The Relationship between Surface Water Quality and Watershed Characteristics

Amir Mansour Mohammadi[1], Mehdi Vafakhah[2*], Mohammad Reza Javadi[1]

1. Faculty of Natural Resources, Noor Branch, Islamic Azad University, Noor, Iran
2., Department of Watershed Management, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran
Email: vafakhah@modares.ac.ir (Corresponding Author)

Abstract: The healthy water resources are necessary and essential prerequisite for environmental protection and economic development, political, social and cultural rights of Iran. In this research, water quality parameters i.e. total dissolved solids (TDS), sodium absorption rate (SAR), electrical conductivity (EC), Na^+, Cl^-, CO_3^{2-}, K^+, Mg^{2+}, Ca^{2+}, pH, HCO_3^- and SO_4^{2-} during 2010-2011 were obtained from Iranian Water Resources Research Institute in water quality measurement stations on Mazandaran province, Iran. Then, the most important catchment characteristics (area, mean slope, mean height, base flow index, annual rainfall, land cover, and geology) were determined on water quality parameters using stepwise regression via backwards method in the 63 selected rivers. The results showed that sodium absorption rate (SAR), total dissolved solids (TDS), electrical conductivity (EC), Na^+ and Cl^- parameters are strongly linked to geology characteristics, while K^+, Mg^{2+} and Ca^{2+} cations is linked to rainfall and geology characteristics. pH and HCO_3^- are related to area, rainfall, land cover and geology characteristics, CO_3^{2-} is related to area, rainfall, rangeland area and geology characteristics and SO_4^{2-} is related to area, rainfall, range and bar land area and geology characteristics. Adaptive Neuro-Fuzzy Inference System (ANFIS) was used for modeling the selected catchment characteristics and water quality parameters. The ANFIS models have a high Nash–Sutcliffe model efficiency coefficient (NSE) and low root mean squares error (RMSE) to estimate water quality parameters.
Keywords: pH; TDS; EC; Water quality; Water cautions; Water anions; Modeling: Mazandaran Province.

1. Introduction

Clean water is an essential prerequisite for environmental protection and economic development, political, social and cultural development of a country [1]. World population growth in recent decades and the increasing demand for food and rising health problems, increased per capita water consumption and pressure on existing water resources, have made resource conservation and food production necessary. Resource conservation and food production in terms of quality and quantity especially soil and water resources are a public duty. Unfortunately in Iran, the entry of fertilizers and pesticides and plant diseases in agriculture, created an imbalance between what is needed and what is consumed. Inappropriate use of chemicals in agriculture, led to increasing pollution of water resources that are passing through the towns and villages which and were infected enough. According to the fact that changes in the environment, under the influence of chemicals need for a strategy and plan for protecting water resources and its pollution control is important for its management [2]. Different factors affect the health of surface waters, so that the water quality at any point in a river, represents the major effects of land cover and existing landuse, weather conditions, rainfall, population density, livestock density, petrology and geology in the watershed [3]. Several indirect methods to simulate natural systems, estimates more accurate, more comprehensive and more complex calculations using a computer has been invented. One of these methods is modeling or simulation. There are many models for predicting water quality parameters including white box and black box models. Among these use of the statistical methods to predict water quality parameters, in terms of taking into account the characteristics of the watershed and lack of complexity of white-box models has attracted the researchers [4]. In recent years, there has been an increasing interest in intelligence models e.g. Artificial Neural Network(ANN), Adaptive Neuro-fuzzy Inference System (ANFIS), Fuzzy Logic and Genetic Algorithm for systems control [5, 6]. During recent years, the evolution of industry, commercial tourism and agriculture in the province is growing. Water quality undoubtedly has a direct impact on the sustainable development of human activities in the province. Therefore factors affecting water quality in rivers and the extent to which these factors affect water quality should be considered. This study aims to identify the most important factors affecting water quality parameters and also to

determine the relationship between water quality parameters and characteristics of the watersheds of rivers using ANFIS in Mazandaran province.

2. Materials and methods

2.1 Study area and data

Mazandaran province with a population density of 127 people per square kilometer in northern Iran is situated on the southern shores of the Caspian Sea. The study area equals 26132.13 km^2 and is located within the latitudes 35° 45′ N to 36° 59′ N and longitudes 50° 10′ E to 54°42′ E. Mazandaran province has 13 major rivers: Babolroud, Tajan, Siyahroud, Haraz, Nekaroud, Sardabroud, Cheshme-e- Gileh, Galandroud, Garmabroud, Chalosroud, Nesaroud, Chalakroud and Safaroud. These rivers flow from the Alborz mountains into the Caspian Sea. The water of the rivers is used in agriculture, industry and for drinking. The study area based on precipitation, temperature characteristics and topography is divided into Hyrcanian temperate and mountain climates. Position of the sampling sites are shown in Fig.1.
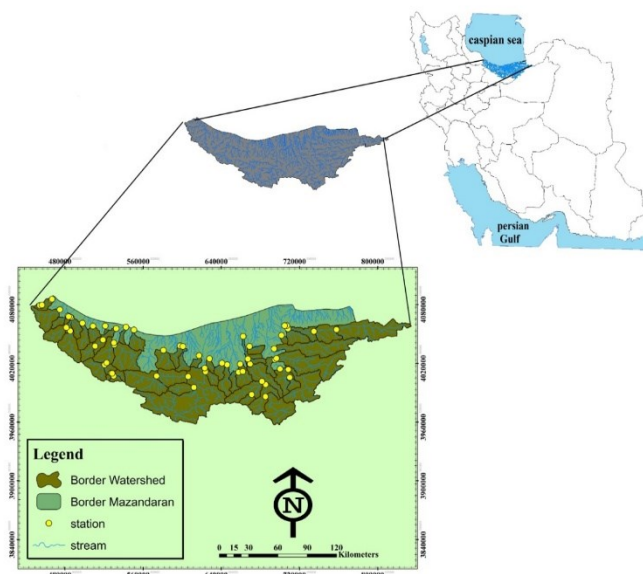

Fig. 1. The sampling sites across Mazandaran province used in the study

In this research, water quality data, including: total dissolved solids (TDS), sodium absorption rate (SAR), electrical conductivity (EC), Na$^+$, Cl$^-$, CO$_3^{2-}$, K$^+$, Mg^{2+}, Ca^{2+}, pH, HCO$_3^-$ and SO$_4^{2-}$ from 63 sampling sites for the period October 2010 to September 2011 collected by Iran Water Resources Research Center (IWRRC) have been used for this study. The sampling sites were selected based on two characteristics: (1) the dam, diversion and direct water utilization did not exist in their upstream (2) data are complete and continuous. Ultimately these 63 sampling sites with continuous and complete data were selected. Water quality parameters have been collected monthly by IWRRC. Samples have been picked in areas with low slope river, a calm and non-muddy river and from the middle of the river. After collecting samples, water quality parameters are measured according to standard methods. After organizing data, their annual mean in order to analyze has been utilized.

2.2 Methodology

2.2.1 Physical characteristics of watershed

At this stage, statistics, information and maps needed were collected. Independent variables, including physiographic variables, mean annual rainfall, base flow index (BFI), land use and geological characteristics were considered. To extract physiographic variables, digital elevation model (DEM) with a scale of 1:50,000 were prepared within Arc/GIS 9.3. Then sampling sites location was determined on DEM. Watershed containing each sampling site was determined and drawn using ArcHydro extension within Arc/GIS 9.3 (Fig. 1). Finally physiographic characteristics, including area, weighted mean slope and weighted mean height were derived. To calculate mean annual rainfall, monthly rainfall data for water-year 2010 (October 1, 2010 to September 30, 2011) were received from Iran Water Resources Research Company (IWRRC). Then the closest weather station to the selected watershed areas was chosen. According to ratio of weighted mean height of watershed area to mean height

of weather station, mean annual rainfall was determined. To calculate BFI, monthly discharges were obtained from IWRRC for 63 sampling sites in water-year 2010. Monthly hydrographs were drawn and amount of the base flow rate was determined by a straight line on the hydrograph. Finally, BFI was determined as ratio volume of water beneath separation line to volume of water beneath recorded hydrograph [7]. To determine land use, land use map was derived using a 2002 land use map with scale of 1:250,000 obtained from the Iran Forest, Ranges and Watershed Management Organization. The land use map was categorized into seven types, including forest, rangeland, bare land, dry farming land, irrigated land, urban and water body. To determine lithology, geological maps with scale of 1:100,000 were obtained from Iran Geological Survey and Mineral Exploration. All geological formations in upstream areas of watersheds were identified and classified in 14 groups, including: Lavas and Granite, Chalk, Shales, Sandstone, Lime, Conglomerates, Marl, Lime & Shales, Conglomerates & Sandstone, Lime & Sandstone, Shales & Sandstone, Lime & Marl, Tuff & Marl and Alluvial formations [8].

2.2.2 Used models

Linear relationship between variables was performed using stepwise regression. Stepwise regression can be performed in three methods, including forward, backward and step by step methods. In backward method, all independent variables were considered into the model, then the effect of each variable elimination is assessed [9].

The Sugeno type with five layers, namely, a fuzzy layer, a product layer, a normalized layer, a defuzzy layer and a total output layer was used in the study [10]. The Sugeno type used in the study was grid partitioning with four different types of membership functions (MFs) named gaussian1 (gauss1MF), gaussian2 (gauss2MF), generalized bell-shaped (gbellMF) and triangular (triMF) [10]. The inputs and outputs data were normalized before modeling between 0.1 and 0.9. Two statistical indices, including Nash-Sutcliffe efficiency coefficient (CE) and root mean squares error (RMSE) were used to evaluate modeling performance. RMSE values change between 0 and 1, values which are closer to zero indicate high accuracy predicted. The CE coefficient varies between negative infinitely and one and values which are closer to 1 indicate a higher performance model.

3. Results and discussion

Table 1 shows statistical summary of water quality parameters rivers of the Mazandaran province during water-year 2010. The highest value of TDS, EC, Cl^-, K^+, Na^+ and SAR were observed in Baleyran stations, pH in Dinarsara station, CO_3^{2-} in Mashalahabad station, HCO^{3-} in Sarokola station, SO_4^{2-} and Mg^{2+} in Khatirkuh station and Ca^{2+} in Pol-e Mergen station. While the lowest value of TDS, EC, K^+, HCO^{3-}, Ca^{2+} and Mg^{2+} were observed in Kelardasht station, pH in Sarokola station, SO_4^{2-} and Cl^- Dinarsra station, Na^+ in Ghalehgardan station and SAR in Ganeksar station.

Table 1. Statistical summary of water quality parameters rivers of the Mazandaran province

Parameters	Mean±Standard diviation	Parameters	Mean±Standard diviation
TDS	367.47±338.62	Cl^-	1.45±3.79
EC	560.86±508.15	Ca^{2+}	2.83±1.72
pH	8.18±0.29	Mg^{2+}	1.46±0.77
CO_3^{2-}	0.054±0.061	K^+	0.045±0.025
HCO_3^-	2.79±0.88	Na^+	1.23±3.45
SO_4^{2-}	1.13±1.85	SAR	0.68±1.64

TDS:Total dissolved solids; EC:Electrical conductivity; SAR: Sodium absorption rate

In general, sampling sites with the highest value for water quality parameters have been located in the western areas of the province. It was observed that sampling sites existing in the western areas of the province have often been located at a lesser distance from the coastline of Caspian Sea and are more greatly affected by human activities. This finding is in agreement with Mirzaee et al.'s (2014) findings which showed that most stations located in the western areas of Mazandaran province has water quality classes between moderate to bad. A considerable amount of literature has been published on water quality using physiographical, meteorological and land use variables [11-13]. However, far too little attention has been paid to geological variables [14]. The highest value of area and weighted mean slope belong to Karehsang and Vaspul stations, respectively. The highest value of weighted average height and rainfall belong to Kelardasht station. The highest value of base flow index belongs to Doab-e Chalous station. In terms of land use, the greatest amount of forest area (100 percent) belongs to Baleyran, Diva, Gavormak and Palande Rudbar stations. The greatest amount of rangeland belongs to Razen, Pol-e Mergen and Dareh Harijan stations. The greatest amount of dry land farming belongs to Sarokola station. In terms of geological formations, the greatest amount of Lavas and Granite group is observed in Razan station, Lime in Charz station, Conglomerates in Paltan station, Marl in Darabkola station, Conglomerates & Sandstone in

Sarokola station, Shales & Sandstone in Vaspul station, and Lime & Marl in Kheyroud Kenar station. Table 2 shows the results of the backward stepwise regression method.

Table 2. Results of stepwise backward method regression procedure via

Regression equation	RMSE	CE	R^2	Significant level
SAR = 0.293 + 0.042 Lime & Marl	1.875	0.072	0.093	0.04
TDS=286.65+364.12 Tuff & Marl + 5.14 Lime & Marl	352.35	0.19	0.047	0.006
EC = 448.11+548.19 Tuff & Marl+7.67 Lime & Marl	531.50	0.18	0.47	0.003
pH = 7.959 + 0.0001 A + 0.0001 P + 0.004 R + 0.143 BL + 0.054 Sandstone + 0.077 Lime & Shales – 0.159 Lime & Sandstone – 0.18 Tuff & Marl – 0.015 Lime	0.223	0.368	0.71	0.002
Na^+= 0.411 + 0.091 Lime & Marl	3.91	0.75	0.31	0.05
K^+= 0.060 – 0.0000227P + 0.025 Tuff & Marl	0.024	0.24	0.52	0.001
Mg^{2+}= 1.996 + 0.0001 P + 1.266 Tuff & Marl	0.60	0.53	0.75	0.00
Ca^{2+} = 3.78 – 0.001 P + 2.075 Tuff & Marl	1.62	0.28	0.56	0.00
Cl^-= 0.566 + 0.098 Lime & Marl	4.32	0.07	0.31	0.043
HCO_3^-= 3.966 + 0.0001 A – 0.001 P – 0.07 AGR + 0.046 DF – 0.009 R – 0.61 Shales + 0.474 Tuff & Marl	0.61	0.51	0.77	0.00
CO_3^{2-}= 0.033 – $2.736*10^{-5}$A + 0.001 R + 0.017 Lime & Shales – 0.021 Lime & Sandstone – 0.042 Tuff & Marl – 0.002 Lime	0.044	0.37	0.68	0.00
SO_4^{2-} = 1.403 – 0.001 P + 0.028 R – 0.764 BL + 2.335 Tuff & Marl	1.42	0.52	0.75	0.00

(A)Area(km^2), (P) Annual rainfall(mm), **(Lime & Marl)** Lime & Marl formations (%), (Tuff & Marl) Tuff & Marl formations (%), (Lime) Lime formations (%), (Lime & Sandstone) Lime & Sandstone formations (%), (Lime & Shales) Lime & Shales formations (%), (Sandstone) Sandstone formations (%), (DF) Dry land farming (%), (AGR) Irrigated land farming (%), (BL) Barren land (%), (R) Rangeland (%).

As can be seen from Table 2, rainfall and land use showed high correlation with the most water quality parameters while among geology formations, only Tuff & Marl group formation which is not widespread in the study area showed high correlation with water quality parameters. The effect of land use on surface water has been investigated in a lot of studies such as that observed by Amiri and Nakane [15] who had reported a significant relation between land use and water quality. Also Na+ and SAR did not show significant correlation with effective parameters. Considering to the results of the backward stepwise regression method in SAR, Na+ and Cl- parameters, Tuff & Marl geology group, in EC & TDS parameters, Tuff & Marl, Lime & Marl geology group, in pH parameter, area & rainfall factors and Barren land & rangeland, Sandstone, Lime & Shales, Lime & Sandstone, Tuff & Marl, Lime, in K+، Mg^{2+} and Ca^{2+}parameters rainfall and Tuff & Marl factors, in HCO3-, area, agriculture, dryland and rangeland and Shales, Tuff & Marl geology group, in CO_3^{2-} parameter area, rangeland and Lime & Shales, Lime & Sandstone, Tuff & Marl and Lime geology groups, in SO_4^{2-} parameter, rainfall factor, rangeland and Barren Land and Tuff & Marl geology group, were chosen as effective variables.

Since there were no differences between results modeling using the four different types of MFs, the results of ANFIS procedure using gbellMF are shown in Table 3.

Table 3. Results of ANFIS procedure using gbellMF

Variables	Modeling		Testing	
	RMSE	*CE*	*RMSE*	*CE*
TDS	0.0117	0.99	0.009	0.99
EC	0.0117	0.99	0.008	0.99
pH	0.00268	0.99	0.19	0.99
CO_3^{2-}	0.0037	0.99	0.393	0.99
HCO_3^-	0.0045	0.99	0.269	0.99
SO_4^{2-}	0.012	0.99	0.0108	0.99
Cl^-	0.0123	0.99	0.011	0.99
Ca^{2+}	0.0137	0.99	0.008	0.99
Mg^{2+}	0.009	0.99	0.007	0.99
K^+	0.018	0.99	0.015	0.99
Na^+	0.011	0.99	0.010	0.99
SAR	0.024	0.99	0.021	0.99

Use of ANFIS for modeling water quality parameters have been used in a lot of previous studies [12, 16]. In general, the ANFIS model has a high efficiency and low error for estimating water quality parameters. Although the major influence of point sources may mask the effect of wider basin attributes on water quality parameters [14] but this effect is little in this study.

4. Conclusions

The purpose of the current study was to model the relationship between water quality parameters and characteristics of the watersheds of rivers using ANFIS in Mazandaran province. This study has found that generally sampling sites with the highest values of water quality are located in the western areas of the province. On the other hand, rainfall and land use showed high correlation with most of the other water quality parameters. Therefore, it is recommended to use the results of this study in order to improve water quality management in the rivers of Mazandran province. The suitable management practices can be used for water pollution control in the rivers of the Mazandran province. In addition, according to high efficiency of ANFIS models, this modeling method was used to model water quality parameters.

5. References

[1] Sargaonkar A, Deshpande V. Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. Environmental monitoring and assessment. 2003;89(1):43-67.

[2] Zhao Y, Nan J, Cui FY, Guo L. Water quality forecast through application of BP neural network at Yuqiao reservoir. Journal of Zhejiang University-Science A. 2007;8(9):1482-1487.

[3] Nasirahmadi K, Yousefi Z, Tarassoli A. Zoning of water quality on Haraz river bases on national sanitation foundation water quality index. Journal of Mazandaran University of Medical Sciences. 2012;22(92):64-71.

[4] Singh KP, Basant A, Malik A, Jain G. Artificial neural network modeling of the river water quality—a case study. Ecological Modelling. 2009;220(6):888-895.

[5] Abghari H, Mahdavi M, Fakherifard A, Salajegheh A. Cluster analysis of rainfall-runoff training patterns to flow modeling using hybrid RBF Networks. Asian Journal of Applied Sciences. 2009;2(2):150-159.

[6] Vafakhah M. Application of artificial neural networks and adaptive neuro-fuzzy inference system models to short-term streamflow forecasting. Canadian Journal of Civil Engineering. 2012;39(4):402-414.

[7] Zarrin H, Sharifi F, Vafakhah M, Mahdian MH. Regional analysis of low flow in Karkheh and Karoon Watersheds. Journal of Applied Sciences. 2009;9(6):1141-1146.

[8] Ahmadi H, Feiznia S. Quaternary formation (Theoretical and applied principles in natural resources). University of Tehran; 2006.p. 557. (in Persian)

[9] Bihamta MR, ZareChahuki MA. Statistics of natural resources. University of Tehran; 2011.

[10] Jang JS. ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics. 1993;23(3):665-685.

[11] Ahangar AG, Soltani J, Abdolmaleki AS. Predicting Mn concentration in water reservoir using Artificial neural network (Chahnimeh1 reservoir, Iran). International Journal of Agriculture and Crop Sciences. 2013;6(20):1413.

[12] Areerachakul S. Comparison of ANFIS and ANN for estimation of biochemical oxygen demand parameter in surface water. International Journal of Chemical and Biological Engineering. 2012;6:286-290.

[13] Ranković V, Radulović J, Radojević I, Ostojić A, Čomić L. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecological Modelling. 2010;221(8):1239-1244.

[14] Rothwell JJ, Dise NB, Taylor KG, Allott TE, Scholefield P, Davies H, Neal C. Predicting river water quality across North West England using catchment characteristics. Journal of Hydrology. 2010;395(3-4):153-162.

[15] Amiri BJ, Nakane K. Modeling the linkage between river water quality and landscape metrics in the Chugoku district of Japan. Water resources management. 2009;23(5):931-956.

[16] Safavi H.Prediction of river water quality by adaptive neuro fuzzy inference system (ANFIS). Journal of Environmental Studies. 2010; 36(3):1-10.