Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Multivariate statistical analysis for the surface water quality of the Luan River, China*

Zhi-wei ZHAO†, Fu-yi CUI

(*State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150090, China*)

†E-mail: hit_zzw@163.com

Received Mar. 24, 2008; Revision accepted July 1, 2008; Crosschecked Oct. 29, 2008

Abstract: In order to analyze the characteristics of surface water resource quality for the reconstruction of old water treatment plant, multivariate statistical techniques such as cluster analysis and factor analysis were applied to the data of Yuqiao Reservoir—surface water resource of the Luan River, China. The results of cluster analysis demonstrate that the months of one year were divided into 3 groups and the characteristic of clusters was agreed with the seasonal characteristics in North China. Three factors were derived from the complicated set using factor analysis. Factor 1 included turbidity and chlorophyll, which seemed to be related to the anthropogenic activities; factor 2 included alkaline and hardness, which were related to the natural characteristic of surface water; and factor 3 included Cl and NO_2-N affected by mineral and agricultural activities. The sinusoidal shape of the score plots of the three factors shows that the temporal variations caused by natural and human factors are linked to seasonality.

Key words: Surface water quality, Cluster analysis, Factor analysis, Temporal variation of scores
doi:10.1631/jzus.A0850094 **Document code:** A **CLC number:** X522

INTRODUCTION

In North China, because of the excessive exploitation of groundwater, surface water resource is the helpless choice for some cities such as Tianjin, Harbin, etc. In recent ten years, the quality of surface water in North China has become worse and worse due to the frequent anthropogenic activities. However, all of water treatment plants in these areas were built in the 1970s or 1980s, failing to produce quality water. A stricter standard for drinking water quality made by the National Standardization Committee and Ministry of Public Health was effective in 2007 (Zhang *et al.*, 2007). Therefore, the reconstruction of these water treatment plants must be considered.

It is well known that surface water quality is largely determined both by the natural processes (such as precipitation rate, weathering processes, and soil erosion) and the anthropogenic influences (including urban, industrial and agricultural activities and exploitation of water resources) (Carpenter *et al.*, 1998; Jarvie *et al.*, 1998). The constant polluting sources of the basin are the municipal and industrial wastewater discharge constitutes (Singh *et al.*, 2004). On the other hand, the surface runoff is a seasonal phenomenon, largely affected by climate in the basin (Singh *et al.*, 2004). Seasonal variations in precipitation, surface runoff, ground water flow, and water interception and abstraction have a strong effect on river discharge and subsequently on the concentration of pollutants in river water (Vega *et al.*, 1998; Shrestha and Kazama, 2007).

Multivariate statistical methods have been widely applied to investigate environmental phenomena in recent years (Laaksoharju *et al.*, 1999; Anazawa *et al.*, 2003; Gűler and Thyne, 2004; Anazawa and Ohmori, 2005). They include cluster analysis, principle component analysis/factor analysis, time series analysis, self-organizing maps and classification and regression trees (CART) strategy, etc.,

* Project supported by the Hi-Tech Research and Development (863) Program of China (No. 2006AA06Z311), and Development Program for Outstanding Young Teachers in Harbin Institute of Technology (No. HITQNJS.2008.044), China

and are a powerful tool for deriving useful information from complicated data about water quality studies (Vega *et al*., 1998; Helena *et al.*, 2000; Bengraine and Marhaba, 2003; Liu *et al*., 2003; Simeonov *et al*., 2003; Simeonova *et al.*, 2003; Simeonova and Simeonov, 2006; Astel *et al.*, 2007; 2008). The multivariate statistical methods have an extensive application in characterization and evaluation of surface water quality and are useful for evaluation of temporal and spatial variations caused by natural and anthropogenic factors (Vega *et al.*, 1998; Reisenhofer, *et al*., 1998; Singh *et al.*, 2004; 2005).

In the present study, a large data matrix, which is obtained during a 10-year (1995~2004) monitoring program from Yuqiao Reservoir, the water source of Tianjin, was subjected to different multivariate statistical techniques to obtain the data of (1) the temporal similarities or dissimilarities, (2) the latent factors explaining the structure of the dataset, (3) the major influence of pollutants (natural and anthropogenic), and (4) the period trend of major contaminants. These data can be the basis for the design of new water treatment plants and improvement of old water treatment plants.

METHODS

Study area

The Luan River, the main water source of Tianjin and Tangshan, China, is a large system in North China. It flows through Hebei Province, the Inner Mongolia Autonomous Region and Liaoning Province, and its basin (115°30′ E~119°45′ E and 39°10′ N~42°40′ N) extends over an area of approximately 44 750 km^2 and has a total length of 877 km. Yuqiao Reservoir (117°32′ E, 40°02′ N), 21.6 m in mean depth, and 2060 km^2 in cover area, the unique water resource of Tianjin, is located in Ji Country, Tianjin, and its purpose is to enhance the water level of the Luan River and to supply water for Tianjin (Fig.1). The capacity of Yuqiao reservoir is 1.559×10^9 m^3. Moreover, the water supply capacity of Yuqiao Reservoir is 0.9×10^9 m^3 annually. The main economic activity of the upstream of Yuqiao Reservoir is agriculture. The geology of the basin is characterized by rock, thin soil, and severe soil erosion.

Parameters and analytical methods

The data obtained from Tianjin Water Bureau contain 19 water quality parameters, and 10 of them were selected for analysis because of their continuity in measurement. The selected parameters including temperature, turbidity, pH, alkalinity, chloride, ammunium nitrogen, nitrite nitrogen, chemical oxygen demand (COD) (manganese), total hardness and chlorophyll, and methods of analysis are shown in Table 1. The water quality data are shown in Table 2, with mean and deviation in order to monitor the distribution of data.

Fig.1 Sketch map of study area in the Luan River

Table 1 Water quality parameters and analytical methods used during 1995~2004 for surface waters of the Luan River, China

Parameter	Analytical method
WT (°C)	Mercury thermometer
Tu (NTU)	Turbidimeter
pH	pH-meter
Alk ($CaCO_3$ mg/L)	Titrimeter
Cl (mg/L)	Spectrophotometric
NH_4-N (mg/L)	Spectrophotometric
NO_2-N (mg/L)	Spectrophotometric
COD_{Mn} (mg/L)	Potassium permanganate
T-har ($CaCO_3$ mg/L)	Titrimeter
C-a (mg/L)	Spectrophotometric

* WT: temperature; Tu: turbidity; Alk: alkalinity; Cl: chloride; NH_4-N: ammonium nitrogen; NO_2-N: nitrite nitrogen; COD_{Mn}: chemical oxygen demand (Mn); T-har: total hardness; C-a: chlorophyl a

Table 2 Water quality characteristics of the Luan River, China

Month		WT (°C)	Tu (NTU)	pH	Alk ($CaCO_3$ mg/L)	Cl (mg/L)	NH_4-N (mg/L)	NO_2-N (mg/L)	COD_{Mn} (mg/L)	T-har ($CaCO_3$ mg/L)	C-a (mg/L)
Jan.	Range	0.06~6.50	1.0~31.9	7.11~8.60	112~252	20~148	0.04~1.43	0.10~0.87	2.4~6.8	126~354	0.15~24.40
	Mean	3.39	4.51	8.20	142.92	27.06	0.18	0.52	3.82	211.35	7.30
	S.D.	0.85	2.98	0.16	15.76	10.37	0.14	0.01	0.5	24.69	6.52
Feb.	Range	0.9~9.5	1.32~21.30	7.72~8.90	124~197	21~38	0.06~1.43	0.2~0.5	3.0~5.6	165~338	0.173~17.750
	Mean	5.34	3.83	8.30	148.18	27.81	0.18	0.20	4.08	213~69	6.45
	S.D.	1.59	1.91	0.18	13.53	3.54	0.11	0.01	0.48	20.57	4.39
Mar.	Range	4.3~14.5	1.74~40.70	7.08~8.90	122~162	20~40	0.03~0.53	0.1~0.5	2.5~5.7	160~286	2.68~17.30
	Mean	9.25	5.19	8.27	142.05	28.70	0.13	0.27	4.07	200.20	8.12
	S.D.	1.88	3.05	0.17	8.50	4.11	0.07	0.01	0.47	19.45	3.65
Apr.	Range	10~20.8	1.5~28.3	7.82~9.00	100~160	19~63	0.02~0.28	0.30~0.67	2.3~6.3	125~400	1.68~25.7
	Mean	15.13	7.01	8.38	137.60	27.25	0.08	0.02	3.82	193.74	7.89
	S.D.	2.14	3.13	0.14	7.88	5.14	0.03	0.01	0.53	21.13	21.71
May	Range	13.4~27.1	1.84~27.70	7.60~9.08	90~230	15~221	0.01~0.21	0.2~0.8	1.5~5.8	134~470	0.15~16.85
	Mean	20.98	7.26	8.36	133.89	28.98	0.09	0.50	3.74	186.13	6.75
	S.D.	2.36	3.53	0.17	11.99	13.35	0.04	0.01	0.51	23.28	3.26
June	Range	20.2~30.3	2.5~56.7	7.57~9.48	95~180	18~108	0.02~0.23	0.60~2.37	2.4~8.3	125~260	0.27~74.08
	Mean	25.11	12.22	8.19	122.44	28.57	0.09	0.87	4.22	172.37	18.44
	S.D.	1.74	7.06	0.27	13.45	7.63	0.04	0.01	0.85	18.17	10.26
July	Range	14.0~32.4	1.5~80.8	7.49~8.98	80~132	19~50	0.02~0.39	0.50~2.69	2.7~8.4	100~255	0.77~129.00
	Mean	27.86	20.63	8.27	111.22	30.55	0.09	1.32	4.68	155.09	29.96
	S.D.	2.03	14.03	0.31	11.07	6.31	0.06	0.01	1.00	22.12	20.31
Aug.	Range	16.8~31.9	2.3~150	7.01~17.60	76~162	18~93	0.02~1.60	0.600~1.677	2.8~9.8	90~245	8.0~198.3
	Mean	27.40	35.17	8.33	107.98	31.62	0.11	0.83	5.35	146.73	39.82
	S.D.	2.08	24.6	0.62	14.62	8.61	0.11	0.03	1.35	23.83	28.56
Sept.	Range	12.7~29.8	3.87~158	7.40~9.48	80~208	20~73	0.02~0.34	0.40~1.72	3.2~9.9	105~220	4.10~153.99
	Mean	23.87	39.98	8.40	108.28	30.77	0.10	0.67	5.79	145.86	44.78
	S.D.	2.61	29.59	0.38	15.66	6.07	0.06	0.02	1.40	19.72	24.71
Oct.	Range	7.1~23.9	4.5~69.8	7.65~9.40	86~204	14~174	0.01~1.00	0.30~0.78	2.2~7.0	86~314	1.46~77.05
	Mean	17.14	18.24	8.32	118.42	30.33	0.09	0.52	4.65	164.39	22.79
	S.D.	2.85	11.20	0.33	16.19	11.88	0.07	0.02	0.88	25.09	15.49
Nov.	Range	0~20.7	2.6~38.8	7.22~9.05	80~170	18~52	0.02~6.04	0.20~0.62	1.8~6.9	120~265	1.31~38.64
	Mean	10.83	10.74	8.26	119.81	26.60	0.17	0.47	4.16	174.01	13.16
	S.D.	3.46	6.31	0.27	15.50	3.38	0.44	0.01	0.70	21.76	6.82
Dec.	Range	0.9~8.1	1.34~12.40	7.40~8.79	112~164	15~88	0.02~0.70	0.1~0.9	1.9~5.5	150~252	2.32~29.30
	Mean	4.00	4.63	8.14	132.06	26.28	0.1	0.42	3.82	194.72	11.87
	S.D.	1.11	1.33	0.16	7.23	4.60	0.07	0.01	0.54	17.05	4.12

* S.D.: standard deviation; WT: temperature; Tu: turbidity; Alk: alkalinity; Cl: chloride; NH_4-N: ammonium nitrogen; NO_2-N: nitrite nitrogen; COD_{Mn}: chemical oxygen demand (Mn); T-har: total hardness; C-a: chlorophyl a

Data treatment

Before the multivariate analysis, *z*-scale transformation was applied to the raw water quality dataset in order to avoid miss classification because of wide differences in data dimensionality (Liu *et al.*, 2003), to eliminate the influence of different units of measurements, and to render the data dimensionless.

The data matrix was normalized for cluster analysis. Kaiser-Meyer-Olkin (KMO) and Bartlett's test were performed to examine the suitability of the data for principal component analysis/factor analysis. High value (close to 1) of KMO generally indicates that principal component analysis/factor analysis may be useful, which is the case in this study: KMO=0.83. Bartlett's test of sphericity indicates whether correlation matrix is an identity matrix, which would indicate that variables are unrelated (Shrestha and Kazama, 2007). The significance level was 0.07 in this study, demonstrating significant relationships among variables.

Cluster analysis

Cluster analysis, which is a wide range of techniques for exploratory data analysis, groups variables into clusters on the basis of similarities (or dissimilarities) such that each cluster represents a specific process. The class characteristics are not known in advance, but may be determined from the data analysis (Singh *et al.*, 2004). The results obtained were justified according to their values in interpreting the data and indicating patterns (Johnson and Wichern, 1992; Adams, 1998). In this paper, the hierarchical cluster analysis (HCA), which was performed on the normalized data matrix by the utilization of the Ward's method and used the squared Euclidean distances as a measure of similarity that reported as D_{link}/D_{max}, was applied to the variables using Minitab 15 (Minitab Inc.) to group the data in temporal pattern.

Factor analysis

Factor analysis (FA), which includes principal components analysis (PCA), enables us to explain the relationships among numerous important variables with a smaller set of independent variables well. PCA has the advantage of the transformation of the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables.

Further rotation of the axis defined by PCA produces new groups of variables called varifactors (VFs). This procedure is known as FA. This technique can not only achieve the aim of data reduction and data grouping, but also determine the relative importance of those processes affecting the environmental quality together with their spatial and temporal variations, and the ecological significance in the new variables. Therefore, FA/PCA is particularly valuable when a chemical, physical, or biological interpretation of the data grouped in VFs is possible. The use of FA/PCA to water quality assessment has increased in recent years, mainly due to the need to obtain appreciable data reduction for analysis and decision (Chapman, 1992; Kucuksezgin, 1996; Chiacchio *et al.*, 1997; Vega *et al.*, 1998; Morales *et al.*, 1999; Helena *et al.*, 2000).

RESULTS AND DISCUSSION

Cluster analysis

Cluster analysis using Ward's method was applied to the average data of every month of one year using squared Euclidian distance as similarity measure. The results were illustrated in Fig.2, where three main groups at D_{link}/D_{max}<0.4 are visible. Group 1 included January, December, February and March, group 2 included April, May, October, and November, and group 3 included June, July, August and September. Clearly, the characteristics of clusters agreed with the seasonal characteristics in North China that the period of April and May are spring; October and November are fall; December, January, February, and March are winter; and June, July, August and September are summer.

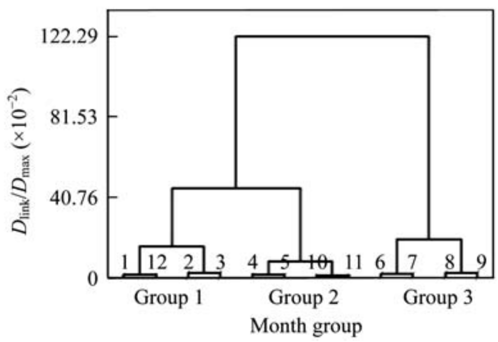

Fig.2 Dendrogram for the relationship among every month of one year about the Luan River, China

Factor analysis

FA was applied to the standardized full dataset. According to the combination of criteria for factor selection eigenvalues higher than 1.0, four most significant factors were taken. Table 3 shows the eigenvalues representing factors, the factor loading (>0.4) which was classified as "strong", "moderate", and "week" corresponding to absolute loading values of >0.75, 0.75~0.50, and <0.4, and the proportion of total variance explained by the factors for the raw data (Unmesh *et al*., 2006). The three factors explained 91.938% of the total variance of the data. Factors 1 and 3 show similar results but with some notable differences, and factor 2 is common.

Table 3 R-mode varimax rotated factor analysis of parameters

Parameter	Analytical result		
	VF 1	VF 2	VF 3
WT	−0.001	−0.094	0.050
pH	0.001	0.001	0.009
Cl	0.021	−0.034	0.636
NH_4-N	0.004	0.005	0.041
T-alk	0.142	0.908	0.373
Tu	0.567	0.137	0.145
NO_2-N	0.147	0.218	0.773
COD_{Mn}	0.002	0.001	0.001
T-har	0.279	−0.416	−0.216
Chlorophyl a	0.646	0.331	0.119
Eigenvalue	8.177	6.316	3.314
Total variance (%)	73.351	14.263	4.324
Cumulative variance (%)	73.351	87.614	91.938

VF: varifactor; WT: temperature; Cl: chloride; NH_4-N: ammonium nitrogen; T-alk: total alkalinity; Tu: turbidity; NO_2-N: nitrite nitrogen; COD_{Mn}: chemical oxygen demand (Mn); T-har: total hardness; C-a: chlorophyl a

Factor 1 accounting for 73.351% of the total variance, which was moderate positively loaded with turbidity and chlorophyll a (Table 3), seemed to be related to eutrophication and soil erosion, both of which are the results of anthropogenic activities. The time series factor scores diagram (Fig.3a) shows a clear sinusoidal pattern, which is indicative of seasonal variation of factor 1. It is clear that the extreme absolute score values of factor 1 present July, August, September and October, the period with more human activities. The results demonstrate that turbidity and chlorophyll a were related to anthropogenic activities.

The strong positively loaded was total alkalinity and the moderate negatively loaded was total hardness in factor 2, which accounted for 14.263% of the total variance (Table 3) and represented the mineral-related hydrochemistry of the surface water. The temporal variation of scores of factor 2 demonstrated that the pollution period of alkaline and hardness was in winter (Fig.3b), in which human activities were reduced, so as the water contamination.

Factor 3 accounted for 4.324% of the total variation, which is relatively small comparison with other factors. Cl and NO_2-N were found to have a moderate factor score. Cl was related to mineral and the contamination of NO_2-N was in relation to the origin in river runoff from the agricultural field along with waste disposal activity. Fig.3c shows that the principle period of contamination of Cl and NO_2-N was summer and autumn, which were the chief period of agricultural activities. Therefore, the contamination resources during this period originate from agricultural pollution to surface water near the agricultural field, runoff of the agricultural field and solid waste disposal activities of cities and towns (Unmesh *et al*., 2006).

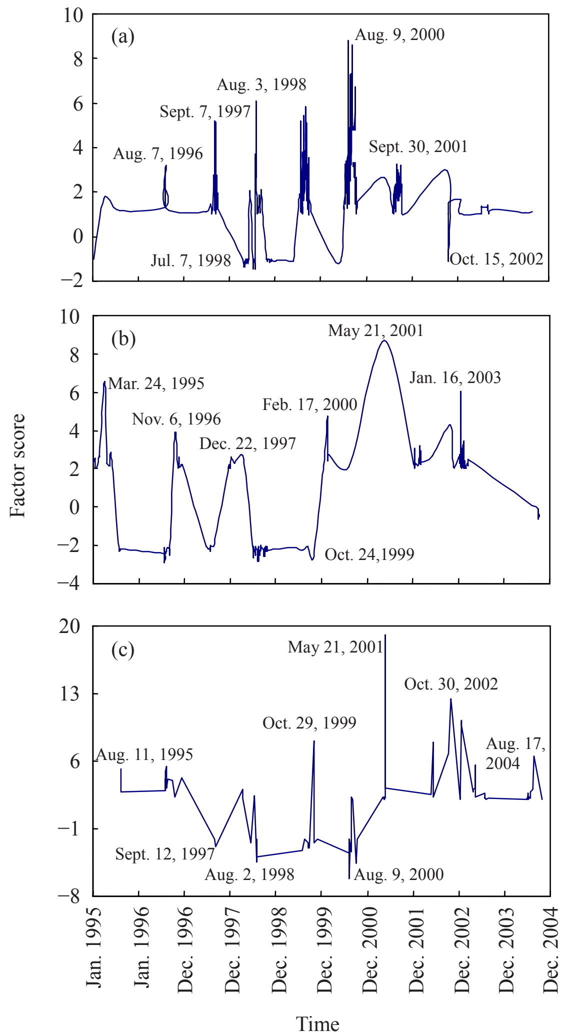

Fig.3 Temporal variation of scores of (a) factor 1, (b) factor 2 and (c) factor 3

CONCLUSION

HCA was applied to the average data of every month of a year. The results demonstrate that, to evaluate the water quality of the Luan River, the months of a year could be divided into 3 groups. Group 1 included January, December, February, and March, group 2 included April, May, October, and November, and group 3 included June, July, August, and September. The characteristics of clusters of the Luan River water quality agreed with the seasonal characteristics of human activities in North China. It implied that HCA was an effective method for the reliable classification of surface water.

FA of data illustrated that turbidity, chlorophyll a, alkaline, hardness, Cl and NO_2-N were the chief contaminants in surface water of Yuqiao Reservoir. The figures of temporal variation of scores of every factor demonstrate the contamination period of turbidity and chlorophyll a present at July, August, September and October of a year because of excessive anthropogenic activities. Alkaline and hardness implied the mineral related hydrochemistry of the surface water and the contamination period was winter, in which human activities were limited. The contaminant of Cl was related to mineral and NO_2-N was related to agricultural activities. And, the contamination period of Cl and NO_2-N was summer and autumn, which is the main period of agricultural activity. It is evident that FA is useful for the analysis of complex data.

References

Adams, M.J., 1998. The Principles of Multivariate Data Analysis. *In*: Ashurst, P.R., Dennis, M.J. (Eds.), Analytical Methods of Food Authentication. Blackie Academic & Professional, London, UK, p.350.

Anazawa, K., Ohmori, H., 2005. The hydrochemistry of surface waters in Andesitic Volcanic area, Norikura Volcano, central Japan. *Chemosphere*, **59**(5):605-615. [doi:10.1016/j.chemosphere.2004.10.018]

Anazawa, K., Ohmori, H., Tomiyasu, T., Sakamoto, H., 2003. Hydrochemistry at a volcanic summit area, Norikura, central Japan. *Geochimica et Cosmochimica Acta*, **67**(18S):17.

Astel, A., Tsakovski, S., Barbieri, P., Simeonov, V., 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*, **41**(19):4566-4578. [doi:10.1016/j.watres.2007.06.030]

Astel, A., Tsakovski, S., Simeonov, V., Reisenhofer, E., Piselli, S., Barbieri, P., 2008. Multivariate classification and modeling in surface water pollution estimation. *Analytical and Bioanalytical Chemistry*, **390**(5):1283-1292. [doi:10.1007/s00216-007-1700-6]

Bengraine, K., Marhaba, T.F., 2003. Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials*, **100**(1-3):179-195. [doi:10.1016/S0304-3894(03)00104-3]

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, **8**(3):559-568. [doi:10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2]

Chapman, D., 1992. Water Quality Assessment—A Guide to Use of Biota, Sediments and Water in Environmental Monitoring (2nd Ed.). Chapman on Behalf of UNESCO, WHO and UNEP. Chapman & Hall, London, p.585.

Chiacchio, U., Librando, V., Magazzu, G., 1997. Monitoring studies of Augusta Bay marine waters. *Environmental Monitoring and Assessment*, **44**(1/3):383-390. [doi:10.1023/A:1005756001153]

Gűler, C., Thyne, G.D., 2004. Hydrologic and geologic factors controlling surface and groundwater chemistry in Indian wells—Owens Valley area, southeastern California, USA. *Journal of Hydrology*, **285**(1-4):177-198. [doi:10.1016/j.jhydrol.2003.08.019]

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J.M., Fernández, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, **34**(3):807-816. [doi:10.1016/S0043-1354(99)00225-0]

Jarvie, H.P., Whitton, B.A., Neal, C., 1998. Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. *Science of the Total Environment*, **210-211**:79-109. [doi:10.1016/S0048-9697(98)00109-0]

Johnson, R.A., Wichern, D.W., 1992. Applied Multivariate Statistical Analysis (3rd Ed.). Prentice-Hall International, Englewood Cliffs, New Jersey, USA, p.642.

Kucuksezgin, F., 1996. Multivariate analysis of water quality parameters in Izmir Bay, Eastern Aegean. *Toxicological and Environmental Chemistry*, **55**:135-144.

Laaksoharju, M., Gurban, I., Skarman, C., Skarman, E., 1999. Multivariate mixing and mass balance (M3) calculations, a new tool for decoding hydrogeochemical information. *Applied Geochemistry*, **14**(7):861-871. [doi:10.1016/S0883-2927(99)00024-4]

Liu, C.W., Lin, K.H., Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *The Science of the Total Environment*, **313**(1-3):77-89. [doi:10.1016/S0048-9697(02)00683-6]

Morales, M.M., Marti, P., Liopis, A., Campos, L., Sagrado, S., 1999. An environmental study by factor analysis of surface sea waters in the gulf of Valencia (Western Mediterranean). *Analytica Chimica Acta*, **394**(1):109-117. [doi:10.1016/S0003-2670(99)00198-1]

Reisenhofera, E., Adamia, G., Barbieria, P., 1998. Using chemical and physical parameters to define the quality of karstic freshwaters (Timavo River, North-Eastern Italy): A chemometric approach. *Water Research*, **32**(4):1193-1203.

Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. *Environmental Modelling & Software*, **22**(4):464-475.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., Kouimtzis, T., 2003. Assessment of the surface water quality in Northern Greece. *Water Research*, **37**(17):4119-4124. [doi:10.1016/S0043-1354(03)00398-1]

Simeonova, P., Simeonov, V., 2006. Chemometrics to evaluate the quality of water sources for human consumption. *Microchimica Acta*, **156**(3-4):315-320. [doi:10.1007/s00604-006-0643-5]

Simeonova, P., Simeonov, V., Andreev, G., 2003. Water quality study of the Struma River Basin, Bulgaria. *Central European Journal of Chemistry*, **2**:121-136.

Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Research*, **38**(18):3980-3992. [doi:10.1016/j.watres.2004.06.011]

Singh, K.P., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: a case study. *Analytica Chimica Acta*, **538**(1-2):355-374. [doi:10.1016/j.aca.2005.02.006]

Unmesh, C.P., Sanjay, K.S., Prasant, R., Binod, B.N., Dinabandhu, B., 2006. Application of factor and cluster analysis for characterization of river and estuarine water systems—A case study: Mahanadi River (India). *Journal of Hydrology*, **331**(3-4):434-445.

Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water byex ploratory data analysis. *Water Research*, **32**(12):3581-3592. [doi:10.1016/S0043-1354(98)00138-9]

Zhang, L., Chen, C.J., Chen, Y.Y., 2007. Sanitary standard for drinking water of China. *Chinese Journal of Public Health*, **23**(11):1281-1282 (in Chinese).