Jurnal Teknologi

Full paper

Environmetric Techniques Application in Water Quality Assessment: A Case Study in Linggi River Basin

Mohd Fadhil Kasim[a,b*], Hafizan Juahir[a], Ismail Tawnie[b], Azman Azid[a], Ahmad Shakir Mohd Saudi[a], Saiful Samsudin[c], Mohd Ekhwan Toriman[a], Bashirah Mohd Fazli[b], Che Noraini Che Hasnam[a]

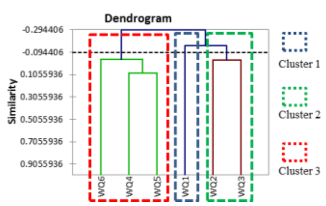
[a]*East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia*
[b]*National Hydraulic Research Institute of Malaysia (NAHRIM), Lot 5377 Jalan Putra Permai 43300 Seri Kembangan, Selangor, Malaysia*
[c]*Faculty of Environmental Study, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*

*Corresponding author: mohdfadhilkasim@gmail.com

Article history

Received :15 May 2014
Received in revised form :
27 January 2014
Accepted :15 March 2015

Graphical abstract



Abstract

In this research, determination of water quality status for Linggi River was carried out by using non-parametric Mann-Kendall analysis. HACA and PCA has been used to classify the river to obtain the clearest picture of the water quality status. The dataset includes six parameters for six monitoring stations (1997 to 2012). Mann-Kendall trend analysis shows significant improvement trend for all parameters studied except for BOD (WQ1 ($P<0.1$) and WQ6 ($P<0.05$)) and SS (WQ4 to WQ6 ($P<0.05$)). This indicates that even though the WQI getting good, a few parameters such as BOD and SS need to be watched and improved by the local authority to make sure the WQI continuously getting better in the future. HACA grouped the six monitoring stations into three different clusters based on their similarities namely less pollution site (LPS), medium pollution site (MPS) and high pollution site (HPS). HACA grouped one station (WQ1) into LPS, two stations into MPS (WQ2 and WQ3) and three stations into HPS (WQ4, WQ5 and WQ6). PCA was used to investigate the origin of each water quality variable based on the clustered region. Three principal components (PCs) were obtained with 75.3% total variation for HPS, 73.4% for MPS and 68.1% for LPS. The major pollution source for HPS are of anthropogenic source (municipal waste, domestic wastes) while for MPS the major source of pollution was from non point source pollution such as animal husbandry and livestock farms. For the LPS, major sources come from the sea tide effect (natural effect). The identification and classification of different region by this study will help the local authorities make better and more informed decisions about the improvement water quality program for the future.

Keywords: Water Quality Index (WQI); Mann-Kendall; cluster analysis; principal component analysis; Linggi River

Abstrak

Dalam kajian ini, penentuan status kualiti air bagi Sungai Linggi telah dijalankan dengan menggunakan analisis Mann-Kendall tak berparameter. HACA dan PCA telah digunakan untuk mengkelaskan Sungai Lingi kepada beberapa bahagian untuk mendapatkan gambaran status kualiti air. Set data kajian adalah dari tahun 1997 hingga 2012 terdiri daripada enam stesen serta enam parameter bagi setiap stesen. Analisis Mann-Kendall menunjukkan trend peningkatan ketara bagi semua parameter yang dikaji kecuali BOD (WQ1 ($P < 0.1$) dan WQ6 ($P < 0.05$)) dan SS (WQ4 ke WQ6 ($P < 0.05$)). Ini menunjukkan bahawa walaupun nilai WQI menunjukkan peningkatan, beberapa parameter seperti BOD dan SS perlu dipantau oleh pihak berkuasa tempatan untuk memastikan nilai WQI secara berterusan semakin baik pada masa hadapan. Analisis HACA mengkelaskan enam stesen pemantauan menjadi tiga kategori berbeza berdasarkan persamaan mereka iaitu kawasan kurang pencemaran (LPS), kawasan pencemaran sederhana (MPS) dan kawasan pencemaran tinggi (HPS). Daripada analisis HACA, satu stesen dikelaskan sebagai LPS (WQ1), dua stesen MPS (WQ2 & WQ3) dan tiga stesen HPS (WQ4, WQ5 & WQ6). Analisis PCA digunakan untuk mengkaji punca utama setiap pembolehubah bagi setiap kategori hasil daripada analisis HACA. Tiga komponen yang utama (PC) diperolehi dengan jumlah variasi sebanyak 75.3% bagi HPS, 73.4% bagi MPS dan 68.1% untuk LPS. Sumber utama pencemaran bagi HPS adalah sumber antropogenik (sisa perbandaran & sisa domestik). Manakala sumber pencemar utama bagi MPS adalah daripada pencemar tidak tetap seperti ladang ternakan. Bagi LPS, sumber utama pencemar adalah kesan daripada pasang surut air laut (kesan semula jadi). Pengenalpastian dan pengkelasan berdasarkan kategori hasil kajian ini dapat membantu pihak berkuasa tempatan bagi membuat keputusan yang lebih baik dan lebih bermaklumat bagi program peningkatan kualiti air pada masa hadapan.

Kata kunci: Indeks Kualiti Air (WQI); Mann-Kendall; analisis kluster; analisis komponen utama; Sungai Linggi

1.0 INTRODUCTION

Surface water pollution by several of contaminants all over the world can be considered as an epidemic problem[1,2,3]. Surface water systems are waters naturally open to the atmosphere, such as rivers, lakes, reservoirs, estuaries, and coastal waters. The quality of a river at any point reflects several major influences, including the natural and anthropogenic inputs[4]. Besides, rivers play a major role in assimilation or transporting municipal and industrial wastewater and runoff from agricultural lands[5]. Therefore, river water quality assessment is of great importance because it directly influences public health (via drinking water) and aquatic life (via raw water). Nowadays, water demand and water pollution are the two major problems in Asian region[6]. Almost 60% of the main river in Malaysia become main source of water supply for domestic, agriculture and industrial use[7].

River is a main water source for human being. Management and treatment of river water quality level require serious attention and monitoring from all parties. The major pollution sources affecting rivers in Malaysia are sewage disposal, discharge from small and medium sized industry that are still not equipped with proper treatment facilities and land clearing consist earthworks activities[8,9]. In Malaysia, especially a variety of river water quality monitoring programme has been implemented. From a study in 2012, 59% out of 473 rivers in Malaysia considered clean, 34% considered slightly polluted and the other 7% considered polluted[10]. The quality of the water supply became critical issue for communities in the area. Water supply needs to be monitored and tested from various aspects and criteria. Any changes that occur in the water supply will give a negative impact on consumer health. Water quality monitoring should be done continuously, immediate action and prevention needs to be done so that the quality of water supplies can be increased from time to time.

Water Quality Index (WQI) has been used to indicate the level of pollution and the corresponding suitability in terms of uses according to the National Water Quality Standards for Malaysia (NWQS) based on water quality data from the monitoring done by the Department of Environment, Malaysia (DoE). NWQS is a standard has been set up by the Department of Environment, Malaysia to classify the level of river water quality. This standard consists five types of class (Table 1). The WQI takes into consideration parameters Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammoniacal Nitrogen (NH3-N), Suspended Solids (SS) and pH[10].

Table 1 Water class and uses guideline from Department of Environment (DOE), Malaysia

Class	Uses
Class I	Conservation of natural environment. Water Supply I – Practically no treatment necessary. Fishery I – Very sensitive aquatic species
Class IIA	Water Supply II – Conventional treatment required. Fishery II – Sensitive aquatic species
Class IIB	Recreational use with body contact
Class III	Extensive treatment required. Fishery III – Common, of economic value and tolerant species; livestock drinking
Class IV	Irrigation
Class V	None of the above

Detection of temporal trends is one of the most important objectives of environmental monitoring[11]. Trend analysis has proven to be a useful tool for effective land use planning, design and management since trend detection provides useful information on the tendency change of land use in the future[11]. Besides, spatial variation of water quality variables also can be useful information of the environmental condition[12] and help researcher identify the pollution sources[13]. Recently, a lot of study has been conducted using cluster analysis to group water quality monitoring stations, suggesting for rapid assessment and only representative stations from each cluster can be used for a reasonable spatial assessment of the water quality[8,14]. Extracted group information from cluster analysis can be used to reduce the number of the sampling site and also can classify the current status of water quality in the river for each sampling site without missing substantial information[8,14].
The Linggi river basin consists of Seremban and nearby town districts. Part of the rivers covers approximately the whole of Seremban town and its outskirts. Linggi water treatment plant supplies 60% and 100% of the water requirements for Seremban and Port Dickson, respectively[15]. Two reports submitted to the government of Negeri Sembilan in 1961[16] and in 1979[17] had shown that the Linggi river is highly polluted and, by WHO standards, can be classified as "heavily-polluted requiring extensive treatment". In 2012, a report by DoE also classified that Linggi River was in the Slightly Polluted condition[10]. The rapid urbanization and industrialization in and around the Linggi River Basin has resulted in increased water quality problems in the state[18].

Therefore, the objectives of this study were; (i) to explore the trend of water quality index of Linggi River using non-parametric test and (ii) determine the spatial distribution in water quality characteristic. The result obtains from this research could be used to plan a lot of study at Linggi River and provide the information for water quality control in Linggi River.

2.0 EXPERIMENTAL

Study Area

The Linggi River is one of the main rivers located at Negeri Sembilan, Malaysia essentially is the hydrological entity. Total area of the watershed is about 128,981 ha. It can be divided into three sub basins which was Upper Linggi (45,412 ha), Lower Linggi (28,061 ha) and Rembau-Siput (55,508 ha) (Figure 1)[19].

Figure 1 Linggi River basin (Sub Basin)

The water quality data in this study were obtained from 6 stations along the main Linggi River (Figure 2). The monitoring station was manned by the Department of Environment, Malaysia (DOE). The details of the selected station are tabulated in Table 2. All this station was identified based on the availability of recorded data from 1997 to 2012. The data were collected from January 1997 until November 2012. WQ1, WQ2 and WQ3 are located in the Lower Linggi sub basin and the other three stations (WQ4, WQ5 and WQ6) are located in the Upper Linggi sub basin. It is worth mentioning that some stations having missing data and not all of those stations were consistently sampled.

There are six water quality parameters (based on the water quality index parameters) were selected in this study. The six parameters are DO, BOD, COD, SS, pH and AN. The descriptive statistics of the measured 15 years data set are summarized in Table 3.

Figure 2 Water quality monitoring stations

Table 2 DOE water quality monitoring station

DOE station code	Study code	Coordinate	
1LI01	WQ1	02° 23 823'	101° 58 951'
1LI02	WQ2	02° 28 908'	102° 00 759'
1LI03	WQ3	02° 30 521'	101° 57 844'
1LI04	WQ4	02° 34 855'	101° 57 443'
1LI05	WQ5	02° 39 121'	101° 55 509'
1LI06	WQ6	02° 42 607'	101° 57 168'

Trend Analysis: Mann-Kendall Non-parametric Test

Temporal trend analysis for all monitored parameter (DO, BOD, COD, SS, pH and AN) for each sampling station is one of thestudy objective and analysed using Mann-Kendal non-parametric test.

The Mann-Kendall trend test[20,21] were applied to examine the performance of a class of non-parametric trend test, which were first proposed by El-Shaarawi (1993)[22]. Mann-Kendall test can be used because only the relative magnitudes of the data is rather than their measured values[23]. The basic principle of Mann-Kendall tests for trend is to examine the sign of all pairwise differences of the observed values[11]. The Mann-Kendall test is based on the statistic S. Each pair of observed values y_i, y_j $(i>j)$ of the random variable is inspected to find out whether $y_i>y_j$ or $y_i<y_j$. The test statistic for the Mann-Kendall test is given as

$$S = \sum_{k-1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_k) \quad (1)$$

Where x_j and x_k are the sequential data values and $j > k$, n is the length of the data set and

$$sgn(x_j - x_k) = \begin{cases} 1 \; if \; x_j - x_k > 0 \\ 0 \; if \; x_j - x_k = 0 \\ -1 \; if \; x_j - x_k < 0 \end{cases} \quad (2)$$

which is the number of positive differences minus the number of negative differences. Variance of S, computed by

$$Var(S) = \left[n(n-1)(2n+5) - \sum_t (t-1)(2t+5) \right] /18 \quad (3)$$

and are asymptotically normal[24], where t is the extent of any given tie and the summation over all ties. In the case that n is larger than 10, the standard normal variable z is computed by using the following equation[25].

$$z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} \; if \; S < 0 \end{cases} \quad (4)$$

Thus, in a two-sided test for trend, at a selected level of significance α, the null hypothesis of no trend is rejected if the absolute value of z is greater than $z_{a/2}$.

Cluster Analysis

Cluster analysis (CA) is a notable method that assembles object into aggregations based on their independent variables or characteristics[26]. Hierarchical agglomerative CA (HACA) is a common method to classify variables or cases (observation/samples) into a cluster by starting with the similar pair of objects and forming higher clusters step by step[27]. The result of the HACA analysis will classify the variable into the cluster with high homogeneity level within the class and high heterogeneity level between class with respect to a predetermined selection criterion[28]. Dendrogram is the illustrated result from HACA analysis, presenting the clusters and their proximity[5]

In this study, HACA has been used to investigate the grouping of the monitoring sites (spatial) into different groups and the achievement of HACA was through the Ward's method using Euclidean distance as a measure of similarity[29]. The Euclidean distance (linkage distance) is reported as D_{Link}/D_{Max}, which represents the quotient between the linkage distance divided by the maximal distance. As the way to standardize the linkage distance represented by the *y*-axis, the quotient is usually multiplied by 100[5,30,31].

Principal Component Analysis (PCA)

Combination of HACA and PCA is the most powerful pattern recognition technique[8,30,31,32]. The combination of this analysis will provide information on the most significant parameters due to spatial and temporal variations that describe the whole data set by excluding the less significant parameters with the minimum loss of the original information[8]. The principle component (PC) can be explained as

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \cdots + a_{im}x_{mj} \quad (5)$$

Where the Z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number and m is the total number of variables. The PCs generated by the PCA are sometimes not readily interpreted; therefore, it is advisable to rotate the PCs by the varimax rotation[8]. Varimax rotations applied to the PCs with the eigenvalues more than 1 are considered significant[33] in order to obtain new groups of variables called varimax factors (VFs). The number of VFs obtained by varimax rotations is equal to the number of variables in accordance with common features and can be included unobservable, hypothetical and latent variables[34]. The VF coefficients having a correlation greater than 0.75 considered as "strong"; 0.75-0.50 as "moderate" and 0.50-0.30, as "weak" significant factor loading[35].

■3.0 RESULTS AND DISCUSSION

Water Quality Analysis

Figure 2 shows the water quality monitoring station taken into consideration in this study. The means concentrations of each parameter are listed in Table 3. The standard deviation for each station for each parameter shows the variation of the water quality-from the monitoring program. The coefficient of variation was computed to allow comparison of the variations of water quality for each parameter among the different stations. It is observed that coefficients of variation for BOD, pH and AN are relatively high at station WQ1 compare to other parameters for each station for all the years from 1997-2012. For the other parameters like DO relatively high at station WQ3, COD (WQ4), SS (WQ6) and WQI (WQ5). These differences in the coefficient of variation could be due to location of the monitoring stations. Stations WQ1 are located at mouth of Linggi River which is nearer to the coastal area and effected by ocean tide. Stations WQ3 are located nearer to the agriculture activity and receives more discharge from the farms compared to other stations. This made high variation for DO. Stations WQ4 are located after Linggi water treatment plant which is surrounded by residential area and palm oil mill and receives schedule discharge from the palm oil mill. Stations WQ5 is located nearer to the residential area and receives various discharges from the residential area. These factors made the coefficient of variation of WQI relatively higher compared to other stations. Station WQ6 are located in the middle of Seremban City which was surround by various type of activities like residential area, wet market, small workshops, food courts and also receives high volume of surface runoff during the raining season. These made the coefficient of variation of TSS high compared to other stations.

Water quality pattern shows that the highest WQI was registered in May 2001 and May 2004 at WQ1 (WQI=90). The highest DO pattern was observed at WQ2 in October 2000 (DO=114.2 %, 9.08mg/L) and the lowest DO pattern was registered in January 2006 located at WQ5 (DO=25.4%, 2.03 mg/L). BOD and COD concentration pattern shows the highest in November 2005 (WQ4) were the concentration was 33.00 mg/L for BOD and 146.00 mg/L for COD. The lowest concentration for BOD and CODnduring the period (BOD=0.00 mg/L, COD=0.50 mg/L) were recorded in January 2012 (WQ4) and February 2008 (WQ3 & WQ4). The highest concentrations for TSS was registered in August 2000 located at WQ6 where the concentration was 5100.00 mg/L and the lowest TSS concentration recoded was in November 2005 located at WQ1 (SS=0.5 mg/L). pH pattern shows the highest pH value was recorded at WQ3 in July 2004 (pH=7.83) and the lowest pH value was recorded at WQ1 in March 2008 (pH=4.71). The highest AN concentration was recorded at WQ6 in August 2008 were the concentration was 9.49mg/L. The lowest concentration for AN was recorded at WQ1 (August 2000, October 2000, January 2002, March 2004), WQ2 (October 2000, December 2000), WQ3 (October 2000, May 2005), WQ4 (October 2000, March 2005), WQ5 (October 2000), WQ6 (October 2000, March 2005, May 2005) were the concentration was 0.5 mg/L.

Table 3 Descriptive summary of DO, BOD, COD, SS, pH, AN and WQI for each station monitored for years 1997-2012

Parameter	Statistic	Station					
		WQ1	WQ2	WQ3	WQ4	WQ5	WQ6
Dissolved oxygen (DO %)	No. of observations	87	88	75	88	88	88
	Minimum	40.80	25.40	27.10	46.70	26.70	41.20
	Maximum	100.00	93.80	100.00	105.90	114.20	98.60
	Mean	76.76	59.30	63.86	77.54	74.18	72.33
	Variance (n-1)	155.23	237.30	344.47	128.46	233.78	131.53
	Standard deviation (n-1)	12.46	15.40	18.56	11.33	15.29	11.47
	Coefficient of variation	0.16	0.26	**0.29**	0.15	0.20	0.16
	Skewness (Pearson)	-0.48	-0.33	-0.08	-0.38	-0.97	-0.57
Dissolved oxygen (DO) mg/L	No. of observations	87	88	75	88	88	88
	Minimum	3.09	2.03	2.11	3.46	2.19	2.84
	Maximum	7.14	7.24	7.71	7.86	9.08	7.71
	Mean	5.48	4.55	4.90	5.95	5.78	5.56
	Variance (n-1)	0.73	1.45	2.08	0.80	1.39	0.85
	Standard deviation (n-1)	0.85	1.20	1.44	0.89	1.18	0.92
	Coefficient of variation	0.15	0.26	**0.29**	0.15	0.20	0.16
	Skewness (Pearson)	-0.39	-0.29	-0.04	-0.36	-0.87	-0.64
Biological oxygen demand (BOD) mg/L	No. of observations	87	88	75	88	88	88
	Minimum	0.50	1.00	1.00	0.00	1.00	2.84
	Maximum	32.50	15.00	20.00	33.00	32.00	7.71
	Mean	3.80	4.26	4.95	6.35	6.79	5.56
	Variance (n-1)	22.34	6.90	10.24	21.29	26.50	0.85
	Standard deviation (n-1)	4.73	2.63	3.20	4.61	5.15	0.92
	Coefficient of variation	**1.24**	0.61	0.64	0.72	0.75	0.16
	Skewness (Pearson)	3.71	1.47	1.92	2.90	2.77	-0.64
Chemical oxygen demand (COD) mg/L	No. of observations	87	88	75	88	88	88
	Minimum	5.00	2.00	0.50	0.50	5.00	3.00
	Maximum	119.00	48.00	64.00	146.00	104.00	99.00
	Mean	36.24	24.32	27.54	30.45	33.05	32.14
	Variance (n-1)	435.86	102.01	152.02	414.34	325.33	302.65
	Standard deviation (n-1)	20.88	10.10	12.33	20.36	18.04	17.40
	Coefficient of variation	0.57	0.41	0.44	**0.66**	0.54	0.54
	Skewness (Pearson)	1.19	0.22	0.70	3.09	1.67	1.57
Suspended solid (SS) mg/L	No. of observations	87	88	75	88	88	88
	Minimum	0.50	15.00	29.00	22.00	17.00	9.00
	Maximum	322.00	620.00	2090.00	2460.00	2950.00	5100.00
	Mean	46.71	134.08	226.12	213.03	266.15	364.56
	Variance (n-1)	3452.58	14695.41	86513.05	98273.02	193372.79	590635.26
	Standard deviation (n-1)	58.76	121.22	294.13	313.49	439.74	768.53
	Coefficient of variation	1.25	0.90	1.29	1.46	1.64	**2.10**
	Skewness (Pearson)	2.93	1.77	4.13	5.01	3.81	4.31
pH	No. of observations	87	88	75	88	88	88
	Minimum	4.71	5.40	5.71	6.14	6.20	5.50
	Maximum	7.76	7.53	7.83	7.63	7.72	7.78
	Mean	6.52	6.76	6.83	6.98	7.11	6.95
	Variance (n-1)	0.53	0.19	0.19	0.12	0.12	0.16
	Standard deviation (n-1)	0.73	0.43	0.43	0.34	0.34	0.40
	Coefficient of variation	**0.11**	0.06	0.06	0.05	0.05	0.06
	Skewness (Pearson)	-0.54	-0.63	-0.20	-0.42	-0.47	-0.82
Ammoniacal Nitrogen (AN) mg/L	No. of observations	87	88	75	88	88	88
	Minimum	0.01	0.01	0.01	0.01	0.01	0.01
	Maximum	4.21	2.54	3.53	4.69	7.01	9.49
	Mean	0.51	0.68	1.03	1.28	1.63	1.48
	Variance (n-1)	0.46	0.36	0.77	1.16	2.23	1.87
	Standard deviation (n-1)	0.68	0.60	0.87	1.08	1.49	1.37
	Coefficient of variation	**1.32**	0.87	0.85	0.84	0.91	0.92
	Skewness (Pearson)	3.69	1.19	0.96	1.21	1.08	2.56
Water quality index (WQI)	No. of observations	87	88	75	88	88	88
	Minimum	49.00	51.00	48.00	35.57	40.00	46.00
	Maximum	90.00	85.66	86.60	83.77	85.29	82.48
	Mean	78.10	71.12	68.49	70.54	67.98	67.89
	Variance (n-1)	64.32	51.55	61.97	65.09	83.85	54.39
	Standard deviation (n-1)	8.02	7.18	7.87	8.07	9.16	7.38
	Coefficient of variation	0.10	0.10	0.11	0.11	**0.13**	0.11
	Skewness (Pearson)	-1.15	-0.52	-0.31	-1.23	-0.64	-0.31

The percentage of concentration for each parameter monitored under this study classified into five different classes based on class set by (DoE) (Table 1: Water class and uses guideline from Department of Environment (DOE), Malaysia) in order to get the view of Linggi River water quality status. Table 4 shows the percentage based on the class set by DoE. It was observed that 71.21% of WQI value recorded during this period of study are within class III. The others are 26.65% within class II, and 2.14% of data within class IV. In other hands, 5.25% DO's concentration within class I, 25.88% within class II, 64.20% within class III, 4.67% within class IV. 8.17% of BOD concentration within class I, 12.65% within class II, 39.11% within class III, 34.44% within class IV and 5.64% within class V. For COD concentration, 4.09% are within class I, 37.16% within class II, 48.64% within class III, 9.14% within class IV and 0.97% within class V. Majority of SS concentration was within class III (37.16%),for other class was 11.09% (class I), 14.59% (class II), 23.35% (class IV) and 13.81% (class V). Majority of pH value was within class I which was 79.96% from all the period study. 14.59% of the data are within class II and the other 5.45% within class III. For AN concentration, 10.51% are within class I, 12.45% of the data are within class II, 34.82% are within class III, 33.07% within class IV and other 9.14% are within class V

Table 4 Percentage for each water quality parameter within Linggi River Basin based on class set by Department of Environment of Malaysia.

Class	Percentage of Parameter						
	DO	**BOD**	**COD**	**SS**	**pH**	**AN**	**WQI**
I	5.25	8.17	4.09	11.09	79.96	10.51	0.00
II	25.88	12.65	37.16	14.59	14.59	12.45	26.65
III	**64.20**	**39.11**	**48.64**	**37.16**	5.45	**34.82**	**71.21**
IV	4.67	34.44	9.14	23.35	0.00	33.07	2.14
V	0.00	5.64	0.97	13.81	0.00	9.14	0.00
Total %	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Temporal Water Quality Trend Analysis

As shown in Table 5, only WQ1 shows no significant trend for DO from 1997-2012. The other station (WQ2-WQ6) shows the upward trend (improvement) (*P value* < *0.05*) which was a good sign in water quality condition from 1997 until 2012. For BOD, three stations (WQ2, WQ3, and WQ5) show no significant trend during the study period. Two of the other three stations (WQ1 (*P value <0.1*) and WQ6 (*P value* < *0.05*)) showing significant upward trend (unhealthy condition) and the other one station (WQ4) shows the significant (*P value* < *0.05*) downward trend (improvement) during the study period. Like BOD, three stations (WQ2, WQ3, and WQ5) show no significant trend in COD. The other three stations (WQ1, WQ4 and WQ6) show the significant (*P value* < *0.05*) downward trend (improvement) during this study period. TSS shows significant (*P value* < *0.05*) trend for WQ4, WQ5 and WQ6 monitoring stations. TSS shows upward trend (unhealthy condition) in this three monitoring station and the other three station (WQ1, WQ2 and WQ3) show no significant trend. All monitoring stations showed significant (*P value* < *0.05*) trend for pH during the study period except for WQ1 (no significant trend). All of that stations (WQ2, WQ3, WQ4, WQ5 and WQ6) show the upward trend. Like pH, all stations (WQ2, WQ3, WQ4, WQ5 and WQ6) showed significant (*P value* < *0.05*) trend except WQ1 for AN. All the station shows the downward trend (improvement) during the study period. For WQI, all monitoring station showed significant trend, (upward) WQ1 (*P value* < *0.1*), WQ2-WQ6 (*P value* < *0.05*)). This is a good sign for water quality status in Linggi River where the water is used as water supply to the surrounding area eventough there were other parameters like BOD and COD shows the upward trend during the study period.

Table 5 Man-Kendall test of trend for water quality analysis for each station during study period (1997-2012)

Station	WQ Parameter	DO	BOD	COD	SS	pH	AN	WQI
WQ1	S	384	505	-1018	95	-59	-163	525
	Z	1.404	1.885	-3.73	0.345	-0.213	-0.594	1.924
	Trend	NT	↑*	↓	NT	NT	NT	↑*
WQ2	S	2270	415	-80	222	1381	-653	1572
	Z	8.178	1.51	-0.285	0.797	4.975	-2.351	5.67
	Trend	↑	NT	NT	NT	↑	↓	↑
WQ3	S	1545	-98	357	-28	1216	-589	1311
	Z	7.063	-0.448	1.63	-0.124	5.559	-2.69	6.003
	Trend	↑	NT	NT	NT	↑	↓	↑
WQ4	S	1351	-668	-500	575	1060	-847	875
	Z	4.866	-2.421	-1.8	2.069	3.818	-3.147	3.157
	Trend	↑	↓	↓	↑	↑	↓	↑
WQ5	S	1179	43	-316	1072	829	-1315	736
	Z	4.246	0.153	-1.136	3.86	2.985	-4.736	2.652
	Trend	↑	NT	NT	↑	↑	↓	↑
WQ6	S	1434	1434	-740	1423	1251	-1188	976
	Z	5.165	5.165	-2.666	5.125	4.506	-4.279	3.519
	Trend	↑	↑	↓	↑	↑	↓	↑

NT=No trend ↑= upward trend (*p<0.05*) ↑*= upward trend (*p<0.1*) ↓= Down trend (*p<0.05*)

Cluster Analysis and Spatial Similarity

Based on the 7 variables (DO, BOD, COD, SS, pH, AN and WQI), the 6 sampling sites are classified into three distinct clusters illustrated as a dendrogram (Figure 3). Thus, the three clusters correspond to relatively less polluted sites (LPS) (cluster 1), moderately polluted sites (MPS) (cluster 2) and highly polluted sites (HPS) (cluster 3), respectively.

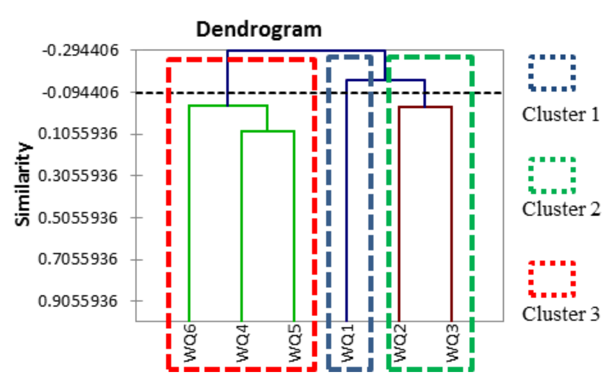


Figure 3 Cluster analysis result for Linggi River

In cluster 1, one sampling sites (WQ1) was included which situated at the downstream of the river, where nearer to the sea and effected by the tide. The mean value of WQI, DO, BOD, COD, TSS, pH and AN in cluster 1 are all in good condition among the three others clusters (Figure 3), reaching to 78.09, 5.50 mg/L, 2.00

mg/L, 33.00 mg/L, 28.00 mg/L, 6.77 and 0.41 mg/L respectively. Cluster 2 covering station WQ2 and WQ3 in the middle of the Linggi River named MPS since their mean concentration of studied parameter (WQI, DO, BOD, COD, TSS, pH and AN) falls in the moderate value of water quality variable (Figure 3). For these sampling site (WQ2 and WQ3) is located at the rural area, where pollutions are mostly derived from local agricultural practices (palm oil plantation). The other 3 monitoring station (WQ4, WQ5 and WQ6) are included in cluster 3 (HPS). Even though these three stations are located at upstream of Linggi River.The monitoring station (WQ6) is located at the urban area, in the middle of Seremban city which is the major city for state of Negeri Sembilan Malaysia. Its mean value of WQI, DO, BOD, COD, SS, pH and AN recorded unhealthy value of 68.80, 5.76 mg/L, 6.54 mg/L, 31.87 mg/L, 281.25 mg/L, 7.01 and 1.46 mg/L among the three cluster (Figure 3). This station may be receiving high volume of pollution such as wastewater and local pollutions mostly from industrial effluents and partially domestic wastewater..

Principal Component Analysis and Source Identification

In this study, PCA was applied to the normalize data set (6 variables) separately for the three different spatial region, HPS, MPS and LPS as resulted from HACA analysis. The input data matrices (variables x cases) for PCA were 6 x 87 for LPS, 6 x 163 for MPS and 6 x 264 for HPS regions.

To identified the main pollution factors, principal component analysis (PCA) was performed separately for the three clusters (LPS, MPS and HPS) as delineated by CA techniques. Three PCs have been found for LPS, MPS and HPS region with eigenvalues larger than 1 summing almost 68.1% (LPS), 73.4% (MPS), and 75.3% (HPS) of the total variance in the data set.. Corresponding VFs, variable loadings and variance explained presented in Table 6.

Table 6 Factor loading of environmental variables on the varimax-rotated PCs for water quality data collected from LPS, MPS and HPS of the Linggi River Basin

Variables	LPS			MPS			HPS		
	VF1	VF2	VF3	VF1	VF2	VF3	VF1	VF2	VF3
DO	-0.170	**0.727**	-0.311	**0.895**	0.045	-0.164	-0.108	-0.651	0.562
BOD	0.683	0.086	0.099	-0.078	**0.795**	0.225	**0.746**	0.300	0.160
COD	**0.768**	0.239	0.109	0.309	**0.771**	-0.009	**0.878**	0.057	-0.081
SS	-0.030	-0.064	**0.924**	-0.129	0.514	-0.600	0.666	-0.485	-0.176
pH	0.191	**0.825**	0.086	**0.871**	0.053	0.200	-0.010	0.029	**0.920**
AN	**0.733**	-0.219	-0.348	-0.034	0.234	**0.835**	0.066	**0.864**	0.053
Eigenvalue	1.711	1.315	1.059	1.823	1.417	1.165	1.858	1.520	1.137
Variability (%)	27.658	22.086	18.336	27.974	25.835	19.602	29.804	24.993	20.453
Cumulative %	27.658	49.744	68.080	27.974	53.809	73.410	29.804	54.797	75.250

Bold value are strong loading (>0.7000)

Less Pollution Site (LPS)

For LPS, between the three VFs, VF1 consist 27.7% of the total variance, showing the strong loading on a chemical pollutants COD and AN and moderate positive loading on BOD. These factors representing the influence of organic pollutant from point sources such as discharge from wastewater treatment plant, domestic wastewater and industrial effluents. These factors also can be from municipal sewage and sewage treatment plants located at the monitoring area. For VF2, explaining 22.1% of the total variance has strong positive loadings on DO and pH, which are related to sea water intrusion where the monitoring station WQ1 situated nearer to the sea. VF3, showing 18.3% from the total of variance, has strong positive loadings on SS.This region is in the downstream of the Linggi River and affected by the activity at upstream of the river and also can be effected from the sea tide event.

Moderate Pollution Site (MPS)

For MPS, VF1 explaining 27.9% of the total variance, has strong positive loading on DO and pH. This also related to the location of monitoring station were WQ2 are nearer to WQI (LPS region). This means that WQ2 also can be affected by the sea tide event that make the loading of pH is positively high. VF2, explaining 25.8% of the total variance, has strong positive loading on BOD and COD. These organic factors, mainly represent the contribution of several types of pollution such as wastewater from agriculture and animal husbandry in livestock farm (non-point source pollution)[36]. VF3, explaining 19.6% of total variance has strong positive loading on AN thus representing the influence from organic pollutants from point source such as discharge from wastewater treatment plantfrom domestic and industrial effluents[8].

Highly Pollution Site (HPS)

In the case of HPS, VF1 explaining the 29.8% of the total variance, has strong positive loadings on BOD and COD considered as an organic factor and can be representing influence from non-point source pollution such as agricultural activities[8,37]. This region also has moderate positive loading from SS representing and related to surface run off were the region was located at Seremban city. VF2, explaining 24.9% of total variance, has strong positive loading on AN thus representing the influence from organic pollutant[8]. As this region is the HPS site, majority of the effluent was came from Seremban city. VF3, explain 20.4% of total variance and has strong positive loading on pH which was related to municipal wastes, oxidation ponds and animal husbandry[8]. As HPS region are located at Seremban city, the major source should be come from the municipal waste.

■4.0 CONCLUSION

The monitoring in the Linggi River is an important part of river monitoring program in Malaysia. The analysis shows the WQI trend are getting improve but the trend for BOD at WQ1 and WQ6 and SS (WQ4 to WQ6) shows the unhealthy trend and need to be taken in consideration for further action/ study. Majority of data set studied was fall between class III (extensive treatment required. Fishery III – Common, of economic value and tolerant species; livestock drinking) of NWQS. From CA analysis, HACA was successfully classified the six monitoring station into three different cluster region namely HPS, MPS and LPS. With this classification, the river can be divided into section base on HACA result for more optimal sampling or monitoring program/ study can be design. The result from this Cluster analysis is very useful in offering reliable classification of surface water for certain area and it can help local authorities reduce the monitoring cost by reducing the monitoring station and also can help the local authorities to classify certain area of monitoring into sub cluster that help them to plan the different approach when making the decision. Application of PCA on the available data based on the region resulted from the HACA analysis shows three parameter responsible for major variations of surface water quality a long the Linggi River were the main source of the variations is come from municipal effluent, industrial effluent wastewater treatment plant, agricultural activity and domestic and commercial areas.

Acknowledgement

The authors acknowledge the Department of Environment (DOE), Malaysian Ministry of Natural Resource and Environment; National Hydraulic Research Institute of Malaysia, Ministry of Natural Resource and Environmnet for their permission to utilise water quality data, advice, guidance and support for this study.

References

[1] R. Noori, M. A. A., A. Ameri, M. Jalili-Ghazizade. 2009. Prediction Of Municipal Solidwaste Generation with Combination of Support Vector Machine and Principal Component Analysis: A Case Study of Mashhad. *Environmental Progress & Sustainable Energy*. 2009(28): 249–258.

[2] R. Noori, A. R. K., A. Farokhnia, M. Dehghani. 2006 Predicting the Longitudinal Dispersion Coefficient Using Support Vector Machine And Adaptive Neuro-Fuzzy Inference System Techniques. *Environmental Engineering Science*. 2009(26): 1503–1510.

[3] R. Noori, M. S. S., A. R. Karbassi, A. Baghvand, H. Taati Zadeh. 2010. Multivariate Statistical Analysis of Surface Water Quality Based on Correlations and Variations in the Data Set. *Desalination*. 2010(260): 129–136.

[4] O. P. Bricker, B. F. J. 1995. Main Factors Affecting the Composition of Natural Waters, In: B. Salbu, E. Steinnes (Eds.). C. Press, Editor., CRC Press: Boca Raton, FL. 1–5.

[5] Shrestha, S., Kazama, F. 2007. Assesment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan. *Environmental Modelling & Software*. 2007(22): 464–475.

[6] Juahir, H., Mohd Fadhil Kasim. 2008. Kajian Awal Impak Aliran Rendah Keatas Kualiti Air Sungai Klang. National Hydraulic Research Institute of Malaysia: Seri Kembangan.

[7] DID. 2001. *DID Anual Report*. Department of Irrigation and Drainage: Kuala Lumpur.

[8] Juahir, H., Zain, S. Z., Yusoff, M. K; Hanida, T. I. T., Armi, A. S. M., Toriman, M. E. 2011**.** Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. *Environmental Monitoring Assessment*. 2011(173): 625–641.

[9] Rosnani, I. 2001. *River Water Quality Status in Malaysia*. In *National Conference on Sustainable River Basin Management in Malaysia*. Kuala Lumpur, Malaysia.

[10] DoE. 2012. *Environmental Quality Report*. Department of Environment, Malaysia: Putrajaya, Malaysia.

[11] Juahir, H., Sharifuddin, M. Zain, Ahmad Zaharin Aris, Mohd Kamil Yusof, and M. A. A. S. a. M. B. Mokhtar. 2010. *Hydrological Trend Analysis Due to Land Use Changes at Langat River Basin.* Environment Asia. 2010(3(Special issue)): 20–31.

[12] Huang. F, W. X. L. L. Z. Z. W. J. 2009. Spatial Variation and Source Apportioment of Water Pollution in Qiantang River (China) Using Statistical Techniques. *Water Resource*. 2009(44): 1562–1572.

[13] Hongmei, Bu, J. W. Y. Z. W. M. 2013. Spatial Characteristic of Surface Water Quality in the Hiacheng River (Liao River Basin) in Northeast China. *Enviromental earth science*. 2013(70): 2865–2872.

[14] Shamsudin, M. S. H. J. Z., S. M & Adhan, N. H. 2011. Surface River Water Quality Interpretation Using Environmetric Techniques: Case Study at Perlis River Basin, Malaysia. *International Journal on Environmental Protection*. 1(5): 1–8.

[15] Abdullah, M. P., & Nainggolan, H. 1991. Phenolic Water Pollutants in a Malaysian River Basin. *Environmental Monitoring and Assesment*. 423–431.

[16] Binnie, D., & Gourley. 1961. *Report on Proposed Water Supply to Seremban and Port Dickson-Linggi Scheme.* Malaysian Govermen: Serembang, Negeri Sembilan, Malaysia.

[17] Binnie, D., & Gourley. 1979. *Report on Development Plan for Seremban and Port Dickson Water Supply*. Malaysian Govermen: Seremban, Negeri Sembilan, Malaysia.

[18] Nather Khan, I., & Firuza, M. B. 2010. Spatial and Temporal Variations of Silica in a Disturbed Tropical River Basin. *Sains Malaysiana*. 2010: 189–198.

[19] Lim, R.P. 1987. *Linggi River Basin: An Approach the Management of Its Water Resource*. Institute of Advance Studies, University of Malaya: Kuala Lumpur.

[20] Mann, H. B. 1945. Non-Parametric Tests Against Trend. *Econometrica.* 1945(13): 245–259.

[21] Kendall, M. G. 1975. *Rank Correlation Methods.* 4th ed. 1975. London: Charles Griffin.

[22] El-Shaarawi, A. H. 1993. Environmental Monitoring, Assessment and Prediction of Change. *Environmentrics*. 1993(4): 381–398.

[23] Gilbert, R. O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold Company.

[24] Hirsch, R. M. S. J. 1984. Non-Parametric Trend Test for Seasonal Data With Serial Dependence. *Water Resource Research*. 20(4): 727–732.

[25] Douglas, E. M., V. R., Kroll, C. N. 2000. Trends in Floods and Low Flows in the United States: Impact of Spatial Correlation. *Journal of Hydrological*. 2000(240): 90–105.

[26] Yi Wang; Peng Wang; Yujun Bai; Zaixing Tian; Jingwen Li, X. S. L. F. M. B.-L. L. 2013. Assesment of Surface Water Quality via Multivariate Statistical Techniques: A Case Study of the Songhua River Harbin Region, China. *Journal of Hydro-environment Research.* 2013(7): 30–40.

[27] Massart, D. L., & Kaufman, L. 1983. *The Interpretation of Analytical Data by Use of Cluster Analysis*. New York: Wiley.

[28] McKenna, J. E., Jr. 2003. An Enhanced Cluster Analysis Program with Bootstrap Significance Testing for Ecological Community Analysis. *Environmental Modelling & Software*. 18(2): 205–220.

[29] Willet, P. 1987. *Similarity and Clustering in Chemical Information Systems*. New York: Research Studies Press, Wiley.

[30] Singh, K. P., Malik, A., Mohan, D., & Sinha, S. 2004. Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variation In Water Quality of Gomti River (India)-A Case Study. *Water research*. 2004(38): 3980–3992.

[31] Singh, K. P., 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.* 2005(35): 3581–3592.

[32] Kannel, P. R., Lee, S., Kanel, S. R., Khan, S. P. 2007. Chemometric Application in Classification and Assessment of Monitoring Locations of an Urban River System. *Analytica Chimica Acta*. 2007(582): 390–399.

[33] Kim, J.O., Mueller, C.W. 1987. Introduction to Factor Analysis: What It Is and How to Do It. *Quantitative Aplications in the Social Science Series*. Newbury Park: Sage University Press.

[34] Vega, M., Pardo, R., Barrado, E., Deban. L. 1998. Assesment of Seasonal and Polluting Effects on the Quality of River Water by Exploratory Data Analysis. *Water Research*. 1998(32): 3581–3592.

[35] Liu, C. W., Lin, K. H., Kuo, Y. M. 2003. Application of Factor Analysis in the Assessment of Groundwater Quality in a Blackfoot Disease Area in Taiwan. *The Science of the Total Environment*. 2003(313): 77–89.

[36] Chung, Y. C., Li, Y. H., Chen, C. C. 2005. Pollutant Removal from Aquaculture Wastewater Using the Biopolimer Chitosan at Different Molecular Weights. *Journal of Environmental Science and Health*. 40(9): 1775–1790.

[37] Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Vousta, D. 2003. Anthemidis, A. Assessment of Surface Water Quality in Northern Greece. *Water Research*. 37: 4119–4124.