# River water quality assessment using APCS-MLR and statistical process control in Johor River Basin, Malaysia

Mohd Saiful Samsudin [1], Azman Azid [1,2,*], Saiful Iskandar Khalit [1,2], Ahmad Shakir Mohd Saudi [3], Muhammad Amar Zaudi [4]

[1]Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia
[2]UniSZA Science and Medicine Foundation Centre, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia
[3]Institute of Medical Science and Technology, University of Kuala Lumpur, 43600 Kajang, Selangor, Malaysia
[4]Department of Environmental Science, Faculty of Environmental Studies, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

## ABSTRACT

The objectives of this study are to determine the most significant parameters of Johor River Basin which contribute to river pollution loading and to discover the potential contamination of pollutants and perform the process capability of water quality. The environmetric techniques and statistical process control have been utilize in this study. PCA extracted eight principal components which explaining 77% of total variance. The APCS-MLR model has revealed $NH_3$-N and $PO_4$ as the main parameter which are main pollutants that give highest contribution towards the river. The control charts have been established for $NH_3$-N and $PO_4$ by using SPC to monitor the level of concentration in a timely manner. Thus, continuous monitoring in the area should be done for better improvement of river quality in the Johor River Basin.

## 1. Introduction

River is the vital natural resource living things especially for the human being (as a substantial role in the economic, social, cultural, religious), the provision of water supply domestic human consumption, irrigation for agriculture, the accommodation of transport, industrial use and a livelihood to the people. The accomplishment in the economic growth and industrialization in Malaysia has convinced to environmental problems with ever increasing land, air and water pollution (Ho, 1996).

Anthropogenic influences as well as natural processes degrade surface waters and impair their use for drinking, industry, agriculture, recreation and other purposes (Carpenter et al., 1998). Concurrently, the process of industrialization in Malaysia has led to economic growth, but at the expense of the environment. Water pollution is a by-product of industrialization of toxic and hazardous waste, which was generated by various industries

and discharged into the aquatic environment (Abdullah, 1995). Changes in the species composition and the decline in the overall health of aquatic within the river basin have been occurred due to the deterioration of water quality (Durell et al., 2004).

Environmetric is an advanced multivariate analysis that is rooted for the assessment of the environmental database (Juahir et al., 2010a; Nasir et al., 2011; Zali et al., 2011; Dominick et al., 2012). Apart from that, environmetric is also exhibited as a division of environmental analytical chemistry that requires multivariate statistical modeling and data treatment known as chemometric analysis (Simeonov et al., 2002; Brodnjak-Vonˇcina et al., 2002; Simeonov et al., 2004; Felipe-Sotelo et al., 2007; Kowalkowski et al., 2006; Pere'-Trepat et al., 2006; Osman et al., 2012; Saim et al., 2009; Gazzaz et al., 2012; Retnam et al., 2013). This quantitative technique is suitable for all aspects of the social and natural environment, including forecasting, mathematical modeling, data analysis and statistics (Juahir et al., 2010b; Nasir et al., 2011). Principal Component Analysis (PCA) is one of the most utilized tools in the environmetric (Shrestha and Kazam, 2007; Krishna et al., 2009; Juahir et al., 2011; Nasir et al., 2011; Dominick et al., 2012). According to

Shrestha and Kazama (2007), data obtained were subjected to different multivariate statistical approaches: (i) to define geogenic and anthropogenic origin, (ii) to identify possible non-point sources of contamination and (iii) to estimate the contributions of possible sources on concentration of determining parameters (Krishna et al., 2009). In fact PCA can identify several pollution factors reasonably, but the interpretation of these factors in terms of actual controlling sources and processes are highly subjective (Liu et al., 2003).

Multiple Linear Regression (MLR) is a multivariate statistical technique that was often applied in a particular study in order to predict relationships between input and output variables without detailing the causes of these relationships (Dominick et al., 2012). On the contrary, MLR has also been carried out in order to measure the relationship between the independent and dependent variable (Guille´n-Casla et al., 2011; Dominick et al., 2012). Consequently, it is recommended to define the percent contribution of different sources for the application of MLR to the PCA scores (Wu et al., 2009). Apart from that, a study conducted by Wang et al. (2009) showed that an analysed data set by using PCA followed by MLR was prominent in order to provide understanding into the precision and quantification of source apportion. The combination of PCA and MLR is known as the Absolute Principle Component Score-Multiple Linear Regression (APCS-MLR) model. This technique was applied in order to determine the contribution of each possible sources defined by PCA (Zhou et al., 2007; Nasir et al., 2011; Su et al., 2011a; 2011b). Hence, it is proven that the utilization of APCS-MLR are capable in the identification of the possible source contributions in each physicochemical (Simeonova et al., 2003). Furthermore, APCS-MLR was applied to calculate the source of contributions after determining the number and characteristics of possible sources by PCA (Zhou et al., 2007).

Statistical Processing Control (SPC) has been used widely in the monitoring, manufacturing processes and service operations (Woodall et al., 2000). Madu (1996) also explained the feasibility of control charts in environmental monitoring. The justification of Maurer et al. (1999) on the empirical analysis of SPC in sediment pollutant analysis using control charts rule supported the statement in Madu (1996). The statement by Maurer et al. (1999) stated that although SPC was primarily developed for the industrial purposes, this methodology can also be used in environmental discipline. The limitation of using SPC in environmental monitoring is depending on the higher sampling frequency in an attempt to identify the developing trends, moreover the use of this tool also requires increased sampling frequency in order to reflect the system or scale that are being monitored (Maurer et al., 1999). Corbett and Pan (2002) encouraged the practitioners to link the theories of SPC in the environmental data to evaluate the environmental performances. Most studies in SPC have been carried out in a small number of areas

especially in environmental discipline. This is because the techniques in modern statistical process control charts and related methods are still remaining unknown to most of the environmental personnel (Corbett and Pan, 2002). In fact, the application of control charts and process capability provide a promising tool in the environmental realm, on the word of Corbett and Pan (2002). Therefore, SPC techniques are important to identify the risk of the environment in the future. Researchers studied on the SPC techniques in environmental data may find that the limitations in SPC are due to insufficient data for control charts to decide (Maurer et al., 1999). This study, however, attempts to provide a profound understanding of the capabilities of each tool in SPC. Despite that, this study also aims to provide an insight on the control charts and process capabilities in order to capture a better analysis and interpretations out of the outcomes particularly concerning the environmental realm. As stated by Besterfield (2009), the SPC has comprised with many tools such as Pareto Diagram, cause and effect diagram, check sheet, process flow diagram, scatter diagram, histograms and control charts. In addition, the control charts are implemented for process stability; where the process is stable if there is no out-of control point and the process is not stable when there is out of control points (Woodall et al., 2000). Theoretically, the phase 1 of the applications in the control charts usually considered any-out of the control points of the chart and in phase 2, the probability of a signal on any one sample is occasionally used if the successive statistics plotted are independent (Woodall et al., 2000). The error of false positive and false negatives is balanced by SPC meanwhile data from a base period is used to construct the control limits. Consistent with Madu (1996), any process involved has its variation and the two causes of variation is known as chance (natural) and assignable cause of variation. The control charts are aiming to identify and eliminate special cause so that only random causes of variation are retained in the system to ensure the stable process (Madu, 1996).

Malaysian rivers can be classified as Class IIB/III Rivers (Abdullah, 1995; Juahir et al., 2010a; Al-Mamun and Zainuddin, 2013). The secondary data of Johor River Basin were acquired from the Department of Environment Malaysia (DOE). The data were selected for evaluation for the year of 2003 to 2007. A total of 30 water quality parameters notably as the dissolve 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 ($PO_4$), Escherichia Coli (E.Coli), coliform and also various types of heavy metal parameters.. The objectives of this study are to determine the most significant parameters of each river basin which contribute to river pollution loading and to discover the potential contamination of pollutants and perform the process capability of water quality. All the data are analysed

using the following software; the software is packaged XLSTAT 2012 and SPC XL software.

## 2. Experimental study

### 2.1. Study area

Johor River is also known as one of the largest catchment area in the southern part of Peninsular Malaysia with a total catchment area of 2751.72 km² (Dorofki et al., 2012). The river flows towards the north of the basin from the Bukit Gemuruh (at an altitude of 109m) and Gunung Belumut (at an altitude of 1,010m) (Kia et al., 2012). The Johor River originates from its source of Layang-Layang River and Sayong River in the upstream area before merger into the Johor River and flows down towards the southeast of the Johor Straits estuarine. The major tributaries that are located at the downstream of the Johor River are the Tiram and Lebam River (DID, 2000). In fact, the main tributaries that diverged the Johor River are Sayong, Linggiu, Tiram and Lebam Rivers as shown in Fig. 1 and Table 1. The water temperature of the river is ranging from 21°C to 32°C. Based on Kia et al. (2012), Johor River is rich with its magnificent beauty of the natural forest and swamps that covers almost the major proportion of the landuse. Whereas, the southern part of the basin is mostly covered by the oil palm and rubber plantations. The main cities that are located adjacent to the Johor River basin are the Kota Tinggi with a total human population of 220000 people. The total catchment area of the Johor River at Kota Tinggi is approximately 1620 km² and the major land use in the Johor River basin are mainly conquered by oil palm plantations, other types of crop cultivations, urbanization, water body and swamps (Kia et al., 2012).

### 2.2. Pre-processing data

Preliminary work was undertaken in the data matrix that included assembly and data transformation. The data below the detection limit were substituted with values equal to half the detection limit. Normal distribution tests were carried out with the support of the W (Shapiro-Wilk) test; the agreement of the distribution of the physico-chemical parameters of water with normal distribution was tested (Sojka et al., 2008; Juahir et al., 2011; Samsudin et al., 2011). Standardization was applied to upturn the influence of variables whose variance is small and conversely. Log scaling is very common in environmental data since some of the variables might show very low or very high values.

### 2.3. Principal component analysis (PCA)

Principal Component Analysis (PCA) used on the normalized data set to observe in contrast of the compositional pattern among the analyzed water

quality parameters (variables) and to recognize the factors that influence each of the parameter (Dominick et al., 2012). The new variable which is knows as Principal Components (PCs) are the linear combinations of the original set of variables (Sousa et al., 2007; Dominick et al., 2012). The PCs can be expressed by Eq. 1 (Dominick et al., 2012):

$$y_{ij} = b_{1i}x_{1j} + bx_{2j} + \ldots + b_{mi}x_{mj}, \tag{1}$$

where y is a component score, b 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 covariance matrix was diagnosed and Eigenvalues are produced which is known as a characteristic root (Vega et al., 1998). This analysis is based on eigenvalue criteria by which a value >1 is deliberate significant, and a new group of variables was produced based on the similarity of the entire data set (Osman et al., 2012). Factor loading gives the correlation between the original variables and the VFs, while the individual transformed observations are called factor scores (Vega et al., 1998). The VF coefficients having a correlation 0.49–0.30 are considered 'weak' significant factor loadings, correlations in the range of 0.74–0.50 are considered 'moderate' and those in the range of >0.75 are considered 'strong' (Liu et al., 2003; Retnam et al., 2013).

### 2.4. Statistical process control

Primarily, the $\bar{x}$ and R chart is one type of a control chart that have a subgroup size of more than two groups that is coupled together in the SPC (Maurer et al., 1999; Douglas, 2009; Besterfield, 2009). Both charts are usually computed in order to determine either the process is stable and predictable. According to Douglas (2009), the control chart often display the average changes over time while the R-chart present the range of subgroups changes over time. The $\bar{x}$ and R charts are used throughout any processes having a subgroup size greater than one where the size may fall between two and ten.

Supposedly, a quality characteristic is normally distributed with mean (μ) and standard deviation (σ), where both the values of μ and σ are known. $\bar{x}$ is normally distributed with mean μ and standard deviation $\sigma\bar{x} = \frac{\sigma}{\sqrt{n}}$.

Besides, the probability is 1 – α that any sample mean may fall between Eq. 2 (Douglas, 2009):

$$\mu + Z_{\frac{\alpha}{2}}a_{\bar{x}} = \mu + Z_{\frac{\alpha}{2}}\frac{\alpha}{\sqrt{n}} \text{ and } \mu - Z_{\frac{\alpha}{2}}a_{\bar{x}} = \mu + Z_{\frac{\alpha}{2}}\frac{\alpha}{\sqrt{n}}. \tag{2}$$

Therefore, this equation could be used as an upper and lower limit. The $\bar{x}$ represent as the average of each sample, while μ is represented as the best estimator and the grand average $\bar{\bar{x}}$ is the process average in Eq. 3 (Douglas, 2009):

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_m}{m}, \tag{3}$$

where $\bar{x}$ an average of the sample and m is the number of subgroups. Meanwhile, $\bar{\bar{x}}$ is represented as the center line on the $\bar{x}$ chart.

Principally, the control chart is only accountable as completed when there is an upper control limit, central line and lower control limit which facilitate in the determination whether the process is stable or not (Besterfield, 2009). Generally, the formulas for constructing the control limits on the $\bar{x}$ chart are described in Eq. 4 (Douglas, 2009) as follows:

$$UCL = \bar{\bar{x}} + A2\bar{R}$$
$$Center\ Line = \bar{\bar{x}} \qquad\qquad (4)$$
$$LCL = \bar{\bar{x}} + A2\bar{R}$$

Nevertheless, the range method is often used in constructing the control limit chart where the process variability is monitored by plotting the values of the sample range R on a control chart.

The range of the sample is the difference between the largest and smallest observations in Eq. 5 (Douglas, 2009):

$$R = xmax - xmin \qquad\qquad (5)$$

where the R1, R2,...., Rm is the range of the m samples. The average range is computed using the following expression in Eq. 6 (Douglas, 2009):

$$\bar{R} = \frac{R_1 + R_2 + \cdots + R_{m.}}{m}, \qquad\qquad (6)$$

whereas, the centerline and control limits of the R chart are given as follows in Eq. 7 (Douglas, 2009):

$$UCL = D4\bar{R}$$
$$Center\ Line = \bar{\bar{x}}. \qquad\qquad (7)$$
$$LCL = D3\bar{R}$$

The constants D3 and D4 are tabulated for various sample sizes in the table of factors in the construction of variables control charts. In fact, the application of $\bar{x}$ and Rconsist of 3 phases; trial control limits, revise trial control limits and new trial control limits (generally are tighter than the first trial control limits). The trial central line is established for the $\bar{x}$ and R charts by using the following Eq. 8 (Douglas, 2009):

$$\bar{\bar{x}} = \frac{\sum_{u=1}^{s} \bar{x}k}{s} \quad \bar{R} = \frac{\sum_{u=1}^{s} \bar{R}k}{s}, \qquad\qquad (8)$$

where $\bar{\bar{x}}$ is the average of the subgroup averages, $\bar{x}k$ is the average of the kth subgroup is the number of subgroups. $\bar{R}$ is the average of the subgroup ranges while $\bar{R}k$ is the range of the kth subgroup.

The trial control limits for the charts are constructed on ±3σ (standard deviations) from the central value. This is presented based on the following Eq. 9 and Eq. 10 (Douglas, 2009):

$$UCL\bar{x} = \bar{\bar{x}} + 3\bar{\bar{x}}$$
$$LCL\bar{x} = \bar{\bar{x}} - 3\bar{\bar{x}}, \qquad\qquad (9)$$
$$UCL\bar{R} = \bar{R} + 3\bar{R}$$
$$LCL\bar{R} = \bar{R} - 3\bar{R}, \qquad\qquad (10)$$

where UCL = upper control limit, LCL = Lower control limit and σ$\bar{x}$ ispopulation standard deviation of the subgroup averages (Douglas, 2009; Besterfield, 2009).

According to Besterfield (2009), in SPC, the most crucial phase is the establishment of the revised central line and control limits. The standard values for the central lines are implemented in order to unleash the best estimate for the available data. $\bar{\bar{x}}$ and $\bar{R}$ are considered as a representative of the process and positioned as the standard values for $\bar{x}_0$ and $R_0$ once the analysis indicates as 'good controls' on the preliminary data (Douglas, 2009; Besterfield, 2009). No out-of control points on both sides of the central line are depicted and no unusual patterns of variation are categorized in better control process. At this stage, only the out-of-control points are analysed for the determination of the process stability (Douglas, 2009; Besterfield, 2009). In the process, if there is an assignable cause the out-of-control points can be discarded. The out-of -control state are most commonly due to the chance caused that may occur as a part of natural variation. Therefore, the data may still remain in the system.

As stated by Douglas (2009), when the data is discarded, the 3rd phase as the new $\bar{\bar{x}}$ and $\bar{R}$ is calculated by the simplified calculations that is displayed as follows Eq. 11 and Eq. 12;

$$\bar{\bar{x}}new = \frac{\sum \bar{x} - \bar{x}b}{s - sb}, \qquad\qquad (11)$$
$$\bar{R}new = \frac{\sum \bar{R} - Rb}{s - sb}, \qquad\qquad (12)$$

where $\bar{x}b$ represent as the discarded subgroup averages, sb is the number of discarded subgroups and Rb represents the discarded subgroup range. The initial 25 subgroups are not plotted with a revised control limit as it is used to report the results for future subgroups (Douglas, 2009; Besterfield, 2009).

## 2.5. Capability index

The process capability analysis is generally an approach that is used in order to assist decision makers in making decision either the process is proficient of complying to the existing environmental legislation or benchmark that have been set for a sufficiently large proportion of time (Corbett and Pan, 2002). The capability index is also known as a measure of the stable and predictable, which has been showcased in the control charts. On the other hand, it is also referred as a measure of the process capability which is termed as the capability ratio that are symbolized by the Process Capability (Cp) which is a necessary complement to a variables control chart. Process Capability Index (Cpk) is used in the capability index in order to measure the centre of the target or nominal value where a minimum value that is normally recommended for Cpk is 1.00 in the control chart (Douglas, 2009; Besterfield, 2009). However, when the Cp value is 1.33 or reach a greater amount, the operating personnel are

responsible to maintain the process centered, stable and predictable (Douglas, 2009). The process capability and tolerance are combined to form a capability index as defined in the following Eq. 13:

$$Cp = \frac{USL - LSL}{6\sigma_0}, \qquad (13)$$

where Cp is the capability index, USL – LSL represents the upper specification limit substitute by the lower specification limit or tolerance and $6\sigma_0$ refers to the process capability. Nevertheless, when the capability index is 1.00, it is considered most commonly categorized as Case (II) situation and if the ratio is greater than 1.00, it will be referred as Case (I) situation which is desirable and if the ratio is less than 1.00, it will be considered as Case (III)

situation which is known as undesirable (Douglas, 2009; Besterfield, 2009).

Consistent with Besterfield (2009), this case situation will differentiate the output of the control processes to the specification limits. In spite of that, the process spreads pertains as the process capability and equal to 6σ meanwhile tolerance is the difference between specifications. Unfortunately, undesirable result may be obtained accordingly whenever the tolerance is established without regard to the process. There are three possible situations that may occur and this explains the situation case (I) when the process capability is less than the tolerance, (II) when the process capability is equal to the tolerance and (III) when the process capability is greater than the tolerance (Douglas, 2009).
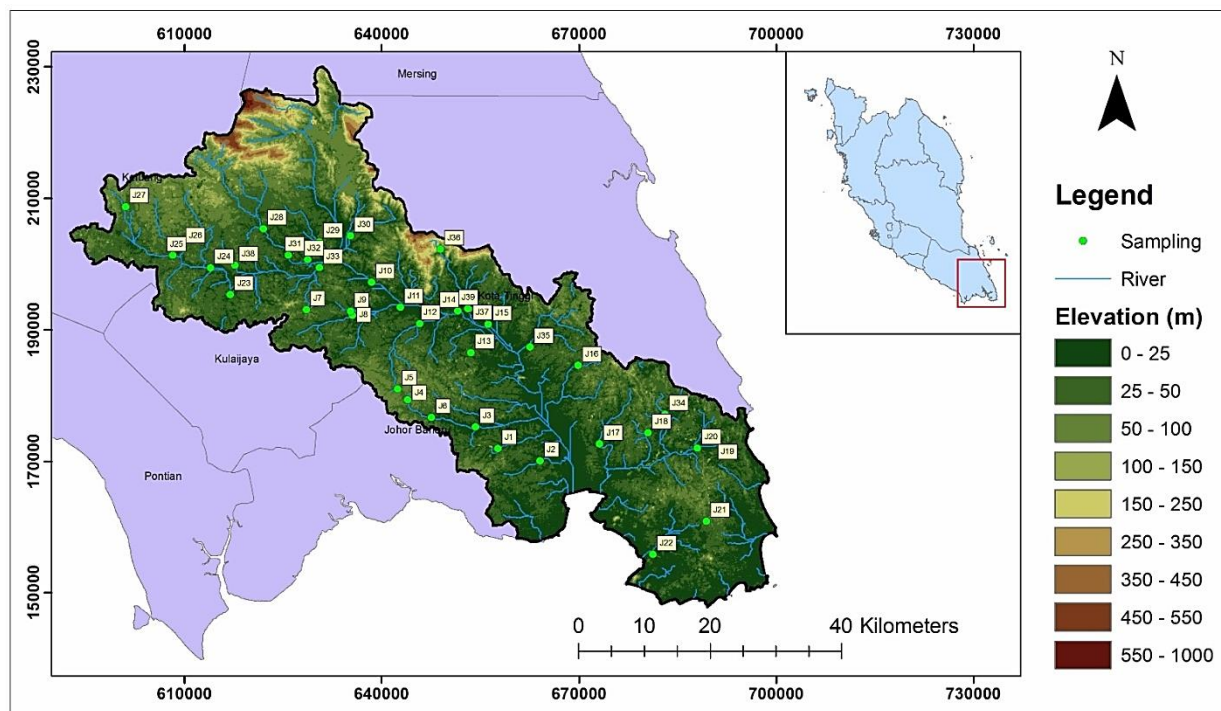


**Fig. 1:** Map of Johor River Basin

## 3. Result and discussion

### 3.1. PCA

PCA of the entire data set in the Johor River Basin (Table 2) elucidated eight PCs with eigenvalues greater than one explaining 77% of the total variance in the water-quality data set. The first factor (VF1) indicates 30.3% of the total variance with strong positive loadings of COND., SAL, DS, TS, Cl, As, K, Mg and Na. These variables are derived from the mineral component that is available in the river. This finding is consistent with Vega et al. (1998) which stated that this group of variables is the common origins of minerals that are more likely induce from the dissolution of limestone and gypsum. Moreover, the seasonal factor such as rain may also affect to the soil infiltration process and transported the pollutants into the river through the surface runoff. On the contrary, VF2 stipulated

11.9% of the total variance with the strong positive loading of BOD, COD, SS and turbidity. This VF represents as the anthropogenic activity that is incorporated from the industrial, domestic and commercial areas. The association among these variable (BOD, COD, suspended solid and turbidity) are corresponded due to the discharge of organic source and sewer pipes effluents which contain pollutants with high concentration of bacteria (Mohd et al., 2011). VF3 accounted for 9.91% of the total variance with strong positive loadings of $NH_3$-N and $PO_4$. This variable is suggested to come from agricultural runoff pollution sources. The presence of $NH_3$-N and $PO_4$ in the water system might be resulted by the runoff and soluble fertilizer that is used in the oil palm plantation industry. This is consistent with Singh et al. (2005) which stated that the usage of nitrogenous fertilizer increases the level of $NH_3$-N and $PO_4$. In fact, nitrogenous fertilizer is widely used in crop cultivations in order to boost the

plants growth by providing enough nutrients for the

plants to perform photosynthesis.

**Table 1:** The information of the monitoring stations on Johor River Basin

| LONGITUDE | LATITUDE | STA No. | RIVER | LOCATION |
|---|---|---|---|---|
| E 103° 54.801' | N 01° 33.294' | J1 | LAYANG | Wastewater Treatment Plant SAJ Sultan Iskandar, Masai |
| E 103° 58.257' | N 01° 32.273' | J2 | SERAI | Keck Seng, Kong-Kong Estate |
| E 103° 52.982' | N 01° 35.071' | J3 | TIRAM | Kg.Sg. Tiram,Ulu Tiram Bridge |
| E 103° 47.418' | N 01° 37.312' | J4 | TIRAM | Bridge near Kilang Getah Skellerup, UluTiram. |
| E 103° 46.599' | N 01° 38.200' | J5 | TIRAM | Bridge in Kulim, Ulu Tiram Estate |
| E 103° 49.358' | N 01° 35.858' | J6 | TIRAM | Ulu Tiram Town's Bridge |
| E 103° 39.104' | N 01° 44.746' | J7 | BKT. BESAR | FELDA Estate at Taib Andak, Kulai |
| E 103° 42.873' | N 01° 44.288' | J8 | SEMANGAR | Sg. Semanggar jalan Sg.Sayong-Bdr. Tenggara Bridge |
| E 103° 42.725' | N 01° 44.624' | J9 | BKT. BESAR | Sg.Bkt.Besar jalan Sg.Sayong-Bdr, Tenggara Bridge |
| E 103° 44.458' | N 01° 47.010' | J10 | JOHOR | Kg. Rantau Panjang near FELDA Bkt. Ramun Jetty |
| E 103° 46.807' | N 01° 44.944' | J11 | JOHOR | Sg.Johor Kg. Semanggar  Bridge |
| E 103° 48.403' | N 01° 43.602' | J12 | TELOR | Kg. Sg. Telor jln Kulai-Kota Tinggi Bridge |
| E 103° 52.587' | N 01° 41.199' | J13 | BERANGAN | Tai Tak, K/Tinggi estate near effluent pool |
| E 103° 51.509' | N 01° 44.636' | J14 | JOHOR | Kg. Sg. Telor Fishing Jetty |
| E 103° 54.022' | N 01° 43.555' | J15 | JOHOR | Kota Tinggi Town Bridge |
| E 104° 01.392' | N 01° 40.155' | J16 | TEMOH | Felda Air Tawar 2, K/Tinggi Bridge |
| E 104° 03.145' | N 01° 33.666' | J17 | LAYAU KIRI | Road to Kg.Baru, Sg.Layau-Ldg. KimLoong K/Tinggi. |
| E 104° 07.106' | N 01° 34.589' | J18 | SEMENCHU | FELDA Estate at Semenchu, K/Tinggi. |
| E 104° 12.508' | N 01° 33.918' | J19 | CHEMANGAR | Sg. Chemangar near Bdr.Penawar Bridge |
| E 104° 11.174' | N 01° 33.333' | J20 | LEBAM | Behind KKS FELDA Adela, K/Tinggi. |
| E 104° 11.896' | N 01° 27.284' | J21 | SENING | FELDA Estate at Sening, K/Tinggi. |
| E 104° 07.520' | N 01° 24.544' | J22 | SANTI | Marine Police, Pengerang Jetty |
| E 103° 32.862' | N 01° 45.993' | J23 | ANAK SG. SAYONG | Kg. Murni-Layang-Layang Road's Bridge |
| E 103° 31.228' | N 01° 48.196' | J24 | SAYONG | Sg.Sayong bridge road to lading MCA, Layang-Layang |
| E 103° 28.143' | N 01° 49.227' | J25 | SAYONG | Sg.Sayong bridge near Layang-Layang town |
| E 103° 28.891' | N 01° 49.942' | J26 | REMIS | Bridge at golf field near Ulu Remis' Estate, Layang-Layang |
| E 103° 24.264' | N 01° 53.233' | J27 | SAYONG | near to railway at Renggam Town |
| E 103° 35.562' | N 01° 51.424' | J28 | PENGGELI | Bridge near to SAJ Bandar Tenggara WTP |
| E 103° 40.170' | N 01° 50.350' | J29 | SEBOL | Sg. Sebol jalan Bdr. Tenggara-K/Tinggi Bridge |
| E 103° 42.718' | N 01° 50.855' | J30 | LINGGIU | Sg. Linggiu jalan Bdr. Tenggara-K/Tinggi Bridge |
| E 103° 37.624' | N 01° 49.243' | J31 | PENGGELI | JPS telemetric station to FELDA Inas |
| E 103° 39.208' | N 01° 48.884' | J32 | ANAK SG. SAYONG | FELDA Estate at Penggeli Timur, Bdr.Tenggara. |
| E 103° 40.171' | N 01° 48.212' | J33 | SAYONG | Bridge near to SAJ Sg. Sayong WTP |
| E 104° 08.513' | N 01° 36.123' | J34 | PAPAN | Sg. Papan bridge near to FELDA Sg.Mas, K/Tinggi |
| E 103° 57.437' | N 01° 41.667' | J35 | SELUYUT | Sg. Seluyut, Kg.Sinaran jetty near to Felda Air Tawar 3, K/Tinggi. |
| E 103° 50.086' | N 01° 49.746' | J36 | PELEPAH | Sg. Pelepah near to  Kota Tinggi Water Fall |
| E 103° 52.409' | N 01° 44.818' | J37 | PANTI | Sg. Panti, Kg.Panti, KotaTinggi bridge |
| E 103° 33.230' | N 01° 48.413' | J38 | BELITONG | Bridge at GSA Sg. Belitong Enam, Layang-Layang |
| E 103° 52.338' | N 01° 44.833' | J39 | PELEPAH | Sg. Pelepah, Kg. Panti, Kota Tinggi Bridge |

Unfortunately, if the plants are over dose with nitrogenous fertilizer, it may stimulate to the transportation of alkaline and nitrates that are contained in the fertilizer into the overlying water. Hence, the people that consume the water may prone to sickness such as hypoxia. The fourth factor (VF4) signifies for 6.52% of the total variance with the strong positive loading of E-coli and coliform. This factor represents as the association of variables in the fecal waste. This is because the presence of E-coli and Coliform often incorporated as a strong indication of sewage discharge into the surface water. This finding is also confirmed by Mohd et al. (2011) which suggested that the E-Coli and Coliform is suspected to be originated from the animal faeces, surface runoffs and discharge from the sewage and wastewater treatment plants. Whereas the fifth factor (VF5) explains 4.24% of the total variance with the strong positive loading of Zn. Initially, Zn is an element that is commonly utilized in the manufacturing industry that includes transportation, construction, machinery and electricity. Moreover, Zn is also used as an activator in the rubber and paint industries. In fact, the white pigment subsists in the water colours or paints are resulted by the touch of zinc oxide in the colour mixtures. In spite of that, VF6 accounts for 3.67% of the total variance

signifying Cd as the strong positive loading. In general, the generation of Cd in the environment is largely influenced by the manufacturing industries (OSHA). It mainly produces Cd for pigments, coating and plating and as a stabilizer for plastic. In fact, Cd also consists in the manufacturing industries as an inevitable byproduct of Zn, Pb and Cu extraction (Lenntech, 2014). VF7 exhibit 3.7% of the total variance with strong positive loading of Fe. Fe is known as one of the most abundant elements in the earth's crust. Nonetheless, Fe is also found as a source metal in the steel and alloy production (Lenntech, 2014). In general, Fe is usually corresponded to the industrial effluents due to its capability to exist in four distinct crystalline forms as described in the periodic table (Juahir et al., 2010b; Nasir et al., 2011; Lenntech, 2014). Hence, these three factors (VF5, VF6 and VF7) indicate as a coherent group of metals association which suggests that these factors are a symbol of anthropogenic activities. The eighth factor (VF8) signifies $NO_3$ as the strong positive loading which exhibit 4.6% of the total variance. Primarily, the sources of nitrogen in water were contributed by various types of pollution sources which this include animal wastes (livestock, birds, mammals and fish), feedlot discharges,

municipal and industrial wastewater, fertilized field,

lawn runoff, septic tanks and vehicle exhausts.

**Table 2:** Loadings of eight Varimax Factors (VFs) of Johor River Basin

| | VF1 | VF2 | VF3 | VF4 | VF5 | VF6 | VF7 | VF8 |
|---|---|---|---|---|---|---|---|---|
| DO | -0.284 | -0.137 | -0.495 | -0.301 | -0.031 | -0.091 | -0.2 | -0.295 |
| BOD | -0.006 | 0.752 | 0.511 | -0.005 | 0.065 | 0.009 | -0.035 | -0.055 |
| COD | 0.01 | 0.92 | 0.257 | -0.018 | -0.028 | 0.019 | -0.094 | 0.002 |
| SS | -0.005 | 0.972 | -0.005 | 0.058 | 0.027 | -0.011 | 0.046 | 0.009 |
| pH | 0.217 | 0.043 | 0.544 | 0.085 | -0.464 | -0.038 | 0.227 | 0.086 |
| $NH_3$-NL | -0.038 | 0.187 | 0.86 | 0.121 | 0.101 | -0.02 | -0.045 | -0.004 |
| TEMP | 0.408 | -0.037 | 0.132 | -0.043 | -0.333 | 0.381 | 0.1 | -0.233 |
| COND | 0.994 | -0.016 | 0.025 | 0.001 | 0.006 | 0.006 | -0.005 | -0.015 |
| SAL | 0.994 | -0.015 | 0.015 | -0.003 | -0.001 | 0.004 | -0.01 | -0.012 |
| TUR | -0.043 | 0.948 | -0.031 | 0.083 | 0.054 | -0.012 | 0.141 | 0.019 |
| DS | 0.994 | -0.015 | 0.023 | -0.009 | 0.006 | 0.006 | -0.006 | -0.013 |
| TS | 0.975 | 0.023 | 0.023 | -0.005 | 0.013 | 0.004 | -0.003 | -0.014 |
| $NO_3$ | -0.077 | 0.006 | 0.096 | -0.069 | -0.023 | -0.083 | 0.027 | <u>0.716</u> |
| Cl | 0.99 | -0.017 | 0.004 | -0.011 | 0.001 | 0.007 | -0.006 | -0.02 |
| $PO_4$ | -0.03 | 0.302 | <u>0.743</u> | -0.041 | 0.122 | -0.029 | -0.033 | 0.055 |
| As | 0.808 | -0.007 | 0.002 | 0.005 | -0.086 | -0.026 | -0.061 | 0.064 |
| Cd | -0.031 | 0.004 | -0.068 | -0.011 | 0.141 | 0.771 | 0.007 | -0.011 |
| Cr | 0.136 | -0.03 | 0.518 | 0.212 | -0.137 | -0.001 | -0.021 | 0.566 |
| Pb | -0.031 | -0.009 | 0.023 | -0.029 | -0.085 | 0.545 | -0.071 | -0.001 |
| Zn | 0.1 | 0.138 | 0.187 | -0.07 | 0.765 | 0.109 | 0.042 | -0.039 |
| Ca | 0.976 | -0.008 | 0.058 | 0 | -0.004 | 0.001 | -0.002 | 0.021 |
| Fe | -0.166 | 0.192 | 0.021 | 0.165 | 0.192 | -0.038 | <u>0.627</u> | 0.186 |
| K | <u>0.713</u> | 0.073 | 0.604 | 0.047 | 0.026 | -0.008 | 0.015 | 0.129 |
| Mg | 0.978 | -0.011 | 0.039 | -0.009 | 0.005 | 0.003 | -0.011 | 0.002 |
| Na | 0.99 | -0.02 | -0.01 | -0.015 | -0.004 | 0.006 | -0.006 | -0.013 |
| OG | -0.012 | 0.003 | 0.027 | 0.298 | 0.082 | 0.121 | 0.149 | 0.456 |
| MBAS | 0.061 | 0.02 | 0.283 | -0.002 | 0.376 | -0.106 | 0.419 | -0.343 |
| E-coli | -0.022 | 0.055 | 0.07 | 0.902 | -0.005 | -0.018 | 0.013 | -0.015 |
| Coliform | -0.028 | 0.064 | 0.077 | 0.916 | -0.048 | -0.005 | 0.026 | 0.057 |
| Eigenvalues | 9.474 | 4.58 | 2.391 | 1.705 | 1.259 | 1.124 | 1.035 | 1.015 |
| Variability (%) | 31.061 | 11.59 | 9.914 | 6.516 | 4.235 | 3.667 | 3.686 | 4.601 |
| Cumulative % | 31.061 | 42.651 | 52.565 | 59.081 | 63.316 | 66.984 | 70.67 | 75.27 |

Note: Values in bold indicate the variables has strong loading >0.75 and value underline indicate the moderate loading

In the environment, $NO_3$ often provides nutrients for the plants in streams, rivers, and reservoirs. In fact, the nitrate levels in the river water often fluctuate due to the seasonal factor and higher nitrate concentration are usually observed during the rainfall season (Ismail, 2011). Therefore, this clearly indicates that the generation of $NO_3$ is resulted by the surface runoff that transported land surface particulates into the overlying water. The summarization of all the possible pollution sources in the Johor River Basin is described in Table 3.

**Table 3:** Summary of possible pollution sources for Johor River Basin

| Parameter | Possible Pollution Sources |
|---|---|
| VF1 : Cond, Sal, DS, TS, Cl, As, K, Mg, Na | Mineral component |
| VF2 : BOD, COD, SS, TUR | Industrial and man activity |
| VF3 : $NH_3$-N, $PO_4$ | Agriculture |
| VF4 : E-coli, Coliform | Sewage |
| VF5 : Zn | Paint/Rubber Industry |
| VF6 : Cd | Rubber/Plastic Industry |
| VF7 : Fe | Abundant element in earth crust |
| VF8 : $NO_3$ | Livestock manure |

### 3.2. APCS-MLR

The results revealed that the APCS-MLR model for Johor River exhibited $R^2$ equivalent to 0.780 indicating a good fit between the measured and predicted concentration as this is described in Table 4. The coefficient of determination ($R^2$) value clearly indicate that there is a strong correlation between the goodness of the receptor modeling approach (APCS-MLR) and the source apportionment of the water variables. Moreover, the APCS-MLR has managed to distinguish VF3 as the highest percentage of contributory with 71.68% of the possible pollution (Table 5). This clearly indicates that $NH_3$-N and $PO_4$ that are influenced by the agricultural runoff are depicted as the major pollution sources in the Johor River Basin. Fundamentally, there are numerous agricultural and livestock activities that have been carried out in the development areas adjacent to the Johor River Basin. The point source and non-point sources of the pollutants include water runoff from the cropland, lawns, gardening and confined livestock that may be the reason of the high contribution of $NH_3$-N and $PO_4$ in the river.

In fact, the Federal Land Development authorities are also situated close to the Johor River. This is to ensure that the land development authorities are engaged directly with the concern related to the river.

**Table 4:** Goodness of fit statistic for regression of WQI

| Statistics | Value |
|---|---|
| $R^2$ | 0.780 |
| Adjusted $R^2$ | 0.777 |
| MSE | 11.958 |
| RMSE | 3.458 |
| AIC | 1319.083 |
| SBC | 1357.505 |

**Table 5:** Percentage of contributions

| Variable | $R^2$ | Diff $R^2$ | MSE | RMSE | % contribution |
|---|---|---|---|---|---|
| All Source | 0.78 | | 11.958 | 3.458 | |
| L-VF1 | 0.779 | 0.001 | 11.974 | 3.46 | 0.1 |
| L-VF2 | 0.763 | 0.017 | 12.864 | 3.587 | 2.7 |
| L-VF3 | 0.328 | 0.452 | 36.464 | 6.039 | 71.7 |
| L-VF4 | 0.761 | 0.019 | 12.979 | 3.603 | 3.1 |
| L-VF5 | 0.752 | 0.027 | 13.425 | 3.664 | 4.4 |
| L-VF6 | 0.708 | 0.071 | 15.812 | 3.976 | 11.3 |
| L-VF7 | 0.752 | 0.028 | 13.465 | 3.669 | 4.5 |
| L-VF8 | 0.766 | 0.014 | 12.714 | 3.566 | 2.3 |
| Total for Diff $R^2$ | | 0.631 | | | 100 |

### 3.3. $\bar{x}$ and R chart for $NH_3$-N and $PO_4$

The trial control limit for the $\bar{x}$ and R chart of the Johor River Basin are shown in Fig. 2. The outcome illustrates that the mean concentration of $NH_3$-N exceeded the UCL. There are four points in the $NH_3$-N mean observations that exceed the control limits of the $\bar{x}$ chart. The four points are points 1, 2, 7 and 8 representing 16% out of all the observations. Meanwhile, the R chart denotes the 1, 2, 8, 10 and 11 as the points that positioned outside the upper control limit indicating 20% of the $NH_3$-N mean concentration. This concludes that the process is not stable for the base period.
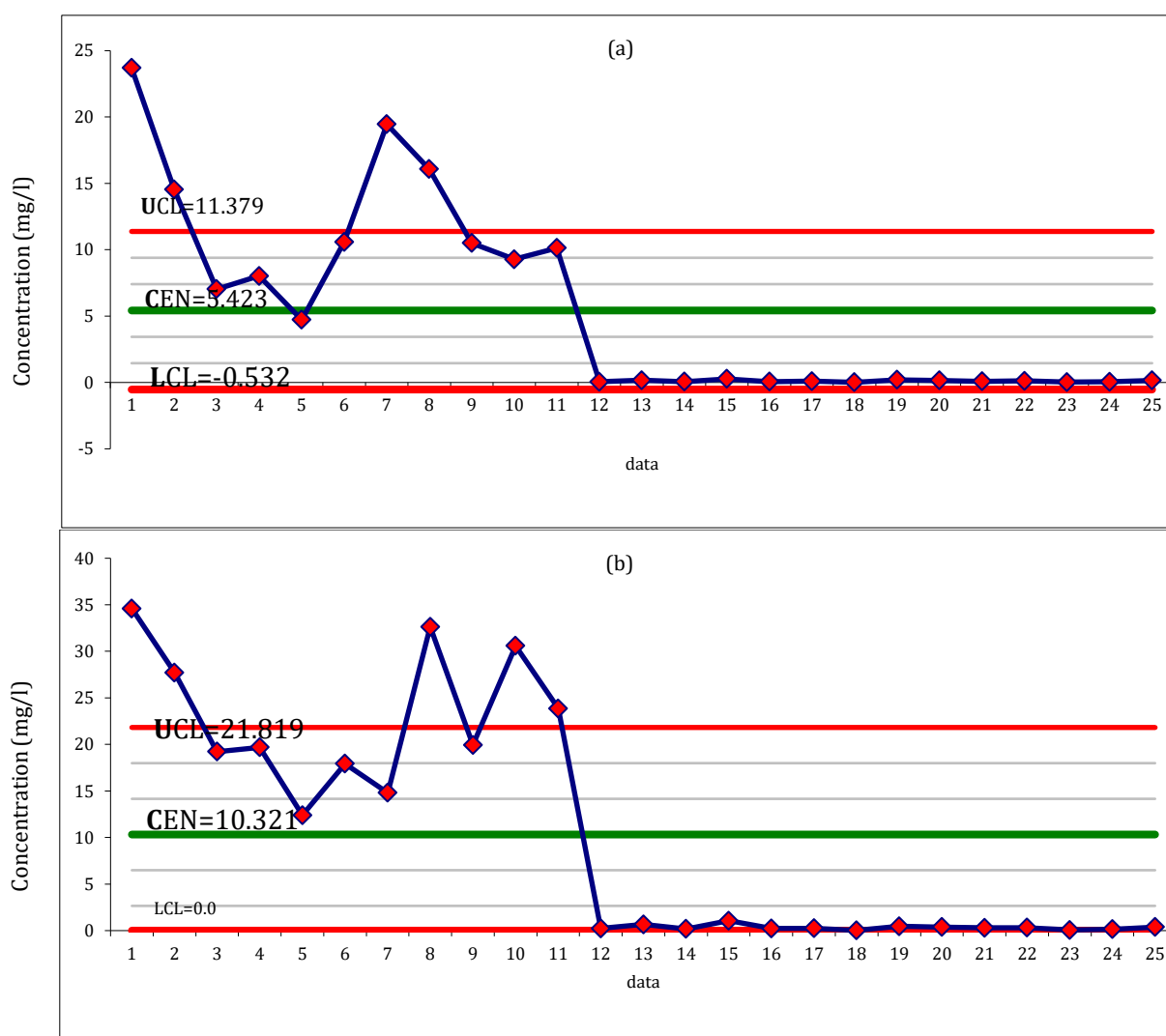


**Fig. 2**: Trial control limit of (a) $\bar{x}$ chart and (b) R chart for the concentration of $NH_3$-N (mg/L)

Based on Fig. 2, the revised control chart is constructed to determine the stability of the process. The entire out of control points is discarded in each chart in order to ensure that the process is stable. This is supported by Corbett and Pan, (2002), which stated that the out of control point is discarded in order to improve the normality so that the data will follow the normal distribution. Hence, the $\bar{x}$ chart shows UCL (0.23055), LCL (-0.05025) and CEN (0.09015) in Fig. 3. Whereas, the R chart depicted the value of UCL is 0.51441, LCL is 0.0 and CEN is 0.24333. The remaining plotted points in Fig. 3 indicate as a stable process whereby it shows that all of the points (observation) are within the control

limits. Thus, it can be applied as the example of the whole operation to create the future prediction and measure the risk of contamination. Subsequently, there are two mean observations that are utilized from the other NH$_3$-N data subgroup in order to determine either the process is stable or not. The results depicted that when the two observations are added in the process, all the points are positioned within the control limits (Fig. 4). Although there is a variation within the control limits, it is still considered as a natural variation of the process. Hence, this signifies that the mean concentration of NH$_3$-N in Johor River is in a control process.
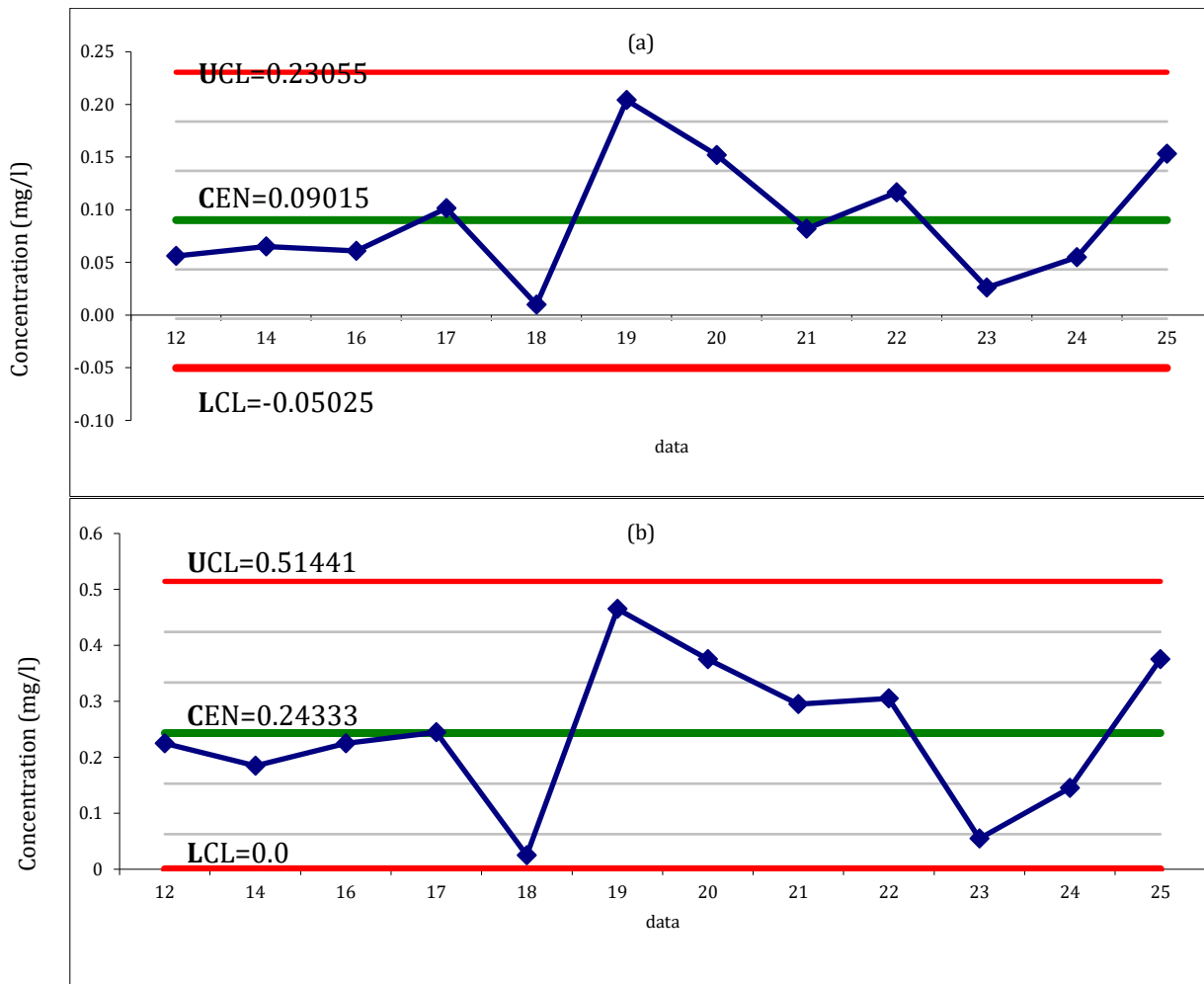


**Fig. 3**: Revised Control Limit (a) $\bar{x}$ chart and (b) R chart for the NH$_3$-N concentration
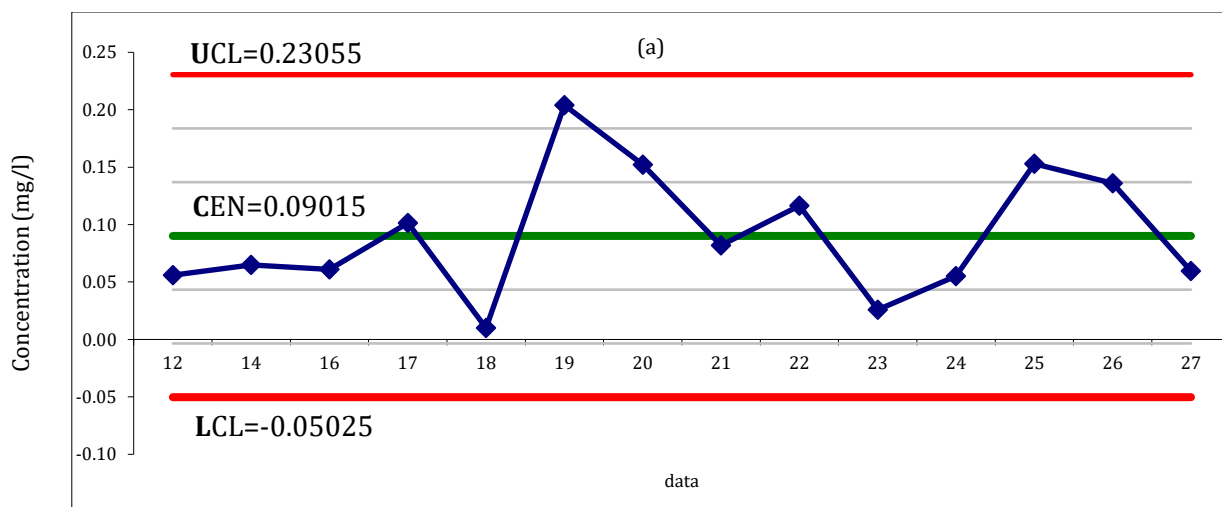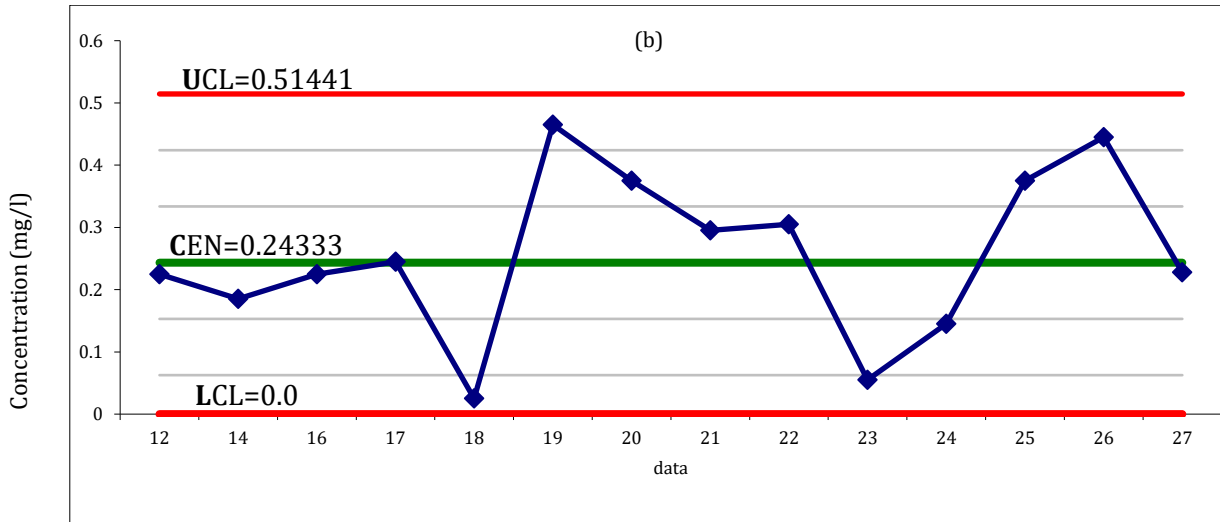
**Fig. 4**: Monitoring of NH$_3$-N concentration using the revised control limit of (a) $\bar{x}$ chart and (b) R chart

The control charts were then applied to PO$_4$ in order to verify whether the process is stable or not in the trial control limit. In the x̄chart (Fig. 5a) point 1, 2, and 3 (mean observations) exceeds the control limit suggesting to a possible assignable cause. This finding indicates that 12% of the mean concentrations positioned outside the upper control limit. Meanwhile, the R chart (Fig. 5b) denotes the same result where the point 1, 2, and 3 exceeded

UCL is indicating 12% of the PO$_4$ mean concentration. The result indicates that the process is not stable for the base period. The out of control point is considered as an assignable cause as it situated outside UCL area. The assignable cause is occurring when there is an undesirable variation which is the cause of the unexpected increase in the PO$_4$ concentration. Hence, this may be the cause of the non-point source pollution.
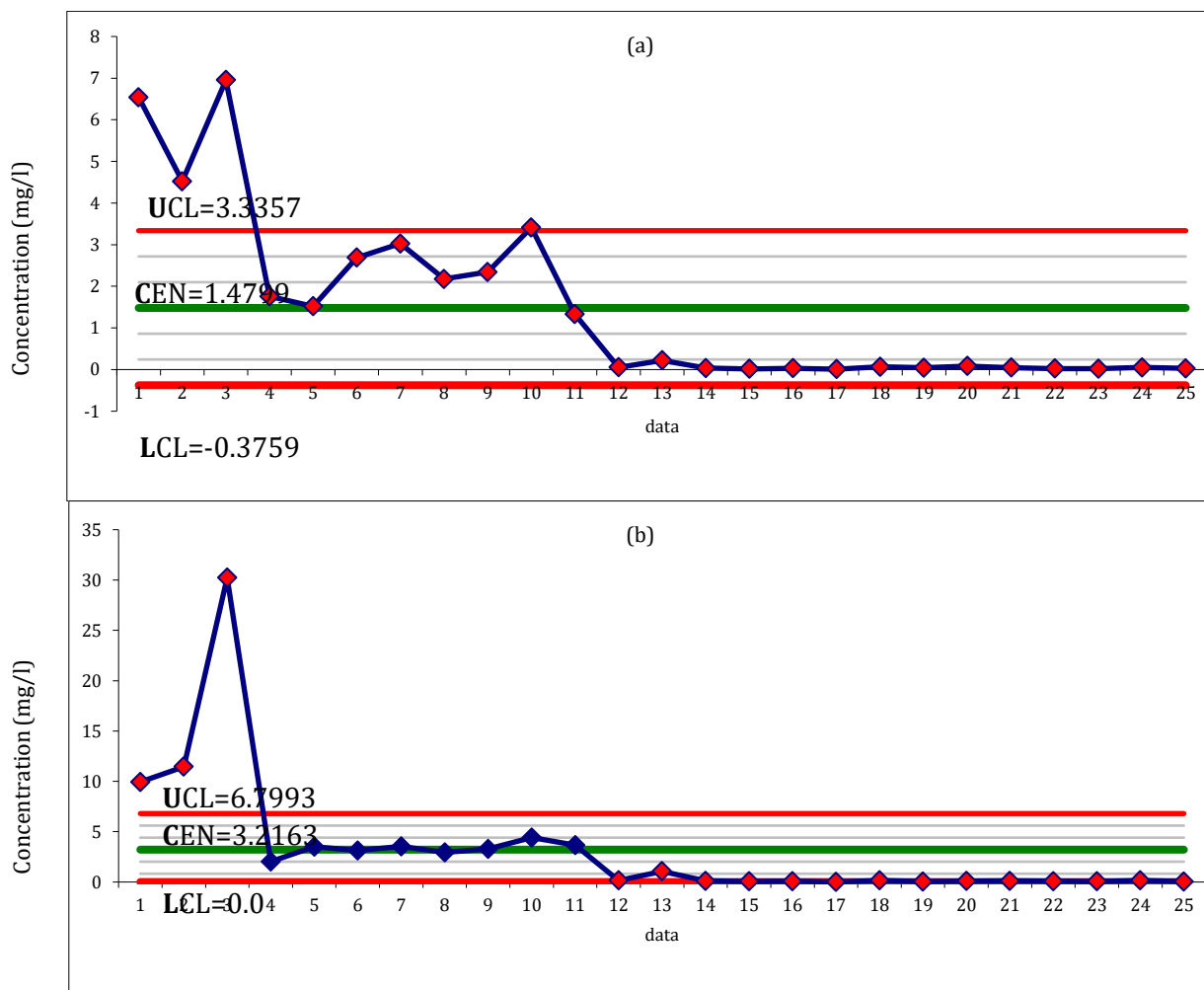


**Fig. 5:** Trial Control Limit of (a) $\bar{x}$ chart and (b) R-chart for the PO$_4$ concentration

Afterward, the out of control points are removed to ensure that the process is stable as shown in Fig. 5. The new control limits charts were computed and the out of control point is discarded. The UCL and LCL are 0.09286 and -0.01437 respectively for the x̄chart whereas the UCL and LCL are 0.19644 and 0.0 respectively for the R chart. The remaining plotted points in Fig. 6 indicate that the process is stable; in fact, it shows that all of the points (observation) are within the control limits. Therefore, it can be used as the representative of the whole process to make the future prediction and measure the risk of pollution. Primarily, the two mean observations from the other $PO_4$ concentration is added to perform the monitoring period using the control limits that have been constructed in Fig. 5. Within this period (Fig. 7), the $PO_4$ mean concentration data values of the other subgroup illustrate that there is no out of control points exceeding the control limits. Thus, this signifies the positive result verifying that the $PO_4$ concentration is in a control process at the Johor River Basin. Data concentration of $NH_3$-N from previous assessment in Fig. 7 found to be within the UCL and LCL and only natural variation occurs. This process is considered to be in statistical control or stable process.

Therefore, the process performance can be predicted by the process capability analysis. The inherent variability of the process is compared with the specification limits in the process capability analysis so that the environmental performance potential can be detected under normal or in control condition (Carbett and Pan., 2002).

### 3.4. Capability index for $NH_3$-N and $PO_4$

Based on Fig. 8, the capability index has been computed to measure the risk to the environment. The capability index Cp is used to measure the potential risk of $NH_3$-N towards the water pollution. The Cp value is found less than 1 which is 0.2622. This shows that the potential risk of $NH_3$-N concentration for unacceptable water pollution is higher. Therefore, the result of the analysis in process capability shows that the process is not suitable in the subsequent large period of time. This is therefore suggested that continuous monitoring should be done by DOE from time to time to ensure that the level of $NH_3$-N concentration in Johor river basin complies with the specification limit that has been set up which is UCL (0.3mg/L) and LCL is (0.1mg/L). The capability index Cp for $PO_4$ indicates less than 1.00 which is 0.2637 in Fig. 9. This will indicate that the potential risk of $PO_4$ concentration for unacceptable water pollution is also higher.
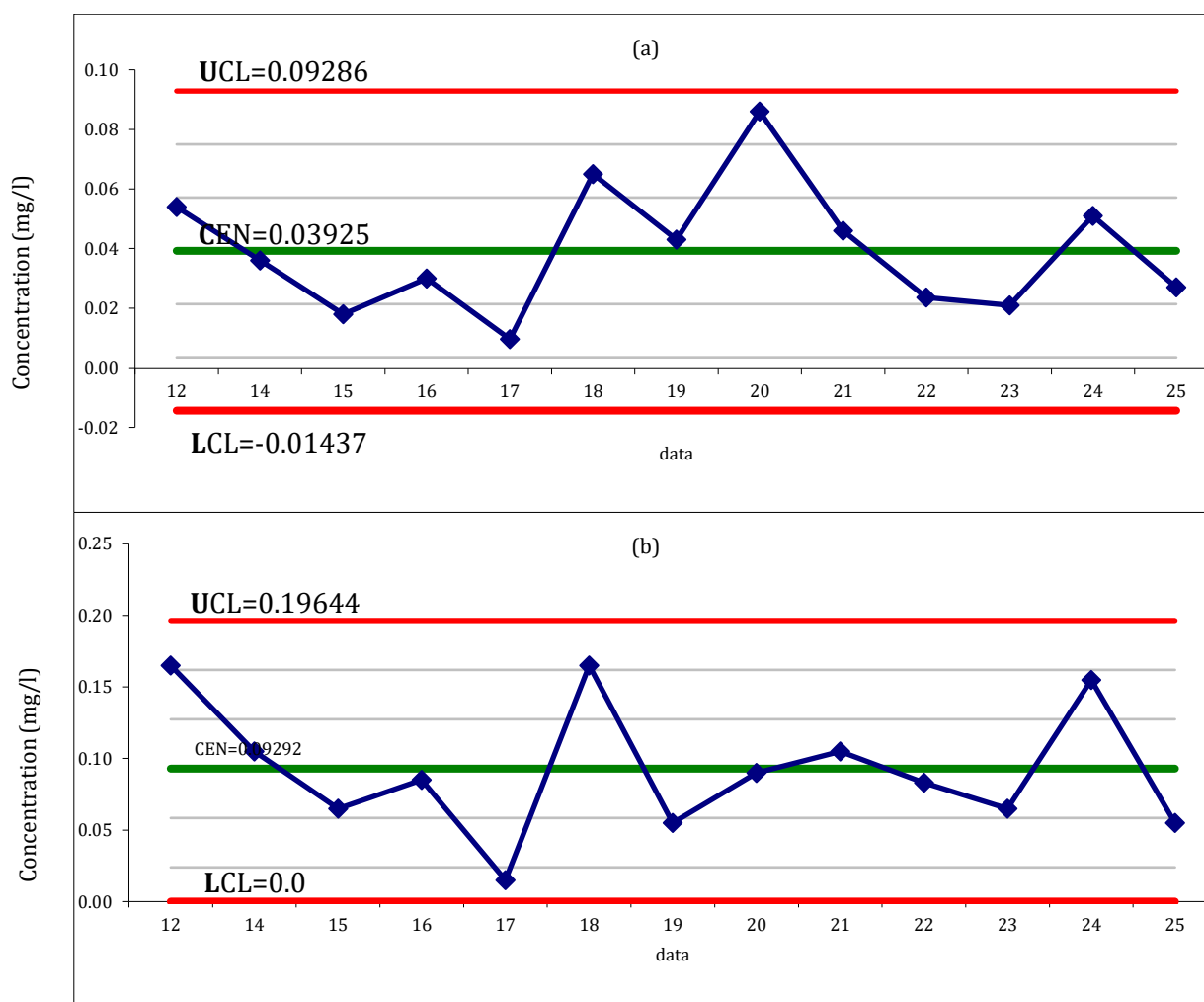


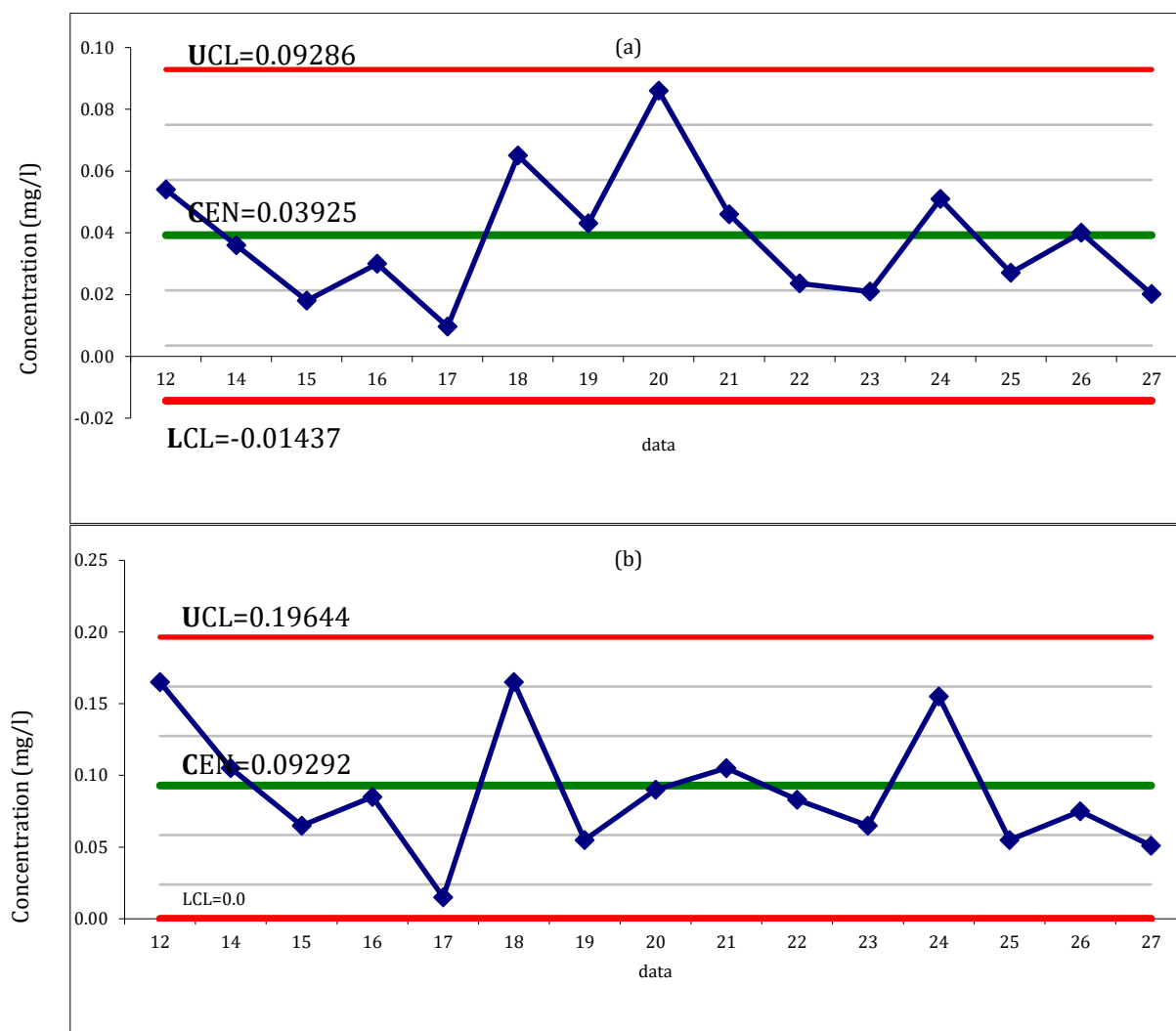**Fig. 6:** Revised Control Limit (a) x̄ chart and (b) R chart for the $PO_4$ concentration

**Fig. 7:** Monitoring of PO₄ concentration data using the revised control limit of (a) $\bar{x}$ chart and (b) R chart

Thus, this result implies that the process is not suitable in the subsequent large period of time. More inspection is needed to control the PO₄ concentration based on the specification limit that has been set up by the DOE which is USL (0.0075mg/L) and LSL (0.005mg/L). This specification limit is referred to the NWQS which has been set up by DOE.
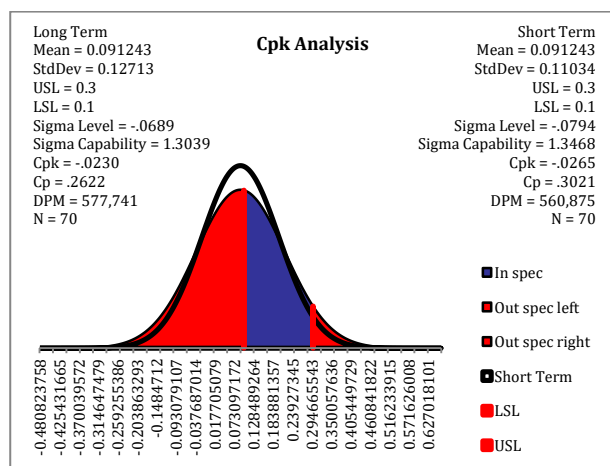


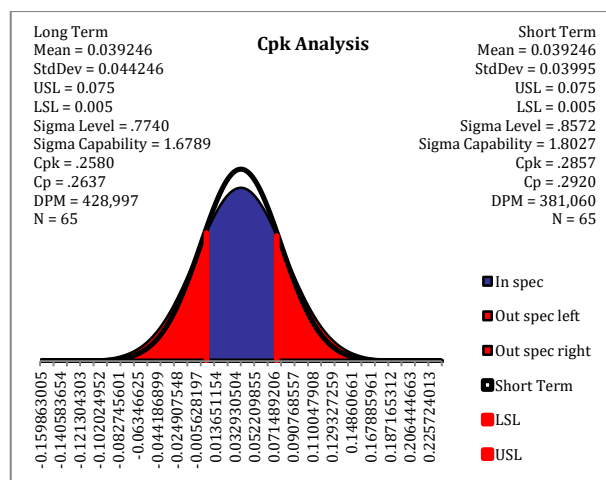**Fig. 8:** Process capability analysis for NH₃N concentration



**Fig. 9:** Process capability analysis for PO₄ concentration

## 4. Conclusion

From this study, the differences of the types of pollution sources and dissimilar of significant possible pollution sources can be approved by using environmetric technic. Both concentrations of NH₃-N and PO₄ have shown a high risk for unacceptable water level by using Process Capability Indices. NH₃-N and PO₄ are found as the main pollutants that give

a higher contribution towards the Johor river basin. According to the previous result, Johor River Basin has face the land alteration towards oil palm plantation and agricultural within the period of 2003-2007. Based on SPC, both concentrations of $NH_3$-N and $PO_4$ has shown the risk of unacceptable water pollution is higher. As a result, continuous monitoring in the area should be done for better improvement of river quality in the Johor River Basin. The important of analysis and modeling of water quality need to take seriously by the authority. From water quality data analysis and modelling, the authority can focus on the most significant parameters which contributed to the river pollution. It is save time and save money budget in water quality sampling and lab analysis of the redundant parameters. The parameters which showed significant towards the water quality at the river basins through the data analysis can be used as reference for the authority in determining which parameters have to monitor at the monitoring stations.

## References

Abdullah AR (1995). Environmental pollution in Malaysia: Trends and prospects. TRAC Trends in Analytical Chemistry, 14(5): 191-198.

Al-Mamun A and Zainuddin Z (2013). Sustainable river water quality management in Malaysia. IIUM Engineering Journal, 14(1): 29-42.

Besterfield DH (2009). Quality control 8th Edition. Pearson Education, Upper Saddle River, USA.

Brodnjak-Vonˇcina B, Dobˇcnik D, Noviˇc M, and Zupan J (2002). Chemometrics characterisation of the quality of river water. Analytica Chimica Acta, 462(1): 87–100.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, and Smith VH (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications, 8(3): 559–568.

Corbett CJ and Pan J (2002). Evaluating environmental performance using statistical process control techniques. European Journal of Operational Research, 139(1): 68-83.

DID (2000). Integrated River Water Management Manual for Malaysia. Department of Irrigation and Drainage, Ministry of Agriculture, Malaysia.

Dominick D, Juahir H, Latif MT, Zain SM, and Aris AZ (2012). Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmospheric Environment, 60: 172-181.

Dorofki M, Elshafie AH, Jaafar O, Karim OA, and Mastura S (2012). Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. International Proceedings of Chemical, Biological and Environmental Engineering, 33: 39-44.

Douglas CM (2009). Introduction to statistical quality control. John Wiley and Sons, Hoboken, USA

Durell GS, Seavey JA, and Higman J (2004). Sediment quality in the lower St. Johns River and Cedar-Ortega River Basin: Chemical contaminant characteristics. Battelle, Duxbury, USA.

Felipe-Sotelo M, Andrade JM, Carlosena A, and Tauler R (2007). Temporal characterisation of river waters in urban and semi-urban areas using physico-chemical parameters and chemometric methods. Analytica Chimica Acta, 583(1): 128–137.

Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, and Juahir H (2012). Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. Marine Pollution Bulletin, 64(4): 688-698.

Guille´n-Casla V, Noelia Rosales-Conrado N, Leo´n-Gonza´lez ME, Pe´rez-Arribas LV, and Polo-Dı´ez LM (2011). Principal component analysis (PCA) and multiple linear regression (MLR) statistical tools to evaluate the effect of E-beam irradiation on ready-to-eat food. Journal of Food Composition and Analysis, 24(3): 456–464.

Ho SC (1996). Vision 2020: Towards an environmental sound and sustainable development of freshwater resources in Malaysia. Geosciences Journal, 40(1): 73-84.

Ismail Z (2011) Monitoring trends of nitrate, chloride and phosphate levels in an urban river. International Journal of Water Resources and Environmental Engineering, 3(7): 132-138

Juahir H, Zain MS, Aris AZ, Yusof MK, Samah MAA, and Mokhtar M (2010b). Hydrological trend analysis due to land use changes at Langat River Basin. Environment Asia, 3: 20-31.

Juahir H, Zain MS, Aris AZ, Yusoff MK, and Mokhtar M (2010a). Spatial assessment of Langat River water quality using chemometrics. Journal of Environmental Monitoring, 12(1): 287-295.

Juahir H, Zain MS, Yusoff MK, Ismail TTH, Samah A MA, Toriman ME, and Mokhtar M (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environmental Mnitoring and Assessment, 173(1): 625–641.

Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WAR, and Moradi A (2012). An artificial neural network model for flood simulation using GIS Johor River Basin, Malaysia. Environmental Earth Sciences, 67(1): 251-264.

Kowalkowski T, Zbytniewski R, Szpejna J, and Buszewski B (2006). Application of chemometrics in river water classification. Water Research, 40(4): 744–752.

Krishna AK, Satyanarayanan M, and Govil PK (2009). Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: A case study from Patancheru, Medak District, Andhra Pradesh, India. Journal of Hazardous Materials, 167(1): 366-373.

Lenntech (2014). Cadmium Cd: Chemical properties of cadmium-Health effects of cadmium-Environmental effects of cadmium, Lenntech B.V., Delft, Netherlands. Available online at: http://www.lenntech.com/periodic/elements/cd.htm#ixzz4lY6Y4epL

Liu CW, Lin KH, and Kuo YM (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science in the Total Environment, 313(1): 77-89.

Madu CN (1996). Managing green technologies for global competitiveness. Praeger, Westport, USA.

Maurer D, Mengel M, Robertson G, Gerlinger T, and Lissner A (1999). Statistical process control in sediment pollutant analysis. Environmental Pollution, 104(1): 21-29.

Mohd I, Mansor MA, Awaluddin MRA, Nasir MFM, Samsudin MS, Juahir H, and Ramli N (2011). Pattern recognition of kedah river water quality data implementation of principal component analysis. World Applied Sciences Journal, 14: 66-72.

Nasir MFM, Samsudin MS, Mohamad I, Awaluddin MRA, Mansor MA, Juahir H, and Ramli N (2011). River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): A case study at Klang River, Malaysia. World Applied Sciences Journal, 14: 73-82.

Osman R, Saim N, Juahir H, and Abdullah MP (2012). Chemometric application in identifying sources of organic contaminants in

Langat river basin. Environmental Monitoring and Assessment, 184(2): 1001-1014.

Pere´-Trepat E, Olivella L, Ginebreda A, Caixach J, and Tauler R (2006). Chemometrics modelling of organic contaminants in fish and sediment river samples. Science of The Total Environment, 371(1-3): 223-237.

Retnam A, Zakaria MP, Juahir H, Aris AZ, Zali MA, and Kasim MF (2013). Chemometric techniques in distribution, characterisation and source apportionment of polycyclic aromatic hydrocarbons (PAHS) in aquaculture sediments in Malaysia. Marine Pollution Bulletin, 69(1): 55–66.

Saim N, Osman R, Spian DRSA, Jaafar MZ, Juahir H, Abdullah MP, and Ghani FA (2009). Chemometric approach to validating faecal sterols as source tracer for faecal contamination in water. Water Research, 43(20): 5023–5030.

Samsudin MS, Juahir H, Zain SM, and Adnan NH (2011). Surface river water quality interpretation using environmetric techniques: Case study at Perlis River Basin, Malaysia. International Journal of Environmental Protection, 1(5): 1-8.

Shrestha S and Kazama F (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. Environmental Modelling and Software, 22(4): 464-475.

Simeonov V, Einax JW, Stanimirova I, and Kraft J (2002). Environmetric modeling and interpretation of river water monitoring data. Analytical and Bioanalytical Chemistry, 374(5): 898–905.

Simeonov V, Simeonova P, and Tsitouridou R (2004). Chemometric quality assessment of surface waters: Two case studies. Chemical and Engineering Ecology, 11(6): 449-469.

Simeonova V, Stratisb JA, Samarac C, Zachariadisb, G., Voutsac D, Anthemidis A, Sofonioub M, and Kouimtzisc TH (2003). Assessment of the surface water quality in Northern Greece. Water Research, 37(17): 4119–4124.

Singh KP, Malik A, and Sinha S (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. Analytica Chimica Acta, 538(1-2): 355-374.

Sojka M, Siepak M, Zioła A, Frankowski M, Murat-Błażejewska S, and Siepak J (2008). Application of multivariate statistical techniques to evaluation of water quality in the Mała Wełna River (Western Poland). Environmental Monitoring and Assessment, 147(1): 159–170.

Sousa SIV, Martins FG, Alvim-Ferraz MCM, and Pereira MC (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environmental Modelling and Software, 22(1): 97-103

Su S, Junjun Zhi J, Lou L, Huang F, Chen X, and Wu J (2011a). Spatio-temporal patterns and source apportionment of pollution in Qiantang River (China) using neural-based modeling and multivariate statistical techniques. Physics and Chemistry of the Earth, 36(9): 379–386.

Su S, Li D, Zhang Q, Xiao R, Huang F, and Wu J (2011b). Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China. Water Research, 45(4): 1781-1795.

Vega M, Pardo R, Barrado E, and 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.

Wang L, Yang Z, Niu J, and Wang J (2009). Characterization, ecological risk assessment and source diagnostics of polycyclic aromatic hydrocarbons in water column of the Yellow River Delta, one of the most plenty biodiversity zones in the world. Journal of Hazardous Materials, 169(1): 460-465.

Woodall WH, Hoerl RW, Palm AC, and Wheeler DJ (2000). Controversies and contradictions in statistical process control/Discussion/Response. Journal of Quality Technology, 32(4): 341-350.

Wu B, Zhao D, Zhang Y, Zhang X, and Cheng S (2009). Multivariate statistical study of organic pollutants in Nanjing reaches of Yangtze River. Journal of Hazardous Materials, 169(1): 1093-1098.

Zali MA, Retnam A, and Juahir H (2011). Spatial characterization of water quality using principal component analysis approach at Juru River Basin, Malaysia. World Applied Sciences Journal, 14: 55-59.

Zhou F, Liu Y, and Guo H (2007). Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. Environmental Monitoring and Assessment, 132(1): 1-13.