

Identification of Water Quality Significant Parameter with Two Transformation/Standardization Methods on Principal Component Analysis and Scilab Software

Jovan Putranda

Department of Civil and Environmental Engineering, Universitas Gadjah Mada, Yogyakarta, INDONESIA
jovan_putranda@yahoo.co.id

Sri Puji Saraswati

Department of Civil and Environmental Engineering, Universitas Gadjah Mada, Yogyakarta, INDONESIA
spswatinz@ugm.ac.id

ABSTRACT

Water quality monitoring is prone to encounter error on its recording or measuring process. The monitoring on river water quality not only aims to recognize the water quality dynamic, but also to evaluate the data to create river management policy and water pollution in order to maintain the continuity of human health/sanitation requirement, and biodiversity preservation. Evaluation on water quality monitoring needs to be started by identifying the important water quality parameter. This research objected to identify the significant parameters by using two transformation/standardization methods on water quality data, which are the river Water Quality Index, WQI (*Indeks Kualitas Air, Sungai, IKAs*) transformation/standardization method and transformation/standardization method with mean 0 and variance 1; so that the variability of water quality parameters could be aggregated with one another. Both of the methods were applied on the water quality monitoring data which its validity and reliability have been tested. The PCA, Principal Component Analysis (*Analisa Komponen Utama, AKU*), with the help of Scilab software, has been used to process the secondary data on water quality parameters of Gadjah Wong river in 2004-2013, with its validity and reliability has been tested. The Scilab result was cross examined with the result from the Excel-based Biplot Add In software. The research result showed that only 18 from total 35 water quality parameters that have passable data quality. The two transformation/standardization data methods gave different significant parameter type and amount result. On the transformation/standardization mean 0 variances 1, there were water quality significant parameter dynamic to mean concentration of each water quality parameters, which are TDS, SO₄, EC, TSS, NO₃N, COD, BOD₅, Grease Oil and NH₃N. On the river WQI transformation/standardization, the water quality significant parameter showed the level of Gadjah Wong River pollution, which are EC, DO, BOD₅, COD, NH₃N, Fecal Coliform, and Total Coliform. These seven parameters is the minimal amount of water quality parameters that has to be consistently measured on predetermined time and location, and also become the indicator of human health and environment health quality. The result of Scilab multivariate analysis was not different with the result from Biplot Add In multivariate analysis, in which the results of water quality significant parameter has been verified with bio-monitoring.

Keywords: water quality monitoring, transformation/standardization, Scilab, Principal Component Analysis.

1 INTRODUCTION

The water quality monitoring is aimed to recognize the water quality dynamic, in order to maintain the continuity of human health/sanitation requirement, and biodiversity preservation (Karr, 1991). The Gadjah Wong River and its tributaries in Special Region of Yogyakarta have been long monitored by its regional Environmental Agency since the declaration of the *Prokasih* Program (Clean River Monitoring Program) in 1995, with at least 35 water quality parameters every 3 or 4 times every year in 13 locations (BLH, 2013). The water quality data is very dynamic, prone to measurement and recording mistakes (Berthouex and Brown, 2002). Therefore evaluation is needed, started with identifying

parameters of water quality that significantly impact the human health and river environment health. Significant parameter becomes the parameter type that needed to be measured consistently in every spatial monitoring, by periodically and constantly if expense limitation and laboratory facility are a hindrance. The quality of water quality monitoring data needed to be maintained since pre-sampling, at the time of sampling, and post-sampling (Resh and McElravy, 1993).

The water quality data has different unit and measurement procedure dealing with the difficulties in evaluating the water quality parameters; therefore it needs the data transformation/standardization method (Dillon and Goldstein, 1984). This research used two

transformation/standardization method, which is the transformation/standardization method with mean 0 variance 1 (McBridge, 2005), and the river WQA transformation/standardization method which has been developed earlier by Saraswati (2015), after previously being cross examined with another transformation/standardization method (Cao et al., 1999). The water quality data transformation/standardization methods that were being used were very affected the multivariate analysis result on water quality; therefore it has the ecological relevance.

The multivariate analysis for water quality data has been used previously by Zang et al. (2009), Zhou et al. (2006), and Fataei (2011). The Principal Component Analysis as one of the multivariate analysis is used to reduce the huge amount of complex data variables into several values representing the entire data variable set, yet still, maintain the character of the data. The river pollution problem needs plenty variables to monitor, including its water quality. The variables are interdependent and usually correlated with each other, yet in the statistical analysis, all variables must be random, and independent with each other. Therefore, the evaluation and monitoring of water quality parameter used the multivariate statistical analysis (Putranda, 2015).

There were an excessive amount of water quality data that would be processed; therefore this research needed the help from computer software, which is Scilab (Baudin, 2010). The Principal Component Analysis used two transformation/standardization methods on the secondary data of water quality in 2004-2013, then aided by Biplot Add In (Lipkovich and Smith, E.P., 2002), being cross examined with result of multivariate analysis on water quality data of 1997-2012, which the biomonitoring result, ex-situ and in-situ, were verified (Saraswati, 2015).

2 DATA AND METHODOLOGY

2.1 Data Set

This research used the secondary data of water quality of Gadjah Wong River, which is the monitoring result of the Environmental Agency of Yogyakarta in 2004-2013 (BLH, 2010). In the period of the 10 years, there were total 35 water quality parameters that have been measured in 9 monitoring locations in the main river, which are Tanen Bridge, Pelang Bridge, IAIN Bridge, Muja-Muju Bridge, Rejowinangun Bridge, Tegalgendu Bridge, Tritunggal Bridge, Wirokerten Bridge, and Wonokromo Bridge (see Figure 1).

The dry season starts from May, while the rainy season begins in November. (Jovan, 2015).

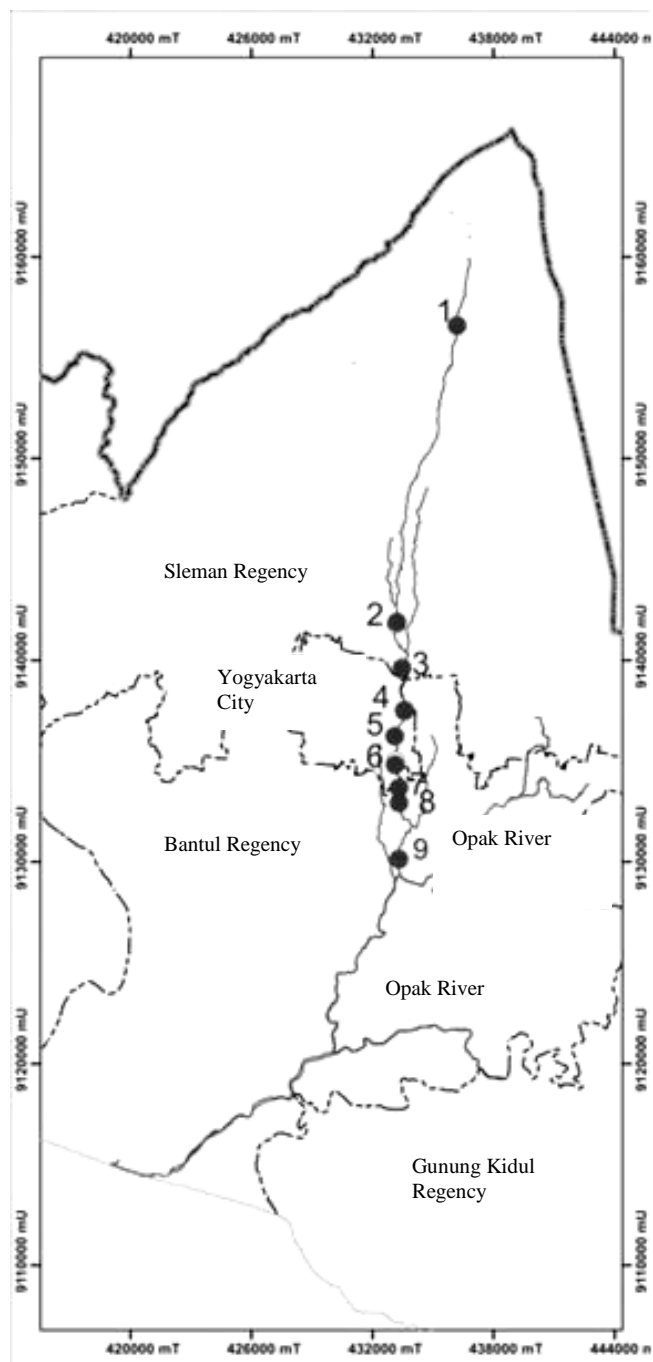


Figure 1. Monitoring location of water quality in Gadjah Wong River (indicated by black dots)

2.2 Data Transformation/standardization

There are two types of transformation/standardization that used in the water quality data analysis.

2.2.1 Standardization mean 0 variance 1

This standardization is the most commonly used in the statistical data processing. This standardization changed every water quality variables until it has

mean 0 and variance 1. The equation of this transformation/standardization is as follows,

$$y_i = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

whereas y_i is result of transformation/standardization variable to i , x_i is water quality variable concentration to i , \bar{x} is data mean, σ is standard deviation.

2.2.2 River WQI Standardization

The river WQI standardization proposed by Saraswati (2015) changed every water quality variable that has score limit of -1 to +1. Score (-) stated the water quality variable is polluted, and score (+) stated that the variable is in good condition or not polluted. Whereas the value of 0 is the score of every water quality variable if the concentration is equal the standard water quality. The river WQI standardization distinguished the water quality variable type according to the water quality variable.

If the water quality decline because of the increasing pollution (for example, DO),

$$y_i = \frac{x_i - Stan_i}{c_i} \quad (2)$$

If the water quality increase because of the increasing pollution (for example, BOD5, COD, and others),

$$y_i = \frac{Stan_i - x_i}{c_i} \quad (3)$$

whereas y_i is variable transformation/standardization result to i , x_i is water quality variable concentration to i , $Stan_i$ is quality standard of water quality variable i , c_i is maximum range, theoretical concentration to quality standard = (theoretical maximum concentration – quality standard) water quality variable to i

If the standard “ $Stan_i$ ” has interval limit (for example, pH),

For $x_i < Stan_{i,mean}$ then,

$$y_i = \frac{x_i - Stan_{i,min}}{Stan_{i,mean}} \quad (4)$$

For $x_i > Stan_{i,mean}$ then,

$$y_i = \frac{Stan_{i,max} - x_i}{Stan_{i,mean}} \quad (5)$$

whereas y_i is result of transformation/standardization variable to i , x_i is water quality variable concentration to i , $Stan_i$ is quality standard of water quality variable to i , $Stan_{i,mean}$ is (maximum concentration of water quality standard + minimum concentration of water quality standard)/2, $Stan_{i,max}$ is maximum concentration of water quality standard span, $Stan_{i,min}$ is minimum concentration of water quality standard span, y_i will have transformation/standardization result interval between (-)1 and (+)1; whereas (-) means pollution has occurred, and (+) means the water quality is good, while 0 is the concentration of water quality equals the concentration of standard water quality.

2.2.3 Standard Water Quality of River

Healthy River can be described from the quality of its water, which is not polluted and not toxic for its biota. Healthy water is the background condition of river water (Lumb et al., 2006). The benchmark of water quality conservation was arranged according to references, local condition on the river water, and bioassay result that ever existed (Saraswati, 2015). In Table 1, the benchmark of water quality conservation is shown with the water quality standard of class I water body, standard water for drink, according to the Government Regulation No.82 Year 2001 on Water Monitoring and Water Pollution Controlling.

2.3 Reliability Test and Validity Test of the Water Quality

Data on the water quality monitoring need to be maintained its data quality assurance in order to be able to be processed further, because it may affect the conclusion on the water quality data which does not depict the real condition on the field. The reliability test and data smoothing were conducted on the raw monitoring data, as in the missing value, censored data, and outlier data, on a parameter, and between the water quality parameters (Saraswati et al., 2013).

2.4 Principal Component Analysis

A significant variable in the PCA is water quality parameter that has a dominant impact on the criteria of water quality condition dynamic data. Determining the parameter was by observing the eigenvector value of component 1 and component 2 on the Principal Component Analysis (Smith, 2002). The chosen parameters were decided with the requirement of the component loading value > 0.5 (Hair et al., 2009); which showed that the component has able to

represent the component analysis in the significantly trusted level. The eigenvector is a matrix that shows multiplier coefficient from origin variable into a PCA score on certain main component. Eigenvalue (explained variance) is the coefficient number that depicts total variances that are explained by each component of the identity matrix (Legendre and Legendre, 1998). The basic value of eigenvalue was used in determining the total number of main component, which is the new variable of water quality.

2.5 Scilab Software

Scilab is a numerical computational package that has been developed since 1990 by the researcher from

INRIA and ENPC. Scilab has functional similarity with MATLAB, yet it is available to be downloaded without license fee (open source). As non-licensed software, Scilab can be used for various Operating Systems (OS), it is easy to preview and modify source code, distributing source code, and used the software for various purposes (Annigeri, 2004). Input data in Scilab is conducted by writing script in the console menu. By writing new function for PCA analysis made Scilab could work. Furthermore, the Biplot Add In is an additional program on Microsoft Excel (Lipkovich and Smith, 2002) that is used for multivariate analysis.

Table 1. Benchmark of conservation on river water quality

No	Water Quality Parameter	Standard Class I, PP No. 82 year 2001	Benchmark on Water Quality Conservation	Min and Max (Theoretical)	ci
1	TSS (mg/L)	50	50	0.001 and ≥ 400	350
2	TDS (mg/L)	1000	1000	1 and ≥ 20000	19000
3	DHL or EC (uS/cm)	-	300	100 and ≥ 700	400
4	pH	6 - 9	6 - 9	6 - 9	7.5
5	DO (mg/L)	6	6	0 and ≥ 10	6
6	BOD5 (mg/L)	2	3	0.5 and ≥ 50	47
7	COD (mg/L)	10	10	5 and ≥ 100	90
8	NO ₂ (mg/L)	10	2	0.01 and ≥ 100	98
9	NO ₃ (mg/L)	0.06	0.02	0.001 and ≥ 1	0.98
10	NH ₃ N (mg/L)	0.5	0.05	0.01 and $\geq 1,25$	1.2
11	SO ₄ (mg/L)	400	75	25 \leq and ≥ 1000	925
12	H ₂ S (mg/L)	0.002	0.002	0.001 and ≥ 0.30	0.298
13	Phosphate (mg/L)	0.2	0.2	0.002 and ≥ 50	49.8
14	Detergent (mg/L)	0.2	0.2	0.001 and ≥ 3.0	2.8
15	Grease & Oil (μ g/L)	1000	20	0 and ≥ 20000	19980
16	Fluoride (mg/L)	0.5	1.2	0.009 and ≥ 6	4.8
17	Hexavalent Chromium (mg/L)	0.05	0.05	0.001 and ≥ 0.25	0.2
18	Fecal Coliform (MPN/100mL)	100	4000	1 and ≥ 20000	16000
19	Total Coliform (MPN/100mL)	1000	20000	5 and ≥ 50000	30000

Table 3. Characteristic of observation data of Gadjah Wong River water quality, 2004-2013

Location	Quality Standard	TDS (mg/L)	TSS (mg/L)	EC ($\mu\text{s/cm}$)	pH	DO (mg/L)	BOD5 (mg/L)	COD (mg/L)	NO ₃ -N (mg/L)	NO ₂ (mg/L)	NH ₃ -N (mg/L)	PO ₄ -P (mg/L)	Fluoride (mg/L)	Sulphate (mg/L)	H ₂ S (mg/L)	Cr ₆₊ (mg/L)	Oil & Grease (mg/L)	Detergent ($\mu\text{g/L}$)	Fecal Coliform (MPN/100mL)	Total Coliform (MPN/100mL)
		1000	50	300	6 - 9	6	3	10	2	0.02	0.05	0.2	1.2	75	0.002	0.05	20	200	4000	20000
1	mean	109.9	20.7	210.8	7.10	6.81	5.95	18.95	2.56	0.02	0.17	0.18	0.49	14.5	0.01	0.00	327	67.20	114461	463569
	SD	25.4	14.5	32.6	0.52	1.05	3.19	12.31	2.73	0.02	0.16	0.19	0.40	6.7	0.01	0.00	452	71.25	208198	756627
	min	37.0	7.0	96.5	6.00	3.90	1.20	3.66	0.01	0.00	0.01	0.01	0.00	0.5	0.00	0.00	0	0.10	4000	7000
	max	152.0	74.0	270	8.07	8.30	15	58	12	0.08	0.70	0.92	1.44	33.0	0.03	0.02	1700	222.80	1100000	2400000
2	mean	159.0	26.3	296.0	7.01	6.06	5.56	21.10	3.05	0.02	0.15	0.28	0.57	20.8	0.01	0.01	1085	72.78	205636	734605
	SD	44.9	16.6	78.4	0.40	1.22	2.90	16.17	2.37	0.02	0.16	0.20	0.44	9.3	0.01	0.01	706	60.24	288930	837515
	min	59.0	2.0	90.0	6.00	3.20	1.60	3.45	0.08	0.00	0.00	0.00	0.00	4.3	0.00	0.00	0	0.10	4000	9000
	max	308.0	68.0	624.6	7.80	7.84	13.09	66	10.6	0.10	0.61	0.79	1.56	47.6	0.03	0.03	2500	247.60	1100000	2400000
3	mean	172.6	37.4	346.0	6.97	6.07	7.04	22.85	3.77	0.04	0.33	0.32	0.53	23.4	0.01	0.01	1587	105.42	326842	1091934
	SD	50.3	22.2	84.4	0.28	1.23	3.54	13.2	2.83	0.04	0.18	0.29	0.39	9.7	0.01	0.01	1041	72.03	366184	838817
	min	6.2	4.0	121	6.40	2.60	1.75	3.45	0.64	0.00	0.06	0.02	0.00	6.1	0.00	0.00	18	0.01	9000	9000
	max	252.0	76.0	511	7.51	8.00	14.78	60	11	0.14	0.84	1.10	1.31	41.2	0.03	0.03	4000	261.90	1100000	2400000
4	mean	202.5	36.9	372.2	7.02	5.89	7.15	24.94	4.29	0.04	0.34	0.43	0.52	28.7	0.01	0.01	1604	106.96	309797	834184
	SD	58.7	19.3	75.7	0.30	1.22	3.87	15.70	3.06	0.03	0.18	0.33	0.37	12.3	0.01	0.01	1086	76.95	373431	791706
	min	112.0	6.0	234	6.60	2.90	1.75	3.45	0.01	0.00	0.03	0.01	0.00	9.6	0.00	0.00	32	0.10	700	11000
	max	338.0	78.2	511.0	7.70	7.47	15.00	68	11.9	0.11	0.76	1.23	1.57	55.0	0.03	0.03	4000	271.60	1100000	2400000
5	mean	223.9	35.6	409.7	7.06	5.70	8.02	24.82	3.60	0.04	0.40	0.40	0.55	33.2	0.01	0.01	1347	82.50	272079	1014197
	SD	68.8	17.8	85.1	0.29	1.36	4.45	16.27	2.51	0.03	0.22	0.31	0.45	13.3	0.01	0.01	907	66.30	318644	921550
	min	119.0	6.0	252.0	6.40	2.50	1.98	3.45	0.17	0.00	0.05	0.01	0.00	12.0	0.00	0.00	0	0.10	4000	9000
	max	386.0	84.0	586.0	7.80	9.00	17.50	58	9.49	0.11	0.85	1.20	1.78	61.2	0.03	0.03	3000	241.80	1100000	2400000

Table 3. Characteristic of data on Gadjah Wong river water quality, observation of 2004-2013 (continued)

Location	Quality Standard	TDS (mg/L)	TSS (mg/L)	EC ($\mu\text{s/cm}$)	pH	DO (mg/L)	BOD5 (mg/L)	COD (mg/L)	NO ₃ -N (mg/L)	NO ₂ (mg/L)	NH ₃ -N (mg/L)	PO ₄ -P (mg/L)	Fluoride (mg/L)	Sulphate (mg/L)	H ₂ S (mg/L)	Cr ₆₊ (mg/L)	Oil & Grease (mg/L)	Detergent ($\mu\text{g/L}$)	Fecal Coliform (MPN/100mL)	Total Coliform (MPN/100mL)
		1000	50	300	6 - 9	6	3	10	2	0.02	0.05	0.2	1.2	75	0.002	0.05	20	200	4000	20000
6	mean	194.0	39.9	381.5	7.07	5.99	6.88	25.27	4.36	0.04	0.33	0.44	0.51	28.9	0.01	0.01	1069	73.78	214637	687033
	SD	44.0	23.5	60.0	0.46	1.14	3.84	16.32	3.12	0.03	0.16	0.30	0.46	10.2	0.01	0.01	715	64.21	298141	815322
	min	127.0	6.0	269	6.00	3.50	1.36	3.66	0.20	0.00	0.06	0.06	0.00	10.9	0.00	0.00	0	0.00	1100	9000
	max	342.0	90.0	520	8.00	8.00	14	60	11.98	0.12	0.81	1.10	1.50	53.5	0.03	0.03	3000	211.40	1100000	2400000
7	mean	219.2	39.2	386.6	7.08	6.24	7.25	27.87	3.97	0.07	0.37	0.41	0.65	32.0	0.01	0.01	1057	92.10	212997	670296
	SD	70.0	19.2	73.0	0.30	1.24	4.62	18.01	3.31	0.04	0.22	0.27	0.47	13.5	0.01	0.01	1127	70.10	221206	749381
	min	125.0	16.0	265.0	6.38	3.20	1.36	3.45	0.14	0.00	0.03	0.02	0.00	9.7	0.00	0.00	0	1.00	7300	7300
	max	380.0	94.0	575.0	7.50	9.50	16	65	11.56	0.14	0.82	1.05	1.69	60.1	0.03	0.03	4000	228.00	1100000	2400000
8	mean	218.5	36.6	401.2	7.10	5.86	6.58	23.90	3.84	0.05	0.29	0.39	0.56	31.5	0.01	0.01	1109	67.12	246066	660289
	SD	61.8	18.9	68.9	0.34	1.16	3.84	16.58	2.75	0.03	0.18	0.30	0.39	12.1	0.01	0.01	1022	59.39	292556	782294
	min	128.0	6.0	272.0	6.35	2.90	1.68	3.45	0.59	0.00	0.00	0.02	0.00	12.7	0.00	0.00	0	0.10	7000	20000
	max	363.0	78.0	551.0	7.70	7.50	15	70	11.23	0.13	0.78	1.12	1.57	57.2	0.03	0.03	4000	220.40	1100000	2400000
9	mean	224.1	42.4	413.1	7.19	6.60	6.01	25.44	3.27	0.04	0.22	0.39	0.53	31.8	0.01	0.01	1195	63.21	237807	627474
	SD	62.1	23.7	68.8	0.28	0.99	3.32	17	2.46	0.03	0.13	0.32	0.39	13.3	0.01	0.01	992	52.65	326885	688270
	min	133.0	6.0	263.0	6.50	4.60	1.75	3.45	0.45	0.00	0.03	0.00	0.00	0.2	0.00	0.00	0	0.00	1500	9000
	max	353.0	96.0	535.0	7.80	8.50	12.70	69	9.44	0.12	0.62	1.19	1.60	55.4	0.03	0.03	4000	192.00	1100000	2400000

3 ANALYSIS RESULT AND DISCUSSION

3.1 Reliability and Validity of Water Quality Data

The locations picked for this research were 9 locations that are located in the main river of Gadjah Wong, also because the measurement data was consistently available. The condition on the data of the 19 chosen water quality parameters for the next processing can be seen in Table 2.

Table 2. The quality of data on water quality parameter of Gadjah Wong River

Data condition	Good Quality Data	Aberrant Data (Outlier)	Censored Data	Missing Value Data
TDS	98%	2%	0%	0%
TSS	95%	5%	0%	0%
DHL or EC	77%	2%	0%	21%
pH	96%	4%	0%	0%
DO	95%	5%	0%	0%
BOD5	91%	4%	5%	0%
COD	95%	2%	3%	0%
NO3-N	98%	1%	1%	0%
NO2	72%	14%	2%	11%
NH3-N	71%	8%	0%	21%
PO43-Flourida	94%	2%	4%	0%
SO4	76%	1%	12%	11%
H2S	59%	1%	0%	40%
H2S	73%	7%	9%	11%
Cr6+	59%	7%	34%	0%
Oil & Grease	83%	6%	0%	11%
Detergent	69%	7%	24%	0%
Fecal Coliform	92%	8%	0%	0%
Total Coliform	80%	3%	16%	0%

Note: GQD = Good Quality Data, AD = Aberrant Data (data outlier), CD = Censored Data, MVD = Missing Value Data

The KMO value and Bartlett's test on Gadjah Wong River were 0.671 and significant of 0. With KMO value above 0.5 and significant below 0.05, this value already meet the requirement for a further analysis. In Gadjah Wong River, it was the PO43- parameter that was issued for further data processing, since it has value of $MSA < 0.5$. The characteristic of water quality data resulted from smoothing is shown in Table 3.

3.2 Water Quality Significant Parameter

3.2.1 Transformation/standardization Method

Table 4 showed the eigenvalue of PCA with transformation/standardization mean 0 variance 1 and transformation/standardization of 18 water quality

parameters. The main component used was 2 components, because of its highest eigenvalue value. The component loadings value of each water quality parameters on both chosen main components are shown in Table 4.

Table 4. Eigenvalue from PCA with transformation/standardization mean 0 variance 1, and river WQI transformation/standardization

Main Component	Transformation/standardization Method	
	mean 0 variance 1	river WQI
Comp.1	3.76	25.41
Comp.2	2.45	18.82
Comp.3	1.66	10.04
Comp.4	1.36	8.77
Comp.5	1.10	6.28
Comp.6	1.03	3.90
Comp.7	0.96	1.94
Comp.8	0.85	1.59
Comp.9	0.80	0.93
Comp.10	0.71	0.61
Comp.11	0.68	0.49
Comp.12	0.64	0.39
Comp.13	0.58	0.29
Comp.14	0.47	0.18
Comp.15	0.41	0.16
Comp.16	0.35	0.13
Comp.17	0.15	0.02
Comp.18	0.05	0.00
Total Eigenvalue	18	79.95

Those results were used to determine the quality parameter that was considered important or significant on the PCA analysis with result data of transformation/standardization mean 0 variance 1 method and river WQI transformation/standardization.

The cross examination result of both methods showed that the eigenvalue value and component loadings of transformation/standardization mean 0 variance 1 method was different from the result of river WQI transformation/standardization method. Each of it produced different total amount and water quality significant parameter type.

On the PCA with transformation/standardization mean 0 variance 1 method, the total amount of eigenvalue value for Gadjah Wong River was equal to its variable amount, meanwhile, the river WQI transformation/standardization method resulted the total amount of eigenvalue value was equal to the total amount of each variable's variance value.

Based on the component loading value requirement, which is more than 0.5, the significant parameters in the Gadjah Wong River according to the transformation/standardization mean 0 variance 1 method was 9 variables, on the Component 1 are TDS (0.872), EC (0.773), BOD5 (0.536), COD (0.619), NH3N (0.507), and SO4 (0.862); and on the Component 2 are TSS (0.740), NO3N (0.688), and Oil & Grease (0.578). From the river WQI transformation/standardization method, the water quality significant parameter was 7 variables, on Component 1 are the EC (2.907), DO (2.505), BOD5 (0.559), COD (1.676), NH3N (1.433), Fecal Coliform (1.559) and Total Coliform (1.675); while on the Component 2 are EC (-2.393), NH3N (-0.783), Fecal Coliform (2.471), and Total Coliform (2.495). The differences between both transformation/standardization methods that were shown by the commonality level are shown in Table 5, as follows:

Table 5. Commonality value, transformation/standardization mean 0 variance 1 and river WQI transformation/standardization

Water Quality Parameter	Transformation/standardization method	
	mean 0 variance 1	River WQI
TDS	1	0.0038
TSS	1	1.0991
EC or DHL	1	17.6735
pH	1	0.5615
DO	1	13.1720
BOD5	1	2.0530
COD	1	9.7778
NO3-N	1	0.2609
NO2	1	0.3779
NH3-N	1	8.4044
PO43-		
Fluoride	1	2.3652
SO4	1	0.0603
H2S	1	0.1988
Cr6+	1	0.4152
Oil & Grease	1	0.7606
Detergent	1	0.1840
Fecal Coliform	1	10.9376
Total Coliform	1	11.6474

The commonality value is the sum of squared factors of the component loadings value. As seen in Table 5 on the transformation/standardization mean 0 variance 1 method, the commonality value for each variable was equal 1, while on the river WQI transformation/standardization method, the commonality value for each variable was equal to the variable variance value.

The Figure 2 (a) shows the scree plot result of PCA multivariate analysis using the Scilab software on the secondary data of water quality of 2004-2013, with

transformation/standardization mean 0 variance 1 method; and the Figure 2 (b) is the result analysis from Biplot Add In with the same transformation/standardization method and on the same year data. The center point of the graphic is the mean value of concentration of each water quality parameters which has been transformed/ standardized, as the equation (1) showed.

Based on the PCA scree plot graphic, 9 water quality parameters deviate significantly from its center point, based on weight to the component 1 and 2 consecutively are TDS, SO4, EC, TSS, NO3N, COD, BOD5, Grease and Oil, and NH3N. The deviation was measured from the measured data difference to bench mark, in form of the mean value of concentration of each water quality parameters. The Large deviation is not necessarily meant that the water quality is "polluted", for the meaning of "polluted" water is if the measured concentration has deviated from the bench mark of the water health quality. The significant weight in consecutively is TDS, TSS, EC, BOD5, COD, NH3N, SO4, NO3N, Grease and Oil, showed by its vector direction and length which moved away from the center point.

Scree plot result of the PCA used the river WQI transformation/standardization method; secondary data of 2004-2013 is shown in Figure 3 (a), while the Figure 3 (b) showed the result analysis of Biplot Add In of the same secondary data. Scree plot result of PCA analysis with Scilab on 2004-2013 data showed the same result with scree plot figure on Biplot Add In result analysis on secondary data of Gadjah Wong River water quality in 1997-2012 (Saraswati, 2015).

As for the status of Gadjah Wong River water quality, which was measured with water quality index by using the significant parameter that was resulted from river WQI transformation/standardization, this has been confirmed with the water quality status that was resulted from biomonitoring on 2012 by Saraswati (2015). Therefore it can be concluded that 7 water quality significant parameters, the EC, DO, BOD5, COD, NH3N, Fecal Coliform and Total Coliform, is the water quality parameters that most influential to the dynamic of "pollution" level on the Gadjah Wong river water. These parameters were significantly deviated away from the bench mark (Table 1) of water quality concentration with chemical physic Bacteriology parameter on the river water health.

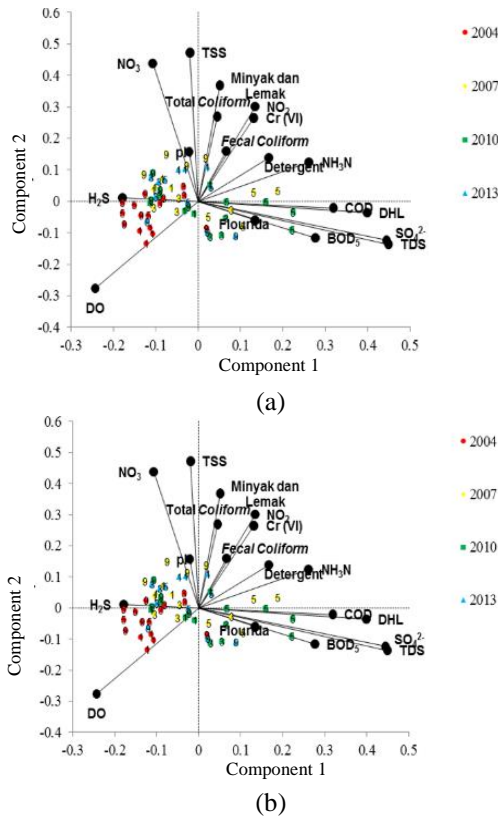


Figure 2. (a)Screeplot of PCA with river WQI transformation/standardization Gadjah Wong River. Scilab; (b)Biplot Add In

Such as the equation (2) - (5), the scree plot center point of Figure 3.a and Figure 3.b are the water quality bench mark value, with the water health criteria that very considers the impact to the biotic aspect in the river. This water quality parameter is the total amount and minimal water quality parameter type which must be measured consistently in its observation field; monitor the water pollution that caused by hydro-climatology natural change, domestic and industry activities. This water quality parameter can detect the impact on human health by using the water quality indicators, Fecal Coliform, and Total Coliform, and detect the impact on water environment health through the indicator of the water quality parameter, DHL or EC, DO, BOD5, COD, NH3N. Based on the scree plot result, the bacteriology indicator showed that the sanitation condition in the river is increasingly worse, compared to that of in 2004. According to the parameter indicators, COD, DO, EC, and NH3N, further the downstream the water quality is increasingly worse, compared to in the upstream; yet the river middle segment (location 3 and 4) in Yogya City is the most polluted one, caused by the organic and inorganic wastes. There is an indication that the pollution moves to upstream, with the increasing settlement and domestic activity in the location.

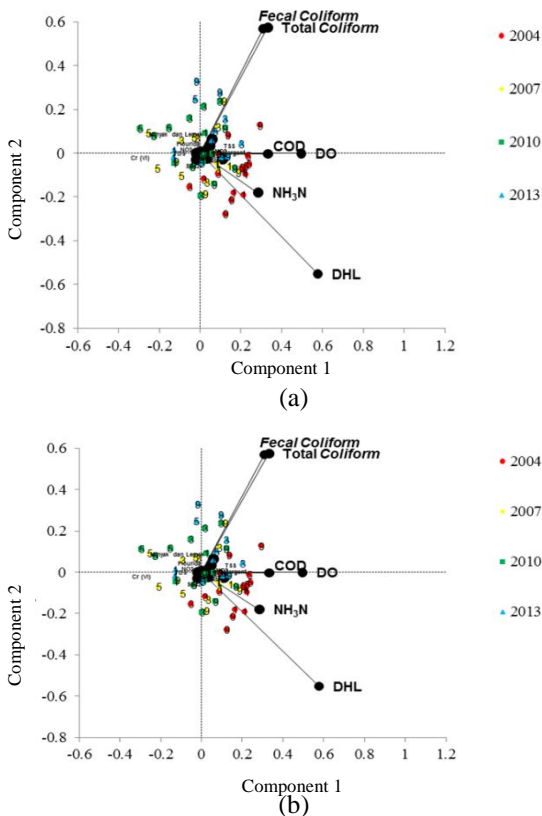


Figure 3. (a)Screen plot of PCA result with river WQI transformation/standardization in Gadjah Wong River. Scilab; (b)Biplot Add In

The scree plot on Figure 3a and 3b, using the water healthy bench mark, the total parameters that significantly deviate were only 7. The significant weight consecutively is EC, DO, COD, Total Coliform, Fecal Coliform, NH3N, and BOD5 as the smallest one. It is shown by each parameter vector length with its direction moving away from the graphic's center point.

4 CONCLUSION

Several conclusions from the research result are as follows,

- a) There were only 18 parameters that are considered reliable and valid, from the 35 water quality parameters that were monitored in 2004-2013.
- b) TDS, SO4, EC, TSS, NO3N, COD, BOD5, Grease Oil and NH3N were the 9 significant parameters that able to explain the dynamic of water quality concentration on each mean concentration of the water quality parameters. These nine water quality parameters did not explain the dynamic of river water pollution level.
- c) EC, DO, BOD5, COD, NH3N, Fecal Coliform, and Total Coliform were the 7 significant parameters that impact the dynamic of Gadjah

Wong river water pollution level. The water quality parameters, Fecal Coliform, and Total Coliform were to detect the human health/sanitation condition, while the EC or DHL, DO, BOD5, COD, and NH3N were to monitor the water healthy environment, which is affected by the hydro-climatology condition change, organic/inorganic waste from the domestic activity, industrial activity, and others. These water quality parameters are needed to be consistently monitored in the entire monitor field.

- d) The PCA study with two transformation/standardization methods on the 2004-2013 data resulted on the same water quality significant parameters with the PCA study on secondary data from 1997-2012 monitoring.
- e) The scree plot result in the Gadjah Wong River according to bacteriology parameter indicator showed increasing trend of higher water quality pollution in each year. Using the parameter indicator of COD, DO, EC, and NH3N, the water quality is worsening in the further downstream, yet the river middle segment in Yogya City is the most polluted. The tendency is moving to the upstream, result from the increasing settlement development and domestic activity in the locations.
- f) Scilab was proven to be quite effective as a statistical data processing tool, because of the available function for PCA calculation on the software, and the easily applied programming language.
- g) The result from multivariate analysis method using the Scilab tool was not different with the result of multivariate analysis using the Biplot Add In.

REFERENCES

Annigeri, S. (2004). *Scilab A Hands on Introduction*. Bhoomaraddi: Department of Civil Engineering B.V., Bhoomaraddi College of Engineering and Technology.

Badan Lingkungan Hidup (BLH) DIY. (2010). *Laporan Analisa Data Kualitas Air Sungai di Provinsi Daerah Istimewa Yogyakarta [Analysis of Water Quality Data of Rivers in Yogyakarta Province]*. Yogyakarta: Badan Lingkungan Hidup (BLH) DIY.

Baudin, M. (2010). *Introduction to Scilab*. The Scilab Consortium.

Berthouex, P. M., & Brown, L. C. (2002). *Statistics for Environmental Engineers*. Washington: Lewis Publishers.

Cao, Y., Williams, D., & Williams, N. E. (1999). Data Transformation and Standardization in the Multivariate Analysis of River Water Quality. *Journal Ecological Applications*, 669-77.

Dillom, W. R., & Goldstein, M. (1984). *Multivariate Analysis*. UK: John Wiley and Sons.

Dillon, W. R. (1984). *Multivariate Analysis Methods and Applications*. UK: John Wiley and Sons.

Fataei, E. (2011). Assessment of Surface Water Quality Using Principle Component Analysis and Factor Analysis. *World Journal of Fish and Marine Sciences*, 159-66.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. (2009). *Multivariate Data Analysis*. Pearson Prentice Hall.

Karr, J. R. (1991). Biological Integrity: A Long-Neglected Aspect of Water Resources Management, *Ecological Applications*, 1(1), 66-84.

Legendre, L., & Legendre, P. (1998). *Numerical Ecology*. Amsterdam: Elsevier Scientific Publishing Company.

Lipkovich, L., & Smith, E. P. (2002). *Biplot and Singular Decomposition Macros for Excel'*. Blacksburg: Department of Statistics Virginia Tech.

Lumb, A., Halliwell, D., & Sharma, T. (2006). Application of CCME Water Quality Index to Monitor Water Quality: A Case of the Mackenzie River Basin Canada. *Environment Monitoring and Assessment*, 113, 411-429.

McBride, G. B. (2005). *Using Statistical Methods for Water Quality Management*. John Wiley & Sons.

Pemerintah Republik Indonesia. (2001). *Peraturan Pemerintah Republik Indonesia Nomor 82 Tahun 2001 Tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air*. Jakarta: Pemerintah Republik Indonesia.

Putranda, J. (2015). *Kajian Kualitas Air Sungai Dengan 2 Teknik Standarisasi Menggunakan Scilab Dan Biplot Add In [Water Quality Study with 2 Standardization Techniques using Scilab and Biplot Add In]*. Yogyakarta: Undergraduate Thesis, Department of Civil and Environmental Engineering, Universitas Gadjah Mada.

Resh, V. H., & Mc Elravy, E. P. (1993). *Contemporary Quantitative Approaches to Biomonitoring using Benthic Macroinvertebrates. Pada D.M. Roosenberg and V.M. Resh. Freshwater Biomonitoring and Benthic Macroinvertebrates*. New York: Chapman and Hall.

Saraswati, S. P. (2015). *Pengembangan Metode Penentuan Kesehatan Perairan Sungai di Daerah Tropis Berbasis Ekohidraulik [Development of Water Health Determination Method in Tropical Area Ecohydraulic Based]*. Yogyakarta: Doctoral Dissertation, Department of Civil and Environmental Engineering, Universitas Gadjah Mada.

Smith, L. I. (2002). *A Tutorial on Principle Component Analysis*.

Suryanto, S. (2015). *Analisa Data Kualitas Air Sungai Menggunakan Metode Statistik Multivariate Principle*

Component Analysis dengan 2 Teknik Standarisasi [Water Quality Data Analysis using Multivariate Principle Component Analysis with 2 Standardization Techniques]. Yogyakarta: Undergraduate Thesis, Department of Civil and Environmental Engineering Department, Universitas Gadjah Mada.

Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., & Yuan. (2009). Assessment of Surface Water Quality using Multivariate Statistical Techniques in Red Soil Hilly Region: A Case Study of Xiangjiang Watershed, China. *Environment Monitoring Assesment*, 123–131.

Zhou, F., Li, Y., & Guo, H. (2006). Application of Multivariate Statistical Methods to Water Quality Assessment of the Water Courses in Northwestern New Territories, Hongkong. *Environmental Monitoring Assessment, Springer Science, published online*.

[this page intentionally left blank]