PHYSICO-CHEMICAL ASSESSMENT ON QUALITY PARAMETER IN SUNGAI PAPAR, SABAH VIA STATISTICAL ANALYSIS

NOR FADHILLAH MOHMAED AZMIN[1*], N SHOFIA A'YUN SYAFIE[1], AZLIN SUHAIDA AZMI[1], MIMI FINA HAMIDON[2], ANI LIZA ASNAWI[1]

[1]*Department of Biotechnology Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia.*
[2]*Jabatan Air Negeri Sabah, Malaysia.*

**Corresponding author: norfadhillah@iium.edu.my*

ABSTRACT: Sg. Papar is one of the rivers in Kota Kinabalu which is mainly used for water supply especially in Papar district. For the past years, many pollution cases concerning Sg. Papar have been reported which originated from various sources including pig farm, agricultural run-off and deforestation. These resulted in a frequent shutdown of the water treatment plants in Papar district leading to water supply disturbance and water supply deficiency in the affected area. The data utilized in this study were obtained from water quality tests performed on river water samples taken from Limbahau water treatment plant recorded from September 2013 to September 2016. Principal Component Analysis (PCA) was used in this study to analyze and correlate the physicochemical parameters with the water treatment plant shutdown. The results revealed that eight parameters (pH, alum, nitrate, TDS, DO, conductivity, colour and chloride) analysed in this study correlate with each other and the parameter that mostly caused the drastic change in the river water and as pollution index is turbidity. This study is critical for understanding the relationship between the water quality paramters and environmental issues.

KEY WORDS: *Water quality parameter; Principal Component Analysis; Chemometrics; Sg. Papar;*

1. INTRODUCTION

The National Water Quality Standards for Malaysia (NWQS) provides an index for the environment assessment of rivers in Malaysia namely Water Quality Index (WQI). This index is used by the Department of Environment (DOE) as a foundation to assess the categories of pollution load and the classess of rivers in Malaysia. This is to ensure that the river water is safe for drinking and daily human activities. Water from rivers are used extensively for drinking, washing, laundry, agriculture, and recreational purposes. In addition, maintaining river water quality based on the standard is important for health and environmental protection [1].

To date the water pollution issues in many rivers in Malaysia are becoming more severe. The major pollutants in river waters are biochemical oxygen demand (BOD), ammoniacal nitrogen (AN) and suspended solids (SS); these pollutants come from various sources including sewage, animal farming and land clearing activities [2]. According to Malaysia Environment Quality Report 2017, the Department of Environment has recorded 5 major type of water pollution sources comprising of sewage treatment plant (49%), piggery (39%), manufacturing industries (9%), agricultural industries (2%), and wet market (1%) [3].

One of the affected and worsening rivers is Sungai Papar, Sabah. Sungai Papar (5.716541 ^{o}N, 115.950587 ^{o}E) located in Papar district 37 kilometers from Kota Kinabalu. Fig. 2 shows the location map of Papar district. Sg. Papar is amongst the main water source and one of the navigable rivers in the district. This river flows from Crocker Range Park, Tambunan, goes through Kimanis, Papar town and finally flows into South China Sea. Sg. Papar's watershed covers 740 square kilometers in two districts of Papar which is Kawang and Pantai Manis and it is 60 km long. Sg. Papar is the main source of water supply in Papar and nearby areas such as Putatan, Penampang and Kota Kinabalu.The Malaysia National Water Quality Standard (NWQS) classfies the river water quality of Sg. Papar is classified into class II [4].

Fig. 2 Location of Papar district

The dependency of majority of the people within Papar to Sg. Papar for their livelihoods affect physical, biological and chemical parameters of the river. Not only that, the effects of hydro-meteorological and climate variability also become one of contributing factors to the variation of water quality parameters of the Sg. Papar. 40% of land usage in Papar is used for agricultural activity, 25% is for industrial and housing purposes, whilst the remaining 35% of the land areas is for forest reservation [5], as illustrated in Fig. 3 below.

Since Papar is well known for its agriculture activity, pollution was reported coming from animal farming and deforestation along th river especially during the replantation season [6]. The dynamic features of Sg. Papar causes the water quality parameters to fluctuate and often changing over time. It worsened during the wet season when river water exhibits muddines i.e. high turbidity caused by the river flows and aeration from upstream region. Extreme turbidity impedes the water pump from operating efficiently especially during wet seasons and the occurance of floods. This leads to water treatment plant (WTP) shutdown until the problem is sorted out. Unfortunately, it is reported that this event occurs almost annually which leads to water supply disturbance and water supply deficiency in the affected area especially in Papar. According to a study [7] at Sukhnag stream, Kashmir water in the stream flows in tremendous velocity in the upper region during flash rains causing soil wastage on the left and right of the embankments of the stream. This cause a great damage to crops, plantation, house, road and communication line in the area. A group of researchers suggested that increase in turbidity caused by human activities results in phytoplankton growth and materials deposited in the river [8]. High value of turbidity normally indicates high level of other parameters associated with water quality such as chemical oxygen demand (COD) or other substrate concentration such as nitrate, sulfate and ammonium [9], [10].

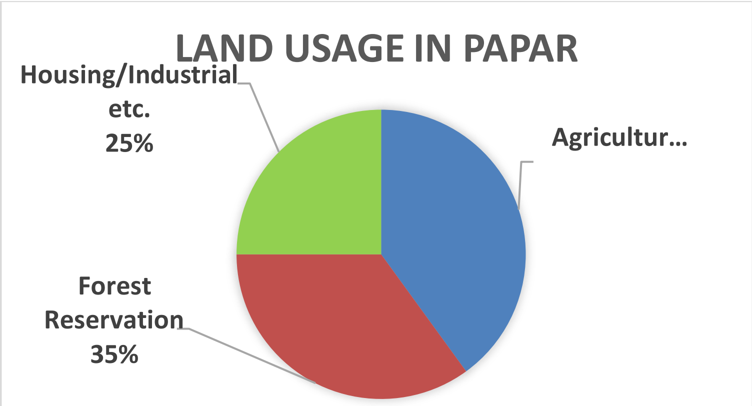

Fig. 3. Percentages of Land Usage in Papar

The PCA method has been used widely in environmental chemometrics including studies in [7], the study of surface water at Illinois river basin [11], and classification of the river water quality over ten years period at Sungai Muda, Kedah [12]. It is the mother of multivariate statistical analysis which can extract the relevant data from a big group number of data [13].

The objective of this paper is to identify the correlation between the physicochemical parameters of Sungai Papar that results to WTP pollution and shutdown. The large dataset recorded during a span of three years (2013 – 2016) is subjected to a multivariate data analysis specifically principal component analysis.

2. MATERIALS AND METHOD

2.1. Study Area & The Water Quality Data

This study focused on the data of river water sample taken from Limbahau water treatment WTP which is located at Kampung Limbahau (5°42'58.2"N 115°56'56.2"E), about 4 kilometer from Papar town. Fig. 4 shows the location of Kampung Limbahau where the nearby area is surrounded with residential area and housing. The land near the river basin is also used for agriculture such as paddy field, rubber and fruit orchard. The archive data of Sg. Papar was provided by Jabatan Air Negeri Sabah and it covers the period of three years (2013 – 2016). The dataset comprises of fifteen water physicochemical parameters: turbidity, electrical conductivity (EC), total dissolved solid (TDS), dissolved oxygen (DO), colour (hazen), pH, ammonia (NH_3^+), nitrate (NO_3^-), iron (Fe), manganese (Mn), chloride (Cl^-), aluminium (Al_3^+), sulfate (SO_4^{2-}), total hardness and alkalinity. The dataset formed a matrix of 127 (observations within the three years period) x 15 (physicochemical parameters) and it is used for the modelling purposes.

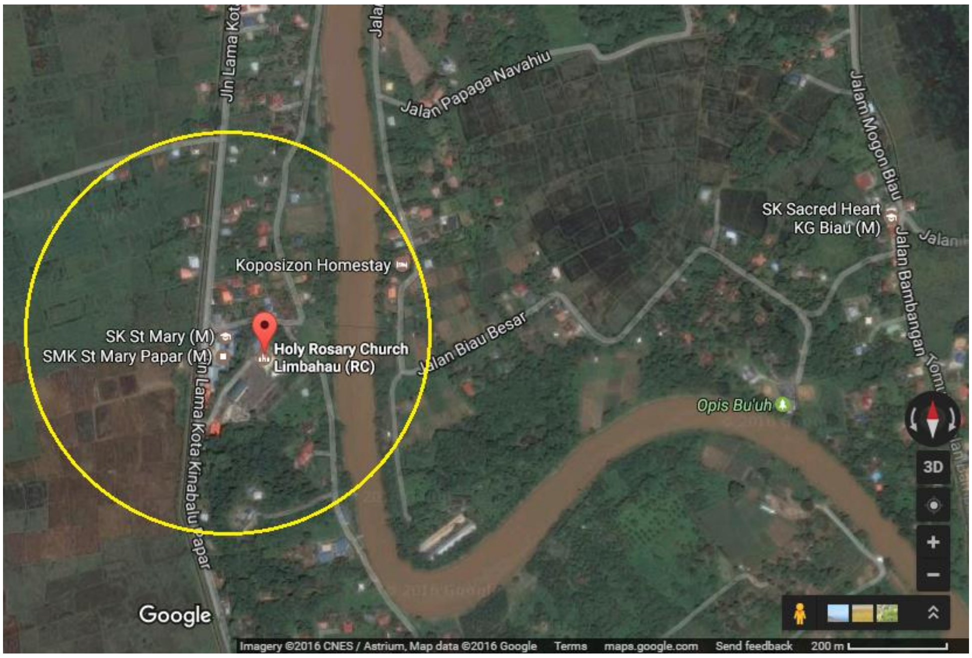

Fig. 4. Location of Kampung Limbahau (retrieved from Google Maps)

2.2 Principal Component Analysis

PCA is a multivariate statistical technique which can be used to reduce the dimensionality of a large dataset while retaining as much as possible the variation present [14]. A data matrix of *m* x *n* (Fig. 5) comprising *n* number of variables which are correlated to various degrees are transformed to a new data set containing *n* new orthogonal, uncorrelated variables called principal components (PCs).

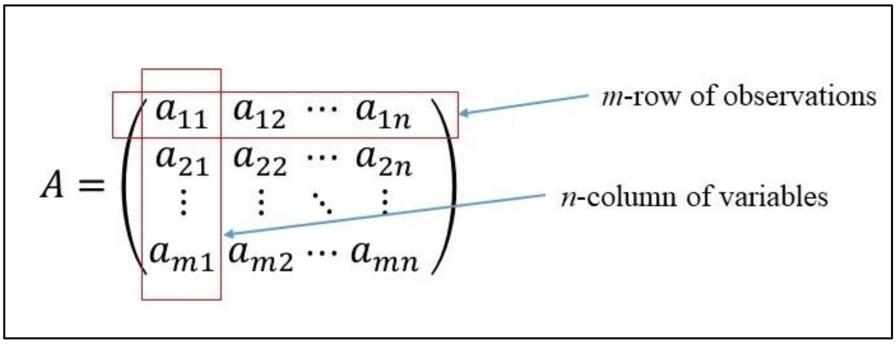

Fig. 5. Data matrix **A** of *m* x *n*

The PCs are linear functions of the original variables. The sum of their variance equal to that of the original variables. The sequences of the PCs are in ordered from the largest variance (PC1), to the next largest variance (PC2), PC3 and so on. This relation is represented in Eq. (1) to and Eq. (2).

$$\text{PC1} = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = \sum_{j-1}^{n} a_{1j}x_j \quad (1)$$

$$\text{PC2} = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = \sum_{j-1}^{n} a_{2j}x_j \quad (2)$$

where x is matrix of,

$$x_i = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

In this study, PCA was performed on the dataset illustrated in Fig. 6 below to extract underlying behaviour and distinctive characteristics of the physicochemical parameters. The ultimate objective is to link the behaviour and characteristics of the physicochemical parameters to the forced shutdown of the WTP. Matlab version R2009b is used to perform the PCA.

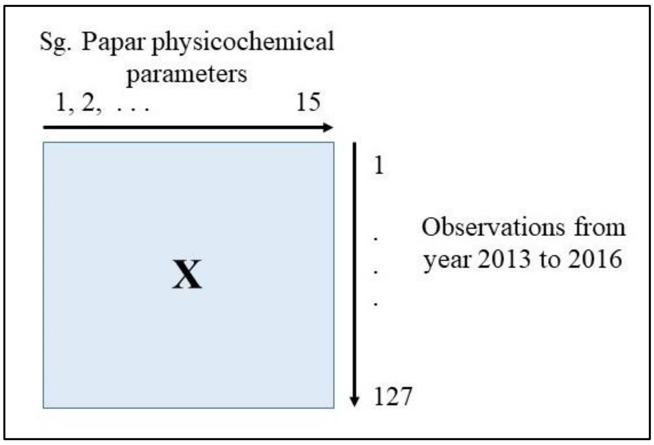

Fig. 6. A data matrix of fifteen physicochemical parameters of Sg. Papar, Sabah from 2013 to 2016.

3. RESULT AND DISCUSSION

A scree plot is the primary step in the analysis because it determines the number of principal components (PC) need to be retained for further analysis. It is a plot of percentage of variance explained versus the number of principal components. In this study, the scree plot of the dataset (Fig. 7) shows that PC1 and PC2 capture 94% of the variation from the 127 observations of the fifteen physicochemical parameters of Sg. Paper river water. What this means is that the correlation of the fifteen original variables is reduced to and can be represented by two latent variables. These latent variables can be further predicted by score plots drawn onto two PCs into two dimensions or a maximum of three PCs in three dimensions.

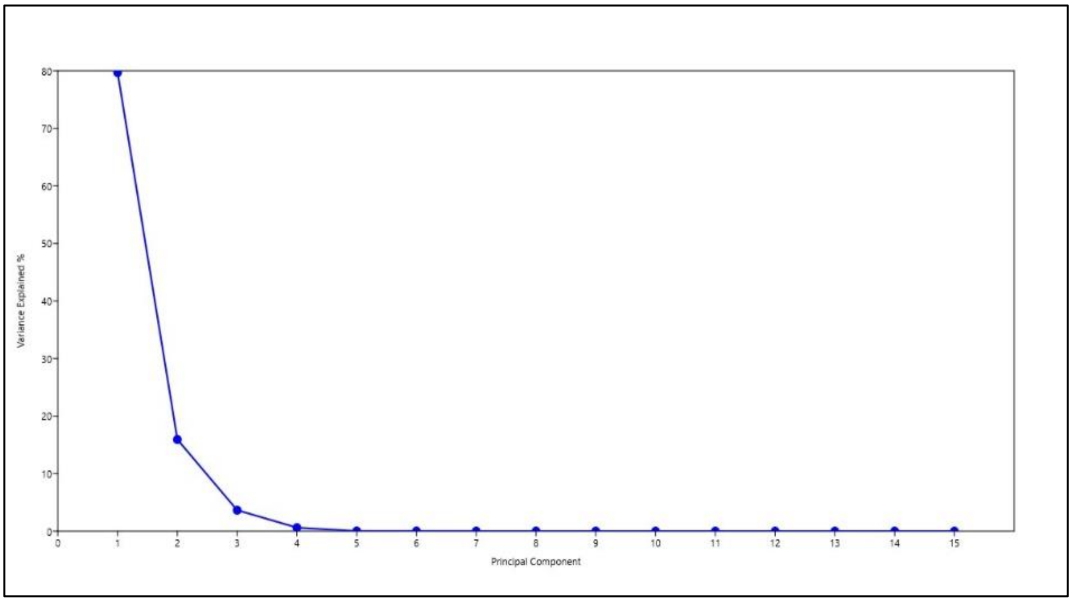

Fig. 7. Scree Plot for the Sg. Papar dataset.

In this study, the biplot is a scatter plot that projects the original observations i.e. the time of the water sample recorded onto PC1 vs. PC2 (Fig. 8) and PC3 vs. PC4 (Fig. 9). Fig. 8 shows that most of the points were positioned near the origin, whilst some were scattered away from the origin along the positive space of PC1. These few points are influenced by their colours despite being recorded in different seasons (wet vs. dry). Positive correlations were observed between EC, TDS and chloride in the positive space of PC2. The positive correlation between EC and chloride is anticipated because electrical conductivity in water is affected by concentration of chloride salt in water. During a rainy season aeration of flow water increase and particulates and minerals were also aerated inside the flowing water. Fig. 9 shows that turbidity and hardness have a positive correlation in the positive space of PC3. Turbidity caused by human activities results in materials deposited in the river and this leads to increase of hardness in the water.

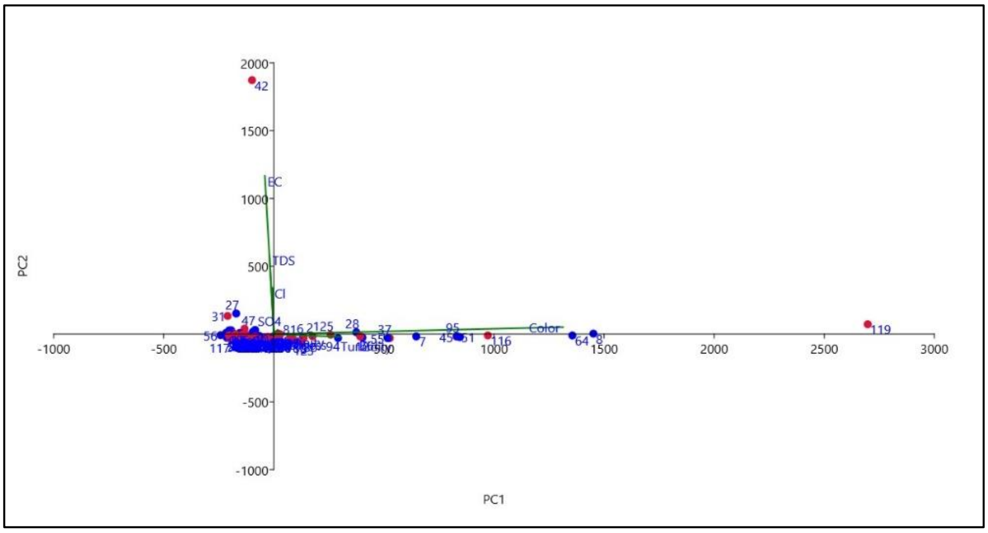

Fig. 8. PCA biplot of the scores along the PC1 and PC2 (wet season (blue) and dry season (red)).

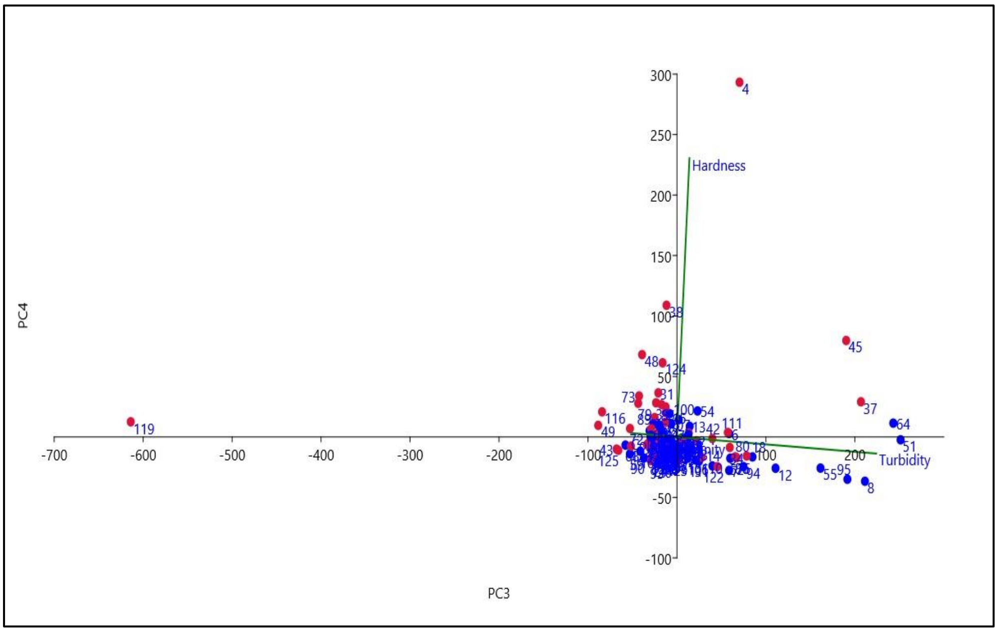

Fig. 9. PCA biplot of the scores along the PC3 and PC4 (wet season (blue) and dry season (red)).

Besides the score plot, the loading plot is also of importance. It answer the question which of the variables are influential. Fig. 10 and Fig. 11 show loading plots of PC1 and PC2 respectively. From both figures, it can be seen that the variables with heaviest loads for PC1 in a descending order are: colour, turbidity and EC whilst for PC2 are EC, TDS and chloride. EC is the common factor for both PCs. The findings in this study can be considered as a pioneer scientific study on Sg. Papar.

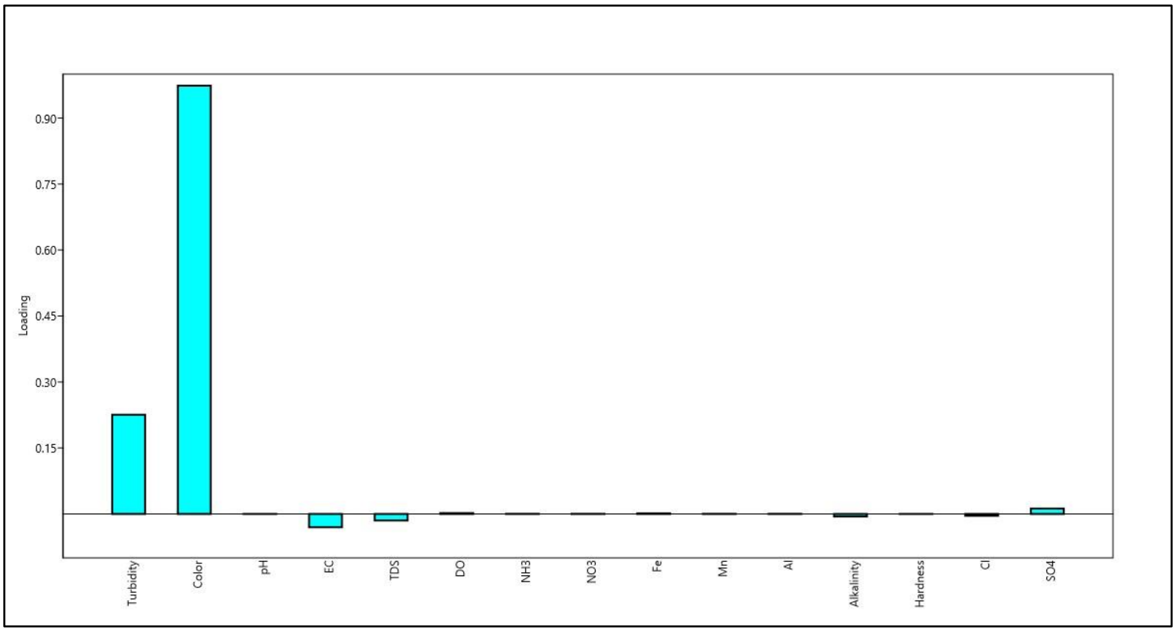

Fig. 10. Loading plot of PC1

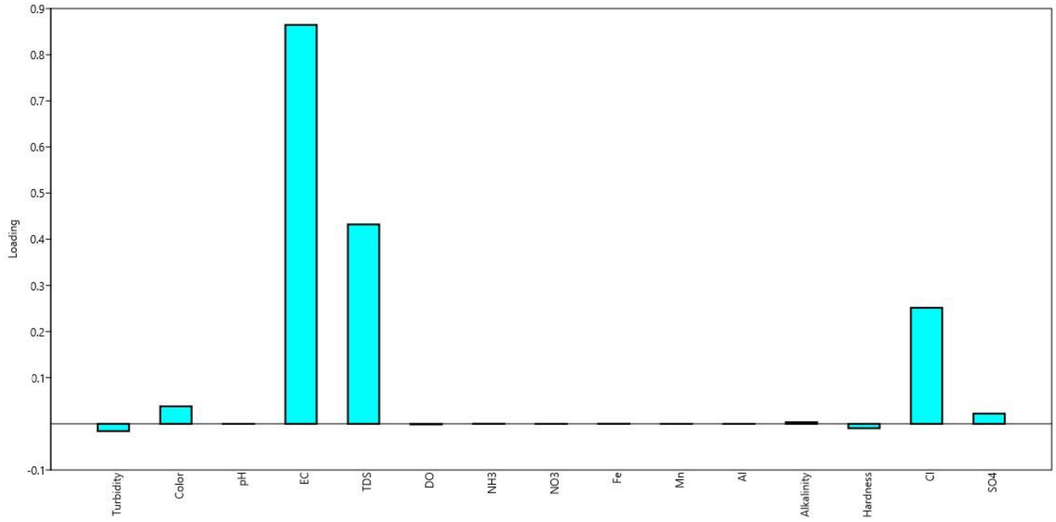

Fig. 11. Loading plot of PC2

4. CONCLUSION

The PCA study on the Sg. Papar physicochemical parameters over the period of three years monitoring (2013 – 2016) shows that two latent variables are sufficient to represent the variation in the fifteen physicochemical variables of the Sg. Papar, Sabah river water. The first latent variable is conjectured to represent the relationship between colour and turbidity whilst the second latent variable is conjectured to represent the relationship between EC, TDS and chloride.

ACKNOWLEDGEMENT

The authors thanked the International Islamic University Malaysia (IIUM) for providing the facilities for this work and the Jabatan Air Negeri Sabah, Malaysia for providing the water quality data.

REFERENCES

[1] Li D and Liu S (2019). Water Quality Evaluation in Water Quality Monitoring and Management. Available: https://doi.org/10.1016/C2016-0-00573-9

[2] Huang YF, Ang SY, Lee KM and Lee TS (2015). Quality of Water Resources in Malaysia. Available: http://doi:10.5772/58)969

[3] Progress of Water Environment Governance in Malaysia. Available: http://wepa-db.net/3rd/en/meeting/20190222/pdf/D2_S3_Malaysia_200219.pdf

[4] Jabatan Pengairan dan Saliran Sabah. (2011). Senarai Laporan JPS.

[5] Papar District Council (2014). Laporan Profil Daerah.

[6] Daily Express (2015, April 9). Local news: 2,000 villagers want Papar pig farm to be closed. Retrieved from Daily Express.

[7] Bhat SA, Meraj G, Yasen S, Pandit AK. (2014) Statistical Assessment of Water Quality Parameters for Pollution Source Identification in Sukhnag Stream: An Infloe Stream of Lake Wular (Ramsar Site), Kashmir Himalaya. Journal of Ecosystem. http://dx.doi.org/10.1155/2014/898054

[8] Nieto PG, Garcia-Gonzalo E, Fernandez JA, Muniz CD. (2014) Hybrid PSO-SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Nothern Spain. Ecological Engineering, 73, 192-200.
[9] Muniz DC, Nieto GPJ, Fernandez AJR, Torres MJ, Taboada J (2012). Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain). Sci Total Environment, 439, 54-61.
[10] Vigil Kenneth (2003). Clean water: An introduction to water quality and water pollution control (2nd ed). Oregon State University Press, Corvallis.
[11] Olsen RL, Loftis JC, Chappell RW (2012). Water quality sample collection, data treatment and results presentation for principal component analysis – literature review and Illinois River watershed case study. Water Research, 46, 1-13.
[12] Azhar SC, Aris AZ, Yusoff MK, Ramlia MF, Juahir H (2015). Classification of river water quality using multivariate analysis. International Conference on Environmental Forensics, 79-84.
[13] Johnson GW, Ehrlich R, Full W, Ramos S (2015). Principal Component Analysis and Receptor Models in Environmental Forensics. Introduction to Environmental Forensics, 610-647.
[14] Ringner M (2008). What is Principal Component Analysis?