KKU ENGINEERING JOURNAL July – September 2015;42(3):235-242 Research Article

KKU Engineering Journal

https://www.tci-thaijo.org/index.php/kkuenj/index

Water quality parameters reduction of the Lamtakong river basin (Thailand) using statistical analyses

Somphinith Muangthong*

Faculty of Engineering and Architecture, Rajamangala University of Technology Isan, Nakhon Ratchasima 30000, Thailand.

Received November 2014
Accepted April 2015

Abstract

Data reduction is very important when dealing with river pollution analysis. Lamtakong River is a small river but it is one of the most significant water resources in Thailand for several community water supply systems, large farming and irrigation areas, and waste water drainage systems. This research aimed to reduce the large number of locations and water quality parameters down to a manageable number. Cluster analysis, a statistical method for sorting groups of similar data into clusters, was applied for grouping of homogeneous locations based on water quality. Of total 15 monitoring sites, 7 sites were chosen for cluster analysis which distinguished them into 2 groups, the upstream and downstream groups. Factor analysis, a statistical method for selecting a few representative variables from many variables, was used to reduce the number of water quality parameters. In this study, the factor analysis technique reduced 17 water quality parameters down to 8 for both the upstream and downstream sections. Five of the reduced parameters are the same for both river sections. The three exceptions demonstrate the different characteristics of the 2 river sections.

Keywords : Factor analysis, Cluster analysis, Dendrogram, Lamtakong river**,** Water quality

1. Introduction

Lamtakong River is one of the uppermost tributary of the Mun River. With a total length of 220 km, it is considered a small river. However, it is very crucial for the support of socioeconomic and ecological assets along its path. Lamtakong River is the main source of water supply for several municipalities, including the cities of Nakhon Ratchsima, Pakchong, Sikew and several townships, serving a combined population of more than 900,000. It also supplies irrigation water to large farming and paddy field areas as well as industrial water supply for several industries. Lamtakong River also acts as waste water drainage systems for the same communities it supplies water to cage fishery is also popular along the river, especially in the reservoir area. These activities cause serious water pollution to the watercourse.

In order to control and prevent the river pollution, the causes of contamination must be determined and understood first. The surface water quality of a river must meet the standards set by the Pollution Control Department (PCD) which classifies a reach of a river into 5 classifications based on water properties, i.e. (i) extra clean, (ii) very clean, (ii) medium clean, (iv) fairly clean, and (vi) does not meet standard. Of the 5 classes, classes 1 and 5 have no specified contaminant level. Classes 3 and 4 are identical except for the specified levels of dissolved oxygen (DO), biochemical oxygen demand (BOD), total coliform bacteria (TCB), and fecal coliform bacteria. Since the Lamtakong originated from the world heritage Khao Yai National Park, its headwater flow is pristine and this reach meets class 1 designation. Along its course, pollutant inputs, from both point sources and nonpoint sources, impair the water quality. The PCD assigned the surface water quality classification for two reaches of the Lamtakong on March 31, 1999 [1]. The first reach, from its confluence with the Mun River upstream-ward to Khonchum diversion dam, about 24 km, was assigned class 4. The second reach, from Khonchum diversion dam up to Ban Bokrached, about 256 km, was assigned class 3. The study of Anantpipatkij during 1996 to 1998, however, found that the Lamtakong was so polluted that it fell into class 5 [2]. Some later studies showed improvement in the water quality but not by much [3-4]. Unquestionably, water quality improvement of the Lamtakong River needs to be achieved through better water resource and quality management [5]. The PCD's monitoring program track 29 water quality parameters at 15 sites along the main and tributaries of the Lamtakong. These numbers are too large for effective analysis and therefore management. Most of the past studies determined the river water quality based only on one or two parameters, e.g. BOD and TCB, and using data from only four monitoring sites downstream of the Lamtakong dam. It is questionable whether using only BOD and TCB from downstream reaches for analysis would render effective and accurate results.

In order to obtain the most representative set of water quality parameters and groupings of homogeneous reaches, multivariate data analyses need to be performed. Traditionally, water quality data are recorded in a table or

*Corresponding author. Tel.: +66 4423 3000
Email address: somphinith.mu@rmuti.ac.th
doi: 10.14456/kkuenj.2015.25

matrix form by parameter and observation (or by monitoring sites) e.g. Table 3. Data analysis starts from this matrix, usually using cluster analysis to analyze observations and factor analysis to select a set of the most representative parameters [6]. The objective of this study was therefore to reduce the number of each of the two data sets collected by the PCD, i.e. 15 monitoring sites and 29 water quality parameters, down to their optimum sizes for water quality management purposes.

2. Materials and methods

2.1 The study area

One of the shortest but most significant river in Thailand is the Lamtakong River. It originates from Khao Yai National Park, a beautiful World Heritage in Dong Phyayen mountain range the gateway from the Central to the Northeast of Thailand. The river flows north then turns eastward to join the Mun River, the longest river in southern part of the Northeast of Thailand. The Lamtakong, with its length of 220 km, passes through several municipalities, including the city of Nakorn Ratchsima, large farming areas, paddy fields, upland cropping areas, and many industrial ares [7]. It supplies fresh water to all involved sectors, including restoring its ecology. More importantly, it collects and conveys the wastes from all sectors away. These wastes degrade the water quality of the river.

The Lamtakong watershed is located in the Northeast of Thailand (Figure 1) between 14º 15´N – 15º 15´N and 101º 15´E – 102º 30´E at the boundary with Central Thailand. The watershed area can be divided into 2 parts: the upper part from the Lamtakong dam upstream-ward to the watershed divide, and the lower part from the dam site downstream-ward to the confluence. The total watershed area covers 3,874 km^2 [5]. The upper subwatershed is steep, rough, and mountainous with the highest elevation of 1,208 m above mean sea level (amsl) covers 1,490 km^2 with the length of the river reach of 116 km. For the lower subwatershed, the river reach begins at the Lamtakong dam site with the elevation of 270 m (amsl) down to the confluence with the Mun River at the elevation of 165 m (amsl). The lower subwatershed is much flatter and covers the river length of 104 km. [8-9]. At about 53 km upstream of the confluence, near Kham Talesor township, a distributary called Lamboriboon departs from the main Lamtakong, flows parallel, and merges back with it at about 3 km. upstream of the confluence.

The Lamtakong River experiences flood in the rainy season and drought in dry season nearly every year. The average annual rainfall of the watershed is about 1,200 mm. with more than 80% from the rainy season. This watershed exhibits the steepest precipitation gradient in Thailand. Its annual mean precipitation decreases from 2,300 mm. at the river head to about 1,000 mm. at the downstream end [9].

The average temperature in the Basin is between 25.9ºC–27.1ºC, with the highest average in April at 29.8 ºC, and the lowest in December at 22.2ºC. The average annual humidity is between 70%–74%. The annual average potential evaporation rate is between 1,768 mm to 1,813 mm. The average wind speed is between 1.2–2.9 knots. Furthermore, Lamtakong Basin contains one of the rare pumped storage hydropower plant in Thailand with capacity of 1,000 MW [10].

2.2 Water quality monitoring

From the 15 water quality monitoring sites undertaken by the PCD, 7 sites were chosen for this study because of their locations (along main water course at reasonable spacing) and continuity in recorded data. Figure 1 and Table 1 show the 7 monitoring locations: LT01, LT02, LT03, LT04, LT05, LT06, and LT07 from downstream to upstream. Similarly, from the 29 water quality parameters monitored by the PCD, 17 most representative parameters were selected for data analysis. This study involved the analysis of water quality data from 1996 to 2012.

The 17 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 solids, total dissolved solids, suspended solids and calcium carbonate ($CaCO_3$). Water quality parameters, their units, and methods of analyses are summarized in Table 2. Table 3 summarizes the mean values of each parameter at each monitoring station.

Table 1 Location of the water quality monitoring stations

Station	Location	Coordinate	
		X	Y
LT01	Lamtakong Bridge, BanYongyeang	199788 N	1660829 E
LT02	Watsamakkee Bridge, Naimuang	189441 N	1658027 E
LT03	Kodchanoun Bridge, Midtrapap	793436 N	1662331 E
LT04	Lamtakong water supply, Krongpai	772363 N	1639488 E
LT05	TOT Academy bridge, Pakchong	764986 N	1632922 E
LT06	Nongsaray bridge, Pakchong	762057 N	1628123 E
LT07	Bokrached bridge, Kanongpra	760902 N	1619516 E

Most of these parameters are of different units as shown in Table 1, therefore, they must be standardized prior to performing any data analysis. The standardized matrix is derived from the water quality matrix by transforming the parameter values in Table 3 into z-score values so that all parameters will have the same mean and variant values of 0 and 1, respectively. The z-score is a new variable obtained by transformation ($z = (x - x_a)/s$), where x is the parameter to be transformed, x_a is the average value, and s is the standard deviation. For example, in Table 3, considering the first parameter, water temperature values are 29.42, 28.57, 27.79, 27.86, 26.19, 26.23, and 26.13. Its average is 27.46 and standard deviation is 1.31. Therefore, the z-score values are 1.50, 0.85, 0.26, 0.31, -0.97, -0.94, and -1.02 for stations LT01 to LT07, respectively. The actual values in Table 3 are to be replaced by the z-score values for all parameters.

2.3 Classification of homogeneous areas

To understand the occurrence and characteristics of the river pollution of the whole Lamtakong River we need to divide the river into sections possessing homogeneous water quality characteristics. The classification of homogeneous areas can be performed using cluster analysis (CA). Cluster analysis is for classifying data into several similar characteristics groups (i.e. clusters). The objective of CA in this study is therefore to identify groups of monitoring sites of the Lamtakong River that are homogeneous in water quality. There are two main techniques of cluster analysis namely hierarchical and non-hierarchical clustering. An example of cluster analysis is provided below.

Suppose we need to group a 9-type vegetation i.e. beech, corn, jackfruit, mango, oak, rambutan, rice, sorghum, and sugarcane. By nonhierarchical method, suppose we want to classify them into 2 groups. Subjectively, we can classify them by cotyledon concept. The first group is monocotyledon i.e. corn, rice, sorghum, and sugarcane. The second group is dicotyledon. Alternatively, if we want 3 groups, we may further classify the dicotyledon into 2 groups by climatic concept, tropical and temperate. The tropical vegetation includes jackfruit, mango, and rambutan, while oak and beech are temperate tree.

The hierarchical clustering method classifies objects into hierarchical orders. If we use the subjective last example: the highest hierarchy is vegetation or plant kingdom, the lowest is individual of 9 items. The classification of the orders in between the lowest and highest can be done either from the highest downward called *divisive* procedure, or from the lowest upward called *agglomerative*. Here the above subjective example can be demonstrated in case of agglomeration. Each of the 9 items is a separate kind of vegetation, therefore they are different from each other, says, for 1 unit. Second step: considering corn, rice, and sorghum as monocotyledon grain crops in one group; sugarcane as monocotyledon sugar crop another group; these 2 groups are different by another 1 unit, making up 2 units. Then in the second step: now considering jackfruit, mango, and rambutan as dicotyledon fruit trees in one group, and beech and oak are timber in another group with differentiation of 1 unit. These two groups are also different in climate - the first is tropical and the second, temperate, making up another 1 unit with 3 units overall. The final step merges monocotyledon group to the dicotyledon with the difference of 1 unit making up overall 4 units. Figure 2 shows the dendrogram, of this example. Dendrogram is a graphic diagram showing items to be grouped on the vertical axis and the length representing the difference among groups on the horizontal axis. It enables clustering for any number of groups. For example, if we want to group them into 3 clusters, we would draw a vertical line across at the length of 2.5 units. We should get 3 clusters as: (i) corn, rice, sorghum, and sugarcane; (ii) jackfruit, mango, and rambutan; (iii) beech and oak.

Table 2 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 Solids	TDS	mg/l	Total dissolved solids dried at 103-105ºC
16	Suspended Solids	SS	mg/l	Suspended solids dried at 103-105ºC
17	Calcium carbonate	$CaCO_3$	mg/l	EDTA Titrimetric method

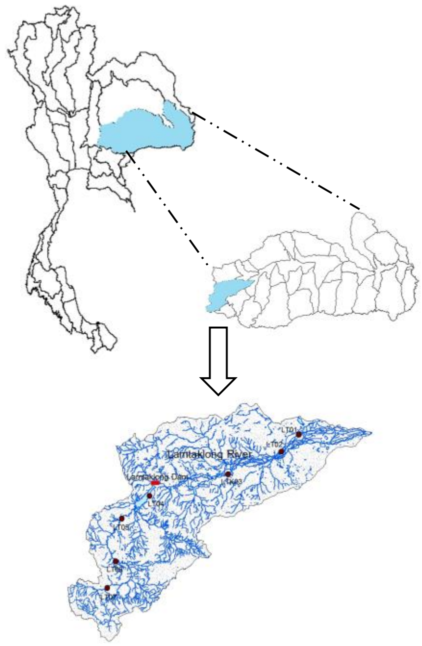

Figure 1 Map of the monitoring sites

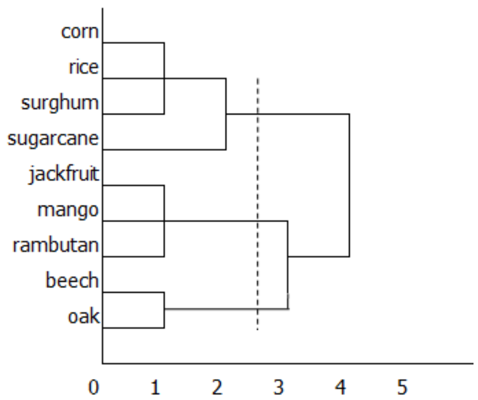

Figure 2 Example of a Dendrogram

In contrast to the above example, the cluster analysis, for grouping monitoring sites by the similarity of their water quality characteristics, must be performed *objectively* from a standardized matrix of the water quality data. Agglomerative procedure, which is simpler than the divisive procedure, was chosen for this study. The standardized water quality matrix was initially turned into a Euclidean distant matrix. This matrix, showing the Euclidean distance of water quality values between each pair of monitoring sites, is a symmetric matrix with 0's on the main diagonal. The Euclidean distance, for a pair of 2 stations, can be calculated by mimicking the distance of two points in three dimensional space, e.g. (x_1,y_1,z_1) and (x_2,y_2,z_2), then the Euclidean distance between points 1 and 2 is $ED_{1,2} = ((x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2)^{1/2}$. Therefore, the variables with subscript 1 and 2 are the water quality parameters of stations 1 and 2, respectively.

In the first step of agglomeration procedure, each individual station belongs to its own cluster. Then for the next several steps, the most similar sites are merged together into bigger clusters. There are three main procedures to identify similarities and to merge any two clusters, i.e. single linkage clustering, centroid sorting, and Ward's method. The single linkage, or nearest neighbour method, is the simplest of all [11]. It can be done manually when clustering a small number of items, such as the 7 monitoring sites selected for this study. The Euclidean distance matrix is the starting point which helps to determine and merge the shortest distance pair. Mooi and Sarstedt [12] gives a good example.

The centroid clustering method starts from the Euclidean distance matrix and merges the pair of shortest distance. Once a cluster contains more than one monitoring sites then the shortest distance between centroids of a pair of 2 clusters is to be the criterion for merging the clusters. The centroid of each cluster can be obtained by averaging all z-score values in the cluster [13].

Ward's method considers the similarities between two clusters by considering the least increment of variance after the two clusters are merged (Ward, 1963). This can be calculated from $V_{rs} = n_r n_s d_{rs}^2/(n_r+n_s)$, where V_{rs} is incremental variance when clusters r and s merged, n_r and n_s are numbers of z-score values of clusters r and s, respectively, and d_{rs} is ED from the centroid method [14]. Ward's method was utilized in this study because it is the most reasonable [15].

2.4 Representative water quality parameters

After subjectively reducing 29 water quality parameters down to 17, it still needed to be further reduced to obtain fewer most representative parameters influencing river pollution. This reduction is to be performed using a multivariate statistical method called factor analysis (FA) [16]. There are two main types of factor analysis, *principal component analysis* (PCA) and *common factor analysis* (CFA) [15]. The principle and result of the two methods are similar however the former is preferable [17]. The latter is mainly for analysis of the underlying structure of water pollution. The former however is for reducing the number of the parameters which is the aim of this study. Since there are 17 parameters, $x_1, x_2, x_3, \ldots, x_{16}, x_{17}$, the PCA model can be written as $\mathbf{X} = \mathbf{bY}$ where $\mathbf{X}$ is a column vector of observed parameters, $\mathbf{b}$ is a square matrix of eigenvectors (weights), and $\mathbf{Y}$ is a column vector of component scores. The element number of **X** must equal to that of **Y**. Normally, both methods should be applied in sequential manner by applying PCA first followed by CFA [16]. The PCA is the main concern in this study. The application of PCA involves 6 steps as follows [18]: (i) preparation of a data matrix consisting of 7 monitoring stations and 17 water quality parameters; (ii) computation of covariance matrix of the 17 parameters; (iii) extraction of eigenvalues and eigenvectors for 17 factors; (iv) selection of the number of factors based on eigenvalues higher than 1 (Kaizer's criterion); (v) rotation of the factor loadings matrix to obtain a simpler structure; and (vi) interpretation of the results.

Water quality data matrix, normally, consists of parameters on the columns and monitoring stations on the rows which is just opposite to Table 3. The units of these parameters are not the same. They need to be standardized, the mean to zero and the variance to 1, so that they are comparable. From the standardized data matrix, the covariance matrix can be created. The covariance matrix composes covariance of each pair of the water quality parameters. It is a 17 by 17 square matrix with 1's on the main diagonal. Since it is a square matrix, we can calculate the eigenvalues and eigenvectors of this matrix. In our case, we get 17 eigenvectors and also 17 eigenvalues. Then the eigenvectors is to be arranged in a matrix form by column

Table 3 Range, mean, and S.D. of water quality parameters of Lamtakong River

Parameters		LT01	LT02	LT03	LT04	LT05	LT06	LT07
WT (ºC)	Range	23.00-35.5	23.2-32	24.00-33.00	24.00-31.70	20.00-32.00	23.50-32.00	23.00-31.20
	mean	29.42	28.57	27.79	27.86	26.19	26.23	26.13
	Std.	3.07	2.26	2.03	1.89	2.01	1.70	1.77
pH	Range	6.82-8.70	6.50-8.00	6.50-8.30	6.50-9.00	6.90-8.40	6.90-8.30	6.90-8.20
	mean	7.58	7.42	7.68	8.07	7.83	7.77	7.65
	Std.	0.48	0.34	0.48	0.69	0.34	0.36	0.35
Turbidity (NTU)	Range	1.80-74.67	2.10-70.00	1.30-71.67	1.20-51	1.00-75.00	1.00-75.20	0.70-67.00
	mean	28.78	15.63	20.92	7.67	19.00	18.76	14.22
	Std.	22.18	13.29	19.63	12.00	22.92	21.22	19.26
Conductivity (µS/cm)	Range	126.0-1280	277-1480	207-390	196-345	175-646	178-600	164-582
	mean	654.71	627.87	304.40	287.60	453.09	438.13	403.44
	Std.	237.38	291.78	38.09	37.82	132.55	127.39	130.38
Salinity (ppt)	Range	0.00-0.50	0.00-0.70	0.00-0.20	0.00-0.20	0.00-0.30	0.00-0.30	0.00-0.30
	mean	0.17	0.16	0.06	0.05	0.11	0.10	0.11
	Std.	0.18	0.20	0.07	0.06	0.11	0.11	0.11
DO (mg/l)	Range	1.40-7.60	0.20-5.10	3.10-7.60	3.0-11.0	3.80-9.00	3.00-7.50	3.90-8.30
	mean	4.47	2.63	5.85	6.85	6.25	5.52	6.28
	Std.	1.73	1.41	1.00	1.56	1.18	1.03	1.02
BOD (mg/l)	Range	0.80-9.70	0.90-9.60	0.10-5.36	0.60-4.90	0.50-6.60	0.30-8.40	0.30-7.30
	mean	5.17	4.27	1.80	1.93	1.99	2.21	1.91
	Std.	2.30	2.45	1.20	1.08	1.36	1.51	1.63
TCB (MPN/100 ml)	Range	1100-160000	9000-240000	60-160000	2-16000	20-160000	400-16000	170-50000
	mean	38447.06	114823.53	17110.88	787.38	26210.00	43144.12	4160.00
	Std.	47655.46	79431.87	39833.27	2744.95	45594.00	52493.38	9058.29
FCB (MPN/100 ml)	Range	20-50000	20-60000	20-30000	2.00-2400	20-16000	20-28000	2.00-5000
	mean	9249.71	20574.71	3163.65	194.59	3464.71	9743.53	582.12
	Std.	12550.47	15343.60	6599.95	437.96	4377.00	8283.87	925.20
TP (mg/l)	Range	0.10-2.99	0.10-2.75	0.03-0.41	0.01-1.22	0.04-1.61	0.05-0.72	0.01-1.30
	mean	1.07	0.72	0.15	0.15	0.30	0.23	0.19
	Std.	0.80	0.60	0.11	0.28	0.30	0.19	0.29
NO_3-N (mg/l)	Range	0.00-3.07	0.00-1.50	0.00-1.70	0.00-0.92	0.00-4.40	0.00-4.90	0.00-4.00
	mean	0.59	0.35	0.35	0.18	0.90	1.09	0.87
	Std.	0.72	0.45	0.40	0.29	0.99	1.24	0.99
NO_2-N (mg/l)	Range	0.00-0.80	0.00-0.18	0.00-0.17	0.00-0.11	0.00-0.30	0.00-0.39	0.00-0.08
	mean	0.10	0.04	0.03	0.01	0.05	0.06	0.01
	Std.	0.14	0.04	0.05	0.02	0.08	0.10	0.02
NH_3-N (mg/l)	Range	0.01-4.60	0.01-4.10	0.00-3.46	0.00-2.46	0.01-2.46	0.01-2.46	0.01-2.46
	mean	1.27	1.07	0.38	0.29	0.39	0.40	0.27
	Std.	1.38	1.08	0.70	0.57	0.65	0.67	0.60
SS (mg/l)	Range	9.00-93.00	2.00-84.00	4-116	1-50	1-199	2-106	1-90
	mean	41.29	27.85	27.26	7.71	27.15	23.30	18.67
	Std.	23.71	19.95	23.30	9.99	29.67	21.76	23.26
TD (mg/l)	Range	314-742	132 -740	124-324	188-310	54-616	138-704	124-686
	mean	439.76	370.24	213.82	186.10	321.47	317.56	275.74
	Std.	108.68	134.00	36.65	34.68	97.20	97.47	94.78
TDS (mg/l)	Range	255-931	193-994	148-273	116-301	105-433	121-402	98-399
	mean	419.82	388.43	200.18	189.28	296.68	281.39	254.34
	Std.	135.06	183.18	27.87	32.63	87.49	75.33	78.18
$CaCO_3$ (mg/l)	Range	101.53-450.50	99-432.50	109-313	66-297	78.40-405	119.40-575	96-560
	mean	190.66	172.18	139.63	133.72	209.08	246.72	257.08
	Std.	93.63	70.71	46.17	48.31	81.30	84.14	89.83

from the highest eigenvalue to smaller ones. For this step, those eigenvectors with very small eigenvalues can be ignored. In our case, we retained only 5 eigenvectors (or principal components). The mathematical explanation is in Martinez and Martinez [14]. Now it is needed to rotate the principal axes so the distribution of strong loading factors to be spread over the 5 components.

3. Results and discussion

3.1 Spatial classification

Cluster analysis was used to group 7 Lamtakong monitoring sites with similar water quality. The result is ilustrated as a dendrogram in Figure 3 showing the observation sites on the Y-axis and similarities on the X-axis. The similarity between a pair of clusters was determined by Ward's method as the smallest of variance increments and indicated by a dlink (linkage distance). To make the illustration easy to grasp, the ratio of linkage distance to its maximum value ($dlink/dlink_{max}$) was used. This ratio is called *length*. Figure 3 shows the first classification into 3 clusters, with the length of 1 unit: cluster A are LT05, LT06, and LT07; cluster B are LT03 and LT04; cluster C, LT01 and LT02. For the next higher hierarchy clusters A and B are merged with the length of 5 units to make cluster 1. Then for the highest hierarchy, cluster 1 fuse with cluster C (called cluster 2 now) at the length of 24 units. Considering linkages in the dendrogram (Figure 3) the grouping of 2 clusters is the most appropriate. Cluster 1 includes LT03, LT04, LT05, LT06, and LT07 and cluster 2 has LT01 and LT02. The positions of these stations in the Lamtakong watershed map (Figure 1) indicate that cluster 1 is the upstream reach with 4 stations upstream of the Lamtakong Reservoir (LT04, LT05, LT06, and LT07) only 1 site (LT03) is located downstream of the reservoir. Cluster 2 includes LT01 and LT02 which are neighboring stations situated in the vicinity of Nakorn Ratchasima Municipality. The study of Sukhaparamate and other on the effects of river pollution on socio-economics of Lamtakong riparian communities divided the riparian area into 2 parts, upstream and downstream [3, 19]. The upstream section is from the Lamtakong Dam downstream-ward to the boundary of Sikhiu District and Sungnern District which covers station LT03. The downstream section contains the river reach within Muang District covering the stations LT01 and LT02. They assumed that the upstream section was unpolluted and the downstream, polluted. They found that due to pollution, the downstream community incurred greater costs for water treatment by 554 million baht than the upstream community in the year 2006. We can deduce from these studies that by grouping LT01 and LT02 into one cluster and LT03 another cluster is reasonable. The upstream section, however, has not only LT03 sites as in Sukhaparamate [13] study but includes LT04, LT05, LT06, and LT07. Therefore the whole region was categorized into 2 clusters, instead of 3, the highly polluted area (LT01 and LT02) and the less polluted area (LT03, LT04, LT05, LT06, and LT07).

3.2 Water quality parameters reduction

Principal component analysis (PCA), which is a branch of factor analysis (FA), was selected to perform water quality data reduction using SPSS software. The whole Lamtakong River can be partitioned into 2 parts: the upstream includes

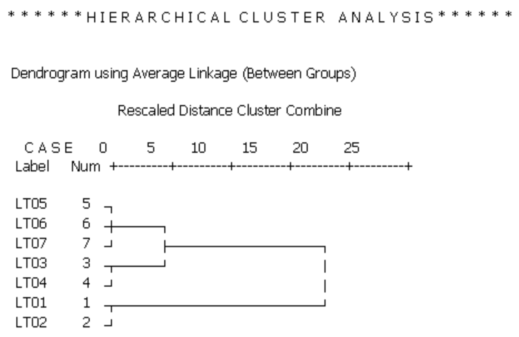

Figure 3 Dendrogram showing clustering of sampling sites according to water quality characteristics of the Lamtakong River

LT03, LT04, LT05, LT06, and LT07, and the downstream covers L01 and LT02. Seventeen water quality parameters from the PCD were used in this study. They were monthly data from all 7 monitoring sites, during 17 years from 1996 to 2012. The analyses were separated into the upstream reach and the downstream reach of the Lamtakong according to the two clusters. PCA produced the first unrotated axes with 17 latent factors. In order to obtain better distribution results, varimax rotation technique was applied to calculate the new factor loadings, eigenvalues, and eigenvectors for all factors. Eigenvalue gives a measure of the significance of the factor. Those with higher eigenvalue are more significant. Eigenvalues of 1.0 or greater are considered significant by Kaiser's criterion [20]. Liu et al., [21] 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.

PCA of the two data sets yielded five principal components for each of both reaches with eigenvalues being higher than 1, explaining 65.20% and 67.90% of the total variance in respective water quality data sets. Table 4 and 5 show the first 5 variable loadings for the upstream part and the downstream part, respectively. Interestingly, equally number of eight strong loadings are for each of both upstream and also downstream parts. Five parameters out of eight are the same for both parts i.e. total solids, total dissolved solid, total coliform bacteria, fecal coliform bacteria and pH. These five parameters show that along the whole Lamtakong receives almost similar contaminant inputs at different degree of strength. This finding coincides with past investigations of Ananpipatkij [2], Wijitkosum [7, 22-23]. All 5 water quality parameters indicate that the main source of pollutants is anthropogenic and less to natural cause.

The first two parameters, total solids and dissolved solids, come from the first factor which is the highest share of total variance. These two parameters derived from the non-point source of forestry and agricultural runoff as well as point sources from waste water treatment plant and industrial wastes. The second factor comprises both types of bacteria. These two bacterial parameters derived from municipal incomplete waste treatment plants as well as large animal farms, for example Chokchai Farm and the like. The last parameter, pH, indicates strong chemical characteristic and highly polluted of the Lamtakong River.

Table 4 Factor loadings of 5 significant components for upstream part

Parameters	VF1	VF2	VF3	VF4	VF5
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
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
TS	**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

Table 5 Factor loadings of 5 significant components for downstream part

Parameters	VF1	VF2	VF3	VF4	VF5
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
TS	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

Apart from the aforementioned 5 parameters, the others are calcium carbonate, suspended solid, and turbidity for the upstream reach, and conductivity, ammonia, and nitrite for the downstream reach. These parameters show that some natural causes of contamination affect the upstream river flow. As mentioned earlier, heavier rainfall occurs at the head of the Lamtakong together with deforestation and large monocrop farming induce high soil erosion [24]. Soil erosion in this area can rise up to the value of 800 tons/hectare annually, which increases suspended load and turbidity of the river flow [10]. Also for the upstream portion, $CaCO_3$ is another dominant parameter. This should be the result of limestone mountains, karsts, and caves spreading throughout the head of the Lamtakong Watershed [25-26]. Limestone is a good source of $CaCO_3$ high rainfall washes it down to small rivulets then to the Lamtakong River.

The other 3 additional significant parameters for the downstream stretch are electrical conductivity and the two types of nitrogen, ammonia and nitrite. These types of nitrogen indicate strong anthropogenic influence, from both agricultural fertilizer and community waste disposals. Wijitkosum [23] noticed eutrophication along the Lamtakong between LT01 and LT02 which should be the consequence of the nitrogen forms. Nitrite in drinking water is very toxic, infants under six months can be very sensitive to nitrite. When study the Lamtakong water quality along the downstream reach, Ramakomut [4] used BOD as quality indicator for modelling. It would be much more appropriated, from this research result, if she could use ammonia as the indicator.

The turbidity parameter drops out from significance in the downstream reach because the Lamtakong Dam block out the suspended load from the upstream part and the rainfall is less erosive than that of upstream [9, 24]. At the same time water quality of the downstream is not any better than the upstream due to the increase in EC.

4. Conclusions

The Lamtakong River can be classified for the purpose of water pollution management into 2 sections, the upstream part and the downstream part, using the cluster analysis technique. The upstream section is much longer being from the western suburb of the city of Nakhon Ratchsima to the upper end of the river. The shorter downstream section is mostly in the area of Nakhon Ratchsima city. The large numbers of water quality parameters, 17 parameters, were reduced to 8 using factor analysis for both upstream and downstream parts. Five of significant parameters are the same for both sections namely, total solids, total dissolved solid, total coliform bacteria, fecal coliform bacteria, and pH. For different significant parameters, the upstream part composes of calcium carbonate, suspended solids, and turbidity, while the downstream are electrical conductivity, ammonia nitrogen, and nitrite nitrogen. The results show that the whole length of the river suffers from both point source and nonpoint source pollutions. Both upstream and downstream contaminants are mainly derived from anthropogenic. The upstream part has some natural inputs but less so in the downstream part.

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. References

[1] Treedej S. Classification of Water Resources for Pong, Chi, Mun, and Lamtakong Rivers. Royal Thai Government Gazett 2542;116(53):47-48.

[2] Ananpipatkij S. Mathematical Model for Water Quality Management in Lam Ta Kong [Master Thesis]. Bangkok: Kasetsart University; 1999. (In thai).

[3] Sukhaparamate S. The Impact Assessment of Water Polution in the Lamtakong River, Muang District, Nakhon Ratchasima [Master Thesis]. Bangkok: Thammasat University; 2009.

[4] Ramakomut S. The Assessment of Lamtakong River Quality within Nakhonratchasima Province using Mathematical Model [Master Thesis]. Khon Kaen: Khon Kaen University; 2010. (In Thai).

[5] Wijitkosum S, Sriburi T. Evaluation of water resource situation in Lam Ta Kong watershed. Journal of Environmental Research 2009;31:21-34,.

[6] Rogerson PA. Statistical Methods for Geography; A Student's Guide. 3rd ed. Los Angelis: Sage; 2010.

[7] Wijitkosum S. The Settlement and Land Use Affecting the Water Quality: A Case Study of Lam Takong Reservoir, Pak Chong District, Nakhon Ratchasima Province [Master Thesis]. Bangkok: Chulalongkorn University; 2001. (In Thai).

[8] RID. Development of Water Resources for the Northeast and the Lamtakong Basin. Bangkok: The Royal Irrigation Department Thailand; 1972. (In Thai).

[9] Jeepech V. Lamtakong. Nakhon Ratchasima: Water Resources Regional Office VI; 2006. (In Thai).

[10] Polprasert C, Liamlaem W, Nitisoravut R. Case Study: Lessons Learned from the Lamtakong Hydropower Project. Germany: Gesellschaft fur Internationale Zusummenabeit (GIZ); 2013.

[11] Sneath PHA, Sokal RR. Numerical Taxonomy. San Francisco: Freeman; 1973.

[12] Mooi E, Sarstedt M. A Concise Guide to Market Research. Berlin: Springer; 2011.

[13] Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 1958;38:1409-1438.

[14] Martinez WL, Martinez AR. Exploratory Data Analysis with MATLAB. New York: Chapman & Hall/CRC; 2005.

[15] Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis: A Global Perspective. 7th ed. New Jersey: Upper Saddle River; 2010.

[16] Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology 2013;9:79-94.

[17] Velicer WF, Jackson DN. Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Multivariate Behavioral Research 1990;25:1-28.

[18] Papatheodorou G, Lambrakis N, Panagopoulos G. Application of multivariate statistical procedures to the hydrochemical study of a coastal aquifer: an example from Crete, Greece. Hydrological Processes 2007;21:1482-1495.

[19] Sutummakid N, Sukhaparamate S. Calculation of Environmental Adjusted Gross Provincial Product of Nakhon Ratchasima. Bangkok: Thammasart University; 2008.

[20] Field A. Discovering Statistics using SPSS for Windows. London: Sage; 2000.

[21] Liu CW, Lin KH, Kuo YM. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science in the Total Environment 2003;313:77-89.

[22] Wijitkosum S. Integrated Water Resource Management in Lam Ta Kong Watershed [Master Thesis]. Bangkok: Chulalongkorn University; 2008. (In Thai).

[23] Wijitkosum S. Is the Lamtakong perishing. Environmental Journal 2010;14:19-30. (In Thai).

[24] Pongboon K. Determination of rainfall erosivity of Lamtakong Watershed Research Station, Nakhon Ratchasima. Journal of Thai Forestry Research 2000;2:5-9.(In Thai).

[25] Latinne A, Weangsothon S, Herbreteau V, Michaux JR. Thai limestone karsts: an impending biodiversity crisis. In The 1st Environment Asia International Conference on Environmental Supporting in Food and Energy Security: Crisis and Opportunity; 2011 March 22; Bangkok, Thailand. Thai Society of Higher Education Institutes on Environment; 2011.

[26] Ponta G, Memon B, LaMoreaux J, Julawong J, Wongsawat S. Karst landforms in the Saraburi Group limestones, Thailand. In: Land L, Docter DH, Stephenson JB, editors. Sinkholes and the Engineering and Environmental Impacts of Karst; 2013 May 6; Carlsbad, New Mexico. National Cave and Karst Research Institute; 2013. p. 299-309.