Assessment of surface water quality using multivariate statistical techniques: A case study of the Nampong River Basin, Thailand

Somphinith Muangthong*

Abstract

This study investigates the spatial water quality pattern of 13 stations located along the main Nampong River. Multivariate statistical methods, namely, the hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA), and the factor analysis (FA), were used to study the spatial variations of the most significant water quality variables and to determine the origin of pollution sources. Sixteen water quality parameters were initially selected and analyzed. Two spatial clusters were formed based on HACA. These clusters are designated as upper part (U/P) of Nampong River and lower part (L/P) of Nampong River regions. Forward and backward stepwise DA managed to discriminate ten water quality variables, from the original 16 variables. PCA and FA (varimax functionality) were used to investigate the origin of each water quality variable due to land use activities based on the two clustered regions. Five principal components (PCs) were obtained with 69.806% total variation for the moderate-pollution source region, while five PCs with 69.327% total variances was obtained for the low-pollution source region. The pollution source for the L/P is of anthropogenic sources (industrial, municipal waste, and agricultural runoff). For the U/P region, the agricultural runoffs are the main sources of pollution. This study concluded the application of multivariate statistical methods to reduce the large number of water quality parameters down to manageable number.

Keywords : Principal Component Analysis, Discriminant Analysis, Nampong River Basin, Thailand

---

Faculty of Engineering and Architecture, Rajamangala University of Technology Isan, Muang, Nakhon Ratchasima, Thailand
* Corresponding author, E-mail: somphinith.mu@rmuti.ac.th Received 17 November 2014, Accepted 14 January 2015

1. Introduction

Rivers are important sources of surface water and a boon to nature and human beings. Surface water quality is affected by both anthropogenic activities and natural processes. Natural processes influencing water quality include precipitation rate, weathering processes and sediment transport, whereas anthropogenic activities include urban development and expansion, and industrial and agricultural practices. These activities often result in the degradation of water quality, physical habitat, and biological integrity of the ecosystem [1]. Increasing exploitation of water resources in the catchment area is responsible for much of the pollution load [2-3]. Pollution of surface water with toxic chemicals and excess nutrients, resulting from storm water runoff, vadose zone leaching, and groundwater discharges, has been an issue of worldwide environmental concern. Therefore, effective and long–term management of rivers requires a fundamental understanding of hydro–morphological, chemical and biological characteristics. However, due to spatial and temporal variations in water quality (which are often difficult to interpret), a monitoring program, providing a representative and reliable estimation of the quality of surface waters, is necessary [4-5].

The application of multivariate methods such as Cluster analysis (CA), principal analysis (PCA), factor analysis (FA) and discriminant analysis (DA) has increased tremendously in recent years for analyzing environmental data and drawing meaningful information [4, 6-11]. Application of different multivariate statistical techniques helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors/sources that influence water systems and offers a valuable tool for reliable management of water resources as well as rapid solution to pollution problems [4-5, 7, 10, 12].

In the present study, a large data matrix, obtained during a 10 year (2003–2012) monitoring program, is subjected to different multivariate statistical techniques to extract information about the similarities and dissimilarities between sampling sites, identification of water quality variables responsible for spatial and temporal variations in river water quality, the hidden factors explaining the structure of the database, and the influence of possible sources (natural and anthropogenic) on the water quality parameters of the Nampong River of Thailand.

2. Methods

2.1 Study area

Nampong River Basin is in the northeast region of Thailand (see Fig. 1) and is one tributary of Chi River Basin. It locate between 16º 15′N – 17º 15′N and 102º 30′E – 103º 15′E. A large part of the basin area is relatively flat with an average elevation of 300 m while the western edge is rougher and rises to heights of

1,300 m. The upper Nampong River Basin yields water for the Ubolratana reservoir that provides downstream areas with hydropower as well as irrigated water for agriculture. Most underlying rocks in the basin are of sandstone or limestone. Surface soils are mainly sandy loam with sandy clay sub-soils and contain more clay in the west and more sand in the east. Stream flow declines very rapidly with the end of the rainy season, indicating poor soil moisture retention. Underlying geology in the basin consists mainly of limestone and sandstone which are notorious for deep leakage.

Nampong River Basin experiences a tropical, semi-arid climate. Southwest monsoon, northeast monsoon, depression from China Sea, and depression from Pacific Ocean are influential to Nampong River Basin. The southwest monsoon brings humidity from Thai gulf causing heavy rain during May to October. There are six rainy season months (May–October). The cool season extends from November to February and is marked by generally dry and cool weather with some thunderstorms from the northeastern monsoon.

Currently Nampong River has the water quality pollution by both natural and anthropogenic process. Thus, the water quality problem in the river causes a problem for the entire river basin ecosystem. Water pollution has a negative impact on many species of animals and plants in the watershed. The wastewater can produce toxins that cause human health problems in the future.

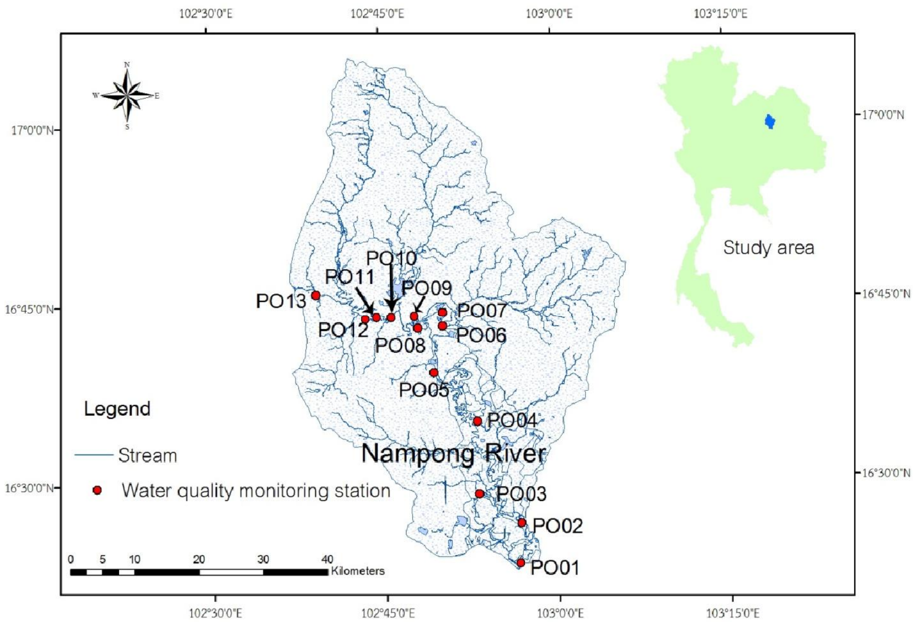

Fig. 1. Map of study area and water quality monitoring stations (listed PO01–PO13) in the Nampong River

Table 1 Water quality parameters, units and analytical methods used during 1996–2012 for surface water of the Lamtakong River

S.N.	Parameters	Abbreviations	Units	Analytical methods
1	Water temperature	WT	C	Thermometer
2	pH	pH	pH	Electrometric (pH meter)
3	Turbidity	Tur	NTU	Turbidity meter
4	Conductivity	Cond	uS/cm	Electrometric (conductivity meter)
5	Salinity	Sal	ppt	Electrometric (conductivity meter)
6	Dissolved Oxygen	DO	mg/l	Azide modification
7	Biochemical Oxygen Demand	BOD	mg/l	Azide modification at 20ºC (5 days)
8	Total Coliform Bacteria	TCB	MPN/100 ml	Multiple tube fermentation technique
9	Fecal Coliform Bacteria	FCB	MPN/100 ml	Multiple tube fermentation technique
10	Total Phosphorus	TP	mg/l	Ascorbic acid
11	Nitrate nitrogen	NO_3-N	mg/l	Cadmium reduction
12	Nitrite nitrogen	NO_2-N	mg/l	Distillation nesslerization
13	Ammonia nitrogen	NH_3-N	mg/l	Distillation nesslerization
14	Total Solid	TS	mg/l	Total residue dried at 103-105ºC
15	Total Dissolved Solid	TDS	mg/l	Total dissolved solids dried at 103-105ºC
16	Suspended Solid	SS	mg/l	Suspended solids dried at 103-105ºC

2.2 Monitored parameters and analytical methods

The data sets of thirteen water quality monitoring stations, comprising 16 water quality parameters monitored wet and dry seasons over 10 years (2003–2012), were obtained from the Pollution Control Department (PCD). The water quality parameters, their units and methods of analysis are summarized in Table 1. Although there are more than 29 water quality parameters available, only 16 parameters were selected water quality monitoring stations all around Nampong River Basin consist of 20 stations. Data analysis allowed the reduction to only 13 water quality monitoring stations. The 13 water monitoring stations were selected because of their continuity in measurement and overlapping coverage of the Nampong River Basin. The selected water quality parameters includes water temperature, pH, turbidity, conductivity, salinity, dissolved oxygen, biochemical oxygen demand, total coliform bacteria, fecal coliform bacteria, total phosphorus, nitrate nitrogen, nitrite nitrogen, ammonia nitrogen, total solid, total dissolved solid, and suspended solid. In this study used water quality data from the data analysis completed by PCD. The water quality measurement stations are shown in the table 2.

Table 2 Water quality measurement stations

Station	Position
PO01	Watunya resort, Pralub
PO02	Promnimit bridge, Koksri
PO03	Nonghin water supply, Moung
PO04	Utumporn temple, Nampong
PO05	Tarmoa-wangchai bridge, Nampong
PO06	KhonKaen sugar factory, Nampong
PO07	Tougteaw Shrine, Nampong
PO08	Nongvay weir, Nampong
PO09	Ban Nongbrownoy, Nampong
PO10	lower Boung Houyjot 100 m., Nampong
PO11	upper Boung Houyjot 100 m., Nampong
PO12	Ban Kambon, Ubonratana
PO13	Ban Bornokkow bridge, Ubonratana

2.3 Data treatment and multivariate statistical methods

The Kolmogorove–Smirnov (K–S) statistics were used to test the goodness–of–fit of the data to log–normal distribution. According to the K–S test, all the variables are log–normally distributed with 95% or higher confidence. Similarly, to examine the suitability of the data for principal component analysis/factor analysis, Kaiser–Meyer–Olkin (KMO) and Bartlett's test were performed. KMO is a measure of sampling adequacy that indicates the proportion of variance which is common variance, i.e., which might be caused by underlying factors. High value (close to 1) generally indicates that principal component/factor analysis may be useful, which is the case in this study: KMO = 0.69. Bartlett's test of sphericity indicates whether correlation matrix is an identity matrix, which would indicate that variables are unrelated. The significance level which is 0 in this study (less than 0.05) indicates that there are significant relationships among variables.

Spearman rank–order correlations (Spearman R coefficient) were used to study the correlation structure between variables to account for non–normal distribution of water quality parameters [4-6, 10]. In this study, temporal variations of river water quality parameters were first evaluated through a season parameter correlation matrix, using Spearman's R. The water quality parameters were grouped into two seasons: wet and dry, and each assigned a numerical value in the data file (dry=1 and wet=2), which, as a variable corresponding to the season, was correlated (pair by pair) with all the measured parameters.

River water quality data sets were subjected to four multivariate techniques: cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) [4-10]. DA was applied to raw data, whereas PCA, FA and CA were applied to experimental data, standardized through z–scale transformation to avoid misclassifications arising from the different orders of magnitude of both numerical values and variance of the parameters analyzed [12-14].

2.3.1 Cluster analysis

Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within–cluster) homogeneity and high external (between clusters) heterogeneity. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set, and is typically illustrated by a dendrogram [11, 15]. The dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between analytical values from the samples [10]. In this study, hierarchical agglomerative CA was performed on the normalized data set by means of the Ward's method, using squared Euclidean distances as a measure of similarity. The Ward's method uses an analysis of variance approach to evaluate the distances between clusters in an attempt to minimize the sum of squares (SS) of any two clusters that can be formed at each step. The spatial variability of water quality in the whole river basin was determined from CA, using the linkage distance, reported as Dlink/Dmax, which represents the quotient between the linkage distances for a particular case divided by the maximal linkage distance. The quotient is then multiplied by 100 as a way to standardize the linkage distance represented on the y-axis [4-6, 11, 14, 16].

2.3.2 Principal component analysis/factor analysis

Principal Component Analysis (PCA) is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. PC provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information [17]. The principal component (PC) can be expressed as:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + ... + a_{im}x_{mj} \qquad (1)$$

Where z is the component score, a is the component loading, x is the measured value of variable, i is the component number, j is the sample number and m is the total number of variables.

Factor analysis (FA) is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables [17]. PCA of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation (raw) generating varifactors [16, 18]. As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations. The FA can be expressed as:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + ... + a_{fm}f_{mi} + e_{fi} \qquad (2)$$

Where z is the measured variable, a is the factor loading, f is the factor score, e is the residual term accounting for errors or other source of variation, i is the sample number and m is the total number of factors.

2.3.3 Discriminant analysis

Discriminant analysis (DA) is used to classify cases into categorical–dependent values, usually a dichotomy. In DA, multiple quantitative attributes are used to discriminate between two or more naturally occurring groups. In contrast to CA, DA provides statistical classification of samples and it is performed with prior knowledge of membership of objects to a particular group or cluster. The DA technique builds up a discriminant function for each group [2, 6], as in the equation below.

$$f(G)i = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \qquad (3)$$

Where i is the number of groups (G), ki is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, wj is the weight coefficient, assigned by DA to a given selected parameters (pj).

In this study, two groups for temporal (wet and dry seasons) and two groups for spatial (two sampling regions) evaluations have been selected and the number of analytical parameters used to assign a measure from a monitoring site into a group (season or monitoring area). The site (spatial) and the season (temporal) were the grouping (dependent) variables, whereas all the measured parameters constituted the independent variables.

3. Results and discussion

3.1 Spatial similarity and site grouping

Cluster analysis was used to detect the similarity groups between the sampling sites. It yielded a dendrogram (Fig. 2.), grouping all thirteen sampling sites of the basin into two statistically significant clusters at (Dlink/Dmax) x 100 < 60. Since we used hierarchical agglomerative cluster analysis, the number of clusters was also decided by practicality of the results, as there is ample information (e.g. land-use) available on the study sites. Cluster 1 (PO01 to PO06) correspond to lower part (L/P) sites. These stations receive pollution from domestic wastewater and industrial effluents located in urban areas. The cluster 2 (PO07 to PO13) correspond to relatively upper part (U/P) sites. In cluster 2, seven stations are situated at the upper part of the river basin. These stations receive pollution from agricultural and farm effluents located in agriculture areas. The results indicate that the CA technique is useful in offering reliable classification of surface waters in the whole region and will make it possible to design a future spatial sampling strategy in an optimal manner, which can reduce the number of sampling stations and associated costs. There are other reports [14, 16, 19] where similar approach has successfully been applied to water quality programs.

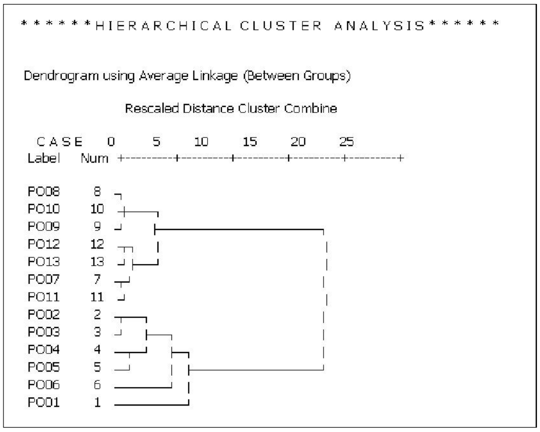

Fig. 2. Dendrogram showing clustering of sampling sites according to water quality characteristics of the Nampong River

3.2 Temporal and spatial variations in river water quality

Temporal variations in river water quality parameters (Table 1.) were evaluated through a season–parameter correlation matrix, which shows that all the measured parameters (16 in number) were found to be significantly ($p < 0.01$) correlated with the season. Temporal variations in water quality were further evaluated through DA. Temporal DA was performed on raw data after dividing the whole data set into two seasonal groups (wet and dry seasons). Discriminant functions (DFs) and classification matrices (CMs) obtained from the standard, forward stepwise and backward stepwise modes of DA are shown in Table 3. In forward stepwise mode, variables are included step–by–step beginning with the more significant until no significant changes are obtained, whereas, in backward stepwise mode, variables are removed step–by–step beginning with the less significant until no significant changes are obtained. The backward stepwise mode, DA gave CMs with 74.3% correct assignations using only eight discriminant parameters (Table 3) with little difference in match for each season compared with the backward stepwise mode. Thus, the temporal DA results suggest that dissolved oxygen, biochemical oxygen demand, fecal coliform bacteria, nitrate nitrogen, nitrite nitrogen, suspended solid, turbidity, and conductivity are the most significant parameters to discriminate between the two seasons, which means that these eight parameters account for most of the expected temporal variations in the river water quality.

Table 3 Classification function coefficients for discriminant analysis of temporal variations in water quality of the Nampong River

Parameters	Wet season coefficient	Dry season coefficient
DO	2.210	2.890
BOD	2.205	2.937
FCB	0.000	0.001
NO_3-N	0.091	7.580
NO_2-N	18.852	44.522
SS	-0.010	0.013
Turbidity	0.023	0.063
Conductivity	0.070	0.082
(Constant)	-13.124	-23.399

Spatial DA was performed with the same raw data set comprising 16 parameters after grouping into two major classes of lower part and upper part as obtained through CA. The sites (clustered) were the grouping (dependent) variable, while all the measured parameters constituted the independent variables. The backward stepwise mode, DA gave CMs with 89.4% correct assignations using only eight discriminant parameters Table 4. DA shows that dissolved oxygen, biochemical oxygen demand, fecal coliform bacteria, total phosphorus, nitrate nitrogen, nitrite nitrogen, suspended solid, total dissolve solid, water temperature, conductivity, and salinity are the discriminating parameters in space.

Table 4 Classification function coefficients for discriminant analysis of spatial variations in water quality of the Nampong River

Parameters	Upper part coefficient	Lower part coefficient
DO	1.809	2.810
BOD	-2.041	-0.137
FCB	0.003	0.001
TP	0.912	-5.149
NO_3-N	-12.801	-47.513
NO_2-N	9.801	5.084
TS	-0.117	-0.080
TDS	0.068	0.144
Water temperature	10.109	8.688
Conductivity	0.011	-0.008
Salinity	7.222	-14.458
(Constant)	-166.561	-134.005

3.3 Data structure determination and source identification

Principal component analysis/factor analysis was performed on the normalized data sets (16 variables) separately for the two different regions, viz., the lower part (L/P) and upper part (U/P), as delineated by CA techniques, to compare the compositional pattern between analyzed water samples and identify the factors influencing each one.

PCA of the two data sets yielded five PCs for the L/P and U/P with Eigenvalues >1, explaining 65.90% and 68.80% of the total variance in respective water quality data sets. An Eigenvalue gives a measure of the significance of the factor: the factors with the highest Eigenvalues are the most significant. Equal numbers of VFs were obtained for two sites through FA performed on the PCs. Corresponding VFs, variable loadings and explained variance are presented in Table 5. Liu et al. [13] classified the factor loadings as 'strong', 'moderate' and 'weak', corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively.

For the data set pertaining to U/P, among the five VFs, VF1, explaining 19.983% of total variance, has strong positive loadings on turbidity, suspended solid and total solid. This factor represents the contribution of non-point source pollution from the forest and agriculture areas. VF2, explaining 15.209% of total variance, has strong negative loadings on pH and DO. This factor represents the contribution of non-point source pollution, indicates the loading of partially decayed organic matters from agricultural areas. VF3, explaining about 12.877% of total variance, has strong positive loadings on conductivity and salinity. VF4, explaining about 12.239% of total variance, has strong positive loading on total coliform and fecal coliform bacterias. This factor represents the contribution of point source pollution from human activities. VF5, explaining 9.019% of total variance, has strong positive loadings on total phosphorus and ammonia nitrogen. This factor represents the contribution of non-point source pollution.

For the data set pertaining to water quality in L/P, among five VFs, VF1, explaining 15.998% of total variance, has strong positive loading on total solid and total dissolved solid. This factor represents the contribution of point source pollution from urban areas. VF2, explaining 14.610% of the total variance, has strong positive loadings on total phosphorus and nitrate nitrogen. VF3, explaining 14.111% of the total variance, has strong negative loadings for pH and nitrite nitrogen. VF4, explaining 12.829% of total variance, has strong positive loadings on turbidity, BOD, and suspended solid. VF5, explaining 12.258% of total variance, has strong positive loadings on total coliform and fecal coliform bacterias. This factor represents the contribution of point source pollution such as toilet and kitchen water from urban areas.

Table 5 Loading of experimental variables (17) on significant principal components for U/P and L/P

Parameters	VF1	VF2	VF3	VF4	VF5
Upper part (five significant principal components)					
Water temp	-0.466	0.546	0.115	0.107	0.113
pH	-0.040	**-0.777**	0.027	0.059	-0.203
Turbidity	**0.722**	-0.028	-0.455	-0.102	0.150
Conductivity	-0.019	0.339	**0.756**	0.040	0.223
Salinity	-0.185	-0.190	**0.854**	0.048	0.105
DO	-0.106	**-0.842**	-0.045	-0.055	-0.055
BOD	0.619	-0.042	0.022	-0.209	-0.053
TCB	-0.103	0.022	0.046	**0.881**	-0.004
FCB	-0.140	-0.042	-0.012	**0.840**	0.070
TP	0.148	0.400	0.093	0.265	**0.669**
NO3-N	-0.074	0.486	0.483	0.087	0.085
NO2-N	-0.137	0.448	0.171	0.558	0.113
NH3-N	-0.045	0.121	0.182	-0.028	**0.878**

Table 5 Loading of experimental variables (17) on significant principal components for U/P and L/P (cont.)

Parameters	VF1	VF2	VF3	VF4	VF5
Upper part (five significant principal components)					
SS	**0.754**	-0.135	-0.330	0.001	0.095
TS	**0.945**	0.088	-0.096	-0.038	0.103
TDS	0.699	0.173	0.334	-0.099	-0.190
Eigenvalue	4.058	2.966	1.718	1.339	1.012
% Total variance	19.983	15.209	12.877	12.239	9.019
Cumulative % variance	19.983	35.193	48.070	60.309	69.327
Lower part (five significant principal components)					
Water temp	-0.490	0.362	0.475	-0.194	0.080
pH	0.094	0.068	**-0.805**	-0.012	-0.061
Turbidity	0.092	-0.011	-0.072	**0.717**	0.020
Conductivity	0.188	0.609	0.439	-0.092	-0.100
Salinity	-0.009	-0.621	0.082	-0.127	0.178
DO	0.082	0.109	-0.453	0.505	0.199
BOD	0.187	-0.156	-0.008	**0.696**	-0.073
TCB	-0.132	0.081	0.088	-0.038	**0.882**
FCB	0.019	0.113	0.027	-0.033	**0.898**
TP	-0.009	**0.697**	0.292	-0.016	0.060
NO3-N	-0.268	**0.685**	-0.058	0.059	0.091
NO2-N	0.065	0.322	**-0.727**	0.028	0.071
NH3-N	-0.104	0.625	-0.149	-0.209	0.324
SS	-0.005	-0.024	0.076	**0.702**	-0.088
TS	**0.942**	-0.013	-0.013	0.173	-0.051
TDS	**0.955**	-0.072	-0.011	0.052	-0.047
Eigenvalue	3.495	2.031	1.903	1.566	1.105
% Total variance	15.998	14.610	14.111	12.829	12.258
Cumulative % variance	15.998	30.608	44.719	57.548	69.806

Bold values indicate strong and moderate loadings.

4. Conclusions

In this case study, different multivariate statistical techniques were used to evaluate spatial and temporal variations in surface water quality of the Nampong river basin. Hierarchical cluster analysis grouped thirteen sampling sites into two clusters of similar water quality characteristics. Although the principle component analysis/factor analysis did not result in a significant data reduction, it helped extract and identify the factors/sources responsible for variations in river water quality at two different sampling sites. Varifactors obtained from factor analysis indicate that the parameters responsible for water quality variations are mainly related to organic pollution and nutrients in upper part areas, and organic pollution and nutrients in lower part areas in the river basin. Discriminant analysis gave the best results both spatially and temporally. For two different sampling sites of the basin, it yielded an important data reduction, as it used only ten parameters (dissolved oxygen, biochemical oxygen demand, fecal coliform bacteria, total phosphorus, nitrate nitrogen, nitrite nitrogen, suspended solid, total dissolve solid, water temperature, conductivity, and salinity) and eight parameters (dissolved oxygen, biochemical oxygen demand, fecal coliform bacteria, nitrate nitrogen, nitrite nitrogen, suspended solid, turbidity, and conductivity). Therefore, DA allowed a reduction in the dimensionality of the large data set, delineating a few indicator parameters responsible for large variations in water quality. Thus, this study illustrates the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, and in water quality assessment, identification of pollution sources/factors and understanding temporal/spatial variations in water quality for effective river water quality management.

5. Acknowledgements

The authors wish to thanks the Office of the Higher Education Commission for providing financial assistance. The authors sincerely thank the Pollution Control Department, Thailand for providing the database.

6. Reference

[1] A. Qadir, R.N. Malik and S.Z. Husain. Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. Environmental Monitoring and Assessment 140(1–3), 2007, pp. 43–59.

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

[3] L. Salvati. The spatial nexus between population growth and land degradation in a dry Meditereanean region: a rapidly changing pattern? International Journal of Sustainable Development & World Ecology 19(1), 2012, pp. 81-88.

[4] S. Shrestha and F. Kazama. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modelling & Software 22, 2007, pp. 464–475.

[5] S. Shrestha and S. Muangthong. Assessment of surface water quality of Songkhram River (Thailand) using environmetric techniques. International Journal of River Basin Management 12(4), 2014, pp. 341-356.

[6] D.A. Wunderlin, M.D.M. Diaz, M.V. Ame, S.F. Pesce, A.C. Hued and M.A. Bistoni. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Research 35, 2001, pp. 2881-2894.

[7] R. Reghunath, T.R.S. Murthy and B.R. Raghavan. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research 36, 2002, pp. 2437-2442.

[8] F. Zhou, Y. Liu and H.C. Guo. Application of multivariate. statistical methods to the water quality assessment of the watercourses in the northwestern New Territories, Hong Kong. Environmental Monitoring and Assessment 132(1-3), 2007, pp. 1-13.

[9] H. Boyacioglu and H. Boyacioglu. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environmental Geology 54, 2007, pp. 275-282.

[10] S. Shrestha, F. Kazama and T. Nakamura. Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. Journal of Hydroinformatics 10, 2008, pp. 43–54.

[11] H. Zhang and C.Y. Jim. Species adoption for sustainable forestry in Hong Kong's degraded countryside. International Journal of Sustainable Development & World Ecology 20(6), 2013, pp. 484-503.

[12] V. Simeonov, P. Simeonova and R. Tsitouridou. Chemometric quality assessment of surface waters: two case studies. Chemical and Engineering Ecology 11(6), 2004, pp. 449-469.

[13] C.W. Liu, K.H. Lin and Y.M. Kuo. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science in the Total Environment 313, 2003, pp. 77-89.

[14] V. Simeonov, J.A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sofoniou and T. Kouimtzis. Assessment of the surface water quality in Northern Greece. Water Research 37, 2003, 4119-4124.

[15] J.E. McKenna Jr. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling & Software 18(3), 2003, pp. 205-220.

[16] K.P. Singh, A. Malik and S. Sinha. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. Analytica Chimica Acta 538, 2005, pp. 355-374.

[17] B. Helena, R. Pardo, M. Vega, E. Barrado, J.M. Fernandez and L. Fernandez. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research 34, 2000, pp. 807-816.

[18] G. Brumelis, L. Lapilla, O. Nikodemus and G. Tabors. Use of an artificial model of monitoring data to aid interpretation of principal component analysis. Environmental Modelling & Software 15(8), 2000, pp. 755-763.

[19] J.H. Kim, R.H. Kim, J. Lee, T.J. Cheong, B.W. Yum and H.W. Chang. Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. Hydrological Processes 19, 2005, pp. 1261-1276.