Assessment ecosystem of the Lamtakong River Basin (Thailand) using multivariate statistical techniques

Review of Integrative Business & Economics Research

Somphinith Muangthong
Faculty of Engineering and Architecture, Rajamangala University of Technology Isan, Nakhon Ratchasima, Thailand

ABSTRACT

This study investigates the spatial water quality pattern of 7 stations located along the Lamtakong River. Multivariate statistical methods, hierarchical agglomerative cluster analysis (HACA), discriminant analysis (DA), principal component analysis (PCA), and 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. Seventeen water quality parameters were initially analyzed. Two spatial clusters were formed based on HACA. These clusters are designated as upper part and lower part of Lamtakong River regions. PCA and FA 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 65% total variation for the high-pollution source (HPS) region, while five PCs with 68% total variances was obtained for the moderate-pollution source (MPS) regions. The pollution source for the HPS is of anthropogenic sources (industrial and municipal waste). For the MPS region, the domestic and agricultural runoffs are the main sources of pollution. This study can conclude that the application of multivariate statistical methods can reveal meaningful information on the spatial variability of a large and complex river water quality data.

Keywords: Cluster analysis, Principal component analysis, Lamtakong River Basin, Thailand

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

1. Introduction

Rivers are important sources of surface water for 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 chemical, biological and hydro–morphological 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 component 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 17 year (1996–2012) 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 Lamtakong River of Thailand.

2. Methods

2.1 study area

Lamtakong River Basin is a tributary at the west of the Moon River Basin. The upper part of this basin is high land with Khoyai Mountain and terracing hillsides.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

Lamtakong River is flow to Moon River at Charermprakait district Nakhon Ratchasima province (Water Resources Reginal Office 5). It rises in the southern Buntat Mountain range at an altitude of around 1,350 m. The eastern part connects to Lampraperng and Moonbon dams, the western part connects to plateau between Nakornnayok River Basin and Parsak River Basin and northern part connects to Lamchangkrai River Basin.

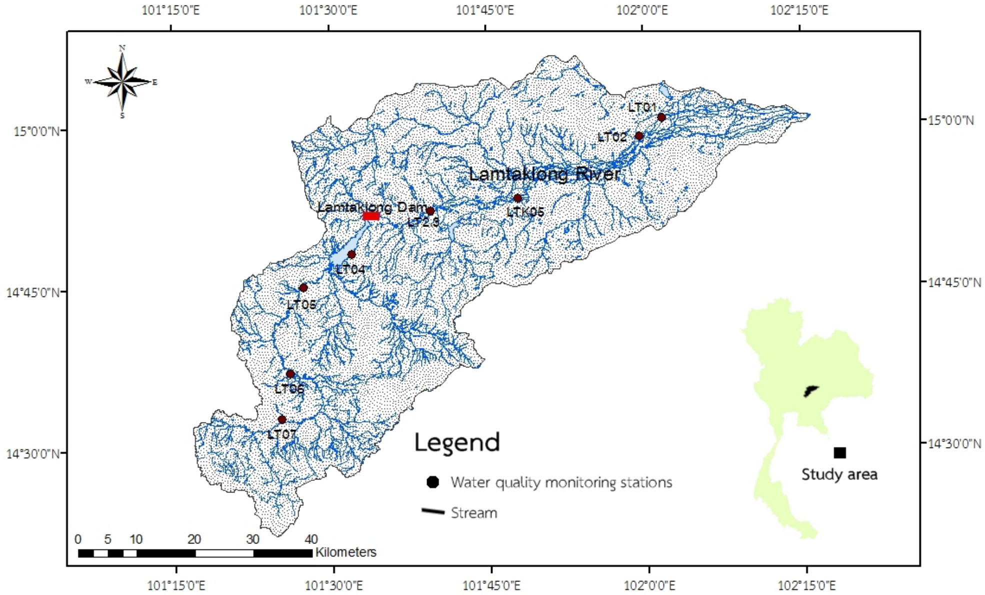

Figure 1 Map of study area and water quality monitoring stations (listed LT01–LT07) in the Lamtakong River

Lamtakong River Basin is located in the northeast of Thailand (see Figure 1) between 14º 15´N – 15º 15´N and 101º 15´E – 102º 30´E. The basin area is 3,425 km^2 and the main stream is drained by the Lamtakong River, length 220 km. Lamtakong River is drained through Pakchong district to Lamtakong reservoir at Muang district Nakhon Ratchasima province. The right side of basin is drainage from high slope area to tributaries in the lower part. Many tributaries at the high slope area are in the upper part of Lamtakong reservoir to Songnern district. Lower part of Lamtakong reservoir is quite low slope and have a rich agricultural land with rice plains fed from several streams.

Currently Lamtakong 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. Having negative effects many types of animals and plants in the river basin. Other more, waste water can be produced toxic. That let to human health problems future.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

2.2 Parameters and analytical methods

The data sets of 7 water quality monitoring stations, comprising 17 water quality parameters monitored wet–dry seasons over 17 years (1996–2012), were obtained from the Pollution Control Department (PCD). Although there are more than 29 water quality parameters available, only 17 parameters were selected due to their continuity in measurement at all selected water quality monitoring stations. Also the water quality monitoring stations all around Lamtakong River Basin consist of 15 stations. Data analysis allowed the reduction to only 7 water quality monitoring stations. The 7 water monitoring stations were selected because of their continuity in measurement and overlapping coverage of the Lamtakong River.

The selected water quality parameters includes water temperature, pH, turbidity, conductivity, salinity, DO, BOD, total coliform bacteria, fecal coliform bacteria, total phosphorus, nitrate nitrogen, nitrite nitrogen, ammonia nitrogen, total solid, total dissolved solid, suspended solid and $CaCO_3$. Water quality parameters, their units and methods of analysis are summarized in Table 1. The PCD has sampled, preserved and analyzed all the water quality parameters as per US. Environmental Protection Agency.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

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
17	Calcium carbonate	$CaCO_3$	mg/l	EDTA Titrimetric method

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–

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

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 analysis may be useful, which is the case in this study: KMO = 0.747. 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 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 season parameter correlation matrix, using Spearman non–parametric correlation coefficients. The water quality parameters were grouped into two seasons: wet season (May–October) and dry season (November–April), and each assigned a numerical value in the data file, 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 [12-15].

2.4 Cluster analysis

Cluster analysis (CA) is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. CA 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 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, 16]. 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. 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

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

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, 5, 6, 11, 14, 16].

2.5 Principal component analysis/factor analysis

Principal Component Analysis (PCA) is designed to transform the original variables into new, uncorrelated variables (axes), called the principal component (PC), which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data can be accounted for as concisely as possible. 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 PC can be expressed as:

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

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) follows PCA. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules, and constructing new variables, also called varifactors (VF). PC is a linear combination of observable water quality variables, while 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 VFs [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:

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

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.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

2.6 Discriminant analysis

Discriminant analysis (DA) is used to classify cases into categorical–dependent values, usually a dichotomy. If DA is effective for a set of data, the classification table of correct and incorrect estimates will yield a high correct percentage. 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. Furthermore, DA helps in grouping samples sharing common properties. The DA technique builds up a discriminant function for each group, which operates on raw data and this technique constructs a discriminant function for each group [2, 6], as in the equation below:

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

Where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, w_j is the weight coefficient, assigned by DA to a given selected parameters (pj). The weight coefficient maximizes the distance between the means of the criterion variable. The classification table, also called a confusion, assignment or prediction matrix or table, is used to assess the performance of DA. This is simply a table in which the rows are the observed categories of the dependent and the columns are the predicted categories of the dependents. When prediction is perfect, all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications.

In this study, two groups for temporal (two 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). DA was performed on each raw data matrix using standard, forward stepwise and backward stepwise modes in constructing discriminant functions to evaluate both the spatial and temporal variations in river water quality of the basin. 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 River water quality for irrigation

Irrigated agriculture is dependent on an adequate water supply of usable quality. The quality of water for irrigation purposes have often been neglected because good quality water supplies have been readily available. The ideal situation is to have several supplies from which to make a selection, but normally only one supply is

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

available. In this case, the quality of water for irrigation purposes based on Pollution Control Department (PCD) of Thailand are DO > 4 mg/l, BOD< 2 mg/l, TCB< 20,000 MPN/100 ml, FCB < 4,000 MPN/100 ml, NO_3-N < 5 mg/l and NH_3-N < 0.5 mg/l.

As identified based on the quality of water for irrigation purposes of PCD are analyzed. The DO is higher than water quality standard as compared to PCD. The BOD and TCB are lower than water quality standard at the stations LT04 to LT07. BOD is higher than the water quality standard at the stations LT01 and LT02, land-use between LT01 and LT02 are urban area at Muang district during Lamtakong bridge, BanYongyeang and Watsamakkee bridge, Naimuang. FCB and NO_3-N are lower than the water quality standard. NH_3-N are lower than water quality standard as compared to PCD, except ammonia nitrogen at the stations LT01 and LT02 which are higher than the water quality standard. However, the water quality in Lamtakong River is suitable for irrigation purposes.

3.2 Spatial similarity and site grouping

Cluster analysis was used to detect the similarity groups between the sampling sites. It yielded a dendrogram (Figure 2), grouping all seven sampling sites of the basin into two statistically significant clusters at (Dlink/Dmax) x 100 < 60. Since we used hierarchical agglomerative cluster analysis (HACA), 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 (LT01 at Lamtakong bridge, BanYongyeang and LT02 at Watsamakkee bridge, Naimuang) correspond to lower part (L/P) sites. These stations receive pollution from municipal and industrial waste water effluent located in city area. The cluster 2 (LT03 at Kodchanoun bridge, Midtrapap; LT04 at Lamtakong water supply, Krongpai; LT05 at TOT Academy bridge, Pakchong; LT06 at Nongsaray bridge, Pakchong; and LT07 at Bokrached bridge, Kanongpra) correspond to relatively upper part (U/P) sites. In cluster 2, five stations are situated at the upper part of the river basin. These stations receive pollution from domestic and agricultural runoffs. 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.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

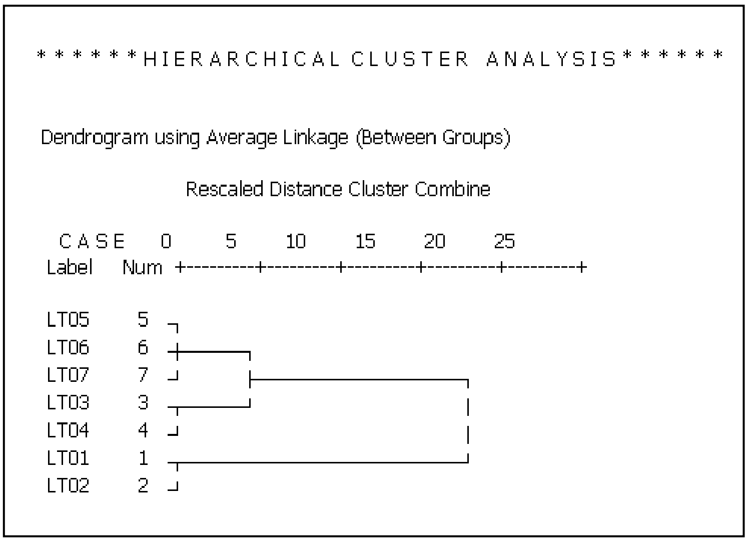

Figure 2 Dendrogram showing clustering of sampling sites according to water quality characteristics of the Lamtakong River

3.3 Temporal and spatial variations

Temporal variations in river water quality parameters (Table 1) were evaluated through a season–parameter correlation matrix, which shows that all the measured parameters 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). Discriminant functions (DFs) and classification matrices (CMs) obtained from the standard, forward stepwise and backward stepwise modes of DA are shown in Table 2. 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 87.8% correct assignations using only nine discriminant parameters (Table 2) with little difference in match for each season compared with the backward stepwise mode. Thus, the temporal DA results suggest that water temperature, turbidity, conductivity, DO, BOD, NO_3-N, NO_2-N, NH_3-N, and TDS are the most significant parameters to discriminate between the two seasons, which means that these nine parameters account for most of the expected temporal variations in the river water quality.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

Table 2 Classification function coefficients for discriminant analysis of temporal variations in water quality of the Lamtakong River

Parameters	Wet season coefficient	Dry season coefficient
Water temp	7.260	7.971
Turbidity	0.082	0.016
Conductivity	0.018	0.032
DO	5.838	5.256
BOD	-0.765	-1.766
NO_3-N	2.427	4.038
NO_2-N	-8.820	0.666
NH_3-N	-2.479	-1.597
TDS	0.030	0.008
(Constant)	-121.716	-136.333

As identified by DA, box and whisker plots of the selected parameters showing seasonal trends are given in Figure 3. The average total dissolve solid (Figure 3a), dissolved oxygen (Figure 3b), 5–day biochemical oxygen demand (Figure 3c), and turbidity (Figure 3d) are higher in wet as compared to dry. In the study period, these might have been due to the rain in the basin during this season. A clear inverse relationship between water temperature (Figure 3e), conductivity (Figure 3f) and nitrogen in form NO_2-N, NO_3-N, and NH_3-N (Figure 3g to Figure 3i) are observed, which is attributed to the seasonality effect. The inverse relationship between water temperature and nitrogen is a natural process because warmer water becomes more easily saturated with nitrogen and thus retains less nitrogen. Similar temporal variations in concentration of nitrate nitrogen were also reported by [19].

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

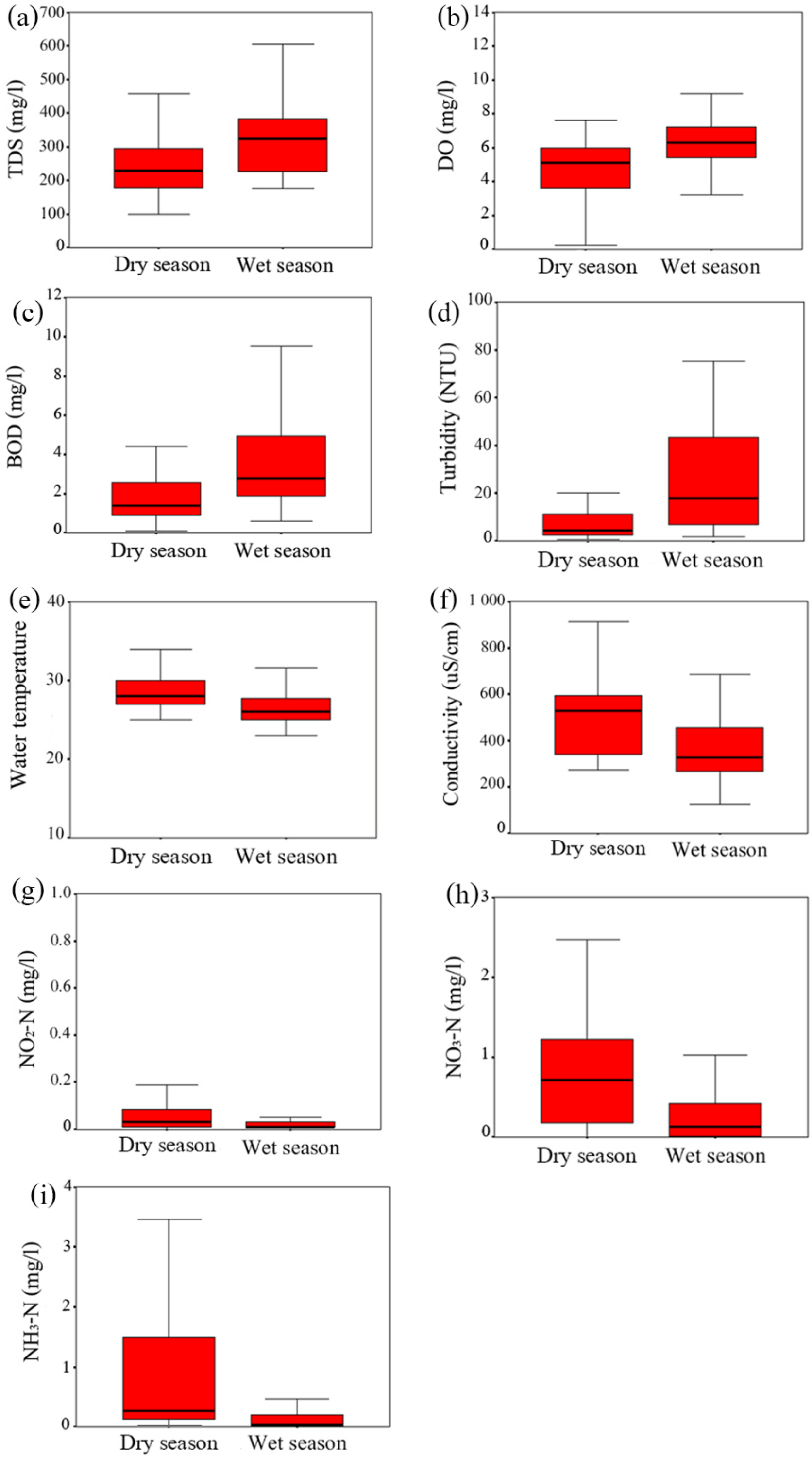

Figure 3 Temporal variables: (a) total dissolved solid, (b) dissolved oxygen, (c) 5–day biochemical oxygen demand, (d) turbidity, (e) water temperature, (f) conductivity, (g) nitrate nitrogen, (h) nitrite nitrogen, and (i) ammonia nitrogen in water quality of the Lamtakong River

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

Spatial DA was performed with the same raw data set comprising 17 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. Discriminant functions (DFs) and classification matrices (CMs), are shown in Table 4. DA shows that water temperature, dissolved oxygen, 5–day biochemical oxygen demand, total coliform bacteria, fecal coliform bacteria, total phosphorus, nitrate nitrogen, ammonia nitrogen, total solid, and $CaCO_3$ are the discriminating parameters in space.

Table 3 Classification function coefficients for discriminant analysis of spatial variations in water quality of the Lamtakong River

Parameters	Upper part coefficient	Lower part coefficient
Water temperature	8.608	7.952
DO	4.726	6.158
BOD	2.315	5.963
TCB	9.908E-05	8.281E-05
FCB	0.000	0.000
TP	0.169	-3.493
NO_2-N	-15.442	-23.841
NO_3-N	-2.523	-4.108
TD	0.083	0.072
$CaCO_3$	0.030	0.046
(Constant)	-162.066	-140.862

Box whisker plots of discriminating parameters identified by spatial DA (backward stepwise mode) were constructed to evaluate different patterns associated with spatial variations in river water quality (Figure 4). The NH_3-N (Figure 4a), DO (Figure 4b) and $CaCO_3$ (Figure 4c) are higher in the lower part (monitoring stations LT01 and LT02), as they receive discharge from domestic wastewater and industrial effluents located in Muang, Nakhon Ratchasima city areas. The trends for water temperature (Figure 4d), TCB (Figure 4e), BOD (Figure 4f), FCB (Figure 4g), TD (Figure 4h), TP (Figure 4i), and NO_3-N (Figure 4j) suggest a high load of dissolved organic matter from the agricultural effluents located at the upper part areas of the monitoring stations. This results in anaerobic conditions in the river, which, in turn, results in formation of total solid, total phosphorus, and organic acids.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

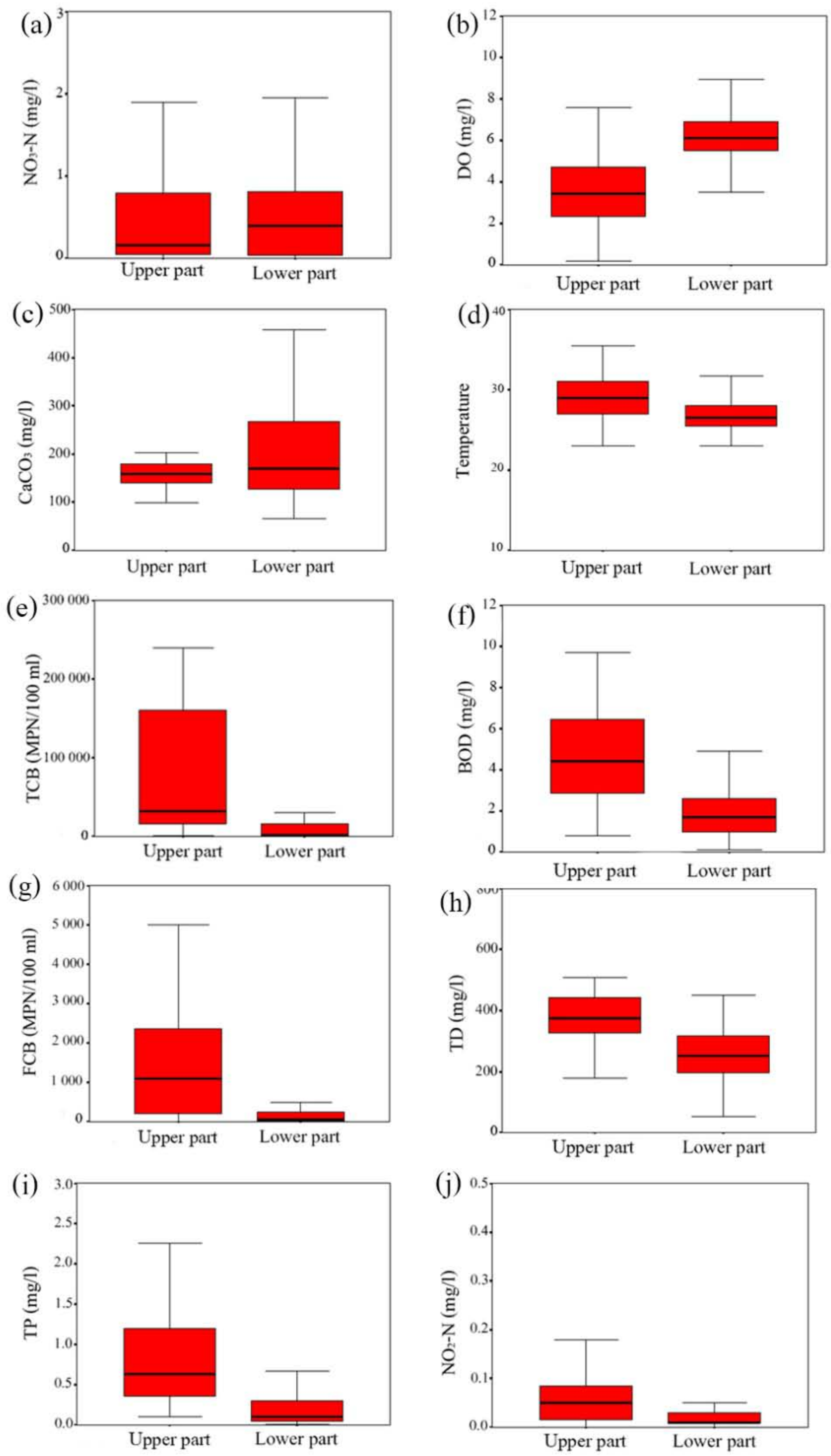

Figure 4 Spatial variations: (a) nitrate nitrogen, (b) dissolved oxygen, (c) $CaCO_3$, (d) water temperature, (e) total coliform bacteria, (f) 5–day biochemical oxygen demand, (g) fecal coliform bacteria, (h) total solid, (i) total phosphorus, and (j) nitrite nitrogen in water quality of the Lamtakong River

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

PCA/FA was performed on the normalized data sets 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. The input data matrices (variables x cases) for PCA/FA were [17x50] for L/P and U/P as presented in Table 4.

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

Parameters	**Upper part coefficient**	**Lower part coefficient**
Water temperature	8.608	7.952
DO	4.726	6.158
BOD	2.315	5.963
TCB	9.908E-05	8.281E-05
FCB	0.000	0.000
TP	0.169	-3.493
NO_2-N	-15.442	-23.841
NO_3-N	-2.523	-4.108
TD	0.083	0.072
$CaCO_3$	0.030	0.046
(Constant)	-162.066	-140.862

PCA of the two data sets yielded five PCs for the L/P and U/P with Eigenvalues >1, explaining 65.20% and 67.90% of the total variance in respective water quality data sets. Eigenvalue gives a measure of the significance of the factor: the factors with the highest Eigenvalues are the most significant. Eigenvalues of 1.0 or greater are considered significant [19]. 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.

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

Parameters	**VF1**	**VF2**	**VF3**	**VF4**	**VF5**
(a)Upper part (U/P) five significant principal components					
Water temp	-0.680	0.115	0.054	-0.063	0.207
pH	-0.109	-0.001	-0.201	-0.273	**0.767**
Turbidity	0.237	0.077	-0.211	**0.768**	-0.105
Conductivity	0.181	0.288	0.628	0.421	-0.154

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

Salinity	0.194	0.565	-0.113	-0.290	-0.571
DO	0.274	-0.182	-0.573	-0.053	0.092
BOD	0.663	-0.183	-0.166	0.204	0.087
TCB	-0.101	**0.832**	0.119	0.083	0.067
FCB	0.044	**0.807**	0.097	0.196	-0.101
TP	0.584	-0.096	0.072	0.216	0.535
NO_3-N	0.072	0.464	0.366	0.223	0.318
NO_2-N	-0.015	0.665	0.218	-0.158	-0.027
NH_3-N	-0.393	0.029	0.610	0.092	0.199
SS	0.317	0.021	-0.044	**0.828**	0.010
TD	**0.828**	0.187	0.046	0.299	0.034
TDS	**0.813**	0.182	-0.102	0.017	0.054
$CaCO_3$	0.090	0.205	**0.776**	-0.277	-0.109
Eigenvalue	3.854	3.284	1.468	1.354	1.126
% Total variance	22.670	19.319	8.638	7.967	6.622
Cumulative % variance	22.670	41.939	50.626	58.593	65.215
(b) Lower part (L/P) five significant principal components					
Water temp	-0.222	0.100	0.608	-0.019	0.515
pH	-0.078	0.020	-0.066	0.155	**0.896**
Turbidity	0.244	0.408	-0.241	0.255	-0.519
Conductivity	**0.798**	-0.168	0.367	0.174	0.069
Salinity	0.665	-0.249	0.139	0.340	-0.268
DO	-0.041	0.689	-0.349	0.289	-0.019
BOD	0.387	0.487	-0.124	-0.227	-0.063
TCB	-0.057	**-0.849**	-0.101	0.003	0.083
FCB	-0.037	**-0.738**	0.122	0.230	-0.122
TP	0.410	0.213	-0.298	-0.318	0.338
NO_3-N	-0.183	-0.121	0.458	0.517	0.031
NO_2-N	0.046	0.031	0.024	**0.808**	0.034
NH_3-N	0.202	-0.011	**0.730**	0.222	-0.074
SS	0.137	0.386	-0.644	0.104	0.060
TD	**0.838**	0.357	-0.143	-0.062	-0.138
TDS	**0.875**	0.234	-0.136	-0.171	-0.089
$CaCO_3$	0.300	-0.157	0.511	0.424	0.324
Eigenvalue	4.084	3.050	1.795	1.559	1.055
% Total variance	24.026	17.941	10.556	9.168	6.205
Cumulative % variance	24.026	41.957	52.523	61.691	67.896

Remark: Bold values indicate strong and moderate loadings.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

For the data set pertaining to U/P, among the five VFs, VF1, explaining 22.67% of total variance, has strong positive loadings on total solid and total dissolved solid. This factor represents the contribution of non-point source pollution from the forest and agriculture areas. VF2, explaining 19.32% of total variance, has strong positive loadings on total coliform bacteria and fecal coliform bacteria. This factor represents the contribution of point source pollution, indicates the loading of partially decayed organic matters from urban areas. VF3, explaining about 8.64% of total variance, has strong positive loadings on calcium carbonate. This VF3 represent the seasonal impact hardness. VF4, explaining about 7.97% of total variance, has strong positive loading on turbidity and suspended solid. This factor represents the contribution of non-point source pollution from agricultural areas. VF5, explaining 6.62% of total variance, has strong positive loadings on pH.

For the data set pertaining to water quality in L/P, among five VFs, VF1, explaining 24.023% of total variance, has strong positive loading on total solid and total dissolved solid. This factor represents the contribution of non-point source pollution from urban areas. VF2, explaining 17.94% of the total variance, has strong negative loadings on total coliform bacteria and fecal coliform bacteria. This factor represents the contribution of point source pollution such as toilet and kitchen water from urban areas. VF3, explaining 10.56% of the total variance, has strong positive loadings for ammonia nitrogen. VF4, explaining 9.17% of total variance, has strong positive loadings on nitrite nitrogen. VF5, explaining 6.21% of total variance, has strong positive loadings on pH.

4. Conclusions

In this case study, different multivariate statistical techniques were used to evaluate spatial and temporal variations in surface water quality of the Lamtakong River Basin. Hierarchical cluster analysis grouped seven sampling sites into two clusters of similar water quality characteristics. Based on obtained information, it is possible to design a future, optimal sampling strategy, which could reduce the number of sampling stations and associated costs. 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 (point source: domestic wastewater) and nutrients (non-point sources: agriculture and orchard plantations) in upper part areas, and organic pollution and nutrients (point sources: industrial and municipal wastewaters) 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 nine parameters (water temperature, turbidity, conductivity,

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

DO, BOD, NO_2-N, NO_3-N, NH_3-N, and TDS) and ten parameters (water temperature, DO, BOD, TCB, FCB, TP, NO_2-N, NO_3-N, TD, and $CaCO_3$. 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.

REFERENCES

[1] Qadir A, Malik RN, Husain SZ. 2007. "Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan". *Environmental Monitoring and Assessment* 140(1-3): 43-59.

[2] Singh KP, 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: 3980-3992.

[3] Salvati L. 2012. "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): 81-88.

[4] 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: 464-475.

[5] Shrestha S, Muangthong S. 2014. "Assessment of surface water quality of Songkhram River (Thailand) using environmetric techniques". *International Journal of River Basin Managemen*t 12(4): 341-356.

[6] Wunderlin DA, Diaz MDM, Ame MV, Pesce SF, Hued AC, Bistoni MA. 2001. "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: 2881-2894.

[7] Reghunath R, Murthy TRS, Raghavan BR. 2002. "The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India". *Water Research* 36: 2437-2442.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)

[8] Zhou F, Liu Y, Guo HC. 2007. "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): 1-13.
[9] Boyacioglu H, Boyacioglu H. 2007. "Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey". *Environmental Geology* 54: 275-282.
[10] Shrestha S, Kazama F, Nakamura T. 2008. "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: 43-54.
[11] Zhang H, Jim CY. 2013. "Species adoption for sustainable forestry in Hong Kong's degraded countryside". *International Journal of Sustainable Development & World Ecology* 20(6): 484-503.
[12] Simeonov V, Simeonova P, Tsitouridou R. 2004. "Chemometric quality assessment of surface waters: two case studies". *Chemical and Engineering Ecology* 11(6): 449-469.
[13] Liu CW, Lin KH, Kuo YM. 2003. "Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan". *Science in the Total Environment* 313: 77-89.
[14] Simeonov V, Stratis JA, 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: 4119-4124.
[15] Singh KP, 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: 355-374.
[16] Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L. 2000. "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis". *Water Research* 34: 807-816.
[17] Brumelis G, Lapilla L, Nikodemus O, Tabors G. 2000. "Use of an artificial model of monitoring data to aid interpretation of principal component analysis". *Environmental Modelling & Software* 15(8): 755-763.
[18] Kim JH, Kim RH, Lee J, Cheong TJ, Yum BW, Chang HW. 2005. "Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea". *Hydrological Processes* 19: 1261-1276.
[19] Fukasawa E. 2005. Determination of origin of nitrate nitrogen in Fuefuki river using stable isotope method. Bachelor thesis, Department of Ecosocial System Engineering, University of Yamanashi, Japan (in Japanese)

ISSN: 2304-1013 (Online); 2304-1269 (CDROM)