ORIGINAL PAPER

# Evaluation of the water quality at Tahtali dam watershed in Izmir-Turkey by means of statistical methodology

Hatim Elhatip · M. Ali Hınıs · Necati Gülbahar

Published online: 12 April 2007 Springer-Verlag 2007

Abstract Sustaining the human ecological benefits of surface water requires carefully planned strategies for reducing the cumulative risks posed by diverse human activities. Municipal governments in Izmir City play a key role in developing solutions to surface water management and protection. Therefore, several methodologies are carried out to develop new solutions for protecting available water resources. In the present study, the major chemical loads of surface water at the Tahtalı dam reservoir were evaluated statistically by using monthly averaged values of the measured parameters. Here, the SPSS and NCSS statistical programs were applied during the statistical analyses. Analyses were carried out in two stages. In the first part of the statistical analyses, the mean, median, minimum, maximum, 25th and 75th quartiles were calculated. In second part, the data were investigated by using statistical median test, normality test, parametric and nonparametric correlation and regression analyses. These methods were performed on water quality data taken from four sample sites representing the recharge and discharge areas at the Tahtalı dam. The Median test is applied to check if medians of water quality data from four sample sites (Dereboğazı, Bulgurca, Menderes and Gölcükler) are equal or not. Commonly a non-parametric test (distribution-free test) is used to compare two independent groups of sampled data. Since there are more than two groups in independent group comparison, Kruskal–Wallis test is

H. Elhatin  $(\boxtimes)$ 

M. A. Hınıs · N. Gülbahar Engineering Faculty, Civil Engineering Department, Aksaray University, 68100 Aksaray, Turkey

applied instead of Mann–Whitney  $U$  test. Finally, the main objective of using statistical analyses in third research is to estimate the types of pollution sources, the level of pollution and its environmental impacts on the Tahtalı dam reservoir by means of statistical methodology.

Keywords Statistical methodology · Contamination sources · Tahtalı dam watershed · Water contamination · Turkey

#### 1 Description of the research area

The Tahtalı dam is in the southwestern part of İzmir city, southwest Turkey (Fig. [1](#page-1-0)). This dam was constructed to supply about 4,055 l/s of drinking water from the Tahtalı River to İzmir city (Gülbahar [1999](#page-9-0)). The construction of the Tahtalı dam was completed in 1996. The topographical conditions were suitable for supplying clean water from the reservoir to İzmir city by gravity flow. A general view of the Tahtalı dam site is given in Fig. [2](#page-1-0).

The maximum reservoir capacity is  $307,000$  hm<sup>3</sup>, whereas the minimum capacity is  $56,000$  hm<sup>3</sup> (Kilinc and Gülbahar [1997\)](#page-9-0). The watershed of the Tahtalı dam covers a surface area of  $515 \text{ km}^2$ , extending along the Cuma plain. The watershed is the water collection area of the Tahtalı River, which runs in the NE–SW direction, through the Cuma plain. About 50% of the area is covered with forests and 30% of the land is used for agricultural purposes, whereas, the remaining represents urban area (Atış [1999](#page-9-0)).

# 2 Hydrochemical evaluation of the Tahtalı watershed

Human activities near the Tahtalı dam site have had direct and indirect effects on the rates of contamination of surface

Engineering Faculty, Environmental Engineering Department, Aksaray University, 68100 Aksaray, Turkey e-mail: hnaim@superonline.com

<span id="page-1-0"></span>

Fig. 1 Location map of the Tahtalı dam site

and groundwater in the Tahtalı watershed area. Direct effects include dissolution and transport of excess quantities of fertilizers with associated materials and hydrologic alterations related to irrigation and drainage. Indirect effects include changes in water–rock reactions in soils and aquifers caused by increased concentrations of dissolved oxidants, protons, and major ions.

Agricultural activities have directly or indirectly affected the concentrations of a large number of inorganic chemicals in groundwater, for example  $NO_3$ ,  $N_2$ ,  $Cl^-, SO_4^{2-}$ ,  $H^+$ ,  $K^+$ ,  $Mg^{2+}$ ,  $Ca^{2+}$ , Fe, Cu, B, Pb, and Zn, as well as a wide variety of pesticides and other organic compounds. For reactive contaminants like  $NO<sub>3</sub>$ , it is suggested that a combination of hydrochemical and environmental-tracer analytical approaches might be required to resolve changing inputs from subsequent alterations as causes of concentration gradients in groundwater.

The results of hydrochemical analyses are given in Table 1.The main water type of the Tahtalı lake is Ca–Mg–



Fig. 2 General view of the Tahtalı dam site

HCO3. The river water type was directly related to the hydrogeochemical properties of the study area.

# 3 Material and methods

#### 3.1 Statistical evaluation of hydrochemical data

Water quality data were collected from four sites shown in Fig. [3](#page-2-0) for three years between 1996 and 2000 and some of the data were obtained from DSİ (General Directorate of State Hydraulic Works). The results were evaluated by using monthly averaged values of the parameters statistically. The SPSS and NCSS statistical programs were used in statistical analyses. Statistical analyses were carried out to approve the water quality changes regarding to relationship between different concentrations. Thus, the statistical analyses were set forth in two parts: in the first part, the summary statistics such as mean, median, minimum, maximum and 25th and 75th quartiles were calculated (NCSS [2004\)](#page-9-0).

In second part, the data were investigated by using statistical methods such as median test, normality test, parametric and non-parametric correlation analyses and regression analyses. These methods were performed on water quality data taken from four sample sites to percieve

Table 1 Hydrochemical analysis of the Tahtalı lake site (mean annual values)

Parameters (units)	1996	1998	2000
$T$ (°C)	9.4	9.7	8.1
PH	7.7	7.8	8.0
$EC$ ( $\mu$ S/cm)	470	510	542
$DO$ (mg/l)	6.6	5.3	4.6
$Na^+$ (mg/l)	14.0	17.0	19.0
$K^{+-}$ (mg/l)	5.50	6.0	7.0
$Ca^{2+}$ (mg/l)	51.00	62.10	70.12
$Mg^{2+}$ (mg/l)	14.30	15.10	24.30
$HCO3- (mg/l)$	303	342	360
$SO_4^{2-+}$ (mg/l)	16.80	18.30	19.20
$Cl^{-} (mg/l)$	20.90	24.50	24.90
Cu (mg/l)	0.005	0.006	0.008
Fe (mg/l)	0.170	0.223	0.430
$Zn$ (mg/l)	0.014	0.016	0.01
$Pb$ (mg/l)	0.006	0.008	0.020
$Cr$ (mg/l)	0.020	0.037	0.080
$B$ (mg/l)	0.09	0.18	0.42
$NO3$ (mg/l)	12.01	16.21	16.65
$NO2$ (mg/l)	0.26	0.32	0.50
$PO4$ (mg/l)	0.10	0.15	0.19
$NH4$ (mg/l)	0.10	0.14	0.17

<span id="page-2-0"></span>



any relations among them. Median test is applied to check if medians of water quality data from four sample sites (Dereboğazı, Bulgurca, Menderes and Gülbahar) are equal or not. The median of data from two or more samples can be used to help make decisions about the equality of the centers of samples. Comparing medians, rather than means, can yield a substantially different conclusion as to whether two distributions differ in terms of some measure of central location (Wilcox [1991\)](#page-9-0). Since, the median is minimally affected by the magnitude of observations, being determined solely by the relative order of observations, this resistance to the effect of a change in value or presence of outlying observations is often a desirable property (Dennis and Hirsch [2002](#page-9-0)). Commonly a non-parametric test (distribution-free test) is used to compare two independent groups of sampled data. Since there are more than two groups in independent group comparison, Kruskal–Wallis test is used instead of Mann–Whitney  $U$  test. This nonparametric test makes no assumptions about the distribution of the data (e.g., normality).

There are several methods to test normality. Among many methods, three tests are chosen for the present study, these are; the Kolmogorov–Smirnov (K–S) test, the Anderson Darling test and the Shaphiro–Wilk test. The K– S test determines if two data sets differ significantly. The K–S test has the advantage of making no assumption about the distribution of data. K–S test is non-parametric and distribution free test, which means the critical values do not depend on the specific distribution being tested. Data were compared with normal distribution values and K–S test statistic is the maximum difference between the cumulative proportions of the observed distribution and expected cumulative–normal distribution.

In the present study, decisions were based on the comparison of tests statistics with 5% significance level of critical values. The Anderson–Darling test is also used to test if a sample of data came from a population with normal distribution. It is a modification of the K–S test and gives more weight to the tails than does the K–S test. The Anderson–Darling test makes use of the specific distribution in calculating critical values.

The application of each statistical methods in the present study is carried out to; check the normality of periodical field measurements, and discern the relationship of the water quality parameters in different areas regarding to different source of contamination. Thus, correlations were carried out in two groups; first group represents the correlations of each water quality data among sample sites and second group represents the correlations of sample sites parameters among water quality data (NCSS [2004](#page-9-0)).

First group of correlation was put forward to show the correlation of each water quality data between four sites, for example, correlation between the BOD of Dereboğazı with the BOD of other sites. Second group of correlation analyses was performed to show any dependency of water quality data with other water quality data for each site. For example, for Dereboğazı site the correlation of BOD with NO3, NO2, PO4, NH4 parameters and as well as the correlation of COD with Cu, Pb, and Fe were investigated. During correletion between sites, three correlation coefficients were investigated; a parametric Pearson correlation coefficient, nonparametric Kendall's Tau\_b and nonparametric Spearman's rho correlation coefficients (SPSS [2004\)](#page-9-0).

Pearson's correlation coefficient is a measure of linear association. Pearson's correlation coefficient is not an appropriate statistic for measuring their association, if the relationship is not linear between two variables. Since regression requires the normality of the data, in correlation and regression analyses, linearity and normal distribution is assumed in this paper. The high population and increase of industrial activities in İzmir city, play an important role in the water contamination problems. Therefore, the main objectives of the statistical application for each method depends on the relationship between different physical and chemical properties of the water quality in the dam site. For example; increase of BOD contents due to the increase of Nutrient Contents (N, P); or increase of COD contents due to the increase of heavy metales (Cu, Pb, and Fe).

Table 2 illustrates the median values of Nitrite, Phosphate, Fe and NH4 parameters are not equal (null hypothesis is rejected) on 5% significant level. Three normality tests (S–W, Shapiro–Wilk; K–S, Kolmogorov–Simirnov; A–D, Anderson–Darling) are applied to four sample sites and results are compared in Table 3.

The table illustrates that for nitrate and nitrite values of all sites, there is not enough data to reject normality. Having the highest test values, normal probability plot of COD and Cu data of Gölcükler site along with 95% of confidence intervals are given in Fig. [4](#page-4-0). Normality tests give the result as normality of the data cannot be rejected and it can be seen from the figure. Data follows the theoretical straight line of normal distribution very closely, hence normally distributed.





Hypotheses: Ho all medians are equal; Ha at least two medians are different



Table 3 Normality test results in each water quality data for four sample sites

<span id="page-4-0"></span>Correlation value of the dependent variable is shown as one in Table [4.](#page-5-0) Dereboğazı site is taken as dependent variable except in COD (2) and Cu (2) correlation analyses. The analyses use Gölcükler site as dependent variable and show the correlation of other sites to Gölcükler site. Here, the plot of histogram for all variables is not normally distributed due to two main reasons; (1) the changes in amount of contaminants in different periods and type of contamination sources in each sampling area; and (2) changes in different nonrelative parameters periodically. Therefore, the main reasons for testing the normality of the field data are; to prove the normality of the relative parameters, find out the periodical changes in the field measurements, and finally to clarify the distribution of periodical changes in water quality in different areas regarding to changes in these relative concetrations. The lower values of Kendall's tau do not mean that tau is less sensitive than Pearson's coefficient, but simply that a different scale of correlation is being used (Dennis and Hirsch [2002](#page-9-0)). Strong correlations were observed between sites and high significant levels are indicated by asteriks (\*) and (\*\*) for 0.05 and 0.01 levels, respectively.

COD (2) analysis shows a very strong correlation between Gölcükler and Menderes sites and Gölcükler and Bulgurca sites. Second part of the correlation analyses was performed for water quality data within each site to show the dependency of pollution parameters. The correlation in this analysis is performed into two groups; correlation of BOD with  $NO_3$ ,  $NO_2$ ,  $PO_4$ ,  $NH_4$  and correlation of COD with Cu, Pb, F for each site and results are given in Tables [5](#page-6-0) and [6,](#page-6-0) respectively. Table [5](#page-6-0) shows a negative correlation between BOD and  $NO<sub>3</sub>$ , however, there is a positive correlation between BOD and NH4 values for Derebogazi and Bulgurca sites. Table [6](#page-6-0) shows correlations between COD and heavy metals. Regression analyses were carried out for the sites of significant correlations and linear

models were constructed. Linear correlations were formed as;

$$
Site 1 = A + B * Site 2 + C * Site 3 + D * Site 4 + ...
$$

The model is a multiple linear regression model which is the extension of simple linear regression model with multiple explanatory variables. A is the intercept  $B$  is the slope coefficient for the first explanatory variable (Site 2), C is the slope coefficient fort he second explanatory variable (Site 3), etc.

Some of the linear correlation models with constant values including  $R^2$  values are given in Table [7](#page-6-0). Since there is a significant correlation between BOD and  $NH<sub>4</sub>$  at Dereboğazı and Bulgurca sites, a regression betweeen BOD and NH4 were sought and given below;

$$
BOD_{\text{Bulgurca}} = -1.4581 + 100.38 * NH_{\text{4Bulgurca}} R^2 = 0.8798
$$
  

$$
BOD_{\text{Derebogaz}I} = 0.2592 + 48.884 * NH_{\text{4Derebogaz}I} R^2 = 0.6660
$$

For comparison, maesured values and regression line of BOD and NH4 parameters of Derebogazı site is given in Fig. [5](#page-6-0). Water quality parameters were investigated statistically and strong relations were shown among four sites. The sample points were choosen as continuous of the river watershed. Dereboğazı is at the outlet of Tahtalı dam and Bulgurca is at the inlet of Tahtalı dam. Menderes and Gölcükler sites were remote points which feeding the inlet of Tahtalı dam. There are industrial activities reported near Menderes site and agricultural activities reported near Gölcükler site (Atış [1999\)](#page-9-0).

As a result of the study, the pollution on the sites and the effect of pollution to downstream can be estimated from the correlation coefficients in Tables [4,](#page-5-0) [5](#page-6-0), and [6](#page-6-0). Based on the correlation coefficient analyses, estimated models were constructed and given in Table [7](#page-6-0). The water quality parameters and their effects can be detected and be esti-





<span id="page-5-0"></span>

\* Correlation is significant at the 0.05 level (2-tailed); \*\* correlation is significant at the 0.01 level (2-tailed)

mated depending on the upstream parameters of the sites (Brooks and Thomas [1991\)](#page-9-0).

# 4 Results and discussion

The statistical application is carried out for 3 years (36 months) with mean monthly values for evaluating the monthly changes in different parameters; approving the application of relevant statistical methodology in such environmental problem; and finally clarifying the changes of water quality parameters. The mean monthly values of water quality data for four sample sites were given in Fig. [6](#page-7-0).

The following graphical presentation illustrate the periodical analyses of NO 3, NO 2, PO 4, NH 4, BOD, COD, Fe, Cu, Pb, phosphates and TDS contents for each site, contamination in terms of statistical methodology. Result of the analyses showed that influences of pollution due to industrial and agricultural activities is significantly taken place at Gölcükler and Menderes sites, where high fluxes of heavy materiales Fe, Cu, Pb and high nutrients contents. Here, some of elements are added to the hydrosphere directly as components of fertilizers or other additives, whereas other elements were associated indirectly with fertilizer loadings or land disturbance through enhanced rates of leaching of natural sources in soils.

Results showed that high nitrate concentrations might occur due to leaching of  $NO_2$ ,  $NO_3$ ,  $NH_4$  and  $PO_4$  from fertilizers and biocides during irrigation of agricultural land around the dam site, in January, February, and April. Moreover, Leakage from overloaded sewage networks are also sources of nutrients and heavy mineral pollution. The dam lake is highly contaminated by extreme NO<sub>2</sub>, NO<sub>3</sub>, NH 4, heavy minerals and phosphate loads originating from irrigation drainage, sewage, and deteriorating septic treatment system.

The main sources of nitrate in the study area are: nitrate from the application of fertilizers in irrigation and from untreated wastewaters directly discharged to the lake. The Periodical changes in heavy metal concentrations; Fe, Cu, and Pb against water temperature and TDS showed that the indirect effects have important implications for geochemical studies of water–rock interactions and can represent sources or sinks for a variety of problematic contaminants of such heavy metals and/or toxic trace elements (Elhatip et al. [2004](#page-9-0)). Moreover, the median and quartiles of data in box plot were given in Figs. [7](#page-8-0) and [8.](#page-9-0) In the box plot green and red dots show the mild and severe outliers, respectively. The Figures illustrate the median and quartiles of BOD, COD, PO 4, nutrients and heavy mineral contents in each variable around the dam site.

<span id="page-6-0"></span>**Table 5** Correlation of BOD with  $NO_3$ ,  $NO_2$ ,  $PO_4$  in each site. BOD is taken as dependent variable

<b>Sites</b>	Pearson				Kendall's Tau b				Spearman's rho						
	BOD NO <sub>3</sub>		NO <sub>2</sub>	PO <sub>A</sub>	NH <sub>4</sub>	BOD NO <sub>3</sub>		NO <sub>2</sub>	PO <sub>4</sub>	NH4	BOD NO <sub>3</sub>		NO <sub>2</sub>	PO <sub>A</sub>	NH <sub>4</sub>
Dereboğazı 1		$-0.61*$	$-0.127$ 0.388		$0.816**$ 1		$-0.351$	$-0.015$ 0.25		$0.708**$ 1		$-0.532$	$-0.028$ 0.321		$0.853**$
Bulgurca		$-0.136$			$-0.453$ $-0.419$ $0.938**$ 1			$-0.148$ $-0.113$	$-0.445$ 0.215			$-0.229$	$-0.17$	$-0.59$	0.235
Menderes		$-0.573$		$-0.509$ $0.735*$ 0.179			$-0.161$	0.081	$-0.081$	$-0.081$		$-0.203$	0.076	$-0.038$ $-0.051$	
Gölcükler		$-0.621*$			$-0.118$ $-0.294$ $-0.256$		$-0.55*$	$-0.11$		$-0.278$ $-0.147$		$-0.683*$	$-0.041$	$-0.429$ $-0.251$	

\* Correlation is significant at the 0.05 level (2-tailed); \*\* correlation is significant at the 0.01 level (2-tailed)

Table 6 Correlation of COD with Cu, Pb, Fe in each site. COD is taken as dependent variable

<b>Sites</b>	Pearson				Kendall's Tau b				Spearman's rho			
	<b>COD</b>	Cu	Pb	Fe	<b>COD</b>	Cu	Pb	Fe	<b>COD</b>	Cu	Pb	Fe
Dereboğazı		$0.615*$	$-0.194$	0.529		0.277	$-0.032$	0.308		0.375	0.11	0.502
Bulgurca		0.021	0.597	0.005		0.143	0.333	0.048		0.107	0.4	$-0.036$
Menderes		$-0.2$	$-0.26$	$-0.151$		$-0.357$	$-0.327$	0.143		$-0.476$	$-0.24$	0.071
Gölcükler		$-0.486$	$-0.264$	$-0.566$		$-0.514*$	0.019	$-0.345$		$-0.638*$	$-0.032$	$-0.536$

\* Correlation is significant at the 0.05 level (2-tailed)

Table 7 Regression analyses between sites for water quality parameters

Parameters	Dependent	А Slope	В Dereboğazı	C Bulgurca	D Gölcükler	E Menderes	$R^2$
Nitrate	Dereboğazı	1.103		0.319	0.064	0.346	0.8494
<b>BOD</b>	Dereboğazı	$-1.249$	-	$-1.368$	0.376	3.521	0.9914
NH <sub>4</sub>	Dereboğazı	0.004	-	0.747	$-0.038$	0.375	0.8119
<b>TDS</b>	Dereboğazı	$-22.667$		1.332	2.661	$-0.675$	0.9132
<b>BOD</b>	Bulgurca	$-0.56$			$-0.5173$	2.4384	0.9795
$\rm COD$	Gölcükler	3.475	$-0.296$	0.777	-	0.104	0.9613

#### 5 Conclusions

The statistical application for each method is mainly carried out to approve the water quality changes in the dam watershed. As a result of statistical analysis used in this study, the water quality parameters and their effects are detected and estimated depending on the upstream parameters of the sites. Here, the relationship between different physical and chemical parameters in the present study showed that the Tahtalı dam lake is highly contaminated by extreme N–P loading, especially during dry periods when eutrophication processes usually take place in static water masses. The process usually occurs due to the continuous increase of nitrogen and phosphorus and the decrease of  $O_2$  in the water. This may cause an anaerobic medium and stimulate the algae growth in the lake. Therefore, continuous pollution of the Tahtalı dam reservoirs may consequently reduce the quality of the lake.

Common agricultural practices have caused substantial increases in groundwater recharge fluxes and concentrations of major elements and a variety of other heavy metals. The increases are related to transient agricultural chemical signal in aquifers that have groundwater resi-



Fig. 5 Graphical representation of measured values and regression line of BOD versus NH4 for Derebogazı site

<span id="page-7-0"></span>

Fig. 6 Graphical representaion of water quality at the dam watershed (1996–2000)



Fig. 6 continued

<span id="page-8-0"></span>



dence times of decades or more. Some elements are added to the hydrosphere directly as components of fertilizers or other additives. Others are associated indirectly with fertilizer loadings or land disturbance through enhanced rates of leaching of natural sources in soils. This leads to physical and biological changes and to increased acidity or ionic strength of agricultural recharge.

In addition, high fluxes of  $NH<sub>4</sub>$  and  $NO<sub>3</sub>$  in recharging groundwater have resulted in unnaturally high concentrations of chemical species released from aquifer materials by oxidation reactions in the saturated zone, because the electron demand of  $NO<sub>3</sub>$  in agriculturally contaminated recharge is commonly several times higher than that of dissolved  $O_2$  in uncontaminated recharge.

In cases of increasing N and P loads being greater than normal conditions, dramatic quality changes will occur in the water mass. Moreover, identifying the quality and quantity of water pollutants originating from non-point sources for solving water pollution questions, is a difficult and complex problem. Since the analyses were carried out for 3 years (36 months) data, which may not be enough to determine the environmental impact of pollution to the watershed, it is necessary to collect more data in the coming years at the same stations fort he validity of derived equations. Therefore, the only real solution to the environmental impacts of pollution is the determination of protection zones and implementation of good land use planning to maintain control of the activities in watersheds.

<span id="page-9-0"></span>Fig. 8 Median and quartiles of TDS, Cu, Pb and Fe contents



Acknowledgments The authors would particularly like to thank the anonymous reviewers for very constructive criticism that helped to improve the quality of the paper and clarify important points. The authors would also like to acknowledge the editor G. Christakos for the timely handling the review processes of the paper.

# References

- Atış I (1999) Drinkable water supply for İzmir, present, past and future. Water Congress, Izmir
- Brooks KF, Gregersen HTJ (1991) Hydrology and management of watersheds. IOWA State University, IOWA
- Dennis RH, Robert MH (2002) Statistical methods in water resources, USGS techniques of water resources investigations, Book 4, Chap A3
- Elhatip H, Afşin M, Kuşçu İ, Dirik K, Kurmaç Y, Kavurmacı M (2004) Influences of human activities and agriculture on groundwater quality of Kayseri–Incesu–Dokuzpınar springs, Central Anatolian part of Turkey. Environ Geol 44(4):490–494
- Gülbahar N (1999) İzmir Tahtalı Barajı Su Toplama Havzasnda Koruma ve Kullanma Dengesi, Doktora Tezi. Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü, Kütahya
- Kılınç MY, Gülbahar N (1997) The balanced multiple use and protection of watershed resources, through integrated watershed management for sustainable development. In: XI world forestry congress, pp 12–22 October, Antalya, Turkey, p 19
- NCSS (2004) Statistical analysis software, Kaysville Utah
- SPSS, SPSS 13.0 for Windows (2004) Software for statistical analysis Wilcox Rand (1991) Testing whether independent treatment groups have equal medians. Psychometrika 56(3)