Case study Malaysia: Spatial water quality assessment of Juru, Kuantan and Johor River Basins using environmetric techniques

Masthurah A.[1]; Juahir H.[2]; Mohd Zanuri N.B.[1*]

Received: June 2020 Accepted: October 2020

Abstract

This study investigates spatial water quality assessment of selected river basins in the three different states in Malaysia. Environmetric techniques namely, cluster analysis (CA), principal component analysis (PCA), and discriminant analysis (DA), were applied to study the spatial variations of the most significant water quality variables in order to determine the origin of pollution sources on water quality data of Juru River Basin, Kuantan River Basin and Johor River Basin. 13 water quality parameters were initially selected and analyzed. Three spatial clusters were formed based on CA, and these clusters were designated as high pollution source (HPS), medium pollution source (MPS), and low pollution source (LPS) at the three river basins, respectively. Forward and backward stepwise DA managed to discriminate water quality variables, respectively from the original 13 variables. The result of this spatial analysis assessment is supported by PCA (varimax functionality,) that was used to investigate the origin of each water quality variable due to land use activities. Thus, this analysis makes it possible to observe the significance of the pollutant sources which contribute to river pollution. Five principal components (PCs) were obtained for all HPS, MPS and LPS regions of all the three river basins, respectively. Pollution sources for the three river basins were mainly originated from industrial waste, municipal waste, domestics waste and also from agricultural runoffs. Finally, the environmetric techniques analysis manage to provide convincing result on the spatial variation of water quality in all the three studied river basins and this eventually will allow more effective and efficient river quality management activities.

Keywords: Cluster analysis, Principal component analysis, Discriminant analysis

1- Centre for Marine and Coastal Studies (CEMACS), Universiti Sains Malaysia, 11800 Gelugor Pulau Pinang.
2- East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Kampung Gong Badak, 21300 Terengganu
*Corresponding author's Email: lailazanuri@usm.my

[DOI: 10.18331/SFS2021.7.2.2]

[DOI: 10.18331/SFS2021.7.2.2]

Introduction

Water is an important resource that is necessary for all aspect of human and ecosystem survival (Najah, El-Shafie, Karim, & El-Shafie, 2013). Water is widely used for irrigation, culturing of fishes and drinking. However in recent years, the flow of river contains heavy industrial effluent and municipal sewage effluent with the advent of urbanization (Varunprasath & A. Daniel, 2010). The general term water quality is used to describe the condition of water characteristic including physical, chemical and biological characteristic of the water (Dogan, Sengorur, & Koklu, 2009). . Many rivers are experiencing from deterioration quality of its characteristic condition, which in turn affects people's health, economy and as well as the environment (Department of Environment (DOE), 2003). Surface water is one of the environmental components that are most vulnerable to pollution impact because this surface river water is the place that received all of the waste that are being disposed into the river by anthropogenic activities (Hamirdin, 2000) In Malaysia, river is the main source of drinking water supplies. The contaminated river will result into a limited quantity of clean water and thus will eventually increase the water treatment cost.

Spatial analysis is one of the methods that usually performed for the purpose of evaluating and identifying the most significant water quality parameters that supposed to be concerned due to the land use activities that affect river ecosystem (Razali, Syed Ismail, Awang, Praveena, & Zainal Abidin, 2018) This land use activities will gradually alter the types of pollutant loadings into the river system (Juahir, Zain, Aris, Yusoff, & Mokhtar, 2010). . Spatial analysis can be conducted by using environmetric technique. Environmetric or also known as chemometric is one of the environmental analytical chemistry fields that utilize multivariate statistical approach for the data analysis (Einax, Zwanziger, & Geiss, 1997; Simeonova, Simeonov, & Andreev, 2003). It can be considered to be the most appropriate analysis performance in order to prevent misinterpretation upon analyzing a large environmental data set (Simeonov, Einax, Stanimirova, & Kraft, 2002; Varmuza & Filzmoser, 2009) . Three common environmetric analysis that usually perform in order to classify wide range of data into groups are the hierarchical agglomorative cluster analysis (HACA) and principal component analysis (PCA) which are then furthered by pattern recognition analysis namely, discriminant analysis (DA) (Adam, 1998) . The objectives of this study are (i) to evaluate spatial variations in the river water quality data of Juru, Kuantan and Johor river basins using environmetric techniques and (ii) to identify the pollution loadings variations due to land use and anthropogenic activities in the three studied river basins.

Study areas and methods

Site description

Three river basins namely, Juru River Basin, Kuantan River Basin and Johor River Basin have been selected in this study.

Johor River Basin

Johor which is the second largest state in Peninsula Malaysia drains a catchment about 2636 km^2 and the Johor River Basin become the main river in Johor that located at coordinate 1°27′00″N 104°01′00″E. This river flows through a north-south direction and empties into the Straits of Johor. Water quality of this river basin is gradually decline due to the increasing levels of various pollutants. The contaminants eventually flow into Johor River estuaries which are rich in habitats that provide spawning and feeding areas for fish and poultry (Najah et al., 2013). The Johor River Basin originates from Gunung Belumut and Bukit Gemuruh in the north and flows to the south eastern part of Johor and finally into the Straits of Johor. The maximum length and breadth of this catchment are 80km and 40km, respectively. About 60% of the catchment area is undulating highland rising to a height of 366m while the remainders are lowland and swamps. The highland in the north is mainly jungle. In the south, a major portion had been cleared and planted with oil palm and rubber. A great amount of pollutants from various sources are discharged into this river. The sources of pollutant which are mainly from the sewerage network of Johor Bahru, Pasir Gudang, Ulu Tiram and Kota Tinggi cities, the industrial wastewater from many industries in the surrounding and agricultural wastewater that contain fertilizers and pesticides are discharges into Johor River (Ismail, 2009).

Juru River Basin

The Juru River Basin is about 75 km^2 in area which is originated from Bukit Mertajam Hill (Lim & Kiu, 1995). This river basin is made of 2 main upstream which are western and eastern upstream. The western upstream are Permatang Rawa River and Rambai River while the eastern upstream are made of Kilang Ubi River and Pasir River. Meanwhile, Juru River forms the middle and downstream (Zali, Retnam, & Juahir, 2011). Juru River Basin had been identified as one of the polluted river in Malaysia ('Department of Environmental' (DOE), 2008). This river has been undergoing extensive monitoring on its water quality as this river experience industrial activities from nearby areas where by this river is continuously become the strategic main effluent discharges location from manufacturing industrial at Prai Industrial Estate (Zali et al., 2011). Types of industries that are operated along Juru River Basin are electronics, textiles, food processing, metal products, rubber, chemical plants and transport equipment industries

[DOI: 10.18331/SFS2021.7.2.2]

(Alkarkhi, Ismail, Ahmed, & Easa, 2009) . There is also a vast harbour operation which involves petroleum based activities located at the estuary which may influence the river water quality during intertidal phase changes (Zali et al., 2011). Juru River Basin and its tributaries flow through urbanized areas and are heavily polluted by domestic waste and discharges from pig farms. According to the DOE report, water quality of this river basin was poor and it seems to be no improvement over the years (Department of Environmental (DOE), 1977; DOE-USM, 1992).

Kuantan River Basin

On the other hand, Kuantan River Basin is about 52,903 ha and located mainly in the forest reserve area. Kuantan River Basin is an important water catchment area which is surrounded by dipterocarp forest area that store variety of flora and fauna of tropical moist forest. However, over the last decade, the development of land for agricultural purposes and the production of natural resources such as timber had changed the entity coverage of this Kuantan River basin and indirectly affect the water river quality of the basin. There is a clear land use change, particularly in the area of the origin forest reserve area of Kuantan River Basin. It is reported that more than 9000 ha of forest land around the basin has been transformed into bushes or agricultural land during the period of 1990-2002. Since 1995, there was a degradation of forest areas due to logging activities that had been carried out nearby the basin and there was also an increasing of agricultural area especially in the area of Kenau River. Land use activities in the forest of Kuantan River Basin affect the river water quality due to soil erosion and surface runoff (Fig. 1).

The Data

For the purpose of this study, data of river water quality from three river basins, namely Juru River Basin, Kuantan River Basin and Johor River Basin which consist of a number of monitoring stations were obtained from Department of Environment (DOE), Ministry of Natural Resource and Environment of Malaysia. All the water quality data from each selected station in this study were based on the available data that had been recorded from 2003-2007. Referring to the sample site, 5 sites represent the Juru sub-basin, namely, Juru, Kilang Ubi, Pasir, Rambai, and Ara while 8 sites represent Kuantan sub-basin, namely, Belat, Kuantan, Galing Besar, Galing Kecil, Pinang, Charu, Riau and Kenau. On the other hand, 21 sites represent the Johor subbasin namely, Layang, Serai(Hilir), Tiram, Tiram(Hulu), Bukit Besar, Semanger, Johor, Telor, Berangan, Temon, Layau Kiri, Semenchu, Chemangar, Lebam, Sening,

[DOI: 10.18331/SFS2021.7.2.2]

Santi, Anak Sg. Sayong, Sayong, Penggeli, Sebol and Linggiu.

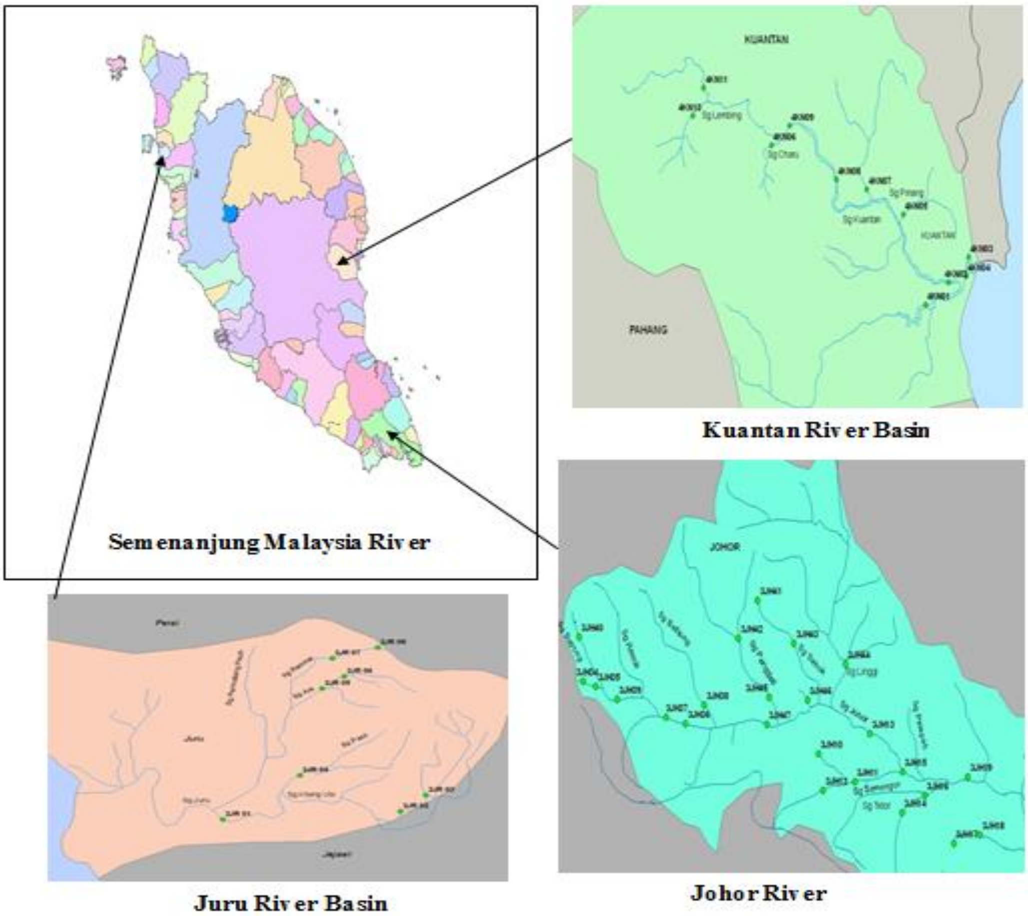


Figure 1: Locations of sampling stations at the three river basins.

Due to the fact that some monitoring stations in these three studied river basins have missing data, only 13 consistent parameters were analyzed and examined among all the 30 river water quality data available. A total of 205 samples in Juru River Basins, 275 samples in Kuantan River Basins and 865 samples in Johor River Basins were used for the analyses. For this study, all the data obtained with 13 water quality parameters; dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), suspended solid (SS), pH, ammoniacal nitrogen (NH_3-N), dissolved solid (DS), total solid (TS), nitrate (NO_3^-), chloride (Cl^-), phosphate (PO_4^{3-}), *Escherichia coli*, and coliform were subjected for the environmetric techniques analysis by using XLSTAT2012 software. The descriptive statistics of the 5 years measured data set of each river basin are summarized in Table 1.

Table 1: Mean values of water quality measurement in Juru, Kuantan and Johor River Basins (2003-2007).

Juru River Basin

Station	STATION1 (2JR01)		STATION2 (2JR02)		STATION3 (2JR03)		STATION4 (2JR04)		STATION5 (2JR05)		STATION6 (2JR06)	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	2.019	1.24	1.609	1.178	2.021	1.757	1.45	0.819	0.818	0.893	1.534	1.123
BOD	11.294	6.824	22.529	22.187	9.824	9.654	42.412	68.252	28.824	26.862	26.412	28.504
COD	65.235	14.643	88.529	69.475	47.706	26.718	116	100.026	84.294	35.773	94.824	90.034
SS	42.294	26.619	21.059	11.694	20.353	36.333	21.059	15.664	39.529	20.657	17	13.465
pH	7.024	0.241	7.062	0.617	6.896	0.3	6.855	0.155	6.82	0.198	6.685	0.199
NH3-NL	6.569	3.323	12.433	10.482	3.567	1.691	2.589	1.359	7.371	4.061	4.345	3.166
DS	438.294	568.025	165.529	75.486	90.059	24.625	141.471	81.971	169.412	45.499	173.588	137.177
TS	480.588	578.033	186.588	84.39	110.412	54.029	162.529	93.763	208.941	42.159	190.588	149.574
NO3	0.084	0.086	0.386	0.465	0.478	0.71	0.209	0.618	0.189	0.589	0.19	0.271
Cl	168.706	282.254	14	6.068	10.824	3.312	29.588	29.687	36.235	12.596	44.765	44.882
PO4	0.607	0.43	1.861	1.691	0.41	0.414	0.482	0.306	0.889	0.52	0.457	0.346
E-coli	15005.88	15696.93	65182.35	87709.97	54082.35	47024.64	93517.65	135244	136518.3	251501.4	216435.3	637100.8
Coliform	96823.53	111981.2	134294.1	148627.7	147352.9	202738.6	228041.2	350454.3	1513318	5374863	2467941	9383452

Station	STATION7 (2JR07)		STATION8 (2JR08)		STATION9 (2JR09)		STATION10 (2JR10)		STATION11 (2JR11)		STATION12 (2JR12)	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	1.056	0.553	4.816	2.111	7.564	0.581	1.939	1.959	1.692	1.024	2.155	1.192
BOD	39.588	34.52	8.882	21.527	1.118	0.322	10.471	10.279	11.059	8.033	5.471	3.22
COD	107.706	59.084	49.882	78.914	17.647	2.989	49.353	24.024	52.529	15.01	46.471	21.41
SS	34.529	11.773	49.471	46.349	21.765	47.404	30.353	32.669	31.176	17.443	74.118	42.672
pH	6.964	0.307	6.995	0.428	7.014	0.289	6.821	0.228	6.812	0.276	7.004	0.48
NH3-NL	28.898	27.459	18.748	51.067	0.126	0.146	5.202	3.686	5.452	4.061	3.814	2.408
DS	339.706	171.002	202.588	364.116	20.729	1.563	126.824	50.501	171.353	73.998	5796.588	6014.504
TS	374.235	175.001	252.059	367.79	41.259	47.96	147.647	60.299	184.941	78.169	5110	5865.725
NO3	0.411	0.657	1.424	1.674	0.295	0.171	0.25	0.228	0.229	0.19	0.245	0.172
Cl	34.529	16.107	9.235	20.113	1.655	1.071	21.297	12.444	37.666	20.355	3262.824	3470.953
PO4	2.74	2.695	2.973	8.055	0.014	0.01	0.638	0.568	0.665	0.654	0.141	0.276
E-coli	74911.77	50633.12	18711.18	30669.93	8188.824	20666.07	198500	304025.1	48364.71	39154.55	87435.29	232043.4
Coliform	145941.2	116420.9	50725.88	62964.98	49498.24	42861.71	579941.2	901395.6	223594.1	222808.4	369417.6	885465.9

Kuantan River Basin

Station	Station 1 4KN01		Station 2 4KN02		Station 3 4KN03		Station 4 4KN04		Station 5 4KN05		Station 6 4KN06	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	5.545	0.887	5.292	0.769	1.858	0.922	3.285	1.178	4.697	0.907	6.407	0.722
BOD	1.72	1.217	1.56	0.852	20.64	35.893	8.4	7.457	1.6	0.849	1.56	1.235
COD	24.44	10.598	26.48	13.494	78.2	76.198	44.24	15.67	24.4	9.108	20.12	6.719
SS	18.96	16.098	26.36	19.787	26.72	14.307	17.48	11.236	8.72	5.326	20.76	41.208
pH	6.624	0.669	6.666	0.399	7.004	0.339	7.184	0.424	6.255	0.739	6.958	0.385
NH3-NL	0.31	1.018	0.299	0.568	4.39	2.23	4.01	1.646	0.141	0.399	0.115	0.22
DS	1298	3362.985	6373.96	6194.675	2006.16	1619.857	1835.72	2738.338	23.8	7.97	20.36	15.268
TS	1316.96	3361.681	6400.32	6185.579	2032.88	1622.493	1853.2	2740.798	32.52	10.565	41.12	41.972
NO3	0.31	0.173	0.189	0.178	0.216	0.372	0.244	0.233	0.236	0.153	0.253	0.178
Cl	621.44	1537.225	3562.96	3584.176	969.8	882.824	968.16	1551.903	4.16	2.31	3.42	6.77
PO4	0.01	0.009	0.091	0.379	0.507	0.362	0.354	0.219	0.092	0.381	0.01	0.019
E-coli	7977.6	17452.94	3924	6674.386	109816	106599	71096	74881.5	2436.8	3640.443	1544.08	2089.109
Coliform	34780.32	75559.09	42088	28976.13	472032	747663.8	242896	331690.2	21512	24900.3	18794.8	18799.77

[DOI: 10.18331/SFS2021.7.2.2]

Station	Station 7 4KN07		Station 8 4KN08		Station 9 4KN09		Station 10 4KN10		Station 11 4KN11	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	4.671	0.867	6.77	0.378	6.834	0.451	6.957	0.54	7.114	0.752
BOD	1.88	0.993	1.36	0.686	1.36	0.686	1.28	0.601	1.2	0.566
COD	22.72	6.168	19.36	4.156	18.96	5.134	19.88	5.03	18.36	2.91
SS	15.72	12.055	27.72	23.376	39.08	98.607	44.04	88.715	18.44	20.644
pH	6.2	0.62	6.994	0.417	6.811	1.007	7.211	0.567	7.282	0.47
NH3-NL	0.081	0.122	0.039	0.094	0.056	0.092	0.038	0.068	0.066	0.122
DS	25.32	11.939	13.12	2.233	12.08	2.331	17.8	6.759	10.08	1.787
TS	41.04	17.192	40.84	24.894	51.16	98.508	61.84	87.774	28.52	21.554
NO3	0.225	0.179	0.209	0.179	0.184	0.177	0.183	0.162	0.153	0.177
Cl	4.46	5.169	1.26	1.124	1.44	1.098	1.98	2.037	1.1	0.825
PO4	0.012	0.024	0.01	0.015	0.01	0.013	0.011	0.019	0.012	0.015
E-coli	3697.2	4356.846	1888	2636.865	1274	1385.537	939.08	1034.752	2195.6	2791.245
Coliform	16406	20597.63	12920	10818.39	12770	12493.84	14100.16	21480.39	14740.8	16558.91

Johor River Basin

Station	3JH03		3JH05		3JH06		3JH07		3JH08		3JH09	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	7.478	0.592	3.957	1.696	4.625	1.003	6.618	1.097	7.526	0.344	4.038	1.66
BOD	1.852	0.848	7.852	6.884	4.667	2.261	2.778	4.969	1.148	0.448	5.519	6.68
COD	24.704	7.413	47.519	22.705	34.667	13.28	25.259	11.781	19.444	6.517	41	23.707
SS	6.019	5.817	93.222	124.249	83.519	75.907	38.944	66.527	28.185	35.02	118.222	213.812
pH	6.931	0.576	6.733	0.421	6.388	0.278	5.963	0.403	5.981	0.576	6.268	0.255
NH3-NL	0.048	0.077	3.814	3.147	1.948	1.446	0.118	0.221	0.177	0.414	1.366	1.052
DS	15.111	2.183	1750.889	3046.837	74.481	29.614	30.185	9.794	30.148	12.128	51.296	12.866
TS	20.481	6.63	1829.63	3026.002	153	75.535	67	66.845	56.63	35.2	164.926	214.692
NO3	0.064	0.056	1.284	2.469	3.926	2.908	0.955	0.483	1.055	0.517	0.941	0.699
Cl	3.212	1.408	919.319	1682.448	10.554	3.072	6.992	3.284	7.323	3.922	10.221	3.573
PO4	0.034	0.08	0.544	0.417	0.124	0.207	0.059	0.181	0.055	0.191	0.097	0.209
E-coli	900.963	1344.163	22069.63	74614.68	13670.74	19988.56	4641.481	5510.192	1955.556	2198.48	21790.37	21867.9
Coliform	15550.44	43002.99	112015.9	304041.7	58585.56	66964.14	23450.37	30849.43	22007.41	36515.21	81993.33	86044.56

Station	3JH10		3JH11		3JH12		3JH13		3JH15		3JH16	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	2.829	1.259	6.929	0.678	6.958	0.599	6.503	0.388	6.322	0.33	7.546	0.294
BOD	16.222	13.409	1.444	0.737	2.37	2.312	1.444	0.497	1.741	1.075	1.37	1.191
COD	98.259	61.706	21.222	5.405	26.481	13.844	21.444	5.801	22.815	5.754	19.222	4.557
SS	25.519	16.509	33.852	21.425	22.148	14.62	82.852	56.95	53.074	27.715	46.852	107.794
pH	7.484	0.335	6.592	0.424	6.846	0.32	6.42	0.51	6.167	0.324	5.968	0.375
NH3-NL	13.427	9.944	0.103	0.143	0.138	0.273	0.126	0.234	0.091	0.126	0.098	0.163
DS	636.222	293.164	29.815	9.952	74.926	42.234	22.963	6.374	24.444	8.917	17	5.761
TS	644.222	329.242	62.185	23.753	94.556	45.735	102.444	63.026	75.111	31.389	62.63	108.948
NO3	2.43	4.817	0.956	0.296	1.217	0.784	0.617	0.276	0.599	0.163	0.448	0.229
Cl	130.281	153.263	5.408	2.275	16.538	10.996	4.611	2.358	5.028	2.106	2.871	1.317
PO4	2.623	1.446	0.05	0.044	0.165	0.118	0.068	0.203	0.03	0.035	0.023	0.032
E-coli	14237.04	17954.12	3581.481	3770.061	3655.556	3141.754	4260.741	5403.092	2812.593	4200.583	3080.741	4468.127
Coliform	103322.2	167688.6	29754.63	33331.31	30612.96	32023.87	33325.93	56649.7	15667.59	15516.66	17769.44	19180.77

Station	3JH18		3JH19		3JH20		3JH22		3JH25		3JH27	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	4.827	1.47	6.297	0.276	5.83	0.663	7.255	0.4	7.35	0.34	6.873	0.55
BOD	5.87	4.634	1.574	1.144	1.389	1.53	1.204	0.436	1.556	1.257	3.222	3.765
COD	52.426	41.961	21.093	7.82	18.907	5.375	18.981	5.186	23.704	16.704	28.111	12.793
SS	22.574	21.379	54.167	38.465	43.019	36.687	18.167	35.682	34.111	80.888	19.222	33.345
pH	7.021	0.54	6.313	0.408	6.219	0.256	5.918	0.321	5.294	0.617	6.348	0.518
NH3-NL	1.972	1.576	0.144	0.462	0.152	0.214	0.312	0.152	0.219	0.321	0.919	1.128
DS	231.537	170.045	21.241	5.51	28.278	15.05	29.204	6.409	31.63	9.511	63.296	41.705
TS	247.37	183.093	73.407	42.537	65.852	34.33	45.259	37.793	64.333	82.156	80.111	54.22
NO3	1.132	0.776	0.569	0.186	0.559	0.17	0.547	0.263	0.776	0.447	0.324	0.142
Cl	46.859	38.496	4.716	1.928	7.27	6.922	7.766	2.741	10.158	3.742	16.152	7.236
PO4	0.386	0.382	0.027	0.033	0.025	0.034	0.036	0.085	0.014	0.022	0.19	0.214
E-coli	8103.704	16175.28	2214.074	1724.755	5935.185	8547.599	2651.852	4399.694	3121.481	5734.126	3565.926	4995.491
Coliform	73255.56	138468.6	14603.7	12690.66	38781.48	77687.42	14847.41	20812.48	12294.07	18362.7	40451.85	76863.23

[DOI: 10.18331/SFS2021.7.2.2]

Station	3JH28		3JH30		3JH32		3JH33		3JH35		3JH36	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	6.186	1.002	4.002	0.904	6.193	0.92	4.745	0.688	4.283	2.189	6.459	0.635
BOD	2.556	2.409	1.815	1.156	12.778	19.846	1.667	0.861	5.481	7.047	1.63	0.777
COD	23.185	5.551	26.259	12.24	79.185	99.005	29.815	16.177	41.481	42.497	24.407	8.987
SS	86.926	215.008	19.722	17.662	63.389	99.536	29.519	37.803	57.741	83.394	84.926	140.109
pH	6.21	0.355	6.344	0.422	6.912	0.47	7.229	0.463	6.555	0.23	5.933	0.482
NH3-NL	0.716	0.515	0.082	0.093	3.404	3.963	0.046	0.063	0.516	1.093	0.069	0.098
DS	229.852	593.082	11899.74	6227.119	260.778	245.42	24996.3	2859.485	72.333	43.95	36.37	8.336
TS	314.185	614.861	11367.59	6583.28	310.333	298.535	24122.07	5521.555	126.333	104.51	113.889	139.725
NO3	0.535	0.261	0.237	0.532	0.986	1.685	0.179	0.852	0.872	0.515	0.813	0.165
Cl	102.182	298.824	6600.507	3462.343	51.333	55.414	14177.78	2132.002	14.216	8.601	9.078	2.574
PO4	0.026	0.036	0.034	0.063	0.649	0.7	0.075	0.235	0.081	0.136	0.019	0.028
E-coli	5610.37	7501.218	1794.815	4298.242	14243.33	18437.5	1166.667	2204.294	8977.037	11434.09	7211.852	11842.3
Coliform	39208.15	54904.39	13141.48	18670.92	90666.3	166029.8	12242.22	41441.05	98326.85	213906.1	68821.3	166677.8

Station	3JH37		3JH40		3JH42		3JH43		3JH44		3JH45	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	6.717	0.642	7.6	0.916	7.495	0.804	6.542	0.517	7.206	0.361	7.233	0.503
BOD	1.481	0.739	1.148	0.448	1.074	0.262	2.889	2.409	1.296	0.457	1.333	0.667
COD	22.481	6.652	18.481	3.236	18.778	2.131	28.741	11.316	19.444	3.457	20.481	5.231
SS	55.926	126.896	21.185	15.072	31.407	24.456	21.407	14.489	58.519	54.254	36.63	27.195
pH	5.636	0.429	5.439	0.571	5.93	0.614	6.357	0.48	6.066	0.406	6.237	0.84
NH3-NL	0.108	0.153	0.041	0.08	0.043	0.059	0.812	1.22	0.103	0.091	0.061	0.086
DS	36	7.712	32.296	13.179	18.444	3.292	79.741	54.769	13.815	12.76	20.296	3.74
TS	89.444	129.798	52.185	23.12	48.481	27.176	96.185	55.503	70.778	58.507	55.111	30.459
NO3	0.952	0.247	0.818	0.438	0.581	0.204	0.858	0.632	0.305	0.255	0.645	0.226
Cl	10.202	3.432	10.484	3.501	4.082	1.528	24.085	13.133	3.709	6.611	5.65	1.657
PO4	0.015	0.023	0.026	0.093	0.071	0.282	0.09	0.221	0.058	0.196	0.097	0.297
E-coli	4809.259	9609.718	1457.037	2533.127	2471.481	2529.387	2033.333	1816.59	1671.111	2591.252	3525.556	3875.578
Coliform	25819.44	48902.18	9708.222	12213.16	22500.74	54093.35	18925.93	22245.56	10420.74	13903.8	16873.33	19451.84

Station	3JH46		3JH47	
Variable	Mean	SD	Mean	SD
DO	3.49	1.778	5.804	0.429
BOD	47.259	85.348	1.852	0.931
COD	354.926	895.858	24.519	10.005
SS	251.778	891.881	66.185	60.527
pH	7.253	0.489	6.123	0.482
NH3-NL	11.396	10.989	0.147	0.218
DS	602.259	520.376	31	5.67
TS	834.333	1085.324	94.074	63.624
NO3	1.696	4.286	0.727	0.324
Cl	154.722	263.425	7.846	2.583
PO4	4.092	6.245	0.072	0.212
E-coli	8166.667	10058.2	5973.333	13920.22
Coliform	59002.96	86295.88	49318.52	115431

Methods of analysis cluster analysis

CA operates on data sets and forms well-defined groups that actually do not exist, but are assign to be clustered together due to the similar level characteristic that occupied by them. The objective of CA is to classify a sample of entities into a smaller numbers usually mutually exclusive groups based on the multivariate similarities among entities (McGarigal, Stafford, & Cushman, 2000) . In this paper, hierarchical agglomerative (HACA) was employed in order to identify the group of river region site (spatial). Ward’s method, using Euclidean distances to measure similarity in HACA is the common

[DOI: 10.18331/SFS2021.7.2.2]

analysis used which aimed at identifying variables that have a high rate of homogeneity level into a group based on the selection criteria set. HACA result is illustrated by a dendrogram, presenting the clusters and their proximity (Juahir et al., 2010). Agglomerative techniques begin with each entity in a class of its own, then fuse (agglomerate) the classes into larger classes. These procedures are well known and they are used widely in ecological research (McGarigal et al., 2000). In this paper, HACA was conducted in order to determine the classification of sampling sites (spatial) in the three river basins; Juru River Basin, Kuantan River Basin and Johor River Basin into groups.

Discriminant Analysis

Discriminant analysis specifies and examines variables that are dominant or well-discriminated among certain data groups. Discriminant function (DF) for each group is created through this DA technique (Johnson & Wichern, 2008) by using below equation, Eq, 1:

$$F(G_i) = k_i + \sum_{j=1}^{n} w_{ij} P_{ij} \quad (1)$$

Where i is the number of groups (G) , k_i is the constant inherent to each group, n is the number of parameters used for the classification a set of data into a given group and w_j is the weight coefficient assigned by discriminant function analysis (DFA) to a given parameter (p_j).

Discriminant analysis (DA) refers to a couple of closely related procedures that have similar objective in discriminating among well-defined group of sampling entities based on a suite of characteristic. This is contrast to cluster analysis (CA), which attempt to organize entities into classes or groups. Cluster analysis often serves as a precursor to DA when prespecified groups do not exist where artificial groups are created by CA, and then ecological differences among the newly created groups are described using DA (McGarigal et al., 2000). In this paper, discriminant analysis was performed in order to examine whether group differ with regards to the mean of variable in predicting the group membership.

In order to conduct this analysis, data from the three assigned region groups in each of the three river basins which had been obtained from CA were selected. The discriminant analysis was performed by using standard, forward stepwise and backward stepwise modes. These three modes were performed in order to determine water quality variables that have high variations according to their spatial distribution among the three studied river basins. In the forward stepwise mode, variables are included gradually starting from the most significant variables until no significant changes are obtained. In backward stepwise mode, the variables are removed gradually starting with the less significant variables until no significant changes are obtained.

[DOI: 10.18331/SFS2021.7.2.2]

Principal component analysis

Principal component analysis gives information upon the most significant variables according to the spatial and temporal variations which distinguish the whole data set by excluding the less significant parameters with minimum loss of original information (Kannel, Lee, Kanel, & Khan, 2007; Singh, Malik, Mohan, & Sinha, 2004; Singh, Malik, & Sinha, 2005) The principal component (PCs) can be expressed as (Eq 2):

$$zij=ai1x1j+aai2x2j+...+...aimxmj \quad (2)$$

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

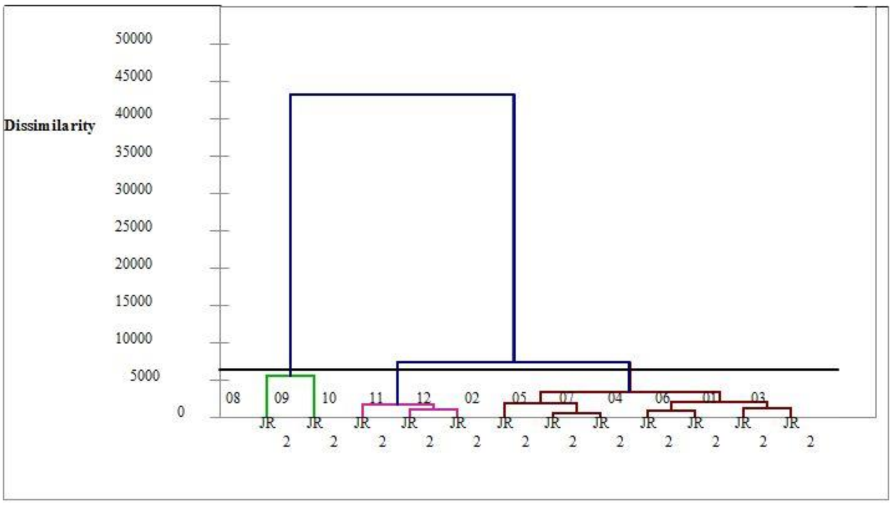
The PCs generated by PCA are sometimes not readily interpreted; therefore, it is advisable to rotate the PCs by varimax rotation. Kim and Mueller (1987) states that, the result of the PCs after varimax rotation with the amount of eigenvalues more than 1 are considered significant for the purpose to acquire new groups of variables called varimax factors (VFs). The VF coefficients which have correlation greater than 0.7 are considered as “strong”;0.5–0.69, as “moderate”; and 0.30–0.49, as “weak” significant factor loadings (Liu, Lin, & Kuo, 2003). In this study, only the VF coefficient that have strong loadings (greater than 0.7) were being considered. Source identification of different pollutants was made on the basis of different activities in the catchment area in light of previous literatures. In this paper, PCA was applied to the data set consist of 13 parameters for each region (HPS, MPS and LPS) of the three studied river basins. Calculations of input data matrices (variables × cases) for this PCA were 13 × 224 for HPS region, 13 × 312 for MPS region and 13 × 807 for LPS.

Results and discussion

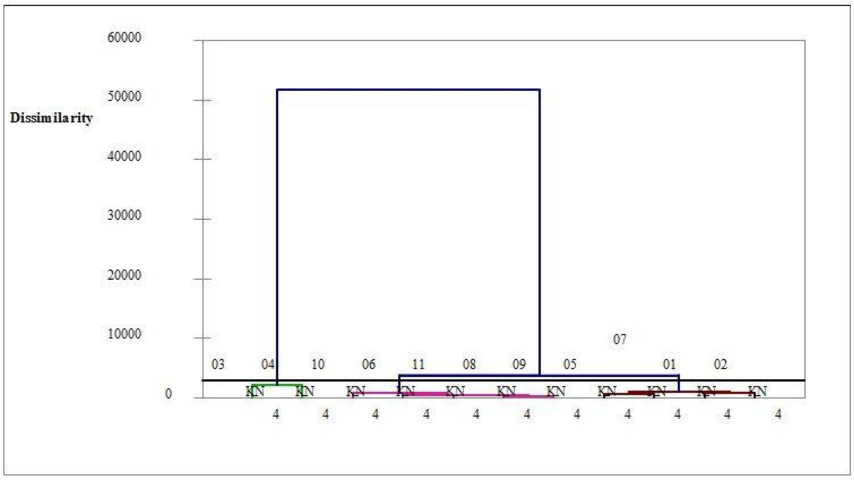
Determination of the sampling station groups Cluster analysis

Under this section, water quality parameters in each data set were examined by using cluster analysis. Cluster analysis was performed towards all of the data sets in the three river basins in order to specify each monitoring stations according to the level of their homogenous characteristics. This analysis was performed towards all of the water quality data in order to assess spatial variation among the sampling stations on each of the three river basins. The cluster analysis resulted into three cluster of sampling stations in Juru, Kuantan and Johor River Basins, respectively (Fig. 2).

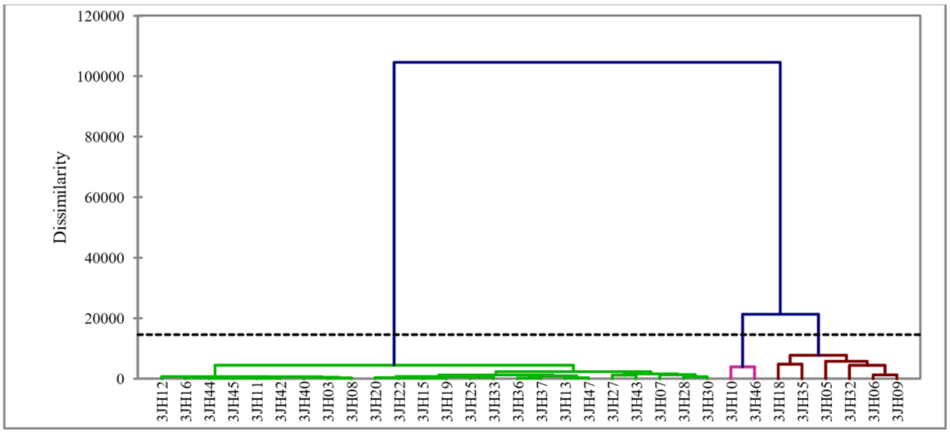
[DOI: 10.18331/SFS2021.7.2.2]



Juru River Basin



Kuantan River Basin



Johor River Basin

Figure: 2 Dendrogram shows sampling stations that had been classified in each three river basins.

The cluster procedure formed three clusters or groups in a very convincing way as the sites in each group have similar characteristics and natural backgrounds. For example, in Juru River Basin, cluster 1 (2JR01, 2JR02,

[DOI: 10.18331/SFS2021.7.2.2]

2JR03, 2JR04, 2JR05, 2JR06, 2JR07) represent the high pollution sources (HPS) region, cluster 2 (2JR10, 2JR11, 2JR12) represent the moderate pollution sources (MPS) region and cluster 3 (2JR08, 2JR09) represent the low pollution sources (LPS) region. The three clusters with three different regions also generated in the other two river basins, Kuantan River Basin and Johor River Basin. Table 2 represents the sampling stations that have been grouped by HACA.

Table 2: Sampling stations that have been grouped into regions.

Regions	River basins /sampling stations		
	Juru River Basin	**Kuantan River Basin**	**Johor River Basin**
HPS	2JR01,2JR02,2JR03, 2JR04,2JR05,2JR06, 2JR07	4KN03,4KN04	3JH10,3JH46
MPS	2JR10,2JR11,2JR12	4KN01,4KN02,4KN075, 4KN07	3JH05,3JH06,3JH09,3JH18,3JH32, 3JH35
LPS	2JR08,2JR09	4KN06,4KN08,4KN09, 4KN10,4KN11	3JH03,3JH07,3JH08,3JH1,3JH12, 3JH13,3JH15,3JH16,3JH19,3JH20, 3JH22,3JH25,3JH27,3JH28,3JH30, 3JH33,3JH36,3JH37,3JH40,3JH42, 3JH43,3JH44,3JH45,3JH47

The result of HACA technique implies that rapid assessment of water quality for the whole stations can be done by monitoring only one station in each cluster that had been assigned by CA. This is because only one monitoring station is already enough to represent the water quality data for the whole group members in each cluster group as every group have their own similar level of data quality characteristics. For example, in Kuantan River Basin, only station 4KN03 in cluster 1 (HPS), station 4KN01 (MPS) in cluster 2 and station 4KN06 (MPS) in cluster 3 need to be monitored in order to represent the water quality assessment of the whole Kuantan River Basins. The result of this analysis proved that cluster analysis (CA) technique is functional upon the classification of river water quality data for optimum future sampling and monitoring strategies.

Among the three river basins, Juru River Basin is proved to be the most polluted river basin as 7 of its monitoring stations are clustered under high pollution sources (HPS) group. This is due to the fact this river basin receives heavy pollution loadings from nearby urbanized area that densely populated by humans and multiple types of industrial located along the basin. In Kuantan River Basin, 2 monitoring stations; 4KN03 at Galing Besar River and 4KN04 at Galing Kecil River are clustered under high HPS group. Galing Besar River is mainly

[DOI: 10.18331/SFS2021.7.2.2]

[DOI: 10.18331/SFS2021.7.2.2]

being polluted by sediment deposition and siltation that resulted from anthropogenic activities. This river is also overwhelmed with various contaminants entering it. The other monitoring stations are not heavily polluted as Kuantan River Basin is mainly located in the forested area. The natural vegetation within this river basin act as filters that prevent the sediment deposition and uptake nitrate and phosphorus which is mainly originate from fertilizer usage. There is also a little disturbance along the river water within this basin which is not resulted into a serious negative impact to the river water quality such as Belat River where by it support mainly residential areas with light industries with low water demand and the existing houses are served by water closets and septic tanks, but there is no evidence of serious water pollution problem although the sullage water is discharged into surface drain (Hill, 1981). For Johor River, station 3JH10 at Bukit Besar River and station 3JH46 at Sayong River are clustered under HPS group. The major land use at the Johor River Basin is oil palm and other crops plantations. There are many oil palm plantation and FELDA land development located in the surrounding area of Bukit Besar River and Sayong River which may influence to the river water quality of the two rivers (Hamza, 2009).

Spatial variation of river water quality

Discriminant Analysis

DA was performed in order to determine water quality variables that have high variation according to their spatial distribution which had been classified by CA into three main clusters for each river basin. The combination of HPS region for Juru,Kuantan and Johor River Basin represent as the dependent variables while the water quality parameters were the independent variables. The same goes for the combination of MPS and LPS region for Juru, Kuantan and Johor River Basin. DA was carried out via standard, forward stepwise and backward stepwise modes. In the forward stepwise mode, variables are included gradually beginning from the most significant variable until no significant changes are obtained. In backward stepwise mode, variables are removed gradually beginning with the less significant variable until no significant changes are obtained. The accuracy of spatial classification all the three regions of the studied river basins were 79.33% (13 discriminant variables) for standard mode, 79.33.74% (9 discriminant variables) for forward stepwise mode and 70.73% (10discriminant variables) for backward stepwise mode. In forward stepwise mode, DO, NH_3N, CI, pH, NO_3^-, BOD, COD and PO_4^{3-} were determined to be the significant variables which indicate that all these nine parameters have high variation upon their spatial distribution in the region of all the three studied river basins. For backward stepwise, *E. coli* exist as the tenth parameter that have high distribution upon the spatial variation. The result of DA which had been illustrated in a classification matrix for each clustered region is shown in Table 3. Box and whiskers plot of some water quality parameters in five years periods (2003-2007) is shown in Fig. 3.

Table 3: Classification matrix for DA of spatial variations in the three studied river basins.

Sampling regions of studied river basins	% Correct	Regions assigned by DA HPS	LPS	MPS
Standard DA mode (13 variables) HPS	**75.34%**	**32**	**2**	**20**
LPS	95.91%	0	635	13
MPS	39.23%	6	59	97
Total	79.55%	38	696	130
Forward stepwise mode (9 variables) HPS	77.58%	32	2	20
LPS	95.91%	0	634	14
MPS	39.23%	5	59	98
Total	79.70%	37	695	132
Backward stepwise mode (10 variables) HPS	**75.34%**	**32**	**2**	**20**
LPS	95.91%	0	634	14
MPS	39.23%	6	60	96
Total	79.33%	38	696	130

The Wilk's Lambda test for standard mode gave a Lambda value of 0.281 and $p<0.0001$. The null hypothesis states that the means of vector of the three clusters region (HPS, MPS and LPS) are equal. The alternatives hypothesis, on the other hand, states that at least one of the means of vectors is different from another. Since the computed p-value is lower than the significance level of alpha=0.05, the null hypothesis should be rejected and the alternative hypothesis should be accepted. The risk of rejecting null hypothesis while it is true is lower than 0.01%. Thus, the clusters are indeed different from each other.

[DOI: 10.18331/SFS2021.7.2.2]

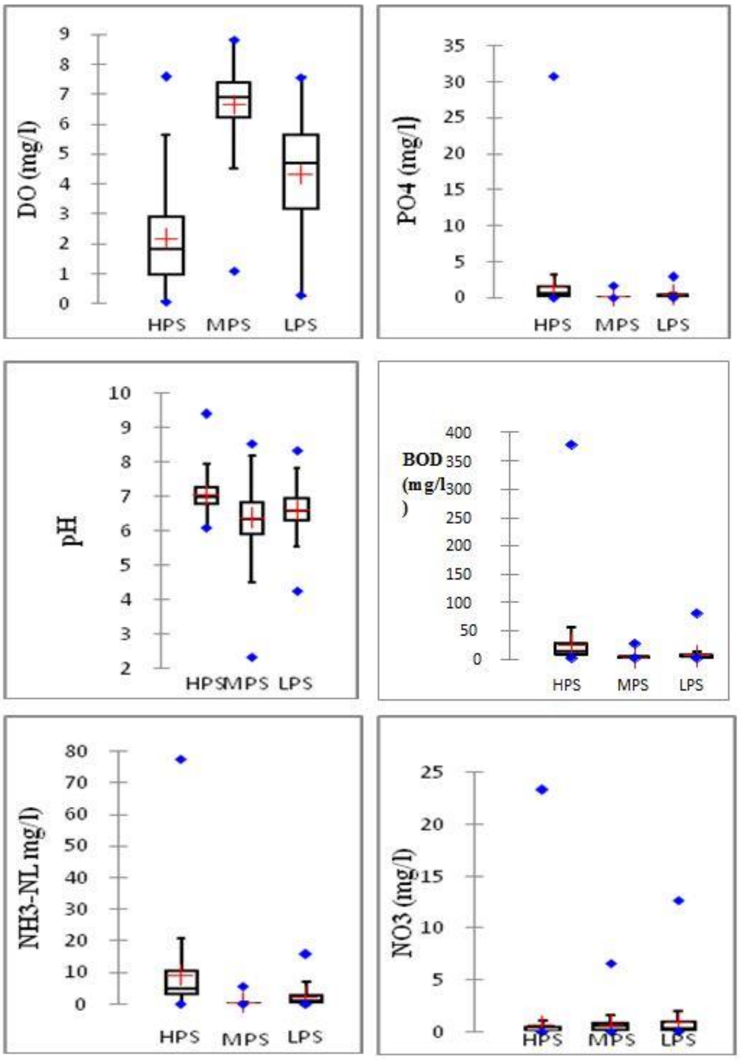


Figure 3: Box and whisker plots of some parameters separated by spatial DA associated with the three river basins. The crosses are mean values, top and bottom of whiskers indicate maximum and minimum values, respectively while horizontal lines of the boxes from top to bottom indicate the third quartile, median, and first quartile, respectively.

Principal Component Analysis (PCA)

Principal component analysis (PCA) was performed on the data set for the purpose of identifying the source of pollutant loading in each clustered region among the three river basins. Five PCs were obtained for each HPS, MPS and LPS regions, with the concerned amount of eigenvalues (larger than 1) sum up the total variance of data set almost 81.9%, 78.8% and 74.1% respectively. Varimax rotation that had been performed through this PCA technique managed to obtain five varimax function (VF) in each HPS, MPS and LPS region. Table 4 shows the details of five VFs obtained together with the amount of variable loading and variance explained for every region groups in the three river basins.

[DOI: 10.18331/SFS2021.7.2.2]

Table 4: Result of VFs that consist of variables loading after varimax rotation for water quality data in HPS, MPS and LPS regions of the three river basins

Variables /unit	HPS region of the three river basins					MPS region of the three river basins					LPS region of the three river basins				
	VF 1	VF 2	VF 3	VF 4	VF 5	VF 1	VF 2	VF 3	VF 4	VF 5	VF 1	VF 2	VF 3	VF 4	VF 5
DO(mg/l)	-7.6E-02	-1.5E-02	-1.6E-01	-2.0E-01	7.1E-01	-4.1E-01	-1.1E-01	-3.89E-01	1.36E-01	4.72E-01	-6.40E-01	-2.37E-01	-1.40E-01	6.00E-02	-3.00E-02
BOD(mg/l)	7.5E-01	1.0E-01	-4.8E-02	3.0E-01	-1.8E-01	8.9E-01	-5.6E-02	-3.50E-02	-1.72E-01	-1.30E-02	-2.70E-02	8.55E-01	9.00E-03	4.80E-02	2.04E-01
COD(mg/l)	9.8E-01	5.5E-02	-1.4E-02	1.2E-01	2.4E-02	8.5E-01	-1.4E-02	-1.00E-03	-2.29E-01	-1.10E-02	1.50E-01	8.64E-01	3.60E-02	-4.60E-02	3.60E-02
SS(mg/l)	9.6E-01	3.2E-02	0.0E+00	1.0E-03	3.9E-02	2.4E-01	-1.9E-02	6.90E-02	-7.41E-01	2.41E-01	-3.10E-02	3.53E-01	4.37E-01	-1.60E-01	-4.56E-01
pH	6.9E-02	2.4E-01	-3.0E-02	4.7E-01	6.0E-01	3.7E-01	1.5E-01	9.80E-02	5.41E-01	3.81E-01	1.69E-01	4.80E-02	7.90E-02	7.57E-01	2.67E-01
NH_3-N (mg/l)	6.9E-02	-3.0E-02	-4.5E-02	8.9E-01	-1.2E-01	7.4E-01	4.0E-03	2.74E-01	2.78E-01	-7.20E-02	-6.70E-02	4.05E-01	1.38E-01	-6.20E-02	6.53E-01
DS (mg/l)	-1.6E-02	9.9E-01	-5.0E-03	2.8E-02	2.2E-02	-1.6E-02	9.9E-01	-1.80E-02	2.80E-02	-3.60E-02	9.83E-01	1.70E-02	-4.00E-02	7.20E-02	-2.00E-03
TS(mg/l)	2.1E-01	9.7E-01	-3.0E-03	2.8E-02	2.9E-02	-1.4E-02	9.9E-01	-1.40E-02	-2.00E-03	-2.90E-02	9.73E-01	2.10E-02	-3.10E-02	6.80E-02	-7.00E-03
NO_3^-(mg/l)	1.0E-02	-7.3E-02	5.0E-03	-1.5E-02	6.5E-01	1.5E-02	-1.2E-01	-2.20E-02	-1.30E-01	8.17E-01	-1.18E-01	5.80E-02	1.24E-01	-7.72E-01	2.70E-01
Cl^- (mg/l)	-4.1E-02	9.9E-01	3.0E-03	-5.1E-02	-1.8E-02	-3.3E-02	9.9E-01	-1.70E-02	2.30E-02	-3.80E-02	9.79E-01	1.50E-02	-4.20E-02	7.90E-02	-6.00E-03
PO_4^{3-} (mg/l)	3.1E-01	-1.6E-02	-2.8E-02	7.6E-01	1.4E-01	7.6E-01	-6.4E-02	9.20E-02	2.95E-01	1.08E-01	2.20E-02	1.77E-01	5.10E-02	-9.00E-03	7.42E-01
E-coli(mg/l)	-2.4E-02	1.2E-02	9.8E-01	-3.6E-02	-5.8E-02	4.1E-02	-3.5E-02	9.47E-01	5.40E-02	-2.10E-02	-4.10E-02	-1.00E-02	8.87E-01	-4.50E-02	3.10E-02
Coliform(mg/l)	-7.0E-03	-1.6E-02	9.8E-01	-1.7E-02	-1.2E-02	5.2E-02	-2.4E-02	9.38E-01	-5.40E-02	-1.60E-02	-3.90E-02	5.00E-02	8.77E-01	2.20E-02	5.70E-02
Eigenvalue	3.339	2.663	2.009	1.404	1.241	3.260	3.001	1.832	1.125	1.036	3.474	2.364	1.626	1.171	1.002
Variability (%)	25.684	20.481	15.451	10.802	9.544	25.075	23.084	14.091	8.653	7.968	26.722	18.188	12.509	9.005	7.705
Cumulative (%)	25.684	46.165	61.616	72.418	81.962	25.075	48.159	62.250	70.903	78.871	26.722	44.910	57.419	66.424	74.129

High Pollution Source (HPS)

For HPS region, among five VFs, VF1 accounts for 25.6% of the total variance, which have strong positive loadings on BOD, COD and SS. In this region, loading of BOD and COD are assumed to be contributed by the direct discharges from nearby pig farm which are not equipped with proper sanitary treatment system (Lim & Kiu, 1995). It is reported that Juru River flow through largely urbanized areas where it had been polluted by domestic waste and discharges from pig farms (Lim & Kiu, 1995). The present of COD in the river basins was assumed to come from the anthropogenic activities that arise from the nearby industrial areas that discharge their industrial waste into these three rivers. The strong loading on SS is possibly originated from the high load of soil runoff and also from wood industry (Zali et al., 2011) nearby these three river basins. VF2, account 20.4% total variance with the positive loading of three variables which are DS, TS AND Cl. The two variables DS and TS can be assumed to be originated from point sources (PS) and non-point sources (NPS) (Ha & Bae, 2001)) as these river basins receive a lot of changes in the land development that also depends on the seasonal variation in studied area. Cl^- on the other hand, is identified to be originated from the mineral salt content in the river. VF3 explaining15.4% of the total variance has strong loading on *E. coli* and coliform which indicate the microorganism parameters in the HPS region of these three rivers. In Juru River for example, source of *E. coli* and coliform are possibly originated from Juru sewage pond located near the river and also from nearby residential areas as human settlements including squatters along the river banks at Juru River are not equipped with proper sanitary systems.VF4, explaining 10.8% of the total variance and has high loading of NH_3-N and PO_4^{3-}. The NH_3-N indicates that the HPS region of these three river basins experienced from pollution that caused by livestock waste

[DOI: 10.18331/SFS2021.7.2.2]

and as well as the agricultural and domestic sewage waste. For example, PO_4^{3-} loads were mainly originated from agricultural runoff such as fertilizers at Juru River flow nearby the Prai Industrial Estate. VF5, explaining 9.5% of the total variance and has strong loading on NO_3^-. The loading of NO_3^- is possibly due to the runoff from agricultural land along the HPS region of these three river basins. This NO_3^- is mainly originated from commonly used nitrogen and potassium fertilizers at the crop planted area of this HPS region.

Medium Pollution Source (MPS)

For MPS, among five VFs, VF1 accounts for 25% of the total variance which include BOD, COD, NH_3-N and PO_4^{3-}. BOD and COD are among organic factors that assumed to be attributed from anthropogenic activities such as farming and timber logging activities that take place along the river basin. The presence of NH_3-N in the river is due to the excessive runoff from the agricultural area nearby the basin regions (Sharip, Zaki, Shapai, Suratman, & Shaaban, 2014). The PO_4^{3-} loading is assumed to be originated from phosphate fertilizer that contain in soils from the agricultural farm area located nearby the MPS region of these three river basins. VF2, explaining 23% of the total variance, has strong loadings on DS, TS AND NO_3^-. Farming and construction were more frequent near these river basins area and had resulted into sediment deposited. Thus, the loading of DS and TS in this MPS region are possibly due to extreme river bank erosion that usually occur during the storm flow which eventually cause the bedload sediment enter the river region (Bolstad & Swank, 1997; Hart, 2006). This assumption is reasonable especially to the river water in Kuantan River Basin which is mainly polluted due to land development through agricultural, timber logging and forest clearing activities. Strong positive loading on NO_3^- is expected to originate from the cultivation area (Vega, Pardo, Barrado, & Debán, 1998) , where crops are planted and the use of inorganic fertilizers such as ammonium nitrate is rather frequent (Juahir et al., 2010) . NO_3^- may also arise from decomposition and degradation of organic matters containing nitrogen (USGS, 2007). The organic matters contained in the municipal waste include urea and protein from the wastewater discharges which enters this MPS region of the three river basins. VF3, explaining 14% of the total variance, has strong loading on *E. coli* and coliform which are related to domestic waste and treatment plant from paper manufacturing industry, rubber and palm oil refineries that located near the river (Qadir, Malik, & Husain, 2008).

Normally, faecal contamination from human occurred when structural and technical flaws in the sewerage system that causing the sewage to be flowed into the river which then leads to the

[DOI: 10.18331/SFS2021.7.2.2]

present of *E. coli* and coliform.VF4, accounts the total variance 8.6%, showing loading on SS that can be attributed from high loads of soil and waste disposal runoff. The last one is VF5 which accounts 7.9% of the total variance and include NO_3^- as the positive strong loading variable. NO_3^- is expected to arise from vegetables farm, oil palm and rubber plantation that are located along the MPS region of these three river basins. The nitrate content in river water is caused by agricultural activity that is commonly associated with the use of chemical fertilizer to facilitate the growth of trees. Thus, when surface runoff occurs during rainy season, waste chemical fertilizer will flow into these basins and caused increasing of NO_3^- content in the river.

Low Pollution Source (LPS)

For LPS region, among five VFs, VF1 represent 26.7% of the total variance, explaining strong loadings on DO, DS, TS and Cl^-. The strong negative loading on DO is cause by the presence of *E. coli* in this LPS region of the three river basins which consumed large amount of oxygen in order to undergo anaerobic fermentation. The negative loading of DO explained that the LPS region in these three river basins had been polluted by municipal waste, oxidation ponds and animal husbandry. DS and TS can be assumed as the sediment accumulation result that happened due to anthropogenic activities at these three river basins such as sand mining operation that is operated at Johor River area. The loading of Cl^- is probably comes from the mineral constituent in the water of this LPS region. VF2 represent the total variance of 18.1% and show the strong positive loading of BOD and COD. The presence of these BOD and COD in this LPS region of the three river basins is believed to be attributed from the influence of point source organic pollutants from sewerage network of the cities located nearby the river. VF3, explain 12.5% of the total variance and has strong loadings on *E. coli* and coliform that signify the contribution of domestic waste to this LPS region. VF4, explaining 9% of the total variance and has strong loading on pH and NO_3^-. The strong loading of pH is expected to arise from several causes such as industrial effluent discharges and other environmental factors. The decrease of pH range into acidic condition are mainly caused by the industrial effluent that release acidic discharges into the river while the significant increase in pH level into alkaline condition are possibly resulted from environmental factors such as the rapid algae growth which remove carbon dioxide from the water during the process of photosynthesis. The NO_3^- loading may additionally derived from agricultural area where inorganic nitrogen fertilizer are in common use such as at vegetable farm near the river. VF5, accounts for 7.7% of total variance, showing strong loadings on NH3NL and PO_4^{3-}. NH_3-N

[DOI: 10.18331/SFS2021.7.2.2]

indicates that the LPS region of the three river basins experienced from pollution that caused by livestock waste and as well as the agricultural and domestic sewage waste while a large amount of PO_4^{3-} loading is possibly originated from the contamination of fertilizer and pesticide discharges from vegetables farm located nearby the river basin.

Environmetric analysis techniques managed to determine spatial variation among the three studied river basins namely, Juru River Basin, Kuantan River Basin and Johor River Basin. Cluster analysis has successfully classified the cluster region namely, HPS, MPS and LPS in each of the three river basins. This classification enables the designation of sampling strategy which can reduce the number of sampling stations and the monitoring cost as on one station in every cluster is enough to represent the accurate rapid assessment of spatial water quality for the whole region among the group. Discriminant analysis on the other hand also gives encouraging results upon discriminating the data of every monitoring stations with discriminant variables assigning high correctly percentage of correlation matrix using forward and backward stepwise modes. Principal component analysis that applied on the data set for each classified region had managed to identify the pollutant loading variation due to land use and anthropogenic activities in the three studied river basins. Generally, this study had showed the ability of environmetric techniques for conducting the analysis and interpretation of a large complex data set for water quality assessment and as well as the identification of pollution sources. This analysis is also useful upon investigating spatial variations of water quality as an effort toward a more effective river basin management. Overall results obtained from this study indicate that anthropogenic activities have significantly influence river water quality variations. If these activities are not controlled, they will consequently generate a great pressure on the river ecosystem and finally become severely polluted river with the loss of critical habitat and overall decrease in the quality of life that inhabit this river ecosystem.

Acknowledgements

The authors would like to express whole gratitude and are grateful to thanks the Department of Environment (DOE) Malaysia for providing the secondary the data for this research project upon the process of completing this manuscript.

References

Adam, M.J., 1998. The Principles of Multivariate Data Analysis. In P. R. Ashurst & M. J. Dennis (Eds.), *Analytical Methods of Food Authentication.* (pp. 350). London, UK: Blackie Academic & Professional.

Alkarkhi, A.F., Ismail, N., Ahmed, A. and Easa, A., 2009. Analysis of heavy metal concentrations in sediments of selected estuaries of Malaysia--a statistical assessment. *Environ Monit Assess*, 153(1-4), 179-185. doi:10.1007/s10661-008-0347-x

Bolstad, P.V. and Swank, W.T., 1997. Cumulative Impacts of landuse on water quality in Southern Appalachian Waterhed 1. *JAWRA Journal of the American Water Resources Association*, *33*(3), 519-533. doi:10.1111/j.1752-1688.1997.tb03529.x

Department of Environment (**DOE**). **1977.** Report on Water Pollution Control: Juru River Basin, Department of Environment, Ministry of Science, Technology and Environment, Malaysia.

Department of Environment (DOE). 2003. Water quality management in Malaysia. *DOE Documents*, Kuala Lumpur.

Department of Environment (DOE). 2008. River Water Quality Report 2008.

Department of Enviroment (DOE)-USM. 1992. Development of Water Quality Criteria and Standards Classification-Juru River, Vol VIII. *Draft Final Report*, Departmental of Environment, Malaysia, Innovation and Consultancy Centre, Universiti Sains Malaysia.

Dogan, E., Sengorur, B., & Koklu, R. 2009. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, *90*(2),1229-1235. doi:10.1016/j.jenvman.2008.06.004

Einax, J. W., Zwanziger, H.W. and Geiss, S. 1997. Chemometrics in environmental analysis. Germany.

Ha, S. R., and Bae, M. S., 2001. Effects of land use and municipal wastewater treatment changes on stream water quality. *Environ Monit Assess*, 70(1-2), 135-151. doi:10.1023/a:1010649705723

Hamirdin, I., 2000. Pengaruh aktiviti manusia terhadap kualiti air permukaan di Lembangan Langat-Semenyih dan Linggi. In H. Mohd Yusof, I. Nor Azizan, & Z. M. Lukman (Eds.), Isu-isu pembangunan di awal abad ke-21. Bangi, Selangor (Malaysia): Fakulti Sains Pembangunan, Universiti Kebangsaan Malaysia.

Hamza, H., 2009. Water quality trend at the upper part of Johor river in relation to raindall and runoff pattern. (Masters), Universiti Teknologi Malaysia, Johor, Malaysia.

Hart, H.M., 2006. Effect of land use on the total suspended solids and turbidity in the little river watershed, Blount County, Tennesse. (Masters), University of Tennessee, Knoxville.

Hill, A.R., 1981. Stream phosphorus exports from watersheds with contrasting land uses in Southern Antario. *JAWRA Journal of the American Water Resources Association,* 17(4), 627-634. doi:10.1111/j.1752-1688.1981.tb01269.x

Ismail, M., 2009. *Application of hydrodynamic water quality model in Sungai johor Estuary.* (Masters), Universiti Teknologi Malaysia, Johor, Malaysia.

Johnson, R.A., & Wichern, D.W., 2008. Applied Multivariate Statistical Analysis. United States: Pearson.

J**uahir, H., Zain, S. M., Aris, A. Z., Yusoff, M. K., & Mokhtar, M. B. 2010.** Spatial assessment of Langat river water quality using

[DOI: 10.18331/SFS2021.7.2.2]

chemometrics. *Journal of Environmental Monitoring,* 12(1), 287-295. doi:10.1039/B907306J

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,* 582(2), 390-399. doi:10.1016/j.aca.2006.09.006

Kim, J. O., & Mueller, C. W. 1987. Introduction to factor analysis: What it is and How to do it (Quantitative application in sosial sciences). United State: SAGE UNIVERSITY PAPER.

Lim, P.E., & Kiu, M.Y. 1995. Determination and speciation of heavy metals in sediments of the Juru River, Penang, Malaysia. *Environmental Monitoring and Assessment,* 35(2), 85-95. doi:10.1007/BF00633708

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. *Sci Total Environ,* 313(1-3), 77-89. doi:10.1016/s0048-9697(02)00683-6

McGarigal, K., Stafford, S., & Cushman, S. 2000. *Multivariate Statistics for Wildlife and Ecology Research* New York, NY: Springer New York.

Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. 2013. Application of artificial neural networks for water quality prediction. *Neural Computing and Applications,* 22(1), 187-201. doi:10.1007/s00521-012-0940-3

Qadir, A., Malik, R. N., & Husain, S. Z. 2008. Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. *Environ Monit Assess, 140*(1-3), 43-59. doi:10.1007/s10661-007-9846-4

Razali, A., Syed Ismail, S. N., Awang, S., Praveena, S. M., & Zainal Abidin, E. 2018. Land use change in highland area and its impact on river water quality: a review of case studies in Malaysia. *Ecological Processes,* 7(1), 19. doi:10.1186/s13717-018-0126-8

Sharip, Z., Zaki, A. T. A., Shapai, M. A. H. M., Suratman, S., & Shaaban, A. J. 2014. Lakes of Malaysia: Water quality, eutrophication and management. *Lakes & Reservoirs: Science, Policy and Management for Sustainable Use,* 19(2), 130-141. doi:10.1111/lre.12059

Simeonov, V., Einax, J., Stanimirova, I., & Kraft, J. 2002. Environmetric modeling and interpretation of river water monitoring data. *Analytical and Bioanalytical Chemistry,* 374(5), 898-905. doi:10.1007/s00216-002-1559-5

Simeonova, P., Simeonov, V., & Andreev, G. 2003. Water quality study of the Struma river basin, Bulgaria (1989–1998). *Central European Journal of Chemistry,* 1(2), 121-136. doi:10.2478/BF02479264

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research,* 38(18), 3980-3992. doi:10.1016/j.watres.2004.06.011

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, 538*(1), 355-374. doi:10.1016/j.aca.2005.02.006

United States Geological Survey (USGS). 2007. Water quality in the upper Anacostia River, Maryland:

[DOI: 10.18331/SFS2021.7.2.2]

Continous and discreate monitoring with simulation to estimate concentrations and yields. Retrieved from Virginia, United States:

Varmuza, K., & Filzmoser, P. 2009. Introduction to Multivariate Statistical Analysis in Chemometrics.

Varunprasath, K., & A. Daniel, N. 2010. Physico-Chemical Parameters of River Bhavani in Three Stations, Tamilnadu, India. *Iranian (Iranica) Journal of Energy & Environment*, 1(4).

Vega, M., Pardo, R., Barrado, E., & Debán, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, *32*(12), 3581-3592. doi:10.1016/S0043-1354(98)00138-9

Zali, M. A., Retnam, A., & Juahir, H. 2011. Spatial Characterization of Water Quality Using Principal Component Analysis Approach at Juru River Basin, Malaysia.

[DOI: 10.18331/SFS2021.7.2.2]