 28 parameters constituted seven PCs explaining about 72.0% of the total variance at the 3BMA Station water quality data set for low flow period as seen in Table 4. f_{L31} has strong positive loadings on TH, EC, M-Al, Ca^{2+}, Na^+, Cl^- and negative loadings on Q; moderate positive loadings on NO_3–N, Mg^{2+}, SO_4^{2-} and moderate negative loadings on Fe. It was formed 29.026% of total variance. f_{L32} accounting for 10.53% of total variance has strong positive loadings on $BOD_{5,}$

Table 4 Low flow period FA/PCA results

Low flow																					
1KMP Station varimax rotated results							2BMP Station varimax rotated results							3BMA Station varimax rotated results							
Variable	Factors						Variable	Factors						Variable	Factors						
	f_{L11}	f_{L12}	f_{L13}	f_{L14}	f_{L15}	f_{L16}		f_{L21}	f_{L22}	f_{L23}	f_{L24}	f_{L25}	f_{L26}		f_{L31}	f_{L32}	f_{L33}	f_{L34}	f_{L35}	f_{L36}	f_{L37}
TH	0.961	−0.136	0.135	0.010	−0.050	−0.067	TH	0.952	−0.236	0.056	0.026	0.165	−0.076	TH	0.945	0.006	−0.136	−0.00	0.071	−0.036	0.003
EC	0.961	−0.125	0.135	−0.051	0.049	−0.116	M-Al	0.948	−0.255	0.013	0.087	0.125	−0.112	EC	0.909	0.144	−0.111	0.155	0.171	0.017	0.176
M-Al	0.948	−0.165	0.230	−0.028	−0.101	−0.037	EC	0.933	−0.162	0-.089	0.131	0.023	−0.275	M-Al	0.906	0.059	−0.181	0.033	0.166	−0.073	0.130
SO_4	0.942	−0.063	−0.188	0.063	0.079	0.128	SS	−0.911	0.181	−0.245	−0.078	0.049	−0.251	Ca	0.889	0.087	−0.204	−0.037	0.004	−0.016	0.115
Ca	0.936	−0.094	0.203	−0.120	−0.132	−0.089	Mg	0.894	0.056	0.290	0.197	0.198	0.182	Na	0.814	0.152	−0.118	0.219	0.288	−0.188	0.165
Na	0.935	−0.100	0.105	0.236	0.033	−0.009	Ca	0.885	−0.377	−0.082	−0.072	0.130	−0.215	Q	−0.799	0.051	0.027	−0.053	0.117	0.180	−0.10
TDS	0.897	−0.102	−0.005	0.183	0.048	−0.290	TDS	0.871	0.090	0.018	0.187	−0.051	−0.441	Cl	0.798	0.205	0.012	0.001	0.057	−0.024	0.297
PO_4	0.850	−0.111	0.355	−0.009	−0.269	−0.164	Na	0.835	−0.141	−0.268	0.419	0.124	−0.122	NO_3–N	0.724	0.112	−0.043	0.260	0.191	0.025	0.266
DO	−0.833	−0.061	0.039	0.003	−0.348	0.221	NO_2–N	0.818	−0.182	0.407	0.169	0.132	−0.102	Mg	0.672	−0.155	0.081	0.075	0.203	−0.059	−0.30
NH_4–N	0.753	0.114	0.437	0.056	0.109	−0.052	PO_4	0.816	−0.239	−0.307	0.395	0.050	0.028	SO_4	0.647	0.016	−0.048	0.385	0.114	0.063	−0.21
Mg	0.738	−0.214	−0.102	0.378	0.200	0.016	Cl	0.815	−0.055	−0.368	0.360	−0.126	−0.155	Fe	−0.613	−0.009	0.431	0.075	0.265	−0.054	0.293
NO_2–N	0.624	−0.326	−0.186	0.323	0.222	−0.401	T	0.791	−0.158	0.109	0.463	−0.339	−0.028	BOD_5	−0.040	0.809	0.064	0.204	0.298	−0.115	0.099
K	0.489	0.266	0.367	0.470	0.170	0.088	NH_4–N	0.673	−0.131	−0.181	0.515	0.399	0.231	T	0.010	−0.782	−0.088	0.121	0.276	−0.094	0.161
F-Strip	−0.006	0.883	0.228	−0.154	0.143	−0.222	COD	−0.121	0.965	−0.199	0.107	0.013	0.052	NH_4–N	0.238	0.729	0.037	−0.036	0.069	0.053	0.144
E-Coli	−0.284	0.853	0.062	−0.049	−0.118	−0.039	Mn	−0.207	0.955	−0.139	0.124	0.033	−0.051	DO	−0.302	0.584	0.104	−0.461	−0.265	0.081	−0.198
T-Coli	−0.203	0.850	0.129	−0.193	−0.074	0.009	F-Strip	−0.257	0.946	−0.091	−0.023	−0.105	0.134	pV	0.139	0.532	0.024	0.440	0.120	−0.142	0.027
COD	−0.009	0.831	−0.258	−0.182	0.252	−0.023	Fe	−0.482	0.811	−0.141	−0.174	−0.210	0.043	F-Strip	−0.093	0.151	0.827	−0.052	0.093	−0.059	−0.150
SS	−0.178	0.785	−0.356	0.001	0.419	0.079	B	0.062	−0.370	0.897	−0.061	0.099	0.111	E-Coli	−0.171	0.039	0.809	0.003	−0.250	0.182	0.041
Fe	−0.117	0.783	−0.436	−0.058	0.300	−0.046	K	0.162	−0.234	0.848	0.028	−0.382	0.073	SS	−0.480	−0.029	0.550	0.188	−0.004	0.077	0.135
NO_3–N	0.183	−0.061	0.898	−0.204	0.022	0.091	NO_3–N	−0.277	0.082	0.687	−0.131	0.110	0.640	T-Coli	0.020	0.513	0.538	−0.056	0.032	−0.126	−0.055
Q	−0.394	0.218	−0.644	−0.231	−0.419	0.309	BOD_5	0.259	0.507	−0.041	0.802	−0.014	0.130	COD	0.049	−0.065	−0.010	0.754	−0.112	−0.121	−0.009
B	0.215	−0.254	−0.038	0.872	−0.158	−0.077	DO	−0.478	0.135	−0.088	−0.683	−0.048	0.109	TDS	0.454	0.078	0.097	0.548	0.197	0.142	−0.020
T	−0.169	−0.497	−0.171	0.727	0.011	0.107	pV	0.146	0.419	−0.456	0.597	0.281	−0.355	B	0.121	−0.068	−0.063	−0.166	0.796	−0.071	−0.059
BOD_5	−0.043	0.338	0.219	−0.093	0.798	−0.073	pH	−0.457	−0.422	−0.140	−0.084	0.758	−0.108	NO_2–N	0.253	0.114	0.020	0.346	0.605	0.014	0.026
Mn	0.470	0.495	−0.294	−0.043	0.529	0.076	SO_4	0.599	0.277	0.086	0.053	0.709	−0.211	Mn	−0.160	−0.097	0.135	0.055	−0.116	0.837	0.151
pH	−0.263	−0.236	0.000	0.039	−0.032	0.880	Q	0.481	−0.058	−0.191	0.319	0.691	0.377	pH	−0.028	0.058	−0.067	−0.161	0.034	0.797	−0.179
							E-Coli	−0.285	0.469	−0.033	−0.053	−0.615	0.544	PO_4	0.456	−0.103	−0.048	−0.078	−0.104	−0.066	0.668
							T-Coli	−0.126	0.088	0.197	0.019	−0.082	0.959	K	0.468	0.225	−0.055	0.233	0.428	0.020	0.494
Variance	38.70	20.28	9.68	7.99	7.12	5.16	Variance	40.35	17.29	10.99	9.78	9.55	9.13	Variance	29.026	10.5	8.262	6.84	6.817	5.633	4.891
%	38.70	58.99	68.68	76.67	83.80	88.96	%	40.35	58.27	69.26	79.04	88.59	97.73	%	29.026	39.5	47.82	54.6	61.47	67.11	72.00

Table 5 Multiple regression models

Station factor	1KMP	Station factor	2BMP	Station factor	3BMA
High flow period					
f_{H11}	$R^2 = 0.978$ Adj. $R^2 = 0.968$ Sign. $= 0.000$ $TH = -0.066 + 0.26EC + 0.041Q - 0.014TDS + 0.132Fe - 0.044E\text{-}Coli - 0.093COD + 0.259Ca + 0.612M\text{-}Al$	f_{H21}	$R^2 = 0.956$ Adj. $R^2 = 0.934$ Sign. $= 0.000$ $Ca = -0.1 + 0.069Na - 0.36\,SO_4 + 0.14COD + 0.21Cl + 1.56TH + 0.95EC - 0.59TDS - 0.809M\text{-}Al$	f_{H31}	$R^2 = 0.999$ Adj. $R^2 = 0.999$ Sign. $= 0.000$ $Ca = 0.096 - 0.008NO_2\text{–}N - 0.476Mn - 0.007SO_4 + 0.004o\text{-}PO_4 + 0.006K - 0.004Cl - 0.02EC + 0.001Na - 0.005TDS + 0.014M\text{-}Al + 1.272TH + 0.008\,NO_3\text{–}N$
f_{H12}	$R^2 = 0.66$ Adj. $R^2 = 0.56$ Sign. $= 0.000$ $o\text{-}PO_4 = 0.57 + 0.015B - 0.01SO_4 + 0.233Na + 0.63K + 0.107F\text{-}Strip$	f_{H22}	$R^2 = 0.867$ Adj. $R^2 = 0.832$ Sign. $= 0.000$ $T\text{-}Coli = -0.037 + 0.75E\text{-}Coli + 0.17SS + 0.058BOI_5 - 0.12Mn + 0.058Fe$	f_{H32}	$R^2 = 0.92$ Adj. $R^2 = 0.91$ Sign. $= 0.000$ $Turb = 0.342 + 0.027Q - 0.837pH + 0.346SS + 0.788Fe$
f_{H13}	$R^2 = 0.55$ Adj. $R^2 = 0.47$ Sign. $= 0.001$ $NO_2\text{–}N = 0.045 + 0.504NO_3\text{–}N + 0.506BOD_5 + 0.409T\text{-}Coli - 0.23NH_4\text{–}N$	f_{H23}	$R^2 = 0.867$ Adj. $R^2 = 0.832$ Sign. $= 0.000$ $o\text{-}PO_4 = 0.047 + 0.48B + 0.66K + 0.19pV - 0.25Fe$	f_{H33}	BOD_5, COD, Pv, Mn
f_{H14}	$R^2 = 0.611$ Adj. $R^2 = 0.596$ Sign. $= 0.000$ $DO = 0.784\text{–}0.79T$	f_{H24}	$R^2 = 0.736$ Adj. $R^2 = 0.70$ Sign. $= 0.000$ $DO = 0.704 + 0.186pH + 0.002NO_2\text{–}N - 0.782T$	f_{H34}	$R^2 = 0.562$ Adj. $R^2 = 0.554$ Sign. $= 0.000$ $DO = 0.776 - 0.665T$
f_{H15}	pH and SS	f_{H25}	NO_3–N, Mg, NH_4–N and Q	f_{H35}	F-Strip and E-Coli
f_{H16}	Mg			f_{H36}	T-Coli
f_{H17}	Mn			f_{H37}	B
				f_{H38}	NH_4–N
Low flow period					
f_{L11}	$R^2 = 0.99$ Adj. $R^2 = 0.98$ Sign. $= 0.000$ $TH = -0.049 + 0.002EC + 0.001M\text{-}Al - 0.003SO_4 + 0.825Ca - 0.003Na + 0.001TDS + 0.004o\text{-}PO_4 - 0.001DO - 0.004NH_4\text{–}N + 0.29Mg - 0.002NO_2\text{–}N - 0.00004K$	f_{L21}	$R^2 = 0.99$ Adj. $R^2 = 0.98$ Sign. $= 0.000$ $Ca = 0.094 + 0.003TDS + 0.002Na + 0.003\,NO_3\text{–}N - 0.001o\text{-}PO_4 + 0.003Cl - 0.001T - 0.001NH_4\text{–}N + 0.33TS + 0.015M\text{-}Al - 0.014EC + 0.001SS - 0.0514Mg$	f_{L31}	$R^2 = 0.975$ Adj. $R^2 = 0.971$ Sign. $= 0.000$ $M\text{-}Al = -0.21 + 0.154Ca + 0.348Na - 0.009Q - 0.165Cl - 0.104NO_3\text{–}N + 0.71Mg - 1.26SO_4 + 0.027Fe + 0.335TH + 0.388EC$
f_{L12}	$R^2 = 0.94$ Adj. $R^2 = 0.92$ Sign. $= 0.000$ $SS = 0.079 + 1.107Fe - 0.93F\text{-}Strip + 0.079E\text{-}Coli + 0.279T\text{-}Coli - 0.242COD$	f_{L22}	$R^2 = 0.929$ Adj. $R^2 = 0.916$ Sign. $= 0.000$ $COD = 0.005 + 0.57Mn + 0.474F\text{-}Strip - 0.059Fe$	f_{L32}	$R^2 = 0.546$ Adj. $R^2 = 0.518$ Sign. $= 0.000$ $BOD_5 = 0.001 - 0.35T + 0.55NH_4\text{–}N + 0.089DO + 0.6pV$
f_{L13}	NO_3–N and Q	f_{L23}	B, K, NO_3–N	f_{L33}	E-Coli, T-Coli, F-Strip and SS
f_{L14}	B, T	f_{L24}	BOD_5, DO, pV	f_{L34}	TDS and COD
f_{L15}	BOD_5 and Mn	f_{L25}	E-Coli, pH, SO_4 and Q	f_{L35}	B and NO_2–N
f_{L16}	pH	f_{L26}	T-Coli	f_{L36}	pH and Mn
f_{L17}				f_{L37}	K and o-PO_4

Table 6 Water quality parameters explained the factors

Station factor	1KMP	Station factor	2BMP	Station factor	3BMA
High flow period					
f_{H11}	98% TH and TDS, COD, E-Coli	f_{H21}	95.6% Ca and COD, Na, TDS	f_{H31}	99% Ca and SO_4, NO_2–N, o-PO_4
f_{H12}	66% o-PO_4 and F-Strip, SO_4	f_{H22}	86.7% T-Coli and SS	f_{H32}	92% Turb. and Q
f_{H13}	55% NO_2–N and NH_4–N	f_{H23}	55.5% o-PO_4 and Fe, pV	f_{H33}	BOD_5, COD, pV, Mn
f_{H14}	61% DO and T	f_{H24}	73.6% DO and pH	f_{H34}	56% DO and T
f_{H15}	pH and SS	f_{H25}	NO_3–N, NH_4–N, Mg and Q	f_{H35}	F-Strip and E-Coli
f_{H16}	Mg			f_{H36}	T-Coli
f_{H17}	Mn			f_{H37}	B
				f_{H38}	NH_4–N
Low flow period					
f_{L11}	99% TH and CO, K, NO_2–N, SO_4	f_{L21}	99% Ca and SS, T, NH_4–N, o-PO_4	f_{L31}	97% M-Al and Fe, Q
f_{L12}	94% SS and COD, F-Strip	f_{L22}	93% COD and Fe	f_{L32}	54% BOD_5 and T, CO
f_{L13}	NO_3–N and Q	f_{L23}	B, K, NO_3–N	f_{L33}	E-Coli, T-Coli, F-Strip and SS
f_{L14}	B, T	f_{L24}	BOD_5, DO, pV	f_{L34}	TDS and COD
f_{L15}	BOD_5 and Mn	f_{L25}	E-Coli, pH, SO_4 and Q	f_{L35}	B and NO_2–N
f_{L16}	pH	f_{L26}	T-Coli	f_{L36}	pH and Mn
				f_{L37}	K and o-PO_4

NH_4–N and strong negative loadings on T; moderate positive loadings on DO, pV according to PCA results. f_{L33} accounting for 8.26% of total variance has strong positive loadings on F-Strip, E-Coli; moderate positive loadings on SS, T-Coli. Whereas, fourth, fifth, sixth and seventh factors accounted for the total variance of 6.84%; 6.18%; 5.63% and 4.89% were correlated with COD and TDS; B^{3+} and NO_2–N; Mn^{2+} and pH; PO_4^{3-} and K respectively.

FA/PCA and MRA were investigated together and M-Al 97% and Fe^{2+}, Q could explain the f_{L31}. This factor represents soil structure which can be interpreted as a mineral component of the river. MRA was applied the data and 54% of parameters in f_{L32} were classified correctly by using BOD_5, DO and T. This factor could explain organic pollution such as municipal wastewater. F-Strip, E-Coli, SS and T-Coli constituted f_{L33} which can be interpreted as domestic wastewater and agricultural activities such as livestock wastes. COD and TDS could explain the f_{L34} as an anthropogenic factor such as industrial wastewater discharge; B^{3+} and NO_2–N explain the f_{L35} which describes the soil erosion and agricultural activities such as irrigation water; Mn^{2+} and pH explain the f_{L36} as industrial wastewater. K^+ and o-PO_4^{3-} explain the f_{L37} as domestic wastewater (Tables 5, 6, and 7).

Table 7 Pollution sources according to factors

Station factor	1KMP	Station factor	2BMP	Station factor	3BMA
High flow period					
f_{H11}	Soil structure	f_{H21}	Soil structure	f_{H31}	Soil structure-agricultural
f_{H12}	Agricultural-atmospheric	f_{H22}	Municipal–agricultural (livestock)	f_{H32}	Soil erosion
f_{H13}	Domestic-dump run-off	f_{H23}	Agricultural	f_{H33}	Industrial
f_{H14}	Seasonal	f_{H24}	Industrial	f_{H34}	Seasonal
f_{H15}	Industrial	f_{H25}	Agricultural-atmospheric	f_{H35}	Domestic
f_{H16}	Mg			f_{H36}	T-Coli
f_{H17}	Mn			f_{H37}	B
				f_{H38}	NH_4–N
Low flow period					
f_{L11}	Soil structure	f_{L21}	Soil structure-soil erosion	f_{L31}	Soil structure
f_{L12}	Municipal–agricultural (livestock)	f_{L22}	Industrial	f_{L32}	Municipal
f_{L13}	Agricultural (irrigation water)	f_{L23}	Agricultural (irrigation water)	f_{L33}	Domestic–agricultural (livestock)
f_{L14}	Soil erosion-seasonal	f_{L24}	Domestic	f_{L34}	Industrial
f_{L15}	Domestic	f_{L25}	Municipal	f_{L35}	Soil erosion-agricultural (irrigation water)
f_{L16}	pH	f_{L26}	T-Coli	f_{L36}	Industrial
				f_{L37}	Domestic

Table 8 Temporal variation discriminant function coefficients

	Period	
	High flow	Low flow
DO	14.098	9.993
EC	1.818	−15.808
F-Strip	4.172	0.362
K	4.197	10.421
Mal	6.983	13.308
Mn	8.002	3.587
NH_4–N	11.035	7.205
NO_3–N	0.123	3.135
pH	6.748	3.213
SO_4	−5.980	−0.118
TDS	6.405	14.786
(Constant)	−12.997	−10.422

Table 9 Temporal variation classification matrix for discriminant analysis

Stepwise DA mode			
% Correct		Period assigned by DA	
		High flow	Low flow
High flow	89.5	145	37
Low flow	74.7	17	109
Total	82.5	162	146

3.3 Discriminant Functions

Temporal variations in water quality were investigated through DA. Temporal DA was performed on standardized data by dividing the data set into two period groups (high–low flow period) at each stations. Discriminant functions obtained from stepwise mode of DA are shown in Table 8. Variables were added step by step in stepwise mode with the more significant until no significant changes and results were obtained. Stepwise mode discriminant functions using 11 discriminant variables presented the classification matrices separating 82.5% cases correctly as seen in Table 9. Stepwise DA shows that DO, EC, F-Strip, K^+, M-Al, Mn^{2+}, NH_4–N, NO_3N, pH, SO_4^{2-}, TDS are the most significant parameters to discriminate between the two different periods. These parameters are responsible for the temporal variations in water quality.

$$z_T = \beta_0 + \beta_1.DO + \beta_2.EC + \beta_3.F\text{-}Strip + \beta_4.K + \beta_5.M\text{-}Al + \beta_6.Mn + \beta_7.NH_4 - N + \beta_8.NO_3N + \beta.\mathrm{pH} + \beta_{10}.\mathrm{SO_4} + \beta_{11}.\mathrm{TDS} \quad (7)$$

Spatial DA was performed on standardized data by dividing the data set into five site groups (1KMP, 2BMP, 3BMA, 4AC, 5US). Discriminant functions obtained from stepwise mode of DA are shown in Table 8. Variables were added step by

Table 10 Spatial variation discriminant function coefficients

Stations					
	1KMP	2BMP	3BMA	4AC	5US
BOD_5	12.233	19.294	21.947	11.239	3.466
Ca	94.876	134.563	71.048	174.739	100.029
DO	90.623	73.136	110.210	90.354	91.382
EC	−17.169	−5.531	17.836	46.814	67.959
E-Coli	7.813	−2.860	7.384	−1.459	2.608
M-Al	37.030	−5.877	61.332	−34.262	23.901
Mg	45.316	65.452	26.595	83.391	46.275
Mn	26.075	14.135	24.839	17.258	21.134
Na	5.143	−2.516	8.682	14.858	5.428
NH_4–N	23.066	17.531	21.523	35.615	22.453
NO_3–N	23.853	27.381	7.040	24.608	18.037
pH	14.842	27.167	0.139	20.641	14.781
SO_4	−2.305	8.104	−19.662	−24.611	−7.051
T	60.526	52.130	69.626	60.003	57.131
TH	−125.494	−152.035	−126.528	−213.181	−186.061
(Constant)	−66.974	−60.726	−78.843	−75.190	−68.376

Table 11 Spatial variation classification matrix for discriminant analysis

Stepwise DA mode						
% Correct		Period assigned by DA				
		1KMP	2BMP	3BMA	4AC	5US
1KMP	91.7	44	2	2	1	3
2BMP	77.1	2	37	1	1	1
3BMA	97.0	0	0	128	0	0
4AC	83.3	1	9	1	40	0
5US	91.7	1	0	0	6	44
Total	90.4	48	48	132	48	48

step in stepwise mode with the more significant until no significant changes and results were obtained. Stepwise mode discriminant functions using 14 discriminant variables presented the classification matrices separating 90,4% cases correctly as seen in Table 9. Stepwise DA shows that BOD_5, Ca^{2+}, DO, E-Coli, M-Al, Mg^{2+}, Mn^{2+}, Na^{+}, NH_4–N, NO_3N, pH, SO_4^{2-}, T, TH are the most significant parameters to discriminate between the two different periods. These parameters are responsible for the spatial variations in water quality (Tables 10 and 11).

$$\begin{aligned} z_S = {} & \beta_0 + \beta_1.BO\dot{I}_5 + \beta_2.Ca + \beta_3.ÇO + \beta_4.EC + \beta_5.E\text{-}Coli + \beta_6.M - Al \\ & + \beta_7.Mg + \beta_8.Mn + \beta_9.Na + \beta_{10}.NH4 - N + \beta_{11}.NO_3N + \beta_{12}.pH \\ & + \beta_{13}.SO_4 + \beta_{14}.T + \beta_{15}.TS \end{aligned} \tag{8}$$

Classification function coefficients for each groups of temporal and spatial DA are given in Tables 8 and 10 respectively. Discriminant coefficients are converted to z scores to eliminate scaling differences among the discriminator variables. These functions can be used to calculate real estimate model (z scores) which may be used for classification of new samples. Among z scores of new sample measurement, the z score which is largest than zero will give us the group which its belongs to.

4 Conclusions

In this study, different multivariate statistical methods were used to investigate the Melen River complex data set obtained during 11 years. FA/PCA, MRA and DA applied the each data set of sample points and results were investigated. FA/PCA identified factors which are responsible for the data structure of each station at high and low flow periods. MRA identified the important and effective parameters according to relation between parameters in factors First factors for each station are explained with soil structure which comes from the dissolution of soil-rock by weathering. All stations are affected the agricultural and industrial pollution due to the intensive agricultural activities such as fertilizer, livestock waste and industrial areas near the river channel in basin. Especially at low flow period, impact of agricultural irrigation water is shown clearly. At rural areas and some cities in basin have no treatment plant for domestic and municipal wastewater. Septic tanks using for domestic wastewater and direct discharge of municipal wastewater are determined at all stations. Wild dump of Düzce city contribution is seen at 1KMP station. DA gave the best results both temporal and spatial analysis. It allowed a determining an indicator parameters responsible for large variations in water quality. DA showed that the most significant parameters for temporal variations are DO, EC, F-Strip, K^+, M-Al, Mn^{2+}, NH_4–N, NO_3–N, pH, SO_4^{2-}, TDS and for spatial variations are BOD_5, Ca^{2+}, DO, E-Coli, M-Al, Mg^{2+}, Mn^{2+}, Na^+, NH_4–N, NO_3–N, pH, SO_4^{2-}, T, TH. A few indicator parameters responsible for the spatial and temporal variations are found out by using DA. Melen River water quality assessment, pollution sources and parameters responsible for the spatial–temporal variations can be identified. And so this study provides the reduction in dimensionality of large data set. In this connection, less data set in monitoring program can be used for identification of water quality because of the measurement cost.

Acknowledgements This study was performed by using water quality data obtained from monitoring programs of Government Water Association in Ankara, Turkey.

References

Amiri BJ, Nakane K (2009) Modeling the linkage between riverwater quality and landscape metrics in the Chugoku District of Japan. Water Resour Manag 23:931–956

APHA-AWWA-WPCF (1999) In standard methods for the examination of water and wastewater, 20th edn. American Public Health Association, Washington

Boyacıoğlu H, Boyacıoğlu H (2006) Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environ Geol 54:275–282

Brumelis G, Lapina L, Nikodemus O, Tabors G (2000) Use of an artificial model of monitoring data to aid interpretation of principal component analysis. Environ Model Softw 15:755–763

Crowther J, Kay D, Wyer MD (2001) Relationships between microbial water quality and environmental conditions in coastal recreational waters: the Fylde Coast, UK. Water Res 35(17):4029–4038

Dixon W, Chiswell B (1996) Review of aquatic monitoring program design. Water Res 30:1935–1948

Dogan E, Sengorur B, Koklu R (2009) Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J Environ Manag 90:1229–1235

Düzce Governorship Environment and Forest Directorship (2007) Düzce Environment State Report, Düzce

Freund RJ, Wilson WJ (1998) Regression analysis—statistical modeling of a response variable. Academic Press

Kannel PR, Lee S, Kanel SR, Khan SP (2007) Chemometric application in classification and assessment of monitoring locations of an urban river system. Anal Chim Acta 582:390–399

Kottı ME, Vlessıdıs GA, Thanasoulıas NC, Evmırıdıs NP (2005) Assessment of river water quality in Northwestern Greece. Water Resour Manag 19:77–94

Kowalkowski T, Zbytniewski R, Szpejna J, Buszewski B (2006) Application of chemometrics in river water classification. Water Res 40:744–752

Liu CW, Lin KH, Kuo YM (2003) Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Sci Total Environ 313:77–89

Maillard P, Santos NAP (2008) A spatial–statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas River Watershed—Brazil. J Environ Manag 86:158–170

Mainstone CP, Parr W (2002) Phosphorus in rivers—ecology and management. Sci Total Environ 282–283:25–47

Mallin MA, Williams KE, Esham EC, Lowe RP (2000) Effect of human development on bacteriological water quality in coastal watersheds. Ecol Appl 10(4):1047–1056

Ouyang Y (2005) Evaluation of river water quality monitoring stations by principal component analysis. Water Res 39:2621–2635

Ouyang Y, Nkedi-Kizza P, Wu QT, Shinde D, Huang CH (2006) Assessment of seasonal variations in surface water quality. Water Res 40:3800–3810

Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D (2006) Application of factor and cluster analysis for characterization of river and estuarine water systems—a case study: Mahanadi River (India). J Hydrol 331:434–445

Pekey H, Karakaş D, Bakoğlu M (2004) Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses. Mar Pollut Bull 49:809–818

Sengörür B, İsa D (2001) Sakarya Nehri'ne Ait Su Kalite Gözlemlerinin Faktör Analizi. Turk J Eng Environ Sci, Tübitak 25:415–425

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River Basin, Japan. Environ Model Softw 22(4):464–475

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofonioub M, Kouimtzis TH (2003) Assessment of the surface water quality in Northern Greece. Water Res 37:4119–4124

Singh KP, Malik A, Sinha S (2005) Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques—a case study. Anal Chim Acta 538:355–374

Singh KP, Malik A, Mohan D, Sinha S (2004) Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. Water Res 38:3980–3992

Sliva L, Williams DD (2001) Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res 35(14):3462–3472

Tinsley EAH, Brown SD (2000) Handbook of applied multivariate statistics and mathematical modeling. Academic, San Diego

Vega M, Pardo R, Barrado E, Deban L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res 32(12):35

Wang X (2001) Integrating water-quality management and land-use planning in a watershed context. J Environ Manag 61:25–36