Proceeding of Engineering Technology International Conference (ETIC 2015)
10-11 August 2015, Bali, Indonesia

Water Quality Assessment of Muar River Using Environmetric Techniques and Artificial Neural Networks

Putri Shazlia Rosman

Faculty of Engineering Technology
Universiti Malaysia Pahang, Tun Razak Highway, 26300 Kuantan Pahang, MALAYSIA

***Corresponding author email, putrishazlia@ump.edu.my**

ABSTRACT

The pollution discharge has influence the chemical composition of Muar River where studied was carried out using the Environmetric Techniques and the Artificial Neural Networks (ANNs) model. Environmetric method, the hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA) and the factor analysis (FA) to study the spatial variations of water quality variables and to determine the origin sources of pollution. ANNs model was used to predict linear relationship between water quality variables, the most significant variables that influence Muar River as well as sources of apportionment pollution. HACA observed three spatial clusters were formed. DA managed to discriminate 16 and 19 water quality variables thru forward and backward stepwise. Eight principal components were found responsible for the data structure and 67.7% of the total variance of the data set in PCA/FA analysis. ANNs analysis, strong relationship correlation was observed between salinity, conductivity, DS, TS, Cl, Ca, K, Mg and Na ($r = 0.954$ to 0.997), moderate relationship observed between COD and *E.coli* ($r = 0.449$) and Cd and Pb ($r = 0.492$) and others variables have no significant correlation. pH was the most significant variables (51.6%) and Fe was less significant variables (-0.52%). The major sources of pollution of the river were due to natural degradation / natural process that affecting the pH value of the river. Other pollution contribution was from anthropogenic sources such as agricultural runoff, industrial discharge, domestic waste, natural erosion, livestock farming and present of nitrogenous species. The ANNs showed better prediction in identified most significant variable compare to Environmetric techniques. ANNs is an effective tool in decision making and problem solving for local/global environmental issues.

KEYWORDS

Environmetric techniques, artificial neural networks, hierarchical agglomerative cluster analysis, discriminant analysis, principal component analysis

INTRODUCTION

Rivers has become an important to people in daily use for drinking, bathing, transportation, recreational activities, agriculture purpose etc. and to support life of many endangered species animals and aquatic. Muar River has riches with natural river ecosystem where the local community people can gain their income by providing boat ride for the tourist to see the beauty of the riverside of Muar River.

According to Malaysia Environmental Quality Report (2009), about 521 rivers were categorized as polluted rivers. From polluted rivers, about 152 rivers were polluted by high BOD level, 183 rivers were polluted by high ammoniacal Nitrogen concentration and 186 rivers were polluted by Suspended solid concentration. The present status of Muar River Basin for overall Water Quality Index (WQI) is slightly polluted that is 81 based on the Water Quality Monitoring Program that has been carried out by Department of Environment with 17 stations in year 2002 until 2008 (DOE, 2009)

The problem that occurs in Muar River is from the agriculture activities. The agricultural land uses remain dominant feature of the entire basin where the water quality in the Muar River basin has become considerably poor and polluted due to diffuse pollution originating from agriculture activities. Other than that water pollution in Muar River derived from numerous sources of pollution that may adversely affect water quality. These major sources are from agriculture wastewater and minor sources pollutions derived from domestic and industrial wastewater.

However, Muar River Basin was categories as slight polluted rivers in Malaysia Environmental Quality Report 2009 by Department of Environment. About 11 rivers stated in Muar River basins were under class II and class III. The rivers are Gemas River, Gemencheh River, Labis River, Merbudu River, Merlimau River, Muar River, Palong River, Senarut River, Serom River, Simpang Loi River and Tenang River.

GENERAL DESCRIPTION OF ENVIRONMETRIC TEACHNIQUE AND ANN MODEL

Environmetric is also known as chemometrics that uses multivariate statistical modeling and data treatment (Simeonov et al. 2000). Environmetric methods is used in exploratory data analysis tools for classification (Brodnjak-Voncina et al.2002; Kowalkowski et al. 2006) of sampling stations or samples identification of pollution sources (Massart et al.1997, Vega et al. 1998; Shresthaa and Kazama 2007) and drawing information from environmental data. Environmetric method has become important tools to evaluate and reveal complex relationships in environmental science area through variety of environmental technique application. Application of environmetric techniques, viz, hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA) / factor analysis (FA) were employed to identify spatial variations of most significant water quality variables and to determine the origin of pollution sources at the Muar River Basin. Discriminant analysis is operates on raw data by using standard, backward and forward stepwise modes in order to construct a discriminant function (DF). Discriminant function for each group was build up from DA technique. DFs are calculated using equation:

$$\mathrm{f}\,(G_i) = k_i + \sum_{j=1}^{n} w_{ij}\, P_{ij} \tag{1}$$

The principle component (PC) can be expressed as

$$z_{ij}=a_{i1}x_{1j}+a_{i2}x_{2i}+\ldots+a_{im}x_{mj} \tag{2}$$

The basic concept of FA is expressed as

$$z_{ij}=a_{f1}f_{1i}+a_{f2}f_{21}+\ldots+a_{fm}f_{mi}+e_{fi} \tag{3}$$

ANNs has become common use in many researchers study. ANNs are frequently to predict various ecological processes and phenomenon related to water resources (Talib *et al*., 2009). The basic structure of ANN model is usually comprised of three distinctive layers, the input layer, where the data are introduce to the model and computation of the weighted sum of the input is preformed, the hidden layer or layers, where data are processed, and the output layer, where the results of ANN are produced (Singh et al., 2009).

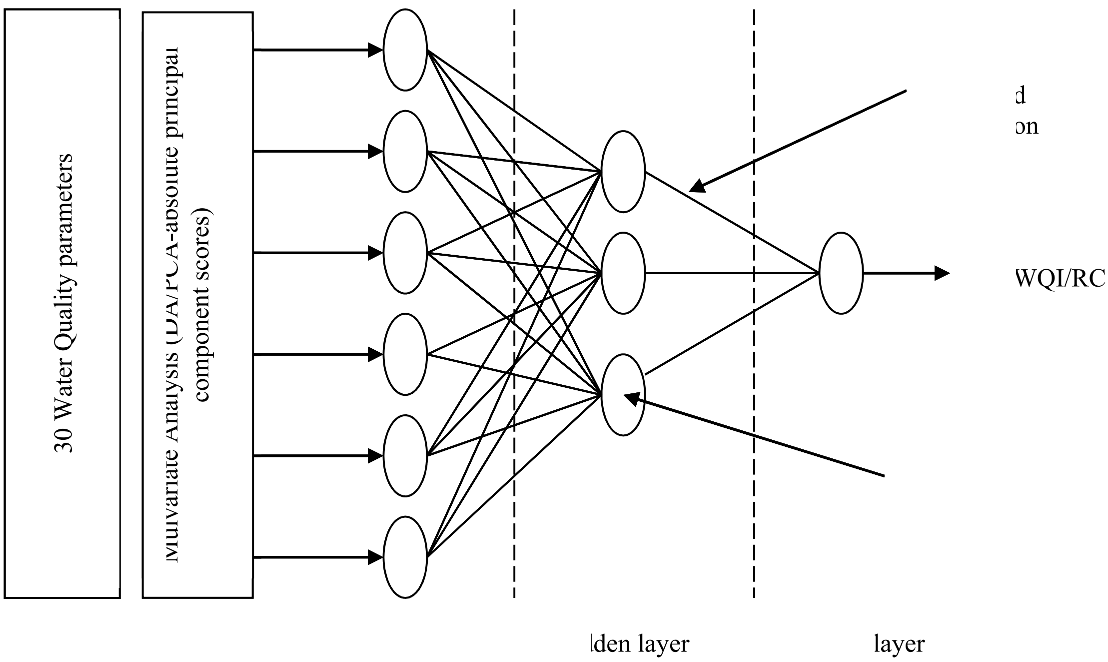

Figure 1. Examples of the ANN network configuration of 30 water quality variables

STUDY AREA AND DATA ACQUISITION

The Muar River is the major water way in State of Johor. The total of catchment area approximately 6,149 km^2 which consists of four states, State of Johor with 3,641 km^2, State of Negeri Sembilan with 2,427 km^2, state of Pahang with 122 km^2 and state of Melaka with 8 km^2. The length of main trunk of Muar River is about 288.20 km with the width is estimated to be 1.6 km. The Muar River has two major tributaries with the principal ones being the Gemencheh River and Segamat River shown in Figure 5. In 2003, about 85% of the total land area within the basin is devoted to agricultural use and about 15% of total land areas are devoted to forests conservation.

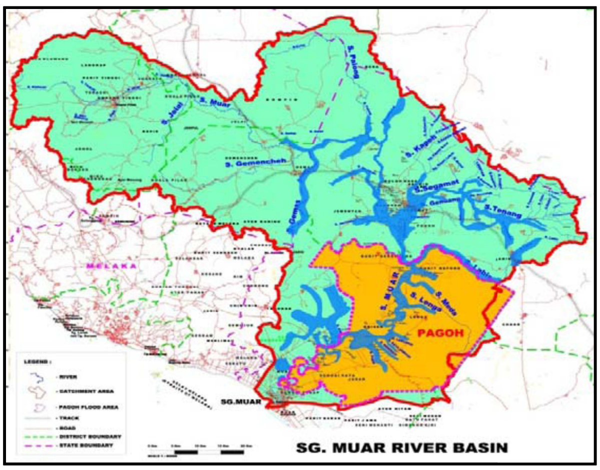

Figure 2. The tributaries of Muar River Basin.

The water stations (WS06, WS07, WS08 and WS11) located at Muar River in Negeri Sembilan, while water stations (WS09, WS12, WS16, WS21, WS23, WS24, WS25, WS26, WS27, WS29 and WS32) located at Muar River in Johor.

DOE Station no.	Study code in map	Study Code	Grid Reference		Location
2525626	S1	WS06	02 38 726	102 35 480	Muar River, Negeri Sembilan
2625612	S2	WS07	02 38 353	102 32 362	Muar River, Negeri Sembilan
2627645	S3	WS08	02 42 330	102 30 134	Muar River, Negeri Sembilan
2722643	S4	WS11	02 45 877	102 16 004	Muar River, Negeri Sembilan
2427610	S5	WS09	02 29 693	102 45 162	Muar River, Johor
2527611	S6	WS12	2 33 321	102 45 837	Muar River, Johor
2627613	S7	WS16	02 38 720	102 45 956	Muar River, Johor
2627631	S8	WS21	02 40 220	102 44 085	Muar River, Johor
2328609	S9	WS23	02 20 814	102 49 773	Muar River, Johor
2228608	S10	WS24	02 16. 594	102 48. 718	Muar River, Johor
2228607	S11	WS25	02 14 838	102 48 591	Muar River, Johor
2227606	S12	WS26	02 12 137	102 46 558	Muar River, Johor
2225605	S13	WS27	02 10 381	102 42 752	Muar River, Johor
2126604	S14	WS29	02 07 765	102 38 739	Muar River, Johor
2126603	S15	WS32	02 08 992	102 35 987	Muar River, Johor

Table 1. Water quality stations by DOE at the study area.

The 30 water quality variables involves are Dissolve oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solid (SS), Ammoniacal- Nitrogen (NH_3-N), Dissolve Solid (DS), nitrate (NO_3), chlorine (Cl), total solid (TS), zinc (Zn), potassium (K), magnesium (Mg), phosphate (PO_4^-), temperature (T), pH, calcium (Ca), sodium (Na), iron (Fe), salinity (Sal), conductivity (COND), turbidity (TUR), oil and grease (OG), cadmium (Cd), lead (Pb), copper (Cr), mercury (Hg), arsenic (As), MBA/BAS, Escherichia coli *(E.coli)*, and coliform.

RESULT AND DISCUSSION

Correlation matrix between variables

By measured simple correlation coefficient (pearson) matrix (r), the degree of a linear association between variables indicates suspended solid and turbidity are strongly and significantly correlated each other (r = 0.820). Strongly relationship correlations also can be observed between salinity, conductivity, dissolve solid, total solid, chloride, calcium, potassium, magnesium and sodium (r = 0.954 to 0.997) which responsible for water mineralization. Moderate relationship between COD and *E.coli* (r = 0.449) and between Cd and Pb (r = 0.492) were indicates significant anthropogenic contribution as sources of contaminants. While DO, BOD, pH, NH_3, temperature, NO_3, PO_4, As, Cr, Zn, Fe, oil and grease, MBAS and coliform showed no significant correlation with any other variables.

HACA: Classification of sampling station on historical water quality data

The result HACA analysis shown in Figure 3 where sampling stations were places in three cluster. The similarities of natural background and characteristics of these stations causes the clustering procedure generated 3 cluster/group in a very convincing ways. Cluster 1 (stations WS06, WS07, WS08 and WS11) correspond to regions of low pollution sources (LPS), cluster 2 (stations WS16, WS21) represents the moderate pollution sources (MPS) and cluster 3 (WS09, WS12, WS23, WS24, WS25, WS26, WS27, WS29 and WS32) represents the high pollutions sources (HPS) from sub basin Muar River.

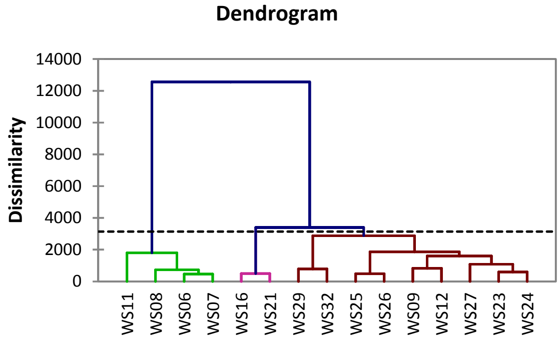

Figure 3. Dendogram showing different cluster of sampling sites located at Muar River Basin on water quality parameters.

Spatial variations of river water quality

Water quality parameters were treated as independent variables while the groups (C1, C2 and C3) were treated as dependent variables. The DA method carried out was via standard, backward stepwise and forward stepwise. The accuracy of spatial classification using standard, backward stepwise and forward stepwise mode DFA were 80.67% (29 discriminant variables), 80.41% (19 discriminant variables) and 80.41% (16 discriminant variables), respectively (Table 3).

Data structure determination

Eight PCs were obtained in this study with eigenvalues larger than 1 where almost 69.7% of the total variance in the data set. Among the eight VFs coefficient that having strong correlation were VF1 accounts for 30.3% of the total variance, shows strong positive loadings on conductivity, salinity, dissolve solid, total solid, chloride and mineral salt such as calcium, potassium, magnesium, and sodium.

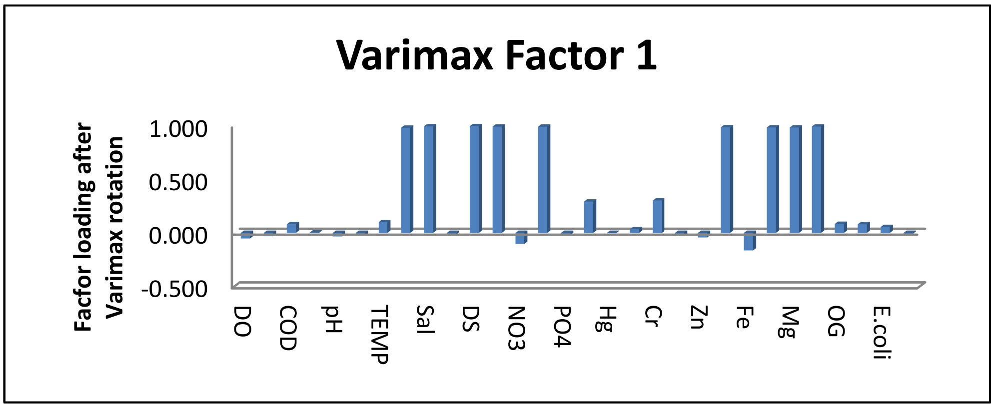

Figure 4. Graph of factor loading after Varimax rotation of principal components analysis / Factor analysis for Varimax Factor 1 on 30 water parameter

VF2 explain about 8.6% of the total variance, show strong positive trace metal such as cadmium. VF3 explaining 7.9% of the total variance has strong loading of suspended solid and turbidity while VF4 explains about 6.3% of the total variance and has strong positive loading on pH. The VF5 explain about 4.6% of the total variance, show strong positive loading of COD and *E.coli*. VF6 and VF7 were explaining 4.2% and 4.0% of variance, respectively, and have a strong loading of ammoniacal nitrogen and nitrate. However, VF8 shows weak factor loading of Ferum and coliform in Muar River.

Artificial Neural Network in prediction of water quality samples.

30 parameters water quality from 776 stations were trained, tested and validated by using ANN on spatial distribution

		R^2	Different (R^2AP)- R^2	Contribution %
All parameter		0.901848		
Conductivity, Salinity, DS, TS, Cl, Ca, K, Mg, Na	D1	0.894891	0.006957	2.493
Cd	D2	0.898588	0.003260	1.168
SS,TUR	D3	0.858842	0.043006	15.409
pH	D4	0.757977	0.143871	51.547
COD, *E.coli*	D5	0.890012	0.011836	4.241
COD, NH_3	D6	0.848334	0.053514	19.173
NO_3	D7	0.883738	0.018110	6.489
Fe	D8	0.903298	-0.001450	-0.520
Total			0.279104	

Table 2: Contribution percentage sources identification from ANNs analysis

The result showed pH with 51.55% is the major contribution parameter into Muar River followed by 19.17% of COD and Ammoniacal- Nitrogen (NH_3-N). About 15.41% of Suspended Solid and Turbidity, 6.49% of nitrate, 4.24% of COD and *Escherichia coli* and parameter such as conductivity, salinity, dissolved solids, total solid, chloride, calcium, potassium, magnesium an sodium are contribute about 2.49% in the river. The cadmium parameter contributes about 1.17% into the river. However, we found that iron (Fe) contributes about -0.52% meaning the present amount of ion in the river was much lower from all parameter above.

CONCLUSION

The HACA could helps to reduce the number of sampling station, associated time and cost. DA gives encouraging results for spatial variations in discriminating the fifteen monitoring station with nineteen and sixteen discriminant variables assigning 80% cases correctly using backward stepwise and forward stepwise modes. The PCA/FA results in eight variables responsible for major variation in along Muar River. The main sources of variations come from agriculture farming, livestock waste, natural erosion, excessive of fertilizer, domestic/ commercial waste and the present of nitrogenous species in the river. However, ANNs models were developed to predicting the high and low contributions of variables to the river. pH was identified the most contribution variables into Muar River with the highest percentage among other variables. However, pH has no significant relationship to other variables as result of analysis ANNs. High pH can be determine due to high amount of nitrogenous species in the river where these species. Compare to Environmetric techniques, ANNs model gives better result interpretation of the variables.

ACKNOWLEDGE

The author sincerely thank to The Department of Environment Malaysia for providing database of water quality Muar River and assistance provided by Faculty of Environmental Studies, Universiti Putra Malaysia.

REFERENCE

1. Adam, M.J. 1998. The principle of multivariate data analysis. In P.R. Ashurst & M. J. Dennis (Eds.), *Analytical methods of food authentication* (p.350). London: Blackie Academic & Professional.

2. Bowers, J.A., Shedrow, C.B., 2000. Prediction stream water quality using artificial neural networks. WRSC-MS-2000-00112. Savannah River Site, Aiken, SC 29808, US Department of Energy.

3. Brodnjak-Voncina, D., Dobenik, D., Novic. M., & Zupan, J. 2002. Chemometrics characterization of the quality of river water. *Analytica Chimica Acta,* 462, 87-100. doi:10.1016/S0003-2670(02)00298-2.

4. Buck, O., Niyogi, D. K., & Townsend, C.R. 2003. Scale-dependence of landuse effects on water quality of streams in agricultural catchments. *Environmental Pollution, 130*. 287-299. doi: 10.1016/j.envpol.2003.10.018

5. Department of Environment Malaysia (DOE) .2009. *Malaysia Environmental Quality Report, 2009*. Putrajaya: Ministry resources and Environment Malaysia.

6. Diamantopoulou, M.J., 2005. Tree-bole volume estimation on standing pine trees using cascade correlation artificial neural network models. *Agricultural Engineering International: the CIGR Ejournal. Manuscript IT* 2006.

7. Diamant.B.Z., 1978.Preventive control of water pollution in developing countries. Water pollution Control in Developing Countries, Bangkok Thailand, pp. 10

8. Dolling. O.R.,and E.A.Varas, "Artificial neural networks for streamflow prediction", *Journal of Hydraulic Research*. 40(5), 2002, pp.547-554.

9. Garcia-Bartual. R., "Short term river flood forecasting with neural networks", *Proceedings of iEMSs 2002*, Switzerland, 2002.*World Applied Sciences Journal* 13 (4): 829-836, 2011

10. Hafizan, J., Sharifuddin, M.Z., Mohd Ekhwan, T., Mazlin, M., Hasfalina, C.M., 2004. Application of artificial neural network models for predicting water quality index. Jurnal Kejuruteraan Awam. 16(2): 42-55.

11. Hafizan, J., Sharifuddin, M. Z., Kamil, Y.M., Tengku Hanidza, T.I., Mohd Armi, A.S., Ekhwan, T.M., and Mokhtar, M. 2010. Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. Environ Monit Assess. Doi10.1007/s10661-010-1411-x.

12. Huang. W., C. Murray, N.Kraus, and J. Rosati, "Development of a regional artificial neural network prediction for coastal water level prediction". *Ocean Engineering* vol.30, pp: 2275-2295

13. Kim, J.O., & Mueller, C.W. 1987. Introduction to factor analysis : What it is and how to do it. Quantitative applications in the social sciences series. Newbury Park: Sage University Press.

14. Kowalkowski, T., Zbytniewski, R., Szpejna, J., & Buszewski, B. 2006. Application of chemometrics in river classification. *Water Research*, 40, 744. doi:10.1016/j.watres.2005.11.042.

15. Liu, C. W., Lin, K. H., & Kuo, Y. M. 2003. Applications of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. The Science of the Total Environment, 313, 77-89. Doi:10.1016/S0048-9697(02)00683-6.

16. Massart D. L. and Kaufman L. 1983. *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*. Wiley, New York.

17. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1997). *Handbook of chemometrics and qualimetrics : Data handling in science and technology* (Parts A and B, Vols 20A and 20B). Elsevier:Amsterdam.

18. Patrick. A.R., W.G. Collins, P.E. Tissot, A. Drikitis, J. Stearns, P.R. Michaud, and D.T. Cox, "Use of the NCEP MesoEta data in a water level prediction artificial neural network", in the *Proceedings of 19th Conf. on weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, 2002.

19. Pradhan, U.K., Shirodkar, P.V. and Sahu, B. K., 2009. Physico-Chemical Characteristics of the Coastal Water Off Devi Estuary, Orissa and evaluation of its Seasonal Changes Using Chemometric Techniques. *Current Science*, Vol 96, No.9

20. Reghunath. R., Murthy, S.T. R., & Raghavan, B. R. 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. *Water Research*, 36. 2437-2442. doi:10.1016/S0043-1354(01)00490-0.

21. Samantray, P., Mishra, B.K., Panda, C.R., and Rout, S.P., 2009. Assessment of Water Quality Index in Mahanadi and Atharabanki Rivers and Taldanda Canal in Paradip Area, India. *J Hum Ecol*, 26(3) : 153-161 (2009).

22. Senthil. D. K., Satheeshkumar.P. and Gopalakrishnan. P. 2011. Ground Water Quality Assessment in Paper Mill Effluent Irrigated Area - Using Multivariate Statistical Analysis. *World Applied Sciences Journal* 13 (4): 829-836 .ISSN 1818-4952

23. Shrestha, S., & Kazama, F. 2007. Assessment of surface water quality using multivariate statistical technique: A case study of the Fuji river basin, Japan. Environmental *Modeling & Software*, 22, 464-475. doi:10.1016/j.envsoft.2006.02.001.

24. Simeonov, V., Stefanov, S., & Tsakovski, S. 2000. Environmental treatment of water quality survey data from Yantra River, Bulgaria. *Mikrochimica Acta*, 134, 15-21. doi:10.1007/S006040070047.

25. Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, Th.:2003, 'Assessment of the surface water quality in Northern Greece', *Wat. Res*. 37, 4119-4124.

26. Singh, K.P., Malik, A. and Singh, V. K., 2005. Chemometric analysis of hydro-chemical data of an alluvial river- a case study. *Water, Air and Soil Pollution*, 170:383-404. doi:10.1007/s11270-00509010-0

27. Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality-A case study. *Ecol. Model*. 220, 888-895.

28. Sobri, H., Amir Hashim, M.K, and Nor Irwan, A.N., 2002. Rainfall-runoff modeling using artificial neural network. *Proceedings of 2nd World Engineering Congress. Kuching*.

29. Sudheer, K.P., Nayak, P.C, and Ramasastri, K.S., 2003. Improving peak flow estimates in Artificial Neural Network river flow models. Hydrological Processess., 17: 667-687

30. Talib, A., Abu Hasan, Y, and Abdul Rahman N.N. 2009. Predicting Biochemical Oxygen Demand As Indicator Of River Pollution Using Artificial Neural Networks. *18th World IMACS/MODSIM Congress, Cairns, Australia* 13-17 July 2009.

31. Thirumalaiah. K., and M.C. Deo, "Hydrological forecasting using artificial neural network", *Journal of Hydrologic Engineering*, 5(2), 2000, pp. 180-189

32. Tokar. A.S., and P.A. Johnson, "Rainfall-runoff modeling using Artificial neural network", *Journal of Hydrology Engineering*, 4(3), 1999, pp. 223-239

33. Vega. M., Pardo, R., Barrado, E., & Deban, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32,3581-3592. doi:10.1016/S00431354(98)00138-9.

34. Wright, N.G., and M.T. Dastorani, "Effects of river basin classification on artificial neural networks based ungauged catchment flood prediction", in the *Proceedings of the International Symposium on Environmental Hydraulics*. Phoenix, 2001.

35. Wright. N.G., M.T. Dastorani, P. Goodwin, and C.W. Slaughter, "A combination of artificial neural networks and hydrodynamic models for river flow prediction", in the *Proceedings of Hydroinformatics* 2002, 2002.