

Application of multivariate statistical analysis to incorporate physico-chemical surface water quality in low and high flow hydrology

Kazem Nosrati¹

Received: 12 March 2015 / Accepted: 17 July 2015 / Published online: 9 August 2015
© Springer International Publishing Switzerland 2015

Abstract Multivariate statistical methods, such as principal components analysis (PCA), discriminant analysis (DA) and general linear models (GLM) were applied to incorporate physico-chemical surface water quality in low and high flow hydrology in Northern Iran, based on analysis of the 7-day low flow index and existing water quality data. In view of this, 7-day low flows were calculated for 15 water years (1991–2006) at 15 monitoring stations. Eleven water quality parameters were extracted during the low flows from the water quality data and compared to water quality during high flows. Significant differences in water quality were noted for some monitoring stations and the pattern and magnitude of the statistically significant responses (t test, $p < 0.05$) varied among sites. PCA, was applied to the data sets of the two low and high flow periods, and resulted in three effective factors explaining 77.8 and 67.4 % of the total variance in surface water quality data sets of the two periods, respectively. The main factors obtained from PCA indicated that the parameters influencing surface water quality are mainly related to natural, point and non-point source pollution in the study area. DA provided an important data reduction as it used only three parameters, i.e. magnesium (Mg^{2+}), calcium (Ca^{2+}) and bicarbonate (HCO_3^-) affording 60 % correct assignments, to discriminate between the two low and high stream flow periods. General regression models revealed that surface water quality parameters were explained by low and high flow and specific discharge. The results of

this study can be useful for water managers for effective surface water quality management under climate change.

Keywords Low flow index · Surface water quality · General linear models · Iran

Introduction

Extreme events such as floods and droughts are complex natural hazards that affect some areas of the world every year and have significant impacts on water quantity (low and high flows). As well as impacting the quantity of water within rivers, low and high flows resulting from droughts and floods can affect water quality and aquatic biology through various physical, chemical and biological processes (Caruso 2002; Hrdinka et al. 2012) and can also aggravate water pollution and therefore, can impact human health and aquatic ecosystems through water quality deterioration.

The issue of the effects of extreme weather conditions, a possible result of climate change, on stream flow quantity (high and low flows) has been extensively investigated in recent years (e.g. Arnell 1999; Hanson and Weltzin 2000). Recently, investigations on the effects of climate change on water quality have also been carried out, mostly focusing on droughts (Elsdon et al. 2009; Van Vliet and Zwolsman 2008; Zwolsman and Van Bokhoven 2007). Mimikou et al. (2000) showed that water quality simulations under future climatic conditions entail significant water quality impairments because of decreased stream flows. Wilbers et al. (2009) demonstrated that the drought period of 2003 in the Dommel River, a tributary of the Meuse River in the Netherlands, did not significantly affect water quality.

✉ Kazem Nosrati
k_nosrati@sbu.ac.ir

¹ Department of Physical Geography, Faculty of Earth Sciences, Shahid Beheshti University, G.C., 1983969411 Tehran, Iran

In Iran the availability of water resources is critical during certain periods. River flows are strongly seasonal characterized by low and high natural flow during summer and winter, respectively. The high frequency of droughts in the area makes it necessary to improve management strategies for water quality and quantity during dry periods. Surface water is not only a major source of drinking water in Iran, but also supplies public water utilities and accounts for almost all of the water supply to rural households. Therefore, knowledge of low and high flow quality and quantity in streams is important for maintenance of the quantity and quality of water resources. Although low flow hydrology and climate change impacts on low flow are recognized in regional scales of Iran (e.g. Eslamian et al. 2010; Modarres 2008; Nosrati et al. 2004, 2015; Nosrati and Shahbazi 2008), little is known about the low flow impacts on surface water quality (Nosrati 2011). Thus, it is important to determine the climatic, and in particular drought impacts on surface water quality.

Different multivariate statistical techniques are widely applied to evaluate water quality through data reduction, classification and relationship (e.g. Machender et al. 2014; Nosrati and Van Den Eeckhaut 2012). Although multivariate statistical techniques such as principal components analysis, factor analysis, cluster analysis and discriminant

analysis are widely applied to evaluate surface and groundwater quality, to our knowledge, however, there have been limited attempts to establish surface water quality parameters to incorporate into general linear mixed models to assessing low and high flow hydrology and physico-chemical surface water quality. The aim of this study is to evaluate the effects of low and high flows on surface water quality using discriminant analysis and general linear mixed models in Sari-Neka Basin, Northern Iran.

Materials and methods

Study area

The present study investigates the effects of low and high flows on water quality at 15 stations in Sari-Neka Basin in Mazandaran Province in the north of Iran (Fig. 1; Table 1). The study area is located in karstic region in the eastern part of Mazandaran province. The study area ($35^{\circ} 56' - 36^{\circ} 52' N$ and $52^{\circ} 56' - 54^{\circ} 45' E$) is geographically divided into two parts: the coastal plains, and the mountainous areas. In plain, intensive large-scale agricultural activities is practiced using irrigation. The Alborz Mountains, south of the

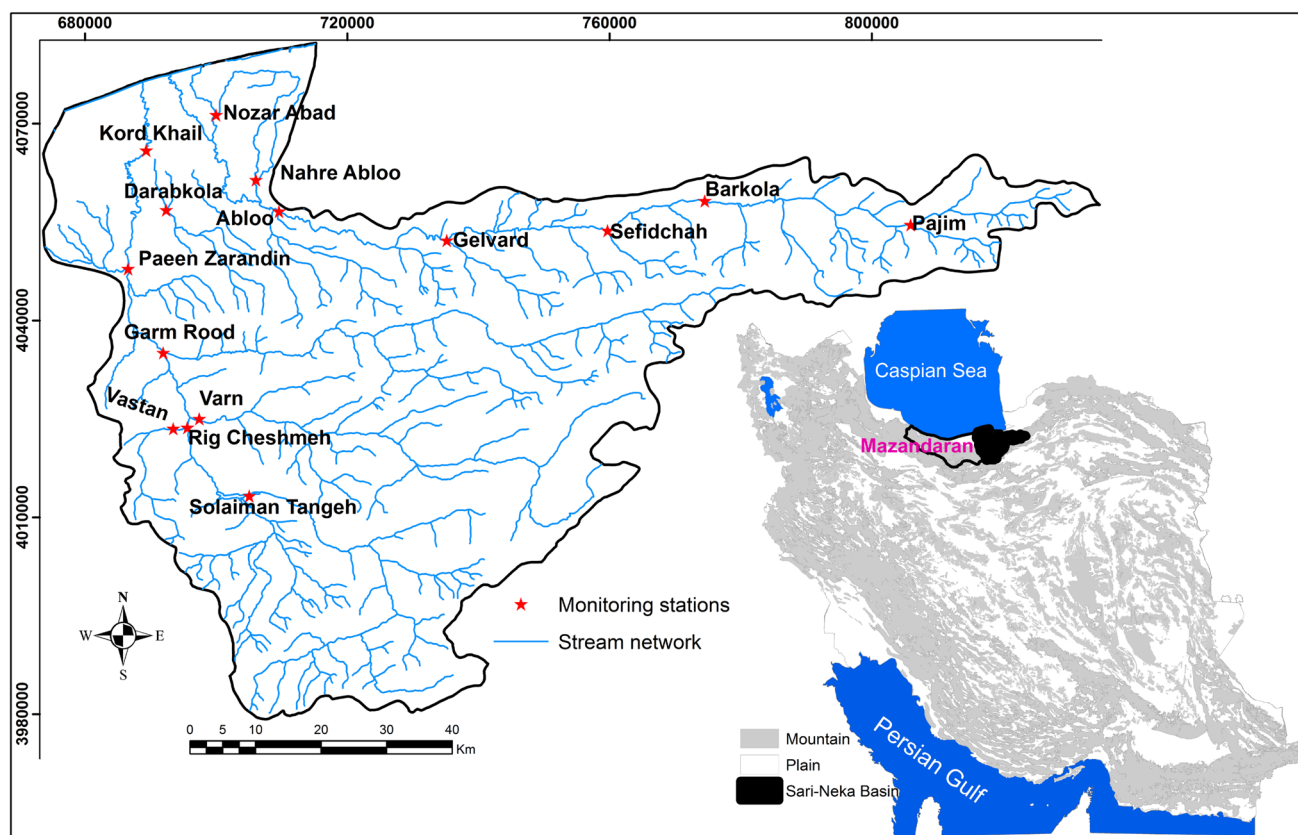


Fig. 1 Location map of the study area and monitoring stations

Table 1 The details of selected monitoring stations in Mazandaran Province

Station	River	Longitude	Latitude	Elevation (m)	Catchment area (km ²)
Sefidchah	Neka	53°54'	36°35'	1000	1043
Gelvard	Neka	53°36'	36°35'	600	1518
Abloo	Neka	53°19'	36°38'	70	1962
Darabkola	Neka	53°14'	36°33'	140	55
Rig Cheshmeh	Tajan	53°10'	36°21'	200	2715
Nahre Abloo	Neka	53°19'	36°39'	75	145
Nozar Abad	Neka	53°15'	36°49'	-10	1992
Kord Khail	Tajan	53°07'	36°43'	-10	4028
Garm Rood	Zalem Rood	53°10'	36°26'	200	894
Varn	Chahar Dangeh	53°12'	36°21'	230	1186
Vastan	Lajim Dareh	53°10'	36°20'	280	126
Solaiman Tangeh	Tajan	43°14'	36°15'	400	1277
Paen Zarandin	Neka	52°20'	36°37'	80	263
Pajim	Neka	54°44'	36°36'	900	1328
Barkola	Neka	54°04'	36°38'	1330	825

study area consists of year-round emerging springs and forested hills. The region climate is a semitropical climate with an average temperature of 25 °C in summer and 6 °C in winter. The mean annual rainfall in the region ranges from more than 1000 mm in the west to 300 mm in the east of the province. Around 33 % of all precipitation falls as snow in the mountainous area. Summer rainfall is also important in the region. The snowmelt period usually begins in mid-April and concludes in late April to early May. The population of the study area has been steadily growing during the last 50 years. 53.2 and 46.8 % of the population are living in urban and rural areas, respectively (SCI 2007).

Mazandaran Province is an important region for agricultural production, which directly depends on river water resources. Low flow frequency analysis will provide essential information regarding the risks of industrial development and water quality management during times of low flow, i.e. summer season, such as water pollution by pesticides and other industrial waste constituents. Such pollutants can also be harmful to fisheries downstream in Caspian Sea and agricultural activities, which are the main source of rice production in the country. All rivers in the region originate in the Alborz Mountains.

Data collection and treatment

Hydrological data series for 1991–2006 were obtained for 15 gauging stations in the region from the archives of Water Resources Researches Organization, Iran. Two data sets were considered in the analysis:

1. natural daily river discharges that had no data gaps for the 1991–2006 period;

2. eleven water quality parameters including sodium adsorption ratio (SAR), electrical conductivity (EC), total dissolved solids (TDS), pH, bicarbonate (HCO_3^-), chlorine (Cl^-), sulfate (SO_4^{2-}), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), and potassium (K^+).

Low flow has been characterized by the reduction in stream flow that may occur over 1 year or over several consecutive years (Smakhtin 2001) and can be assessed using low flow indices, i.e. lowest annual flow for a given duration (e.g. 7 days), particularly low flows that occur in the same season each year (Tallaksen et al. 1997). In order to evaluate surface water quality during low flow period, first, the annual 7-day minimum discharge series for each gauge were computed. Then, the 7-day average of each water quality parameters associated with the annual 7-day minimum discharge was calculated. The 7-day low flow periods were coincident with water quality data. Moreover the highest annual stream flow and associated water quality parameters for 15 water years (1991–2006) at each gauge were selected to compare to low flow periods.

Basic stream flow characteristics, including the specific discharge, q (mean daily discharge divided by basin area), and the discharge were determined based on low and high stream flow series (1991–2006). This allows determination of significant relationship between the water quality parameters and hydrological characteristics at the monitoring stations between low and high stream flow.

Statistical analysis

The normality and homogeneity of variance of the associated water quality parameters values were tested by two-

tailed Kolmogorov–Smirnov and Levene tests, respectively. These statistical analyses were followed by a *t* test for the identification of significant differences between low and high stream flow periods. Only those variables for which the *t*-test statistics for low and high stream flow categories were significant ($p < 0.05$) were retained for further analysis.

Principal components analysis (PCA) was used as the method of factor extraction for this study because it requires no prior estimates of the amount of variation in each surface water quality variable explained by the factors. PCA was performed on standardized variables to eliminate the effect of different measurement units on the determination of factor loading. Factor loadings are the simple correlations between the water quality variables and each factor. In our study, principal components (main factors) with eigenvalues >1 were selected and subsequently subjected to a varimax rotation to minimize the number of variables that have high loadings on each factor. In addition, communalities of every single variable for factor model were calculated to estimate the portion of variance in each variables explained by the rotated principal components. A high communality for a surface water quality variable indicates a high proportion of its variance is explained by the factors. In contrast, a low communality for a surface water quality variable indicates much of that attribute's variance remains unexplained. Less importance should be ascribed to surface water quality variables with low communalities when interpreting the factors (Nosrati et al. 2015).

Standard, forward and backward stepwise discriminant analysis (DA) was performed for retained surface water quality parameters to select water quality indicators that were most discriminating between the low and high stream flow periods. In standard mode, all variables enter simultaneously into the model. In forward stepwise mode, variables move into the model in successive steps; at each step the variable with the largest significant value will be chosen for inclusion in the model. The stepping will terminate when no other variable has a significant value, whereas, in backward stepwise mode, all variables are included into the model, and then are removed variables step by step; with the smallest significant value until no other variable in the model has a significant value. Thus, as the result of a successful discriminant analysis, one would only keep the important variables in the model, that is, those variables that contribute the most to the discrimination between groups. DA was performed on standardized variables to eliminate the effect of different measurement units on the determination of factor loading (Hill and Lewicki 2007).

Pairwise comparisons as discussed above do not allow to fully quantify and to understand the interaction between

the different independent variables. Therefore, the effects of low and high stream flow periods and river basin hydrological characteristics (including the specific discharge and the discharge) on water quality parameters were examined with mixed model analysis. Designs that contain random effects for one or more categorical predictor variables are called mixed-model designs. Random effects are classification effects where the levels of the effects are assumed to be randomly selected from an infinite population of possible levels. The solution for the normal equations in mixed-model designs is identical to the solution for fixed-effect designs. The variables used to build the statistical model consisted of both a categorical variable (low and high stream flow periods: dummy = 1 for low flow and dummy = 0 for high flow) and covariates (including specific discharge and the discharge) as fixed effects. The mixed analysis is able to account for sampling at the same observation point at different moments in time and also allows to identify a monitoring station. This effect accounts for drainage basin characteristics of monitoring stations that were not directly measured and is therefore considered as a random effect within the model. In order to identify the optimal variable to explain variations in water quality parameters, we used backward stepwise general regression model using a minimum significance level of 5 % for model entry. All variables (except categorical data) were subjected to natural logarithmic transformation in order to assure homoscedasticity and linearity between the dependent and the explanatory variables. Statistical analyses were carried out using STATISTICA V. 8.0 (StatSoft 2008).

Results and discussion

Low and high flow effect on water quality

According to the *t* test, water flow differs significantly ($p < 0.05$) between the hydrological low and high stream flow periods during 1991–2006 at the 15 selected monitoring stations (Table 2). Statistically significant differences ($p < 0.05$) were noted for physico-chemical parameters, except for pH and K^+ concentration. However, the comparison showed that the pattern and magnitude of the response varied among stations (Table 2). Differences for water quality parameters at Nahre Abloo and Pajim stations were not detected when comparing the low and high periods (Table 2). SAR, Na^+ and TDS were statistically significant in the most monitoring stations. Although differences were not always statistically significant, the general pattern was that physico-chemical concentrations were lower during the high flow period at most of the monitoring stations. The same results have also been

Table 2 Mean and significant values of surface water quality variables in two low and high stream flow periods, 1991–2006 (italicized indicated significant value)

Variables	Sefidchah			Gelvard			Nozar Abad		
	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value
Discharge (m ³ s ⁻¹)	2.6	0.1	<0.001	9.3	0.8	<0.001	5.2	0.2	<0.001
SAR	0.9	0.7	0.13	0.5	0.5	0.81	0.6	1.0	0.04
K ⁺ (mg L ⁻¹)	2.2	2.1	0.79	1.8	1.4	0.13	1.8	2.3	0.06
Na ⁺ (mg L ⁻¹)	31.3	26.5	0.25	14.9	15.2	0.91	21.4	43.2	0.03
Mg ²⁺ (mg L ⁻¹)	57.0	64.0	0.15	42.1	43.3	0.78	49.2	67.1	0.01
Ca ²⁺ (mg L ⁻¹)	107.5	111.5	0.66	92.5	80.7	0.04	97.6	136.0	0.01
SO ₄ ²⁻ (mg L ⁻¹)	139.6	118.3	0.34	62.3	53.6	0.47	65.6	128.3	0.02
Cl ⁻ (mg L ⁻¹)	35.3	26.9	0.04	17.9	14.9	0.16	18.5	54.4	0.01
HCO ₃ ⁻ (mg L ⁻¹)	236.3	272.3	0.02	213.7	208.0	0.70	252.6	311.1	0.04
pH	7.8	7.7	0.45	7.8	7.7	0.65	7.8	7.8	0.93
EC (μS cm ⁻¹)	640.2	646.1	0.88	464.9	437.9	0.33	539.2	802.8	0.01
TDS (mg L ⁻¹)	418.5	423.3	0.86	303.7	288.1	0.39	352.8	523.5	0.01
Variables	Abloo			Darabkola			Kord Khail		
	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value
Discharge (m ³ s ⁻¹)	21.1	0.5	<0.001	4.9	0.2	0.01	98.4	2.7	<0.001
SAR	0.4	0.7	0.04	0.4	0.8	0.08	0.7	1.9	<0.001
K ⁺ (mg L ⁻¹)	1.7	1.6	0.66	1.5	1.9	0.10	1.7	2.2	0.06
Na ⁺ (mg L ⁻¹)	14.4	22.3	0.04	12.1	30.6	0.08	24.9	80.0	<0.001
Mg ²⁺ (mg L ⁻¹)	39.5	47.6	0.08	39.9	53.9	0.01	22.7	33.7	<0.001
Ca ²⁺ (mg L ⁻¹)	86.1	79.4	0.32	83.1	109.5	0.02	51.0	67.1	<0.001
SO ₄ ²⁻ (mg L ⁻¹)	44.6	74.6	0.02	44.5	72.5	0.15	49.9	97.0	<0.001
Cl ⁻ (mg L ⁻¹)	17.3	20.2	0.30	12.8	36.6	0.16	27.2	97.3	<0.001
HCO ₃ ⁻ (mg L ⁻¹)	206.6	212.9	0.68	205.4	271.5	0.01	221.1	294.1	<0.001
pH	7.8	7.7	0.45	7.6	7.7	0.78	7.8	7.7	0.43
EC (μS cm ⁻¹)	436.5	483.7	0.20	801.0	639.1	0.70	544.1	968.8	<0.001
TDS (mg L ⁻¹)	286.5	316.8	0.21	268.7	417.3	0.01	354.4	628.6	<0.001
Variables	Rig Cheshmeh			Nahre Abloo			Garm Rood		
	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value
Discharge (m ³ s ⁻¹)	24.3	3.4	<0.001	1.4	0.3	<0.001	9.5	1.2	<0.001
SAR	0.7	1.4	<0.001	0.4	0.5	0.06	0.5	0.6	0.34
K ⁺ (mg L ⁻¹)	1.6	2.0	0.06	1.6	2.1	0.30	1.3	1.4	0.77
Na ⁺ (mg L ⁻¹)	25.5	53.8	<0.001	12.8	17.7	0.06	18.1	20.1	0.38
Mg ²⁺ (mg L ⁻¹)	48.3	62.4	0.06	41.6	42.8	0.74	23.3	26.1	0.09
Ca ²⁺ (mg L ⁻¹)	123.0	129.7	0.64	82.6	90.1	0.20	55.8	48.9	0.04
SO ₄ ²⁻ (mg L ⁻¹)	139.9	158.6	0.50	62.8	63.4	0.97	44.3	47.4	0.65
Cl ⁻ (mg L ⁻¹)	27.6	60.6	<0.001	14.6	17.5	0.21	18.6	21.3	0.38
HCO ₃ ⁻ (mg L ⁻¹)	235.1	286.7	0.17	195.4	218.2	0.08	243.0	236.6	0.59
pH	7.8	7.9	0.31	7.8	7.7	0.42	7.8	7.8	0.78
EC (μS cm ⁻¹)	630.0	828.2	0.03	425.3	467.2	0.17	547.1	541.9	0.81
TDS (mg L ⁻¹)	410.1	537.9	0.03	277.9	307.0	0.15	356.0	352.6	0.81

Table 2 continued

Variables	Varn			Vastan			Solaiman Tangeh		
	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value	High flow	Low flow	<i>p</i> value
Discharge (m ³ s ⁻¹)	6.0	1.2	<0.001	3.0	0.2	<0.001	12.8	3.8	<0.001
SAR	1.3	2.0	<0.001	0.4	0.6	0.04	0.5	0.6	<0.001
K ⁺ (mg L ⁻¹)	2.0	2.3	0.45	1.2	2.0	0.14	2.7	3.8	0.52
Na ⁺ (mg L ⁻¹)	48.0	73.7	<0.001	14.7	19.0	0.12	18.2	23.9	0.02
Mg ²⁺ (mg L ⁻¹)	25.5	30.6	0.01	20.9	22.6	0.48	26.4	29.2	0.13
Ca ²⁺ (mg L ⁻¹)	60.4	52.7	0.07	64.8	52.6	0.02	60.0	61.9	0.58
SO ₄ ²⁻ (mg L ⁻¹)	56.5	67.5	0.19	38.2	38.5	0.97	65.1	93.6	0.01
Cl ⁻ (mg L ⁻¹)	59.0	103.6	<0.001	18.8	18.1	0.87	24.9	29.1	0.09
HCO ₃ ⁻ (mg L ⁻¹)	264.6	246.5	0.13	257.5	242.6	0.35	234.9	224.6	0.36
pH	7.7	7.8	0.23	7.5	7.7	0.16	7.7	7.7	0.76
EC (μS cm ⁻¹)	720.5	841.0	<0.001	551.6	530.3	0.51	583.9	643.3	0.01
TDS (mg L ⁻¹)	471.6	547.8	<0.001	356.7	343.3	0.51	378.8	420.2	0.01
Variables	Paeen Zarandin			Pajim			Barkola		
	High flow	Low flow	<i>p</i> -value	High flow	Low flow	<i>p</i> -value	High flow	Low flow	<i>p</i> -value
Discharge (m ³ s ⁻¹)	4.7	0.2	<0.001	3.7	0.5	<0.001	1.1	0.3	<0.001
SAR	0.4	0.6	0.07	0.9	0.8	0.62	1.2	1.1	0.30
K ⁺ (mg L ⁻¹)	3.4	3.1	0.83	3.8	2.1	0.23	2.7	2.0	0.06
Na ⁺ (mg L ⁻¹)	15.0	20.0	0.09	33.3	31.2	0.62	49.0	46.2	0.54
Mg ²⁺ (mg L ⁻¹)	19.6	25.7	0.01	31.1	33.6	0.25	36.7	38.1	0.63
Ca ²⁺ (mg L ⁻¹)	52.4	45.4	0.08	63.1	57.1	0.31	59.6	62.9	0.40
SO ₄ ²⁻ (mg L ⁻¹)	28.2	37.9	0.18	76.4	72.3	0.67	125.8	115.8	0.49
Cl ⁻ (mg L ⁻¹)	17.9	17.1	0.83	46.5	33.3	0.29	52.7	45.5	0.18
HCO ₃ ⁻ (mg L ⁻¹)	233.7	241.9	0.54	254.2	276.7	0.07	249.2	284.1	0.02
pH	7.5	7.8	0.14	7.5	7.4	0.61	7.5	7.4	0.56
EC (μS cm ⁻¹)	461.9	485.4	0.41	668.4	634.0	0.45	773.3	762.0	0.79
TDS (mg L ⁻¹)	304.6	312.1	0.71	438.1	407.9	0.31	532.0	507.4	0.49

reported by previous studies. Zwolsman and Van Bokhoven (2007) and Van Vliet and Zwolsman (2008) demonstrated that water quality was negatively influenced by droughts, with respect to water temperature, eutrophication, major ions and heavy metals; they also indicated that the impact of droughts on water quality will be greater when the water quality is already poor. Prathumratana et al. (2008) proposed TSS, alkalinity and conductivity as sensitive water quality parameters for monitoring impacts of changing climate in the lower Mekong River. In their study, negative significant correlations were generally found between discharge flow and dissolved oxygen (DO), pH and conductivity (from 0.2 to 0.9). Worrall and Burt (2008) observed decreasing dissolved organic carbon (DOC) fluxes and concentrations in the areas that had experienced severe droughts in British rivers. Österholm

and Åström (2008) showed that the severity of individual summer droughts in the Pajuluoma acid sulphate area of Finland had little or no impact on the water quality during subsequent autumn and spring.

The EC indicates the amount of material dissolved in water. According to the WHO guidelines (WHO 1983), the maximum admissible EC concentration is 250 μS cm⁻¹ for drinking water. All monitoring stations in the study area had conductivity values exceeding this maximum permissible limit for potable water but for 66.6 % (n = 10) of the stations, the EC is higher during low flow periods compared to high flow periods. The average TDS is higher during low flow periods than during high flow periods. The recommended (most desirable) values of TDS for potable water is 500 mg L⁻¹ (WHO 1983). 33.5 % (n = 5) of the monitoring stations have TDS above 500 mg L⁻¹ during

low stream flow. Variation in TDS may be related to land use and pollution (Gaillardet et al. 1999; Nosrati and Van Den Eeckhaut 2012) and can be used to indicate the influence of human activities on water chemistry (Han and Liu 2004).

Determination of water quality factors

For the two low and high flow periods separately, PCA was performed on the normalized data sets to identify the factors replacing the most important variables. Factors with eigenvalues of 1.0 or greater are considered significant and factors with the highest eigenvalues are the most significant. The results of principal component analysis showed that the first three principal components (PCs) with eigenvalues >1 , accounted for $>77\%$ of variability in water quality in low flow period (Table 3). Communalities for water quality indicate these three factors explained $>90\%$ of variance in SAR, Na^+ , EC and TDS; $>80\%$ in Mg^{2+} , Ca^{2+} and Cl^- ; $>60\%$ in SO_4^{2-} , HCO_3^- and pH; $<35\%$ in K^+ (Table 3). A high communality estimate suggests that a high portion of variance was explained by the factor; therefore, it would get higher preference over a low communality estimate. Thus, K^+ was the least important attribute due to the lowest communality estimates in low flow period.

For the low flow period data set, PC1 explained the largest proportion (51.96 %) of total variance. PC1 had a strong positive loading (>0.75) on SAR, Na^+ , Cl^- , EC and TDS, and a moderate positive loading (0.5–0.750) on HCO_3^- (Table 3). Factor 1 represents the salinity of water which can be explained by natural and anthropogenic processes. The leaching of soil material, mixing of existing salts in soil, and high evaporation and evapotranspiration rates resulted in very high concentrations of ions that contribute to an increase of TDS and to a further deterioration of the water quality. Salinity is the total amount of inorganic solid material dissolved in any natural water, and water salinization refers to an increase in TDS and in the overall chemical content of the water. There are many natural sources such as atmospheric deposition, interactions between soil or rock and water, and salt water intrusion that can contribute to sodium and chloride concentrations. Chloride can be also enriched in natural waters due to the weathering of granites and magmatic rocks. HCO_3^- exhibits moderate positive loadings on both factors PC1 and PC2. This means that the variability of the HCO_3^- in the study area is affected by two distinct processes. In order to explanation the spatial distribution of factor PC1, factor score coefficients were calculated for variables. These coefficients represent the weights that are used when computing factor scores from the variables. HCO_3^- had a lower factor score coefficient (0.04) for PC1 compared to

PC2 (0.18). Thus, it can be concluded that HCO_3^- is more important parameter in PC2.

PC2 explained a significant proportion (15.96 %) of the total variance, had strong positive loading on Mg^{2+} and Ca^{2+} , and had moderate positive loading on SO_4^{2-} and HCO_3^- (Table 3). Factor 2 represents the natural hydro-geochemical evolution of water by groundwater-geological interaction which can be explained by the dissolution of rocks and minerals in sediments by chemical weathering. This factor explains the erosion from upland area during rainfall events. The dissolution of limestone and dolomite is possible source of Ca^{2+} and Mg^{2+} . The SO_4^{2-} sources in surface waters include: (1) atmospheric deposition (Wayland et al. 2003), (2) sulfate-bearing fertilizers and (3) bacterial oxidation of sulfur compounds (Sidle et al. 2000). Natural processes such as the dissolution of carbonate minerals and dissolution of atmospheric and soil CO_2 gas could be a mechanism supplying HCO_3^- to the groundwater, which recharge the low flow.

PC3, explaining 9.85 % of total variance, had a strong positive loading on pH and a moderate negative loading on K^+ (Table 3). The negative loading of K^+ on PC3 indicates that the source of this parameter can be related to anthropogenic pollution sources, the result of different pollution sources such as effluents of domestic origin, septic tanks, fertilizers and pesticides application in agriculture.

The results of principal component analysis showed that the first three principal components (PCs) with eigenvalues >1 , accounted for $>67\%$ of variability in water quality in high flow period. Communalities for water quality indicate these three factors explained $>70\%$ of variance for 7 variables with the exception being for K^+ , HCO_3^- , Cl^- , and EC (Table 3). Thus, those four variables were the least important attribute due to the lowest communality estimates. For high flow period, PC1 explained the largest proportion (40.1 %) of the total variance had a strong positive loading on SAR, Na^+ , Cl^- , and TDS, and a moderate positive loading on SO_4^{2-} (Table 3). Dissolution of gypsum and sodium sulphate minerals could increase SO_4^{2-} concentration in water. However there is no relationship between SO_4^{2-} and Ca^{2+} or Na^+ indicating that the excess of SO_4^{2-} in this period mostly result from the leaching of fertilizers, pesticides and increasing air pollution. Chlorine may be derived from pollution sources such as effluents of industrial and domestic origin, fertilizers and septic tanks, indicating anthropogenic pollution sources (Ritzi et al. 1993).

PC2 explained significant proportion (16.33 %) of the total variance, had strong positive loading on Mg^{2+} and Ca^{2+} , and had a moderate positive loading on SO_4^{2-} (Table 3). PC2 represents a hydrochemical processes that lead to high Mg^{2+} and Ca^{2+} concentrations. Associations between Mg^{2+} and Ca^{2+} suggest dissolution of calcite and

Table 3 Proportion of variance using varimax rotation and communality estimates of water quality parameters for low and high flows

Variables	PC1	PC2	PC3	Communality estimates
<i>Low flow (three significant principal components)</i>				
SAR	0.95	-0.01	0.03	0.90
K ⁺ (mg L ⁻¹)	0.07	-0.04	<i>-0.56</i>	0.32
Na ⁺ (mg L ⁻¹)	0.97	0.13	-0.01	0.96
Mg ²⁺ (mg L ⁻¹)	0.13	0.91	-0.07	0.85
Ca ²⁺ (mg L ⁻¹)	0.14	0.91	0.10	0.85
SO ₄ ²⁻ (mg L ⁻¹)	0.43	<i>0.64</i>	0.07	0.61
Cl ⁻ (mg L ⁻¹)	0.93	0.06	0.01	0.86
HCO ₃ ⁻ (mg L ⁻¹)	<i>0.53</i>	<i>0.55</i>	-0.29	0.67
pH	0.05	0.01	0.82	0.68
EC (μS cm ⁻¹)	0.88	0.41	-0.02	0.94
TDS (mg L ⁻¹)	0.87	0.41	-0.03	0.93
Eigenvalue	5.72	1.76	1.08	
% Total variance	51.96	15.96	9.85	
Cumulative % variance	51.96	67.92	77.77	
<i>High flow (three significant principal components)</i>				
SAR	0.94	-0.02	0.01	0.88
K ⁺ (mg L ⁻¹)	0.15	-0.05	<i>0.64</i>	0.43
Na ⁺ (mg L ⁻¹)	0.97	0.02	0.06	0.94
Mg ²⁺ (mg L ⁻¹)	0.14	0.87	-0.01	0.78
Ca ²⁺ (mg L ⁻¹)	-0.07	0.89	0.03	0.79
SO ₄ ²⁻ (mg L ⁻¹)	<i>0.59</i>	<i>0.59</i>	-0.08	0.71
Cl ⁻ (mg L ⁻¹)	0.76	0.02	0.22	0.62
HCO ₃ ⁻ (mg L ⁻¹)	0.46	0.26	0.35	0.40
pH	-0.08	-0.01	-0.84	0.71
EC (μS cm ⁻¹)	0.47	0.21	-0.16	0.29
TDS (mg L ⁻¹)	0.89	0.22	0.12	0.85
Eigenvalue	4.41	1.80	1.21	
% Total variance	40.10	16.33	10.96	
Cumulative % variance	40.10	56.43	67.39	

Bold and italic values indicate strong (>0.75) and moderate (0.5–0.750) loadings, respectively

dolomite affected by erosion and deposition from upland area. SO₄²⁻ exhibits moderate positive loadings on both PC1 and PC2. The factor score coefficients of SO₄²⁻ in factors PC1 and PC2 are 0.11 and 0.24 respectively. Thus, it can be concluded that SO₄²⁻ in PC2 is more important parameter.

PC3, explaining 10.96 % of the total variance, had a strong negative loading on pH, had a moderate positive loading on K⁺ (Table 3); and represents a hydrochemical processes that lead to high K⁺ concentrations. The negative correlation with pH indicates that introduction of H⁺ into the water is not due to natural dissolution from soil or rock and the source of this parameter can be related to atmospheric pollution.

Overall, the intensive agriculture practiced in the study area affects all parameters included in the analysis in two low and high flow periods. Irrigation with local surface and

groundwater induces a groundwater cycle that increases the salinity in the upper aquifer through irrigation return flow, thereby increasing the concentrations of all ions present in solution. Application of PCA in our study area shows that the dominant factors are explained by the following processes: soil–groundwater interactions, and agricultural, industrial and atmospheric pollution.

Identification of water quality indicators

Surface water measurements that were not significantly affected by low and high stream flow types within the study area (pH and K⁺) were excluded from further consideration as possible candidates to identify surface water quality indicators. DA was done with the two low and high stream flow periods as grouping variable and the 9 retained water quality parameters as independent variables to remove

Table 4 Chi-square tests with successive roots removed for discriminant analysis applied to the two low and high streamflow periods of monitoring stations

Roots removed	Eigenvalue	Canonical-R	Wilks' Lambda	Chi-square	df	p value
Standard mode	0.11	0.31	0.90	46.05	11	<0.0001
Forward mode	0.10	0.31	0.91	43.8	7	<0.0001
Backward mode	0.08	0.27	0.93	34.7	3	<0.0001

Table 5 Classification functions resulting from discriminant analysis applied to the two low and high stream flow periods of monitoring stations

Variables	Standard mode F = 4.3 p < 0.0001			Forward mode F = 6.5 p < 0.0001			Backward mode F = 12.5 p < 0.0001		
	Low flow	High flow	P value	Low flow	High flow	p value	Low flow	High flow	p value
SAR	-0.37	0.38	0.327						
Na ⁺ (mg L ⁻¹)	0.27	-0.28	0.564	-0.20	0.21	0.007			
Mg ²⁺ (mg L ⁻¹)	-0.32	0.33	<0.001	-0.35	0.35	<0.001	-0.30	0.30	<0.001
Ca ²⁺ (mg L ⁻¹)	0.24	-0.25	0.006	0.21	-0.22	0.007	0.25	-0.25	<0.001
SO ₄ ²⁻ (mg L ⁻¹)	0.01	-0.01	0.966	0.12	-0.13	0.123			
Cl ⁻ (mg L ⁻¹)	-0.16	0.16	0.254						
HCO ₃ ⁻ (mg L ⁻¹)	-0.30	0.31	0.038	-0.12	0.13	0.081	-0.19	0.20	<0.001
EC (μS cm ⁻¹)	0.07	-0.07	0.242	0.07	-0.07	0.225			
TDS (mg L ⁻¹)	0.18	-0.19	0.381						
Constant	-0.73	-0.76		-0.73	-0.76		-0.72	-0.75	

redundant variables. A significant result was obtained, independent of the discriminant function removal method used (Table 4). Thus, the set of water quality parameters used clearly allow discriminating between the two low and high stream flow periods.

The Mahalanobis distances between two categories are significantly different. However, F values of the backward mode (F = 12.0, p < 0.0001) are significantly higher as compared to standard (F = 4.4, p < 0.0001) and forward mode (F = 6.5, p < 0.0001).

The standard and forward stepwise discriminant analysis mode yielded classification matrices assigning ca. 63 % of the cases correctly including 9 and 6 parameters respectively, (Tables 5, 6). However, in the backward stepwise mode, 60 % of the low and high stream flow periods were correctly classified by a model using only three discriminant parameters, Mg²⁺, Ca²⁺ and HCO₃⁻ (Tables 5, 6). Thus, the DA results suggest that Mg²⁺, Ca²⁺ and HCO₃⁻ are the most significant parameters to discriminate between the two low and high stream flow periods.

Water quality in relation to hydrological characteristics

The resulting variance components and mixed model contains a fixed part and a random part. In the models, the random part contains a site-effect, i.e. I expect that part of the variance that cannot be explained by the independent

Table 6 Classification matrix resulting from discriminant analysis applied to the two low and high stream flow periods of monitoring stations

Classification	% Correct	Samples assigned by DA	
		High flow	Low flow
<i>Standard mode</i>			
High flow	66.7	160	80
Low flow	60.4	95	145
Total	63.6	255	225
<i>Forward mode</i>			
High flow	68.0	164	76
Low flow	58.1	100	140
Total	63.1	264	216
<i>Backward mode</i>			
High flow	64.5	155	85
Low flow	54.5	109	131
Total	59.6	264	216

variables and their interactions is due to the fact that differences between drainage basin characteristics of monitoring stations are not entirely accounted for by the properties that were included in the statistical analysis. Therefore, it would be possible that changes in land and resource use will have a comparable or greater effect on water quality than changes in hydrological characteristics.

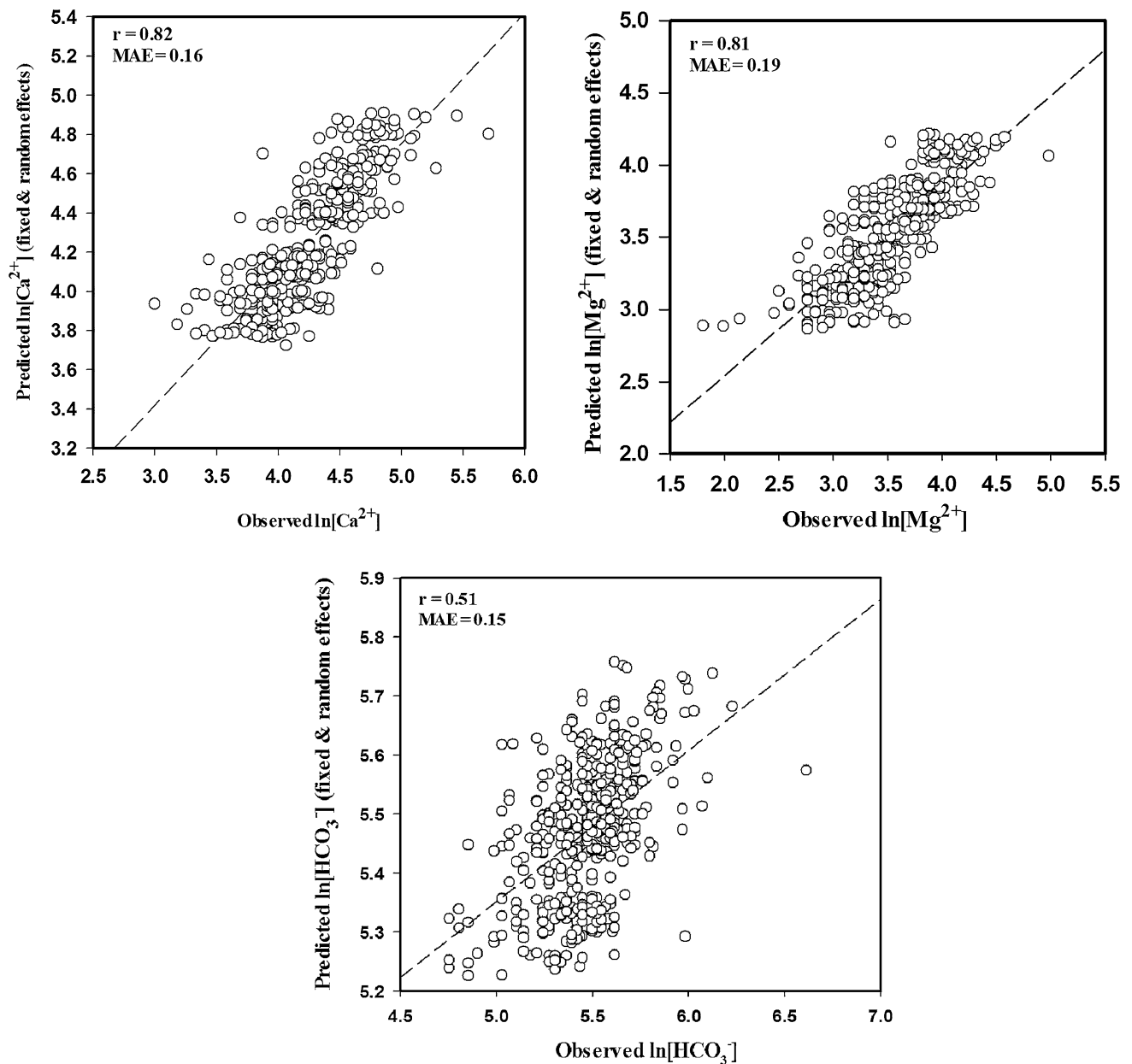


Fig. 2 Scatterplot of observed versus predicted data points (fixed + random effects) for Ca^{2+} , HCO_3^- and Mg^{2+} . *MAE* mean absolute error; *r* Pearson correlation coefficient ($p < 0.05$)

The intraclass correlation coefficient for stations effects as random effect on water quality parameters was computed as the ratio of the estimated variance component for the station to the total error variance, indicating that 21, 18 and 24 % of the non-explained variation in Ca^{2+} , HCO_3^- and Mg^{2+} is accounted for by the station effect, respectively. Root mean squared error (RMSE) of the model for Ca^{2+} , HCO_3^- and Mg^{2+} were 18.9, 52 and 10.2 mg L^{-1} , respectively. The observed and predicted values for the natural logarithm of Ca^{2+} , HCO_3^- and Mg^{2+} were plotted in Fig. 2.

Table 7 gives information about the significance of effects and models performance for the prediction of the natural logarithm of water quality when no random effect (station effect) is considered. Backward stepwise general regression analyses for all water quality data showed that 7-day low flow ($D_{\text{streamflow}} = 1$ for low flow and $D_{\text{streamflow}} = 0$ for high flow) and specific discharge (q) entered as significant parameters (Table 7).

The predicted values for $\ln[\text{Ca}^{2+}]$, $\ln[\text{HCO}_3^-]$ and $\ln[\text{Mg}^{2+}]$ are plotted against the observed values for $\ln[\text{Ca}^{2+}]$, $\ln[\text{HCO}_3^-]$ and $\ln[\text{Mg}^{2+}]$ in Fig. 3. RMSE of

Table 7 Backward stepwise general regression analysis for water quality parameters with hydrological characteristics (n = 480)

Regression parameters					Regression statistics			
Dependent variable	Predictor	B	t value	p value	R ²	Adjusted R ²	F	p value
ln[Ca ²⁺]	Intercept	4.17	143.2	<0.001	0.033	0.028	7.5	0.001
	ln[q]	-0.05	-3.8	<0.001				
	D _{streamflow}	0.06	2.3	0.022				
ln[HCO ₃ ⁻]	Intercept	5.41	385.7	<0.0001	0.095	0.093	46.8	<0.0001
	ln[q]	-0.04	-6.8	<0.0001				
ln[Mg ²⁺]	Intercept	3.40	131.6	<0.0001	0.114	0.112	57.8	<0.0001
	ln[q]	-0.08	-7.6	<0.0001				

D_{streamflow}, dummy = 1 for low flow and dummy = 0 for high flow; q specific discharge, B raw regression coefficient

the model for Ca²⁺, HCO₃⁻ and Mg²⁺ were 31.1, 57.1 and 15.7 mg L⁻¹, respectively. These regression models identified drought and specific discharge as the main factors. The t statistics shows that the specific discharge is the most important variable in the models (Table 7). Therefore, water quality parameters are perhaps influenced by factors controlling the seasonal supply of physico-chemical water quality. Dakova et al. (2000) found the best statistically significant correlation between hydrobiological indexes and discharge during the low flow period. Elsdon et al. (2009) detected minimal differences in water quality between land uses during a period of extensive drought.

Global climate change projections indicate changes in rainfall, causing increased frequency and severity of low flow in some regions (Sheffield and Wood 2008). Low flow conditions are determined by a suite of natural and anthropogenic factors and are an integral part of every river regime (Smakhtin 2001). The basis for estimating low flows is, therefore, of crucial importance for protection of water quality. Decreases in water levels due to drought can affect catchment functioning (including partitioning, storage, and release of water), throughout the following year or even for several years if the drought occurs in a larger area. A 1-year drought not only causes water level decreases, but also results in many other changes. When water levels decrease, solutes become more concentrated as the amounts of water decrease in rivers. This pattern is consistent with the results of the present study, as shown by the increase of Mg²⁺, Ca²⁺, HCO₃⁻, SO₄²⁻, Cl⁻, EC, TDS and Na⁺ concentrations during low flow (Table 2). The concentration increase is hypothesized to be associated with evaporation from rivers and the ground surface, as well as the increase of residence and contact of waters with soils during recharge and during discharge of groundwater into rivers (Caruso 2002; Murdoch et al. 2000). These results demonstrate that water quality degrades under low flow conditions,

and in the context of a climate change increase in drought conditions, leads to an increase of at risk situations related to potential health impacts (Delpia et al. 2009).

Conclusions

Differences in water quality were detected between low and high stream flow periods but the differences are not the same for each constituent. Climate change resulting in more intense and frequent droughts could cause considerably lower stream flows and consequently have effects on surface water quality, mainly increasing constituent concentrations. Also the chemical characteristics of groundwater could have a principal role in influencing low flow chemistry. The results of the mixed models containing a site-effect showed that part of the variance that cannot be explained by the independent variables and their interactions is due to the fact that differences between drainage basin characteristics of monitoring stations are not entirely accounted for by the properties that were included in the statistical analysis. It means that the significant differences in water quality can be explained by anthropogenic influences such as land-use changes. The intensive agriculture practiced in the study area affects all parameters included in the analysis. Irrigation with local surface and groundwater increases the concentrations of all ions present in solution. Urban, municipal and industrial land use types also can affect the quality of surface water. Consequently, the quality of surface water is affected by many factors including precipitation and characteristics of the catchment area. This study confirms that multivariate statistical techniques can be adopted for analysis and interpretation of complex data sets of water parameters in surface water quality assessment and in the identification of important factors. Further research should investigate the response of water quality

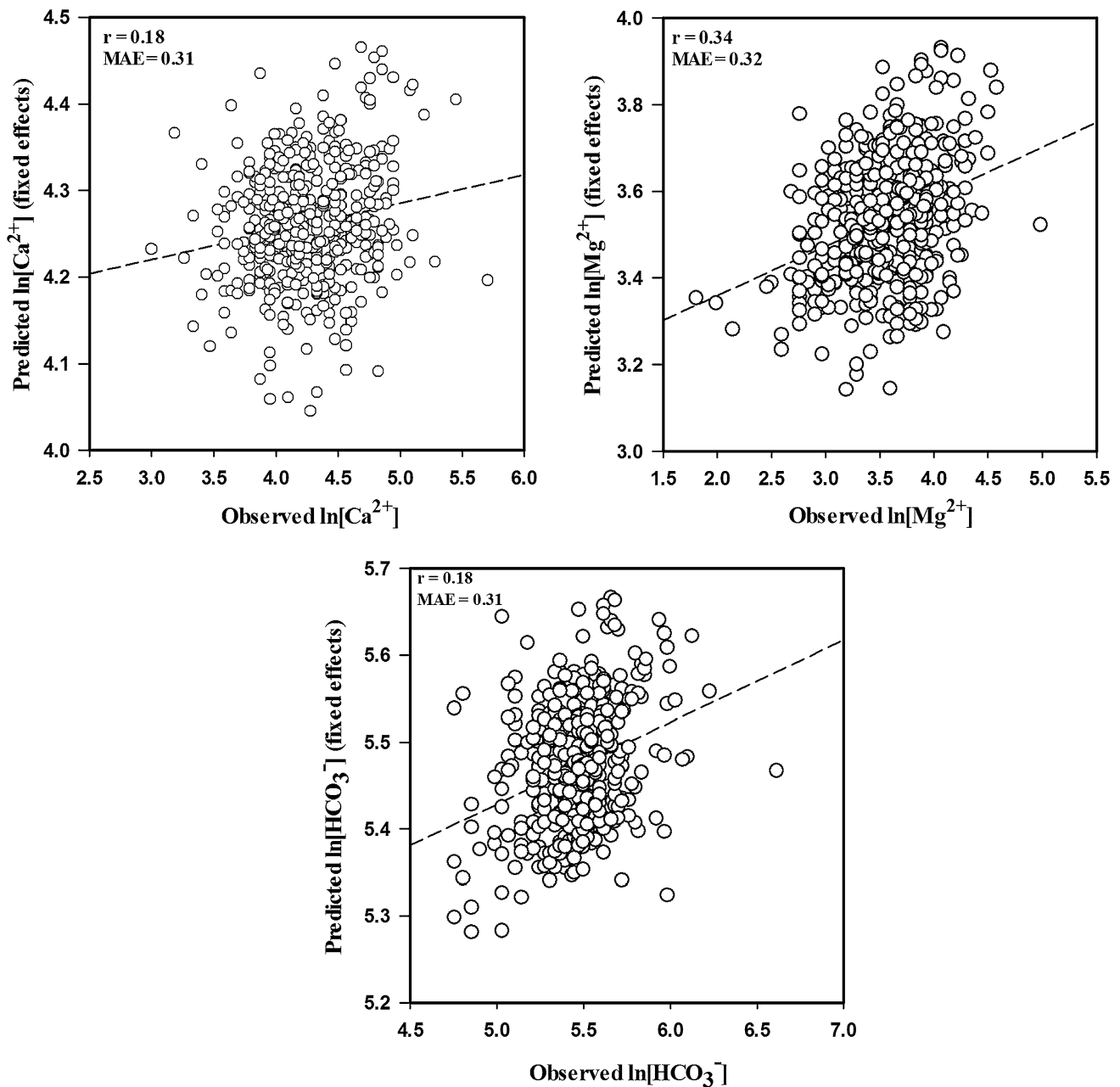


Fig. 3 Scatterplot of observed versus predicted data points (fixed effects only) for Ca^{2+} , HCO_3^- and Mg^{2+} . *MAE* mean absolute error, *r* Pearson correlation coefficient ($p < 0.05$)

parameters to drought conditions under land use changes and future climate change scenarios.

Acknowledgments This project was funded by a grant from the research council of Shahid Beheshti University, Tehran, Iran (Grant number 600.3683).

References

- Arnell NW (1999) Climate change and global water resources. *Global Environ Change* 9:S31–S49
- Caruso B (2002) Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago. *N Z J Hydrol* 257:115–133
- Dakova S, Uzunov Y, Mandadjiev D (2000) Low flow—the river’s ecosystem limiting factor. *Ecol Eng* 16:167–174
- Delpa I, Jung AV, Baures E, Clement M, Thomas O (2009) Impacts of climate change on surface water quality in relation to drinking water production. *Environ Int* 35:1225–1233
- Elsdon TS, De Bruin MBNA, Diepen NJ, Gillanders BM (2009) Extensive drought negates human influence on nutrients and water quality in estuaries. *Sci Total Environ* 407:3033–3043
- Eslamian S, Ghasemzadeh M, Biabanaki M, Talebizadeh M (2010) A principal component regression method for estimating low flow

- index. *Water Resour Manage* 24:2553–2566. doi:[10.1007/s11269-009-9567-2](https://doi.org/10.1007/s11269-009-9567-2)
- Gaillardet J, Dupré B, Louvat P, Allegre C (1999) Global silicate weathering and CO₂ consumption rates deduced from the chemistry of large rivers. *Chem Geol* 159:3–30
- Han G, Liu CQ (2004) Water geochemistry controlled by carbonate dissolution: a study of the river waters draining karst-dominated terrain Guizhou Province, China. *Chem Geol* 204:1–21
- Hanson PJ, Weltzin JF (2000) Drought disturbance from climate change: response of United States forests. *Sci Total Environ* 262:205–220
- Hill T, Lewicki P (2007) STATISTICS: methods and applications. StatSoft, Tulsa
- Hrdinka T, Novický O, Hanslík E, Rieder M (2012) Possible impacts of floods and droughts on water quality. *J Hydro-Environ Res* 6:145–150. doi:[10.1016/j.jher.2012.01.008](https://doi.org/10.1016/j.jher.2012.01.008)
- Machender G, Dhakate R, Reddy MN, Reddy IP (2014) Hydrogeochemical characteristics of surface water (SW) and groundwater (GW) of the Chinnaeru River basin, northern part of Nalgonda District, Andhra Pradesh. *India Environ Earth Sci* 71:2885–2910
- Mimikou M, Baltas E, Varanou E, Pantazis K (2000) Regional impacts of climate change on water resources quantity and quality indicators. *J Hydrol* 234:95–109
- Modarres R (2008) Regional frequency distribution type of low flow in North of Iran by L-moments. *Water Resour Manage* 22:823–841
- Murdoch PS, Baron JS, Miller TL (2000) Potential effects of climate change on surface water quality in North America. *J Am Water Resour Assoc* 36:347–366
- Nosrati K (2011) The effects of hydrological drought on water quality. In: Peters NE, Krysanova V, Lepistö A, Prasad R, Thoms M, Wilby R, Zandaryaa S (eds) *Assessment of water quality under changing climate conditions*. IAHS Publ, Wallingford, pp 51–57
- Nosrati K, Shahbazi A (2008) Low flow estimation using hybrid method in Northeastern of Iran. *Iran J Nat Res* 60:829–841
- Nosrati K, Van Den Eeckhaut M (2012) Assessment of groundwater quality using multivariate statistical techniques in Hashtgerd Plain. *Iran Environ Earth Sci* 65:331–344
- Nosrati K, Eslamian S, Shahbazi A (2004) Investigation of effect of climate change on hydrological drought. *J Agric* 6:49–56
- Nosrati K, Laaha G, Sharifnia A, Rahimi M (2015) Regional low flow analysis in Sefidrood Drainage Basin, Iran using principal component regression. *Hydrol Res* 46:121–135. doi:[10.2166/nh.2014.087](https://doi.org/10.2166/nh.2014.087)
- Österholm P, Åström M (2008) Meteorological impacts on the water quality in the Pajuluoma acid sulphate area, W. Finl Appl Geochem 23:1594–1606
- Prathumratana L, Sthiannopkao S, Kim KW (2008) The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River. *Environ Int* 34:860–866
- Ritzi RW, Wright SL, Mann B, Chen M (1993) Analysis of temporal variability in hydrogeochemical data used for multivariate analyses. *Groundwater* 31:221–229
- SCI (2007) Statistical Center of Iran, Iran statistical year book (2006–2007). www.amar.org.ir
- Sheffield J, Wood EF (2008) Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Clim Dyn* 31:79–105
- Sidle W, Roose D, Shanklin D (2000) Isotopic evidence for naturally occurring sulfate pollution of ponds in the Kankakee River Basin, Illinois-Indiana. *J Environ Qual* 29:1594–1603
- Smakhtin V (2001) Low flow hydrology: a review. *J Hydrol* 240:147–186
- StatSoft (2008) STATISTICA: [data analysis software system], Version 8.0 for Windows update. StatSoft, Inc, Tulsa
- Tallaksen LM, Madsen H, Clausen B (1997) On the definition and modelling of streamflow drought duration and deficit volume. *Hydrolog Sci J* 42:15–33
- Van Vliet M, Zwolsman J (2008) Impact of summer droughts on the water quality of the Meuse river. *J Hydrol* 353:1–17
- Wayland KG, Long DT, Hyndman DW, Pijanowski BC, Woodhams SM, Haack SK (2003) Identifying relationships between base-flow geochemistry and land use with synoptic sampling and R-mode factor analysis. *J Environ Qual* 32:180–190
- WHO (1983) Guidelines to drinking water quality. WHO, Geneva
- Wilbers GJ, Zwolsman G, Klaver G, Hendriks AJ (2009) Effects of a drought period on physico-chemical surface water quality in a regional catchment area. *J Environ Monit* 11:1298–1302
- Worrall F, Burt T (2008) The effect of severe drought on the dissolved organic carbon (DOC) concentration and flux from British rivers. *J Hydrol* 361:262–274
- Zwolsman J, Van Bokhoven A (2007) Impact of summer droughts on water quality of the Rhine River: a preview of climate change? *Water Sci Technol* 56:45–55