Surface water quality assessment using factor analysis

Hülya Boyacioglu
Dokuz Eylul University, Faculty of Engineering. Department of Environmental Engineering, Tinaztepe Campus Buca, 35160, Izmir, Turkey

Abstract

In this study, the factor analysis technique is applied to surface water quality data sets obtained from the Buyuk Menderes River Basin, Turkey, during two different hydrological periods. Results show that the indices which changed the quality of water in two seasons and locations differed. During low-flow conditions, water quality was strongly affected by agricultural uses. On the other hand, the main pollution source changed from agricultural uses to urban land uses in high-flow periods. Therefore major water pollution threats in the basin were urban and agricultural land uses which are defined as non-point sources. This technique is believed to assist decision makers in identifying priorities to improve water quality that has deteriorated due to various land uses.

Keywords: factor analysis, factor scores, Buyuk Menderes River, water quality

Introduction

Water quality monitoring has one of the highest priorities in environmental protection policy (Simeonov et al., 2002). The main objective is to control and minimise the incidence of pollutant-oriented problems, and to provide water of appropriate quality to serve various purposes such as drinking water supply, irrigation water, etc.

The quality of water is identified in terms of its physical, chemical and biological parameters (Sargaonkar and Deshpande, 2003). The particular problem in the case of water quality monitoring is the complexity associated with analysing the large number of measured variables (Saffran, 2001). The data sets contain rich information about the behaviour of the water resources. The classification, modelling and interpretation of monitoring data are the most important steps in the assessment of water quality.

Multivariate statistical methods including factor analysis have been used successfully in hydrochemistry for many years. Surface water, groundwater quality assessment and environmental research employing multi-component techniques are well described in the literature (Praus, 2005). Multivariate statistical approaches allow deriving hidden information from the data set about the possible influences of the environment on water quality (Spanos et al., 2003).

Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable (Yu et al., 2003). There are three stages in factor analysis (Gupta et al., 2005):

- For all the variables a correlation matrix is generated
- Factors are extracted from the correlation matrix based on the correlation coefficients of the variables
- To maximise the relationship between some of the factors and variables, the factors are rotated.

☎ +90 232 412 71 31; fax: +90 232 453 11 43;
e-mail: hulya.boyacioglu@deu.edu.tr
Received 16 January 2006; accepted in revised form 24 April 2006.

A first step is the determination of the parameter correlation matrix. It is used to account for the degree of mutually shared variability between individual pairs of water quality variables. Then, eigenvalues and factor loadings for the correlation matrix are determined. Eigenvalues correspond to an eigenfactor which identifies the groups of variables that are highly correlated among them. Lower eigenvalues may contribute little to the explanatory ability of the data. Only the first few factors are needed to account for much of the parameter variability. Once the correlation matrix and eigenvalues are obtained, factor loadings are used to measure the correlation between the variables and factors. Factor rotation is used to facilitate interpretation by providing a simpler factor structure (Zeng and Rasmussen, 2005).

This study evaluated the possibility that a smaller group of water quality parameters/ locations might provide sufficient information for water quality assessment. Factor analysis was applied to a surface water quality data set collected from Buyuk Menderes Basin, Turkey using 'the *Statistical Package for the Social Sciences Software-SPSS 10.0 for Windows*'. Water quality monitoring was conducted at 21 stations in the study area during low- and high-flow periods. The selected parameters for the estimation of surface water quality characteristics were: electrical conductivity (EC), total dissolved solids (TDS), sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), sulphate (SO_4^{2-}), nitrate-nitrogen (NO_3-N), Kjeldahl Nitrogen, biochemical oxygen demand (BOD_5) and chemical oxygen demand (COD). COD measurements were performed using the potassium dichromate method.

Study area

The Buyuk Menderes River Basin is located in Western Anatolia and covers Uşak, Aydın and Denizli Provinces with a total land area of about 25 000 km^2 (Fig. 1). The basin is endowed with one of the most fertile soils in the country and the economy of the region is heavily dependent on agricultural production. In addition, rapid industrialisation and population growth over the past few decades have created additional stress on the environmental conditions in the region (Boyacioglu et al., 2004). The