KMITL Sci. Tech. J. Vol. 12 No. 1 Jan. - Jun. 2012

Evaluation of Surface Water Quality using Multivariate Statistical Techniques: A Case Study of U-tapao River Basin, Thailand

Saroj Gyawali[1], KuaananTechato[1*], Chumpol Yuangyai[2] and Sathaporn Monprapusson[3]

[1] Faculty of Environmental Management, Prince of Songkla University, Hatyai, Songkhla, Thailand
[2] Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[3]Faculty of Social Science, Srinakharinwirot University, Bangkok, Thailand

Abstract

Multivariate statistical techniques, such as cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA), and factor analysis (FA) were applied to evaluate temporal and spatial variations of water quality and to identify potential pollution sources of U-tapao River Basin (URB). A large set of water quality data were collected from 21 monitoring stations of river during five years (2007-2011) and analyzed for 12 parameters. Hierarchical cluster analysis grouped 21 sampling sites into three clusters, relatively less polluted (LP), medium polluted (MP) and highly polluted (HP) sites, and based on the similarity of water quality characteristics. From DA, five significant parameters including temperature, pH, dissolved oxygen, fecal coliform bacteria and ammonia were identified with correctly assign about 69.6% for the temporal variation and four significant parameters temperature, pH, dissolved oxygen and ammonia with correctly assign about 63.3% for the spatial variation. PCA/FA, applied to analyze the data sets of the three different groups obtained from cluster analysis, resulted latent factors accounting for 75.16%, 76.01% and 70.51% of the total variance in water quality data sets of LP, MP and HP areas and also accounting for 72.97% and 71.51% of the total variance in water quality data sets of wet and dry seasons respectively. The PCA/FA assisted in extracting and recognizing the factors responsible for spatial and temporal variations and indicated that the parameters for water quality variations are mainly related to organic pollutants and showed that agriculture and urban activities were the major pollutant sources.

Keywords: Multivariate statistical techniques, Surface water quality, U-tapao River Basin, Water pollution

*Corresponding author: Tel +66-818418962
E-mail: uhugua@hotmail.com

1. Introduction

Water quality has become one of the major environmental concerns worldwide and is influenced by natural and anthropogenic disturbance, such as wastewater, runoff effluents, land reclamation, atmospheric deposition and climate change [1]. In recent years, more and more attention has been paid to surface water quality because of its strong linkage with human well being [2-3]. The quality of river at any point reflects several major influences, including the lithology of the basin, atmospheric inputs, and climatic conditions [4] and governed by both natural process and anthropogenic effects [5-9]. So, wastewater from agricultural, industrial and urban activities and often natural processes such as erosion and weathering degrades water quality and impair their use for drinking, industrial, agriculture, recreation or other purposes [10]. Clean river water is a vital commodity for the well-being of human societies, and damage of inland aquatic system was one of the most serious environmental problems of the last century [11]. Rivers, due to their role in carrying-off the domestic and industrial wastewater and run-off from agricultural land in their vast drainage basin are among the most vulnerable water bodies to pollution [12]. Discharges from municipals and industries are considered as a point source while surface runoff is as a non-point source due to its characteristics that are highly influenced by spatial and seasonal changes [13-15].

Spatial and temporal variability in water chemistry in rivers and streams is directly related to different factors. Rivers and streams are highly heterogeneous at different spatial scales. The spatial heterogeneity within the stream is due to local environmental conditions that change through time and differences in local channel form, while degree of temporal variability of surface water chemistry varies as a function of stream/river type and depends on the chemical parameter of interest [16]. Due to spatial and temporal variations in water quality, regular monitoring programs are recognized to be the essential step to characterize and control the surface water pollution [17]. However, many monitoring programs result in large and complicated data sets consisting of physical, chemical, biological and microbiological properties, which are difficult to analyze and interpret because of latent interrelationships among parameters and monitoring sites [3]. For this reason, multivariate statistical methods like cluster analysis (CA), discriminat analysis (DA), principal component analysis (PCA) and factor analysis (FA) have been widely applied to interpret and derive useful information from complicated data about water quality studies [14-21].

In recent years, many studies related to these methods have been carried out. For instance, multivariate statistical methods, such as FA was used by Charkhabi and Sakizadeh [22] to identify the sources of pollution and temporal variation of water quality of Anzali wetland, Northern Iran . Similarly, Zhou et al [3] used PCA and CA to classify the sampling sites and to identify the latent pollution source. And, Zhao and Cui [23] used CA and FA to find the temporal variation of water quality of Luan River, China, taking 19 parameters of 10 stations. Further, multivariate methods, like, CA , DA and PCA/FA were used by Shrestha and Kazama [4] to analyze the water quality dataset, including 12 parameters at 13 sites of the Fuji river basin from 1995-2002 to obtain temporal and spatial variations and to identify potential pollution sources. Again, Huang, Ho and Du [24] used CA, DA and PCA to analyze the costal water quality from 22 inshore sampling sites to extract latent information about the similarities or dissimilarities among the monitoring periods or sites and identify pollution sources leading to spatiotemporal variations in water quality in Macau peninsula. Obviously, previous studies provided valuable insights into the application of CA, DA and PCA techniques to environmental management and protection. However, comprehensive application of CA, DA and PCA/FA to analysis of the river water quality along U-tapao River Basin regarding spatial–temporal variation and sources identification has not been conducted.

Therefore, in this study, the data sets obtained during 2007-2011 in U-tapao River Basin in southern Thailand, were analyzed with CA, DA and PCA/FA. The objectives of this study were: (1) to extract information about the similarities or dissimilarities between monitoring sites, (2) to identify significant parameters responsible for temporal and spatial variations in river water quality, (3) to find the influence of possible pollution sources on water quality parameters. The results of this study are expected to be helpful to optimize river-monitoring plan and provide a valuable tool in developing assessment strategies for effective water quality management as well as rapid solutions on pollution problems. In addition, this study is aimed to provide information and scientific understanding to policy makers, environmentalists and researchers dealing with these kinds of river systems and yet have not been the subject of scientific investigation.

2. Materials and Methods

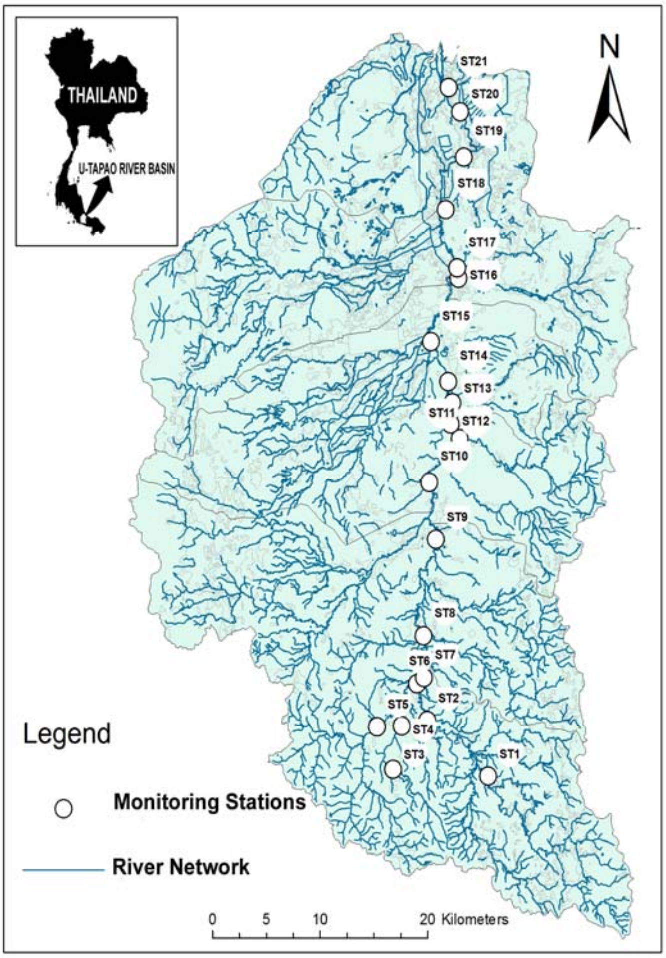
2.1 Study area

U-tapao is a sub-basin of Songkhla lake basin (SLB) which is located in southern part of Thailand. The basin is about 60 km long from north to south, and 40 km wide from west to east, and total coverage is about 2,305 square kilometers. The longitude and latitude of basin is 100º 10' through 100º 37’ E and 6º 28' through 7º 10' N respectively (Figure1). Basin has a tropical monsoon climate and it is governed by two monsoons; the southwest monsoon and the northeast monsoon with average rainfall of approximately 1800 mm per annum varying between 1600 and 2400 mm and temperature of the area varies between 24ºC and 32º C throughout the year. In the basin, more than 75% of area is covered by agricultural land use and about 13 % by forest and forest land is located mostly in mountainous areas, whereas agricultural and grassland areas are sparsely distributed throughout the basin. U-tapao is one of the most important rivers of SLB which has 10 tributaries including major and minor ones. The river serves as a major source of domestic and industrial water supply for Hatyai and Songkhla cities. During its course of 90 kms, it receives pollution load from both point and non-point sources.

2.2 Parameters and monitoring stations

The secondary data of 12 water quality parameters : water temperature (TEMP), pH, dissolved oxygen (DO) biological oxygen demand (BOD), suspended solid (SS), electrical conductivity (EC), turbidity (TUR), fecal coliform bacteria (FCB), ammonia (NH_3), nitrate (NO_3), nitrite (NO_2) and total phosphorous(TP) at 21 monitoring stations over 5 years (2007–2011) were collected from the Environmental Office -16, Songkhla and this organization is the authentic organization of collecting and maintaing data of southern region of Thailand. The water quality parameters, their units and descriptive statistics are mentioned in Table 1. Thai metrological department has categorized southern part of Thailand into two seasons. So, all data were divided into two parts, i) dry season (February, March, April and May) and ii) wet season (June, July, August, September, October, November, December and January) for seasonal variation analysis.

In this study, station 1 to 9 represent upstream region of basin and most of areas are less affected from human activities, out of which, station 1-3 are the least effected regions. Station 10 to 17 represent midstream region of basin and most of areas were affected by almost all types of pollution from residential, agricultural and industrial activities. Most of rubber processing and agricultural based industries are located along the station 12 to 17. Station 18 to 21 represent downstream region of basin and these are highly affected from agriculture based pollutants and station 20 and 21 are highly affected by agricultural, as well as shrimp and pig farming activities. Overall, the most of industries are located on the banks of river with about 10 % at the upstream, 65% at midstream and 25% at downstream region. The main commercial city, Hatyai is located in midstream region whereas traditional city, Songkhla is located in downstream region of the basin.



ST-1	Sapan langwat huayku
ST-2	Office of Water Work ,Sadao
ST-3	Mitr Sampan Community
ST-4	Sanepong School Bridge
ST-5	Ban Nam Hua Bridge
ST-6	Safe Skin Industry
ST-7	Ban Huathanon Bridge
ST-8	Ban Takian Pao Bridge
ST-9	Ban Tha Pho Ok Bridge
ST-10	Muang Kong Temple Bridge
ST-11	Ban Prao Bridge
ST-12	Siam Fiber Board Cooperation
ST-13	Ban Klong Pom Bridge
ST-14	Public Works Bridge, Klong Pla
ST-15	Bangsala Temple Bridge
ST-16	Hatyai University Bridge
ST-17	Utapao Canal Water Gate
ST-18	Ta Sae Temple Bridge
ST-19	Narong Nok Temple Bridge
ST-20	Wat Kutao Bridge
ST-21	Songkla Laguna Bridge

Figure 1 Map of study area and surface water quality monitoring stations (listed ST-1 to ST-21) in the U-tapao river basin (URB).

Table 1 Mean, standard deviation, minimum and maximum values of 12 water quality parameters (WQP) of U-tapao river of 21 monitoring stations from year 2007 to 2011

WQP	Mean	SD	Min	Max
TEMP (ºC)	28.96	1.75	25.12	34.37
pH	6.891	0.925	2.907	12.531
BOD (mg/L)	3.597	2.431	0.931	16.243
DO (mg/L)	4.109	1.396	0.802	11.237
EC (µs/cm)	766.09	4,249.25	24.00	52,800.00
SS (mg/L)	45.72	45.37	7.00	189.00
TUR (NTU)	40.38	29.23	3.81	151.03
FCB (mpn/100ml)	13,878	31,341	1,300	1,600,000
NH_3 (mg/L)	0.458	1.126	0.010	3.910
NO_3 (mg/L)	1.066	1.185	0.110	4.980
NO_2 (mg/L)	0.622	1.388	0.003	5.521
TP (mg/L)	1.081	2.683	0.010	3.133

2.3 Statistical methods

In this study, multivariate statistical techniques were used to analyze the huge data set of 21 monitoring stations of river basin. The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), assists 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 that influence water environment systems and offers a valuable tool for reliable management of water resources [4]. All mathematical and statistical calculations were implemented using SPSS version-12 and Microsoft Office Excel 2007.

2.4 Data treatment

Prior to multivariate statistical analysis, the normality of the distribution of each variable was checked. The Kolmogorove-Smirnov (K-S) statistics were used to test the goodness of fit of the data to log-normal distribution [4]. According to the K-S test, all the variables were log-normally distributed with 95% or higher confidence. CA and PCA/FA were applied on experimental data standardized through z-scale transformation in order to avoid misclassification due to wide differences in data dimensionality [20]. To examine the suitability of the data for PCA/FA, Kaiser-Meyer-Olkin (KMO) and Bartlett's Sphericity tests were performed [4]. In this study, the KMO is 0.647 (>0.6) and the significance level of Barelett's test was less than 0.05, so the data set was found to be appropriate for factor analysis.

3. Results and Discussion

3.1 Cluster analysis (Spatial similarity and site grouping)

Since one of the objectives of this study was to identify similarities or dissimilarities among monitoring sites and to distinguish each cluster analysis was applied to find out the similarity groups between 21 monitoring sites. Cluster analysis was performed on the mean values of the parameters for each of the stations. It yielded a dendrogram (Figure 2), grouping all 21 sampling sites of the basin into three statistically significant clusters at (Dlink/Dmax) X 100 < 25. The results separated three different groups (clusters) along the U-tapao River consisting all the measured values.

3.1.1 Group A: This group (cluster) could be regarded as relatively less polluted (LP) sites of the river. Most of stations (ST-1, ST-2, ST-3, ST-4, ST-5, ST-7, ST-8, and ST-9) are situated at upstream sites and ST-10 and ST-12 are situated the midstream region. In this area, the urbanization and industrialization level is relatively low. Hence, the impact of human activities on the riverine ecosystem is relatively low. The inclusion of the midstream sampling location in cluster group suggests the self purification and assimilative capacity of the river. Most of this area is underdeveloped. Although, the most of surrounding areas of ST-1, ST-2, ST-3, ST-4 and ST-5 are covered by forest and rubber plantations, some mining and the direct discharge of domestic wastewater from villages somehow contaminated the water. Overall these sites are somehow influenced by agricultural activities; however, less pollutant pressures from industrial and household wastewaters.

3.1.2 Group B: This group (cluster) could be regarded as relatively highly polluted (HP) sites of the river. Most of stations (ST-11, ST-13 and ST-16) are situated at midstream sites whereas ST-6 is situated in upstream and ST-18 and ST-20 are situated in downstream region. Most of sites are located industrial and city area; therefore, these sampling stations received pollutants mostly from domestic wastewater, wastewater treatment plants and industrial effluents. Beside these, untreated sewage of agriculture activities and surface runoff from Hatyai city also increased pollutants.

3.1.3 Group C: This group (cluster) could be regarded as relatively moderate polluted (MP) sites of the river. The stations ST-19 and ST-21 are situated at downstream region whereas for ST-14, ST-15 and ST-17 are situated in midstream region. Most of land uses in this cluster are agriculture, especially paddy fields, where relatively high levels of fertilizers and pesticides are used. Therefore, these inputs reflect the result of soil erosion and leaching. Most of land uses of ST-14, ST-15 and ST-17 are rubber plantation, so these sites receive pollution from non-point sources. The water quality of ST-21 is poor because this site is surrounded by ancient town Songkhla.

The results indicate that the CA technique is useful in offering reliable classification of surface water in the whole region and make it possible to design a future spatial sampling strategy in an optimal method, which can reduce the number of sampling stations and associated costs of sites [12-14].

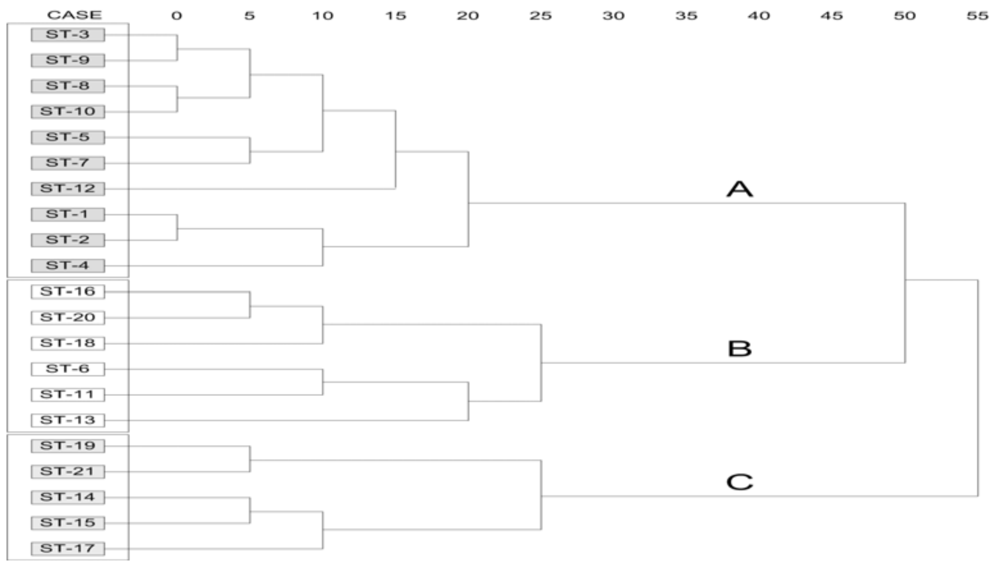


Figure 2 Dendrogram showing clustering of sampling sites according to surface water quality characteristics of URB

3.2 Correlation analysis (Interrelationship within parameters)

The correlation matrix of water quality parameters obtained from U-tapao river was examined. Some parameters showed significant correlationship. DO was negatively correlated to BOD ($r = -0.366$, $p<0.05$), EC ($r = -0.433$, $p<0.05$), TUR ($r = -0.267$, $p<0.01$), NH_3 ($r = -0.680$, $p<0.05$) and TP ($r = -0.527$, $p<0.01$). The negative relationship of DO with other parameters revealed the high organic pollution along with anthropogenic activities in the present study area. Inverse relationships DO with nutrients imply that the organic portion of nutrients play a major role for depletion of dissolved oxygen in the river system. There is negative correlation between TEMP and DO ($r = -0.136$, $p<0.01$), which is natural process; warm water easily becomes saturated with oxygen and thus can hold less DO. The value of DO decreases with increase in the value of most of the polluted parameters; thus, it can serve as a single useful index of river water quality. And, the strong negative relationship of DO with other parameters indicates a common anthropogenic source.

3.3 Discriminant Analysis (Temporal/spatial variations in river water quality)

In this study, temporal variations of river water quality parameters were first evaluated through a season parameter correlation matrix, using Spearman non-parametric correlation coefficients (r_s).The results shows that the parameters; TEMP, pH, BOD, DO, EC, TUR, FCB, NO_3, NO_2 and TP were found to be significantly ($p<0.05$) correlated with the season, except SS. Among these, temperature exhibited highest correlation coefficient ($r_s = 0.73$) followed by DO ($r_s= 0.52$). The season-correlated parameter can be taken as representing the major source of temporal variations in water quality [4].

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 seasons (dry and wet). Discriminant functions (DFs) and classification matrices (CMs) obtained from the standard, forward stepwise and backward stepwise modes of DA are illustrated in Table 2 and 3.The standard and forward stepwise mode DFs, using 12 and 9 discriminant variables, respectively, yielded CMs correctly assigning >69 % of the cases. However, in the backward stepwise mode, the DA produced a CM with close to 69.6 % of the case, assignment using only five discriminant parameters showing that TEMP, pH, DO, FCB and NH_3 were significant parameters of temporal variations in the river water quality.

Like temporal DA, spatial DA was performed with the same raw data set comprising 12 parameters after grouping into three major classes A, B and C obtained through CA. Just as temporal DA, discriminant functions and classification matrices were also obtained from the standard, forward stepwise, backward stepwise modes, showing in Tables 4 and 5. The standard and forward stepwise mode DFs use 12 and 7 discriminant variables, respectively, correctly assigning 63.3% and 63.9% of the cases to the two groups. In the backward stepwise mode, the DA produced a CM with close to 63.9% correct assignment using only 4 discriminant parameters, showing that TEMP, pH, DO and NH_3 were significant parameters of spatial variables.

3.4 Principal component/factor analysis (Source identification)

In case of temporal variation, PCA/FA was executed on 12 variables for two seasons, in order to identify important seasonal water quality parameters. An 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 [4]. Classification of factor loading is thus 'strong', 'moderate' and 'weak', corresponding to absolute loading values of > 0.75, 0.75-0.50 and 0.50-0.30, respectively [20]. Corresponding, variable loadings and explained variance are presented in Table 6 and strong loading values have been highlighted.

Table 2 Classification function coefficients for discrimnant analysis of temporal variation in water quality of URB

Parameters	Standard mode		Forward stepwise mode		Backward stepwise mode	
	Dry	Wet	Dry	Wet	Dry	Wet
TEMP	10.129	9.797	10.389	10.060	10.517	10.185
pH	5.989	6.198	6.541	6.744	5.847	6.025
BOD	-0.198	-0.228	-0.189	-0.279		
DO	3.953	3.838	3.836	3.720	3.898	3.800
EC	0.001	0.001	0.000	0.000		
SS	0.910	0.980				
TUR	0.342	0.337	0.280	0.275		
FCB	-2E-6	-2E-6	-2E-6	-2E-6	-8-E6	-3-E5
NH_3	13.038	14.256	6.155	7.301	8.376	9.322
NO_3	74.426	75.353				
NO_2	-54.398	-55.149				
TP	15.552	15.994	3.128	3.392		
Constant	-245.112	-238.226	-143.204	-185.188	-185.586	-177.281

Table 3 Classification matrix for discriminant analysis of temporal variation of water quality

		Period assigned by DA	
Seasons	**Percent correct**	**Dry**	**Wet**
Standard DA Mode			
Dry	48.4	92	125
Wet	79.5	221	57
Total	69.2	313	182
Forward stepwise DA mode			
Dry	48.4	92	125
Wet	80.2	223	55
Total	69.5	315	180
Backward stepwise DA mode			
Dry	49.8	95	122
Wet	80.9	225	53
Total	**69.6**	**320**	**175**

Table 4 Classification function coefficients for discrimnant analysis of spatial variation

WQP	Standard mode			Forward stepwise mode			Backward stepwise mode		
	Group A	Group B	Group C	Group A	Group B	Group C	Group A	Group B	Group C
TEMP	10.757	11.410	11.235	10.922	11.576	11.404	10.929	11.577	11.397
pH	7.272	7.757	7.437	7.112	7.588	7.275	7.157	7.644	7.341
BOD	-0.317	-0.294	-0.310	0.031	0.055	0.045			
DO	1.888	1.051	1.077	2.090	1.258	1.288	2.026	1.940	1.227
EC	0.001	0.001	0.001	0.000	0.000	0.000			
SS	0.920	0.916	0.927						
TUR	0.343	0.347	0.354						
FCB	-3E-5	-4E-5	-5E-5	-3-E5	-4E-5	-3-E5			
NH_3	15.417	15.063	16.192	10.247	9.940	11.005	9.985	9.739	10.811
NO_3	78.955	79.831	81.353						
NO_2	-59.265	-60.542	-61.598						
TP	17.306	17.638	17.895						
Constant	-251.33	-271.11	-266.30	-186.62	- 205.73	-198.83	-186.24	-205.30	-198.290

Table 5 Classification matrix for discriminant analysis of spatial variation

		Period assigned by DA		
Season	**Percent correct**	**Group A**	**Group B**	**Group C**
Standard Mode				
Group A	82.2	171	11	26
Group B	38.3	38	42	46
Group C	45.6	50	45	65
Total	63.3	259	98	137
Forward stepwise mode				
Group A	82.2	171	9	28
Group B	36.3	37	42	47
Group C	45.5	51	41	68
Total	63.9	259	92	143
Backward stepwise mode				
Group A	81.2	169	7	32
Group B	37.1	36	43	47
Group C	46.1	53	38	69
Total	63.9	258	88	148

For dry season, VF1, which explained 21.07% of the total variance, had strong positive loadings on EC and moderate loading on SS and TUR, and had high negative loadings on NO_3. This factor indicates that during this period, the surface runoff originated from the fields containing high load of solids from waste disposal source [4]. VF2, which explained 19.45% of the total variance, had strong negative loadings on pH and moderate negative loading on TEMP and positive loading on TP. This factor explains forming of organic acids leading to decrease pH. VF3, which explained 16.93% of total variance, had strong positive loading of NO_2. NO_2 has relationship with runoff from agriculture field, so the contamination in this period is associated with agriculture as well as solid waste disposal activities of cities and towns [23]. VF4, which explained 15.05% of the total variance, had strong positive loading on DO and moderate negative loading on NH_3.

For wet season, VF1, which explained 21.05% of the total variance, had strong positive loadings on TP and negative loadings on NH_3.VF2, which explained 18.64% of the total variance, had strong negative loading on NH_3 and moderate positive loading on DO. These factors indicate the decrease of organic pollution in this period. VF3, which explained 17.27% of the total variance, had high positive loading on EC and moderate negative loading on TEMP and moderate positive loading on TUR. This factor explains the erosion from upland areas during rainfall or similar events [4]. VF4, which explained 14.08% of the total variance, had strong positive loading on NO_3. This factor explains the pollution of river due to surface runoff.

From factor analysis of seasonal variation explained that, in dry season the river received pollution mostly from point sources as opposed to wet season, where it received higher amount of nutrients from surface runoff from agriculture areas. The strong positive loading EC and strong negative loading of pH on both seasons is a peculiar thing. Actually, U-tapao river is connected to Songkhla lake (or lagoon) from where salty water enters into river system. The concentration amount of salinity depends upon seasonal pattern and EC and pH are also related with salinity. So, there is high variation of EC and pH value and showed strong positive and negative loading in both seasons.

In case of spatial variation, Principal component analysis/factor analysis was performed on the normalized data sets (12 variables) separately for the three different regions, viz., groups A B, and C as delineated by CA techniques, to compare the composition pattern between analyzed water samples and identify the factors influencing each one PCA/FA of the three data sets yielded four VF for the group A, three VF for group B and four VF for group C with eigenvalues >1, explaining 71.62%, 71.77% and 72.01% of the total variance in respective water quality data sets (Table 7).

For the data set pertaining to group A, among four VFs, VF1, which explained 22.35% of total variance, had strong positive loading on BOD and NO_3 and moderate positive loading on TUR and NH_3. Since, BOD and NO_3 represent organic pollution and this factor represents the contribution of non-point source pollution from rubber plantation areas. In these areas, farmers use the nitrogenous fertilizer, which undergo nitrification processes, and the rivers receive nitrate nitrogen via groundwater leaching [4] and moderate loading of TUR explains the erosion from upland areas.VF2, which explained 19.36% of total variance, had strong positive loading on DO and SS and moderate positive loading on FCB and moderate negative loading on TEMP. Since the inverse relationship between TEMP and DO is a natural process, the warmer water becomes saturated more easily with oxygen and it can hold less dissolved oxygen [16]. Strong positive loading of SS and FCB explain domestic wastewater contaminated by fecal pollution from local livestock, especially pig farming. VF3, which explained 17.70% of the total variance, had positive loading on EC and NO_2. Generally, geological deposits, natural or organic matter decomposition and agriculture runoff are sources of NO_2 [16]. VF4, which explained 15.74% of total variance, had strong negative loading on TP and moderate positive loading pH. This factor represents the erosion effect during cultivation of soil and associated organic matter.

For group B, VF1, which explained 35.04% of total variance, had strong positive loading on pH, EC and NO_2 and moderate positive loading on TUR. The existence of high loading on pH and EC is due to a lot of ions from industrial pollution. This factor explains the anthropogenic activities on the surrounding areas by the physiochemical source of variability [4]. And strong positive loading of NO_2 with positive loading of TUR indicates the relationship of river runoff from the agricultural field along with waste disposal activity [26]. VF2, which explains 22.65% of total variance, had strong positive loading on FCB and moderate positive loading on TP. This factor explains the effects of pollution from domestic waste as well as livestock waste from surrounding areas because FCB is strongly related to municipal sewage and wastewater treatment plants along the river [4]. VF3, which explained18.32% of total variance, had high negative loading on DO and positive loading on NH_3. Negative relationship between NH_3 and DO can be explained such that high levels of dissolved organic matter consume large amount of oxygen [16]. This factor explains the organic pollution from municipal sewage and industrial wastewater [14].

For group C, VF1, which explained 19.82% of total variance, had high positive loading on TP and moderate negative loading on TUR. This factor explains agricultural runoff from phosphorous based fertilizers and the domestic wastewater particularly containing detergents contribute to elevated levels of phosphorous in surface waters [26]. VF2, which explained 18.35% of total variance, had high positive loading on pH and negative loading on NO_2. VF3, which explained 17.97% of total variance, had moderate positive loading on TEMP, DO and EC. This factor explains the downstream dilution effect of water. VF4, which explained 16.67% of total variance had strong positive loading on NH_3. This factor explains the pollution from domestic wastes and stream bed material [26].

From PCA/FA showed that the level of pollution generally increases from upstream to downstream of the river. Overall, there were three types of pollution in the study area: organic pollution, nutrients pollution, and fecal pollution. The Group B sites (HP) influenced by household wastewater presented the highest concentrations of nutrients and extremely high pollution due to discharge of wastewater from industry and domestic. The Group C sites (MP) were influenced from pollution from agriculture runoff and domestic waste from city area, since domestic wastewater discharges from the dense combined sewer system from city, fecal pollution was one of the potential pollution sources for both Group B and Group C. The Group A (LP) sites were influenced especially by agricultural facilities; however, this was less pollutant pressures than industrial and household wastewaters.

Table 6 Loadings of experimental variables (12) on principal components for two seasons

Variables	Dry Season				Wet season			
	VF1	VF2	VF3	VF4	VF1	VF2	VF3	VF4
TEMP	-0.488	0.646	0.248	-0.339	0.256	0.113	-0.738	-0.066
pH	-0.231	**-0.912**	-0.186	0.116	**-0.827**	0.253	0.158	0.128
BOD	0.132	0.144	0.520	-0.230	0.438	-0.480	0.289	0.226
DO	0.198	-0.201	-0.132	**0.876**	-0.204	0.725	0.062	0.020
EC	**0.764**	-0.601	0.132	0.155	-0.102	0.238	**0.957**	-0.127
SS	0.562	0.079	-0.037	0.245	0.031	0.456	-0.034	-0.334
TUR	0.742	0.077	0.117	-0.024	0.423	-0.207	0.626	-0.215
FCB	-0.181	0.012	-0.139	0.411	-0.154	0.329	-0.241	0.048
NH_3	-0.273	0.100	-0.204	-0.730	-0.035	**-0.994**	-0.067	-0.067
NO_3	**-0.753**	-0.152	0.534	0.351	-0.028	-0.006	-0.190	**0.981**
NO_2	-0.177	0.224	**0.943**	0.046	0.639	-0.009	0.051	0.675
TP	0.122	0.745	0.616	0.036	**0.986**	-0.014	-0.074	0.144
Eigenvalue	2.529	2.334	2.033	1.861	2.580	2.237	2.073	1.691
% Total variance	21.075	19.451	16.938	15.506	21.504	18.641	17.278	14.088
Cumulative % variance	21.075	40.526	57.464	72.970	21.504	40.145	57.423	71.511

Table 7 Loadings of experimental variables (12) on principal components for Group A, Group B and Group C data sets

	Group A				Group B				Group C		
	VF1	**VF2**	**VF3**	**VF4**	**VF1**	**VF2**	**VF3**	**VF4**	**VF1**	**VF2**	**VF3**
TEMP	0.294	-0.612	0.144	0.326	0.333	-0.209	0.691	-0.072	-0.738	0.074	0.431
pH	-0.091	-0.149	0.399	0.732	-0.286	**0.772**	-0.074	0.350	**0.759**	-0.103	-0.164
BOD	**0.750**	-0.217	-0.195	-0.016	0.158	-0.713	0.009	0.350	0.328	0.247	0.656
DO	0.514	**0.770**	-0.082	-0.147	-0.463	0.078	0.528	-0.171	-0.013	-0.036	**-0.812**
EC	-0.148	0.033	**0.872**	0.219	-0.352	0.176	0.736	-0.006	**0.998**	0.053	0.009
SS	-0.077	**0.750**	0.395	0.273	0.101	0.472	0.096	-0.387	0.487	-0.102	-0.140
TUR	0.647	0.212	-0.334	0.042	-0.737	-0.252	0.349	-0.155	0.712	-0.230	0.514
FCB	-0.273	0.653	0.133	0.202	0.473	0.244	0.440	0.497	-0.130	**0.989**	-0.030
NH_3	0.686	0.025	0.036	-0.494	-0.347	-0.117	-0.360	**0.794**	-0.279	-0.127	**0.926**
NO_3	**0.845**	-0.425	-0.128	-0.009	0.698	0.363	-0.275	-0.367	-0.572	0.437	-0.270
NO_2	-0.233	0.175	**0.913**	-0.076	0.317	**-0.792**	0.041	-0.345	**0.899**	0.307	0.088
TP	0.032	-0.198	-0.050	**-0.900**	**0.767**	0.142	0.449	0.258	-0.480	0.737	0.445
Eigenvalue	2.683	2.323	2.124	1.890	2.259	2.104	2.037	1.942	4.206	2.718	2.198
% Total variance	22.359	19.361	17.701	15.746	19.826	18.534	17.978	16.179	35.047	22.650	18.320
Cumulative % variance	22.359	41.750	59.422	75.168	19.826	37.361	54.339	70.518	35.047	57.697	76.017

4. Conclusions

In this study, multivariable statistical methods were successfully applied to evaluate temporal and spatial variations in river water quality and source identification at the monitoring sites in U-tapao River Basin. Hierarchical cluster analysis grouped 21 sampling sites into three groups, i.e., less polluted area, moderate polluted area and high polluted area based on their similarity of water quality characteristics. Based on obtained information, it is possible to design an optimal sampling strategy, which could reduce the number of sampling stations and associate costs. Also this analysis allowed the identification of three different zones in the river, with different water quality. From correlation analysis, the negative relationship DO with other parameters reveals the high organic pollution along with anthropogenic activities in the basin. Discriminant analysis gave the best results both spatially and temporally. For the temporal variation analysis, the DA used only five parameters (TEMP, pH, DO, FCB and NH_3) with close to 69.6% correct assignment. It was found that a parameter that can be significant in contribution to water quality variations in river for one season may less or not be significant for another one. For the spatial variation analysis, the DA also used only four parameters (TEMP, pH, DO and NH_3) and correctly assigned about 63.9%. 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. Although the factor/principle component analysis did not result in a significant data reduction, it helped extract and identify the factors/sources responsible for temporal and spatial variations in river water quality. Factor analysis explained in dry season the river received comparatively high amount point source pollution from domestic and industrial sector whereas in wet season, the received pollution from non point source like surface runoff from agriculture and residential areas. Varifactors obtained from factor analysis indicate that the parameters responsible for water quality variations are mainly related to temperature and organic pollution in relatively less polluted areas and organic pollution and nutrients in both medium and highly polluted areas in the basin. Considering the results of the measured physiochemical water parameters and the results of factor and cluster analyses, the agriculture and urban land use were the most contributing factors to the pollution of the river. 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. It is recommended to Environmental Office-16, Songkla to adjust more water quality parameters for effective monitoring system and might reduce the monitoring stations for cost benefit purpose.

5. Acknowledgements

The authors are highly thankful to the staffs of Environmental Office-16, Songkhla and GEO-Informatics Research Center for Natural Resource and Environment, Hatyai, Thailand for all the support provided for the study.

References

[1] Reza, R. and Singh, G. **2010**. Heavy Metal Contamination and its Indexing Approach for River Water. *International Journal of Environmental Science and Technology*, *7*(4), 785-792.

[2] Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V. and Ahiska, S. **2008**. Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake. Turkey. *Environmental Monitoring and Assessment*, *144*, 269-276.

[3] Zhou, F., Liu, Y. and Guo, H. **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*, 1-13.

[4] Shrestha, S. and 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] Pejman, A. H., Bidhendi, G. R., Karbassi, A. R., Mehrdadi, N. and Bidendi, M.E. **2009**. Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *International Journal of Environmental Science and Technology, 6*(3), 467-476.

[6] Liao, S. W., Gau, H. S., Lai, W. L., Chen, J. J. and Lee, C. G. **2007**. Identification of pollution of Tapeng Lagoon from neighbouring rivers using multivariate statistical method. *Journal of Environmental Management*, *88*(2), 286-292.

[7] Mahvi, A. H., Nouri, J., Babaei, A. A. and Nabizadeh, R. **2005**. Agricultural activities impact on groundwater nitrate pollution. *International Journal of Environmental Science and Technology*, *2*(1), 41-47.

[8] Carpenter, S. R., Caraco, N. E., Correll, D. L., Howarth, R. W. and Smith, V. H. **1998.** Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, *8*(3), 559-568.

[9] Jarvie, H. P., Whitton, B. A. and Neal, C. **1998**. Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. *Science of The Total Environment, 210*, 79-109.

[10] Kuppusamy M. R. and Giridhar V. V. **2006**. Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environment International*, *32* (2), 174-179.

[11] Tanriverdi, C., Alp, A., Demirkiran, A. R. and Ucjarde, F. **2010**. Assessment of surface water quality of the Ceyhan River basin, Turkey. *Environmental Monitoring and Assessment*, *167*, 175-184.

[12] Singh, K.P., Malik, A. and 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.
[13] Najafpour, S., Alkarkhi, A.F.M., Kadir, M.O.A. and Najafpour, G.D. **2008**. Evaluation of Spatial and Temporal Variation in River Water Quality. *International Journal of Environmental Research*, *2*(4), 349-358.
[14] Singh, K.P., Malik, A., Mohan, D. and 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.
[15] Vega, M., Pardo, R., Barrado, E. and Deban, L. **1998**. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis.*Water Research*, *32*, 3581-3592.
[16] Varol, M. and Sen, B. **2009**. Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environmental Monitoring and Assessment, 159*, 543-553.
[17] Simeonov, V., Simeonova, P. and Tsitouridou, R. **2004**. Chemometric quality assessment of surface waters: two case studies. *Chemical and Engineering Ecology*, *11* (6), 449-469.
[18] Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M. and Fernandez, L. **2000**. Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis. *Water Research*, *34*, 807-816.
[19] Bengraine, K. and Marhaba, T. F. **2003**. Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials*, *100*, 179-195.
[20] Liu, C.W., Lin, K.H. and Kuo, Y.M. **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.
[21] Sundaray, S.K. **2010**. Application of multivariate statistical techniques in hydrogeochemical studies-a case study:Brahmani–Koel River (India). *Environmental Monitoring and Assessment, 164*, 297-310.
[22] Charkhabi, A. H. and Sakizadeh, M. **2006**. Assessment of Spatial Variation of Water Quality Parameters in the Most Polluted Branch of the Anzali Wetland, Northern Iran. *Polish Journal of Environmental Studies*, *15*(3), 395-403.
[23] Zhao, Z. and Cui, F. **2009.** Multivariate statistical analysis for the surface water quality of the Luan River, China. *Journal of Zhejiang University SCIENCE A*, *10*(1), 142-148.
[24] Huang, J., Ho, M. and Du, P., **2011**. Assessment of temporal and spatial variation of coastal water quality and source identification along Macau peninsula. *Stochastic Environmental Research and Risk Assessmen*t, *25*, 353-361.
[25] Zhang, Y., Guo, F., Meng, W. and Wang, X., **2009**. Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. *Environmental Monitoring and Assessment*, *152*:105-121.
[26] Yang, Y.H., Zhou, F., Guo, H.C., Sheng, H., Liu, H., Dao, X. and He, C.J., **2010**. Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environmental Monitoring and Assessment*, *170*, 407-416.