MULTIVARIATE STATISTICAL TECHNIQUES FOR THE ASSESSMENT OF SURFACE WATER QUALITY INTENDED FOR HUMAN CONSUMPTION

E. M. Smeti and S. K. Golfinopoulos

A The Athens Water Supply and Sewerage Company (EYDAP SA), Quality, Research & Development (R&D) Division, 156 Oropou St., 11146, Galatsi, Athens, Greece.

B University of the Aegean, Department of Financial and Management Engineering, 41 Kountouri Street, 82100, Chios, Greece.

*E-mail: esmeti@eydap.gr

The man-made reservoirs of Marathon, Mornos and occasionally the natural lake Yliki are the supply sources of raw surface water in the greater Athens metropolitan area, with more than 4 million inhabitants. Multivariate statistical techniques, such as principal components (PCA) / factor analysis (FA), cluster analysis (CA), discriminant analysis (DA) and classification trees (CART) were extensively explored and applied to the surface water quality data of the Marathon reservoir, Yliki Lake and Mornos reservoir to interpret the data structure and to evaluate temporal and spatial variations in the water quality. Water samples from Marathon, Yliki and Mornos were routinely analyzed for 16 physico-chemical parameters over a 4 or 5-year period (2003-2008), on a monthly basis. The analyzed parameters were: Turbidity, pH, electrical conductivity, total alkalinity, total hardness, total dissolved solids, dissolved oxygen, chemical oxygen demand, NO₃⁻, NH₄⁺, Cl⁻, Na⁺, SO₄²⁻, PO₄³⁻, Ca²⁺ and Mg²⁺. The results were subjected to a PCA/FA and 3, 4 and 6 latent factors were extracted for Marathon, Yliki and Mornos with 75.6%, 80.4% and 80.4% of the total variance, in each reservoir/lake, being explained respectively. Cluster analysis was used for detecting natural groupings in data. This multivariate statistical technique resulted in two major temporal clusters for each reservoir/lake. Discriminant analysis as well as CART was used for determining which variables are the most efficient in discriminating between the two clusters originated by CA. The high percentage of correct classification of both methods (over 90%) indicated the good accuracy of the models. During the period through which the Marathon reservoir was reinforced by water transferred from the Mornos reservoir, a stepwise-mode DA showed that the discrimination of samples from the Mornos and the Marathon reservoirs was based mainly on Mg²⁺ and Cl⁻.

KEYWORDS: principal component analysis, factor analysis, cluster analysis, discriminant analysis, classifications trees, water quality.

REFERENCES: