International Journal of GEOMATE, Oct., 2017, Vol.13, Issue 38, pp.89-96
Geotec., Const. Mat. & Env., ISSN:2186-2990, Japan, DOI: http://dx.doi.org/10.21660/2017.38.402807

PRIMARY POLLUTANT SELECTION AND DETERMINATION OF WATER QUALITY INDEX IN CLASS DISCHARGE DIVISION BASED ON THREE CLASS MARKOV MODEL

*Mariana Marselina[1], Arwin Sabar[1], Indah R S S Salami[1] and Dyah Marganingrum[2]

[1]Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Indonesia; [2]Indonesian Institute of Sciences, Indonesia

*Corresponding author; Received: 25 Feb. 2017, Revised: 9 March 2017, Accepted: 30 March 2017

ABSTRACT: This study aims to identify how the dominant parameter selection may show pollutants source in every segment of the reservoir, be it reservoir input segment, reservoir middle segment or body reservoir segment, and reservoir outlet segment that turns out to be diverse. Water Quality Index (WQI) value in this study was calculated for each parameter chosen based on the Principal Component Analysis (PCA). In addition, this study showed that pollution in the rainy season was lower that the pollution in the dry season, this was indicated by the WQI values obtained. This study also showed that the reservoir input pollution source was dominated by domestic activities and industry, while the middle segment of the reservoir usually polluted by floating nets activity, and in the reservoir outlet, pollutant accumulation occurred as indicated by the presence of H_2S as the main pollutant.

Keywords: Markov model, Principal component analysis, Primary pollutant, Saguling reservoir, Water quality index

1. INTRODUCTION

Reservoir water quality is not only determined by natural processes, including weather and soil erosion, but also by anthropogenic activities [1], [2]. Research on reservoir water quality needs to consider the impact of human activities, the variations in pollutant concentrations in space and time, and biology, physics and chemistry processes of natural systems [3]-[5].

Saguling is a reservoir located in upstream Citarum watershed, West Java Province, Indonesia. Saguling forms Citarum Cascade Reservoir which sequentially comprises Saguling, Cirata and Jatiluhur Reservoirs. Saguling is located in the upper Citarum River which acts as a trap spot for pollutants discharged into Citarum River and around the residential area of Bandung. The main purpose of Saguling establishment is to generate electricity. However, the displacement of approximately 40,000 people from agricultural land has led to the use of Saguling for the cultivation of floating net cages which has certainly contributed to the water pollution [6].

The location of Saguling reservoir in the upper part of Citarum watershed makes Saguling act as a trap spot for pollutants discharged into the Citarum River and around the residential area of Bandung. This location is described in Figure 1.

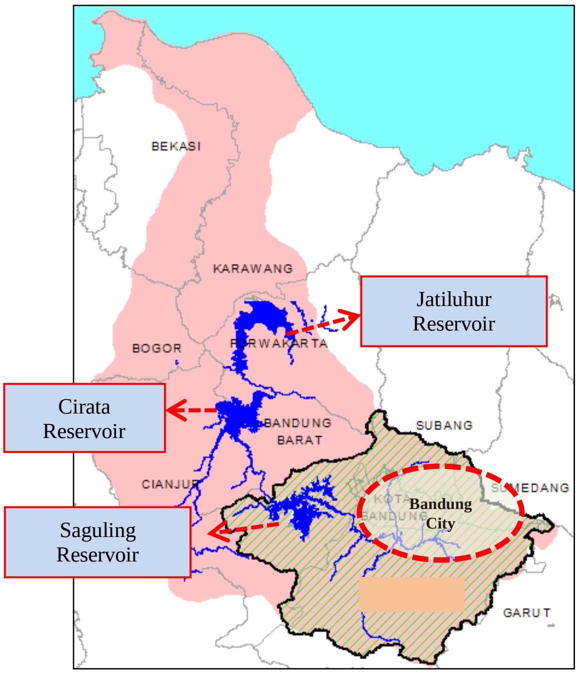

Fig 1. Location of Saguling and Citarum Watershed

Research [7] found that Saguling has been heavily polluted, particularly by domestic sewage and industrial waste. Most of the contaminants in Saguling, such as organic matter, nutrients, and heavy metals, comes from Bandung and that is why Saguling was identified as a very tropical reservoir.

A number of studies on the determination of water quality have been conducted by applying various statistical analyses and techniques, such as factor analysis or principal component analysis to identify the most influential factor in water contamination; whether it is due to natural factor or anthropogenic factor from human activities [8]-[15]

Research [16] on water quality in Pearl River Delta, China in 1952-2009 found that there is a difference between the measurement results obtained in the rainy season and that obtained in dry season in a yearly research period. Concentration levels of BOD, NH^4, SO_4^{2-}, and Cl^- during the dry season is higher than the rainy season.

Research [17] in the Mediterranean river basin by comparing water quality of in two scenarios: rainy season and dry season. During the rainy season, it was found that farm waste runoff and surface erosion dominantly influence nutrient pollution, while during dry season untreated domestic waste entirely become the major pollutant.

Study on water quality in Saguling reservoir in this research was conducted by statistical analysis of the main components analysis to determine the most dominant pollutants in the deterioration of water quality in Saguling, and Water Quality Index (WQI) method was used to determine the pollution parameters in Saguling under three atmospheric conditions: dry, normal and wet years, to see the relationship between pollutant concentration and discharge conditions with water quality throughout the years (1999-2013).

2. MATERIAL AND METHOD

2.1 Determination of Wet, Normal, and Dry Years with Markov Model

Markov model has a formula as follows (Descombes dan Berthod, 2006) :

$$q_i = d_i + e_i \qquad (1)$$

where:
di = deterministic component
ei = random component
Otoregresip deterministic component has a linear formula as follows:

$$d_i = \beta_0 + \beta_1 q_{t-1} + \beta_2 q_{t-2} + \ldots\ldots\ldots + \beta_m q_{t-m} \qquad (2)$$

where d_i is a linear combination of m previous flow, for a limited m. The simplest model has the following formula:

$$d_i = \beta 0 + \beta_1 q_{t-1} + e_1 \qquad (3)$$

The above model assumes that all previous influence to the flow is now reflected in the value of the previous flow. Furthermore, in lag model, the amount of β_0 dan β_1 constants are specified with the exact form of e_i. In the first place, it is considered that the flow follows the normal distribution.

If μ is the middle-value flow, ρ is the correlation of coefficient serial lag and σ is flow diversity, the following equation is established:

$$q_1 = \mu + \rho\,(q_{i-1} - \mu) + e_i \qquad (4)$$

Here, it is assumed that the flow of qi is the sum of the median, some of the differences in previous flow q_{i-1} with the mean and random components e_i. If q_i has a normal distribution, then e_i should also have a normal distribution as well. The flow of q_i has median μ, so the suggested form will give the desired middle value. The flow q_i has diversity as follows:

$$E\,[\mu + (q_{i-1} - \mu) + e_i]^2 - \mu^2 = \sigma^2\rho^2 + \sigma_e^2 \qquad (5)$$

where σ_e^2 is a diversity of the random components ei. The diversity q_i is related with σ_e^2 which expressed in the following formula:

$$s_e^2 = \sigma^2\,(1-\rho^2) \qquad (6)$$

If t_i is a random variable that has a normal distribution, but does not serially rely on 0 median and has 1 value of standard deviation, then:

$$ti\ \sigma\,((1 - \rho 2\,))^{0,5} \qquad (7)$$

also, a normally distributed variable does not depend serially, on median 0 and variance σ2 (1-ρ2). Thus the model is as follows:

$$q_i = \mu + \rho\,(q_{i-1} - \mu) + t_i\,\sigma\,((1 - \rho^2\,))^{0,5} \qquad (8)$$

Simplification of water discharge that entered the reservoir in Markov model was conducted by dividing them into three classes. Based on the water division, matrix stochastic can be created each month that divides the historical data into three classes. The stages of Markov model processing in dividing discharge category is as follows:

a. Frequency distribution analysis

An analysis of the frequency distributions of water discharge per month in the period 1999-2013 was conducted. Then the frequency distribution of historical data was matched with the two frequency distribution models: Normal and LogNormal. Each month, the most suitable distribution of the two alternative types was chosen.

b. Class division

The second phase after the selection of frequency distribution was the distribution of water discharge class. The process investigated in this study was the first order for three classes. Therefore, the amount of discharge was divided into three classes as follows:

1) Dry discharge (represented by 0)
2) Normal discharge (represented by 1)
3) Wet discharge (represented by 2)

Class intervals for each class divisions was obtained by dividing the probability curve of the distribution of the selected population into 3 equal parts, namely 0.333, 0.667, and 1 as shown in Figure 2. Range value of each class was the middle value of each class which is in probability curve 0333, 0.667, and 1. The probability determination of each data was conducted by using Weibull method as follows:

$$\boldsymbol{P\ (Xm)} = \frac{\boldsymbol{m}}{\boldsymbol{N+1}} \tag{9}$$

with:

P (Xm) = probability of a set of values that are expected during the observation period

N = number of observation of X variate

m = serial number of events or rate of events

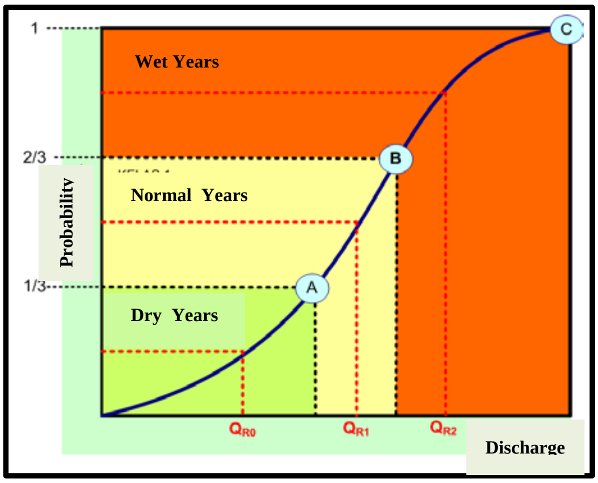

Fig 2. Class divisions based on probability in Markov model

2.2 Monitoring Quality Station in Saguling Reservoir

Monitoring of water quality in Saguling was conducted regularly every three months within one year span, with each taken in March, June, September, and December respectively. There were 11 locations of Saguling quality monitoring stations: Nanjung Station (input), Batujajar, Cipatik Muara, Muara Ciminyak, Cimerang, Tjihaur, Muara Cijere, Cijambu Muara, Muara Tjihaur, Turbine Intake and Tailrace. The monitoring was conducted on forty-four (44) water quality parameters which consist of physical and chemical parameters. Main study location in this research was Nanjung Station, Muara Ciminyak Station, and Intake Turbine station as shown in Figure 3.

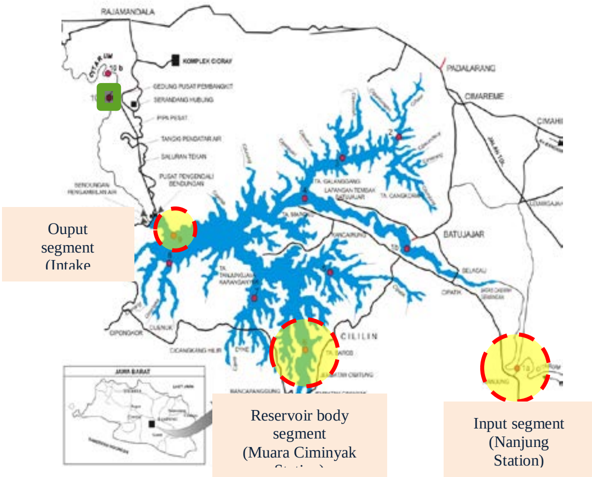

Fig 3. Location of Monitoring Quality Station in Saguling Reservoir

2.3 Principal Component Analysis *(PCA)*

Principal Component Analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components was less than or equal to the number of original variables. This transformation was defined in such a way that the first principal component has the largest possible variance (that was, accounts for as much of the variability in the data as possible), and each succeeding component, in turn, has the highest variance possible under the constraint that it was orthogonal to the preceding components. The resulting vectors were an uncorrelated orthogonal basis set. PCA was sensitive to the relative scaling of the original variables. The Process of PCA can be shown in Figure 4.

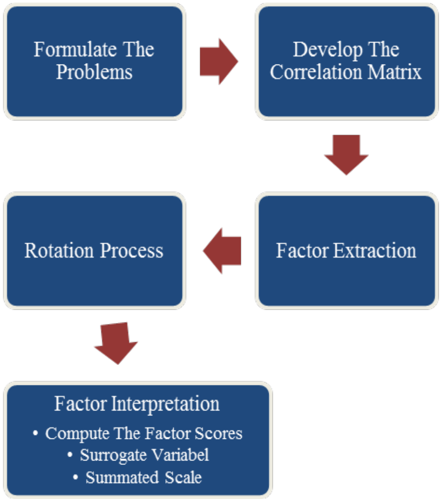

Fig 4. Process of Principal Component Analysis

PCA including factor analysis was one of interdependence analysis between variables. The principle method of factor analysis was to determine the dominant factors in describing a problem by summarizing many data sets to be more concise and no longer correlated. In other words, it was an attempt to extract a number of common factors from variables X1, X2, and so on. Therefore, in the factor analysis, a smaller number of factors were obtained compared to the amount of x variable origin and most of the information from x variable origin stored in a number of factors. The factors formed were unobservable new variables or latent variables. Schematically, the principle method of factor analysis is shown in Figure 5.

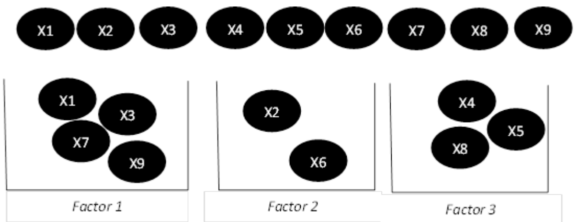

Fig 5. Analytic scheme in factor analysis

2.4 The Measurement of Saguling Water Quality Index

WQI method does not distinguish between physical, chemical, and biological contaminants. All types of contaminants are given equal value [18]. WQI method is quite easy and simple with the following equation:

$$\boldsymbol{WQI} = \frac{\Sigma\left[\frac{Ci}{Pi}\right]}{n} \qquad (10)$$

where:

WQI = quality index (water quality index)
Ci = concentration of the variable i
Pi = quality standards for variable i
n = number of variable

3. RESULTS AND DISCUSSION

Saguling reservoir is in Indonesian territory which is geographically situated between two continents (Asia and Australia) and two oceans (the Indonesian Ocean and the Pacific Ocean). This position placed Indonesia as the region with the influence of El Nino (La Nina) phenomenon in the Pacific Ocean and Dipole Mode phenomenon that occurs in Indonesian Ocean. It also made Indonesia a region with two seasons: rainy season and dry season. The rain analysis results in the Saguling basin area, based on the average monthly rainfall in 1986-2008, showed monsoon type of rain pattern where there was on peak rainfall (unimodal), in this case, occurred in March (Figure 6). Referred to Mohr classification, the wet months occur in October to May and the dry months occur in June-September.

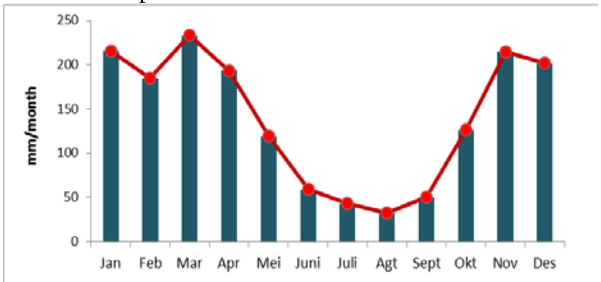

Fig 6. The average monthly rainfall in the basin of Saguling area (1986-2008)

Two season pattern has been a challenge in the management of water resources due to excessive rain during the rainy season, and drought during the dry season. The highest rainfall in rainy season occurs in March and the lowest rainfall in dry season occurs in August. In the rainy season, fairly good water quality is obtained and in the dry season, the water quality cannot be used as a raw water source.

3.1 Determination of the primary pollutants in all monitoring locations

The dominant parameter in Nanjung station, Muara Ciminyak station, and Intake Turbin Station based on Principal Component Analysis that have done to all parameters (44 paramters of water quality) that monitored in each station can be shown in Figure 7.

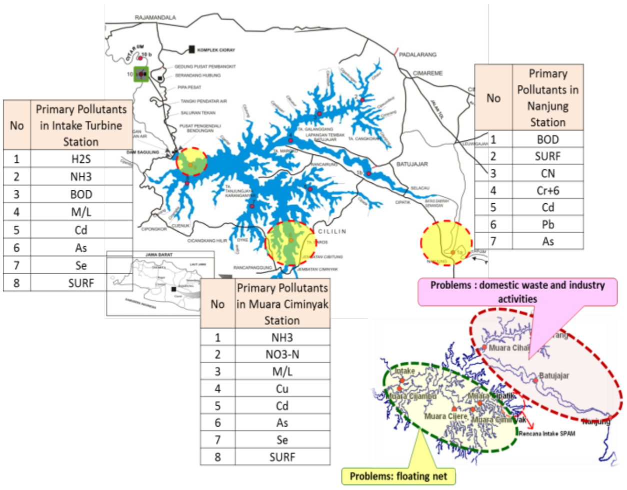

Fig 7. The Major Pollutant in Each Monitoring Post with Main Component Analysis

3.2 Main Parameter Determination in Nanjung Station

Based on Principal Component Analysis, we could see that dominant pollutant in Nanjung Station as input segment consisted of BOD (Biochemical Oxygen Demand), SURF (detergents), CN (cyanide), Cr (Chromium), Cd (cadmium), Pb (Lead), Arsenic (As).

BOD and SURF are pollutant parameters that indicate the existence of domestic waste or household waste. Cyanide, Chromium, Cadmium, Lead, and Arsenic parameters are a heavy metal group that indicate pollution from industrial activities.

3.3 Main Parameter Determination in Muara Ciminyak Station

Dominant pollutant in Muara Ciminyak Station as body reservoir segment consisted of NH3 (Ammonia), NO3-N (Nitrate as Nitrogen), M/L (oils and grease), Cu (Copper), Cd (Cadmium), As (Arsenic), Se (selenium), and SURF (detergent). The presence of NH_3 and NO_3-N were matched with North reservoir which was widely used as the location of Keramba cage (KJA) that enhance nutrients concentration.

3.3 Main Parameter Determination in Intake Turbin Station

Dominant pollutant in Intake Turbine station as output segment consisted of H_2S (hydrogen sulphide), NH_3 (Ammonia), BOD (Biochemical Oxygen Demand), M/L (Oils & Grease), Cd (cadmium), As (Arsenic), Se (selenium), and SURF (detergent). The presence of H_2S showed accumulation of various pollutants from the upstream of the reservoir that were mixed anaerobically. H_2S is a harmful chemical component in water. In natural waters with sufficient aeration (sufficient oxygen), H2S usually can not be found because the sulphide has been oxidized to sulphate (Effendi, 2003).

3.5 Determination of Dry, Normal, Wet Years using Markov Discreet Method

Distribution of research observation year (1999-2013) was conducted by Markov Discreet Method. It was divided into dry, normal, and wet years. The class divisions were performed in Nanjung, Muara Ciminyak, and Turbine Intake Stations. The class divisions in dry, normal, and wet years in each station can be seen in **Table 1**.

Tabel 1. Determination of Dry, Normal, Wet Years according to Markov Discreet Method

Distribution of Observation Year (1999-2013) into Dry, Normal, and Wet Years	Dry Years (0)	Normal Years (1)	Wet Years (2)
Nanjung Station	2003	1999	2003
	2004	2000	2004
	2006	2002	2006
	2008	2009	2008
	2011	2012	2011
Muara Ciminyak Station	2002	2000	2002
	2003	2005	2003
	2004	2007	2004
	2006	2009	2006
	2011	2012	2011
Intake Turbine Station	2002	1999	2002
	2003	2000	2003
	2004	2008	2004
	2006	2009	2006
	2011	2012	2011

3.6 Water Quality Index Method

Analysis of water quality status was conducted by comparing the observation results with the standard values (Rosemond et al., 2009; Pejman et al., 2009; Altansukh and Davaa, 2011; Arias et al., 2010). The evaluation of water quality can be done with spatial and temporal approaches [19]-[21]. Therefore, the availability of periodic and regular data monitoring (time series), was the key for successful evaluation of water quality. In addition to the historical series of data, the selection of appropriate water quality parameters was also a very important variable in the evaluation of water source quality [18], [20].

This research used Air Quality Index (Water Quality Index) approach to analyze water quality. Water quality parameters that used for the calculation of the WQI were the main pollutant parameter that was found from Principal Component Analysis in all monitoring location (BOD, SURF, CN, $Cr^{+}6$, Cd, Pb, As, Se, NH3, and NO_3-N).

WQI show the relative contamination levels to allowed water quality parameters or water quality parameter standards [18]. Water quality parameter standards used in this WQI analysis refer to Government Regulation No. 82 of 2001 on the designation of water class. Saguling is categorized as class I because it is used for drinking water in downstream cascade Citarum.

Source or basis for determining the water quality was based on WQI which was based on research by Altansukh and Davaa that conducted in Tuul River, Mongolia. Determination of water quality based on WQI value can be seen in **Table 2**.

Tabel 2. Determination of Water Quality Based on WQI Value

No	WQI Value	Water Quality		Recommendation
		Class	Status	
1	WQI ≤ 0.30	1	Very clean	No treatment necessary. Suitable for all kind of water usage.
2	0.31 ≤ WQI ≤ 0.89	2	Clean	After treatment, it can be used for drinking and agriculture. Without treatment, can be used for fishery.
3	0.90 ≤ WQI ≤ 2.49	3	Slightly polluted	Unsuitable for drinking and agriculture. If there's no choice, use it after treatment. Without treatment, can be used for livestock, recreation, and sports purposes.
4	2.50 ≤ WQI ≤ 3.99	4	Moderately polluted	Can be used for irrigation and industrial purposes after a proper treatment.
5	4.00 ≤ WQI ≤ 5.99	5	Heavily polluted	After an appropriate treatment can be used for heavy industrial activities without human body contact.
6	WQI ≥ 6.00	6	Dirty	Unsuitable for any purpose. An extensive treatment requires.

In this section, WQI value was determined at each class: dry, normal, and wet year; in Nanjung Station, Muara Ciminyak Station, and Intake Turbine Station. WQI value for each class can be seen in Figure 8.

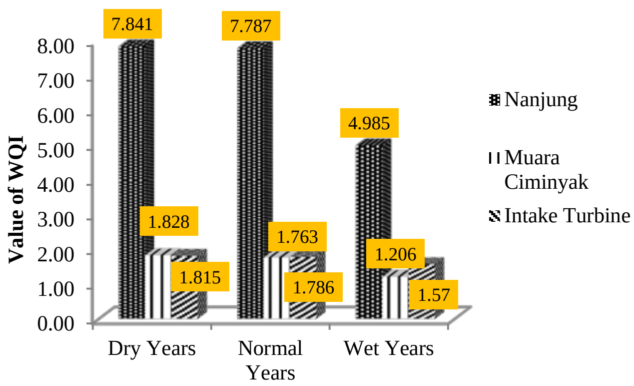

Fig 8. The score of each monitoring station in dry, normal, and wet conditions

Based on Figure 8, the highest WQI value is gained from the dry year, followed by the normal year, then, the wet years. It happened in Nanjung Station, Muara Ciminyak Station, and Intake Turbine Station. This was in accordance with the consistency of flow rate and concentration where it tend to be low in the dry season and the pollution concentration will increase. On the contrary, the reservoir flow rate increased in the rainy season, therefore the pollution concentration will decrease due to the dilution in the reservoir.

4. CONCLUSIONS

Based on Main Component Analysis in the

determination of dominant contaminants in Saguling, the result was that the pollution in Nanjung Station as input segment came from domestic activities. The sources of pollution in Muara Ciminyak Station as body reservoir segment were dominated by activities of floating nets which will increase the concentration of nutrients reservoir, shown by the presence of NH3 and NO3-N. The source of pollution in the Intake Turbine Station as output segment was the accumulation of pollution from the input of Nanjung and Saguling watersheds. It was indicated by the presence of H2S as the residue of contaminant's decay and the presence of NH3, BOD, and SURF which showed pollution from floating nets and domestic activities, and also the parameters of industrial pollution such as Cd, As, and Se.

Pollutant concentration is a function of volume and time in year or in month. The study on water quality in Saguling was conducted with discrete Markov model three class in Nanjung Station, Muara Ciminyak Station, and Intake turbine Station. The study on water pollution in Saguling with discrete approach shows the concentration of pollutants were higher during dry seasons and lower during rainy seasons. The highest contamination occurred in August due to dry season and the lowest contamination occurred in March as this month is included in the rainy season.

Based on the analysis of water quality using Water Quality Index (WQI), it was found that WQI value in Nanjung Station is 7,841 during dry year and 4,984 during wet year, in Muara Ciminyak Station was 1,828 during dry year and 1,206 during wet year, and in Intake Turbin Station was 1,815 during dry year and 1,570 during wet year. Based on WQI values obtained at each monitoring station, water quality in Nanjung Station was categorized as heavily polluted while Muara Ciminyak Station and Intake Turbin Station were categorized as lightly polluted.

Based on the hydrological analysis of Saguling Reservoir, and the reservoir's location between Asian and Australian Continent geographically with the change of two seasons which are rainy season and dry season, it is recommended that further research needs to be conducted using PCA in determining the dominant pollutant parameter in each part of the reservoir in order to determine the location of raw water intake and the calculation of WQI value of each condition, dry, normal, and wet season.

5. REFERENCES

[1]Kazi et al. (2009) : Assessment of Water Quality of Polluted Lake using Multivariate Statistical Technique : A case Study. *Ecotoxicol Environ Saf*, 72, 301-309

[2]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, 3980-3992.

[3]Aubert, A.H., Gascuel-Odoux, C., Merot, P. (2013) : Annual Hysteresis of Water Quality: A Method to Analyse The Effect of Intra and Inter-annual Climatic Condition. *Journal of Hydrology*, 478, 29-39.

[4]Lai et al. (2013) : Development of A Water Quality Modeling System for River Pollution Index and Suspended Solid Loading Evaluation. *Journal of Hydrology*, 478, 89-101

[5]Merten, G. H.; Minella, J. P. (2002) : Kualitas Air di Pedesaan: Tantangan bagi Kelangsungan Hidup Masa Depan. *Agroekologidan Pembangunan Pedesaan Berkelanjutan*, 3, 33-38

[6]Djuangsih, N. (1996) : Eutrofikasi di Waduk Saguling. Laporan Penelitian Hibah Bersaing. PPSDAL – Lembaga Penelitian Universitas Padjadjaran.

[7] Hart, et all. (2002) : Nutrient Budget for Saguling Reservoir, West Java, Indonesia. *Water Research*, 36, 2152–2160

[8] Andrade, E. M; Palacio, H. A. Q; Souza, I. H; Leao, R. A.O; Guerreiro, M. J. Land use effects in groundwater composition of an alluvial aquifer (Trussu River, Brazil) by multivariate techniques. 2008. Environmental Research, Vol 106, pp. 170-178

[9] Boyacioglu, et al. (2006) : Surface Water Quality Assessment by Environmetric Methods. *Environmental Monitoring and*

[10]Guedes, H. A. S., Silva, D. D., Elesbon, A. A. A., Ribeiro, B. M., Matos, A. T., Soares, J. H. P. (2002) : Application of Statistics Multivariate Analysis in Dove River. *Jurnal Pertanian dan Teknik Lingkungan*, 16, 558-563.

[11]Laureano and Navar. (2002). An Assessment of Stream Water Quality of the Rio San Juan, Nuevo Leon, Mexico, 1995–1996. *Journal of Environmental Quality*, 31, 1256-1265.

[12]Peterson et al. (2001). Process Identification by Principal Component Analysis of River Water-Quality Data. *Ecological Modelling*, 138, 193–213

[13]Selle et al. (2013) : Understanding Processes Governing Water Quality in Catchments Using Principal Component Scores. *Journal*

of Hydrology, 486, 31–38.

[14]Vialle, C., Sablayrolles, C., Lovera, M., Jacob, S., Huau, M. C., and Montrejaud-Vignoles, M. (2011) : Monitoring of Water Quality from Roof Runoff : Interpretation Using Multivariate Analysis. *Water Research*, 45, 3765–3775

[15]Vonberg, D., Vanderborght, J., Cremer, N., Pütz, T., Herbst, M., and Vereecken, H. (2014) : 20 Years of Long-Term Atrazine Monitoring in A shallow Aquifer in Western Germany. *Water Research*, 50, 294-306.

[16]Fan, et al. (2012) : Water Quality Management Based on Division of Dry and Wet Seasons in Pearl River Delta, China. Clean - Soil, Air, Water, 40, 381–393

[17]Sebastia, et al. (2013) : Analysis of The Effects of Wet and Dry Seasons on A Mediterranean River Basin: Consequences for Coastal Waters and Its Quality Management. Ocean & Coastal Management, 78, 45-55

[18]Altansukh and Davaa. (2011) : Application of Index Analysis to Evaluate the Water Quality of The Tuul River in Mongolia. *Journal of Water Resource and Protection*, 3, 398-414

[19]Rosemond, Simone de, Duro, Dennis C. Dan Dubé, Monique. (2009) : Comparative Analysis of Regional Water Quality in Canada using The Water Quality Index. *Environmental Monitoring Assessment*, 156, 223–240.

[20]Mahmood Q, Baig SA, Nawab B, Shafqat MN, Pervez A. (2011) : Development of Low-Cost Household Drinking Water Treatment System of The Earthquake Affected Communities in Northern Pakistan. Desalination, 273, 316-320.

[21]Mustapha, A., and Aris, A.Z. (2011). Spatial Aspect of Surface Water Quality Using Chemometric Analysis. *Journal of Applied Sciences in Environmental Sanitation*, 6, 411-426