Spatial and temporal variation of water quality of a segment of Marikina River using multivariate statistical methods

Vanseng Chounlamany, Maria Antonia Tanchuling and Takanobu Inoue

ABSTRACT

Payatas landfill in Quezon City, Philippines, releases leachate to the Marikina River through a creek. Multivariate statistical techniques were applied to study temporal and spatial variations in water quality of a segment of the Marikina River. The data set included 12 physico-chemical parameters for five monitoring stations over a year. Cluster analysis grouped the monitoring stations into four clusters and identified January–May as dry season and June–September as wet season. Principal components analysis showed that three latent factors are responsible for the data set explaining 83% of its total variance. The chemical oxygen demand, biochemical oxygen demand, total dissolved solids, Cl^- and PO_4^{3-} are influenced by anthropogenic impact/eutrophication pollution from point sources. Total suspended solids, turbidity and SO_4^{2-} are influenced by rain and soil erosion. The highest state of pollution is at the Payatas creek outfall from March to May, whereas at downstream stations it is in May. The current study indicates that the river monitoring requires only four stations, nine water quality parameters and testing over three specific months of the year. The findings of this study imply that Payatas landfill requires a proper leachate collection and treatment system to reduce its impact on the Marikina River.

Key words | cluster analysis, landfill, leachate, principal component analysis, rivers, water quality

Vanseng Chounlamany (corresponding author)
Department of Chemistry, Faculty of Natural Science,
National University of Laos,
Vientiane,
Lao PDR
E-mail: *vnchounlamany@gmail.com*

Maria Antonia Tanchuling
College of Engineering,
University of Philippines Diliman,
Quezon City,
Philippines

Takanobu Inoue
Department of Architecture and Civil Engineering,
Toyohashi University of Technology,
Toyohashi 441-8580,
Japan

INTRODUCTION

A major source of pollution of water resources is landfill sites, which are growing due to ever-increasing domestic, commercial, municipal and industrial waste. Landfills produce leachate that can seep into groundwater or be carried into rivers and lakes. Landfill management is a challenging issue in developing countries due to lack of regulation and control as well as poor design, construction and maintenance of leachate management systems.

Experimental studies have been done for specific rivers in selected developing countries to examine river pollution due to landfill leachate by analyzing water quality parameters such as organic, inorganic and heavy metal content (e.g., Kjeldsen *et al.* 2002; Mor *et al.* 2006; Yusof *et al.* 2009). Marikina River in Quezon City, Philippines, is a main river that used to be classified as Class 'A' of the DAO Standards of the Philippines (DENR 1990; Sia Su 2005). Since late 1990s, Marikina River has been seriously polluted due to various point and non-point sources associated with landfills, solid waste and wastewater disposal from nearby communities and runoff from agricultural lands along the river (Sia Su 2005). Payatas landfill, located in Quezon City, is closer to the Marikina River and discharges its leachate to the river through Payatas creek. Payatas landfill was constructed without an appropriate bottom liner and it does not have a proper leachate management system. In the Philippines, studies on impact of landfill sites such as Payatas on nearby rivers are rarely available. A critical need exists for such studies to facilitate river pollution control and landfill management. With this background in mind, an experimental study of Payatas leachate and the

doi: 10.2166/wst.2017.279

water quality of a Marikina River segment and two creeks discharging to the river was conducted. Chounlamany (2015) presents the details of experimental study.

Multivariate statistical methods such as cluster analysis (CA) and principal component analysis (PCA) have been widely used as unbiased methods in the analysis of water quality data to obtain useful information such as a reduced number of latent factors/sources of pollution and assessment of temporal and spatial variations (Vega *et al.* 1998; Singh *et al.* 2004; Shrestha & Kazama 2007; Cho *et al.* 2009; Cai *et al.* 2012; Wang *et al.* 2013; Feher *et al.* 2016).

The objectives of this study are to apply the CA and PCA to a set of water quality data involving a segment of Marikina River and assess information about the similarities and dissimilarities among the sampling stations due to spatial and temporal variations; and identify the water quality characteristics due to natural or anthropogenic influences of point sources and non-point sources such as Payatas landfill leachate and creeks carrying domestic waste. It is hypothesized that leachate from Payatas has a significant effect on the water quality of the Marikina River, especially during the dry season.

METHODS

Description of study area

The Marikina River has depths ranging up to 20 m, spans from 70 to 120 m, and a total area of nearly 75.2 hectares and is 27 km long. The climate of the Philippines is characterized by relatively high temperature, humidity, rainfall and typhoons. Metro Manila region has two seasons based on rainfall, the dry season from November to April and the wet season over the remainder of the year (Flores & Balagot 1969). The Marikina River is subjected to pollution from a variety of points (e.g., Payatas landfill, creek outfalls) and non-point sources (e.g., domestic waste disposal from communities along the river banks, agricultural runoff, etc.). A network of sampling stations as shown in Figure 1 was

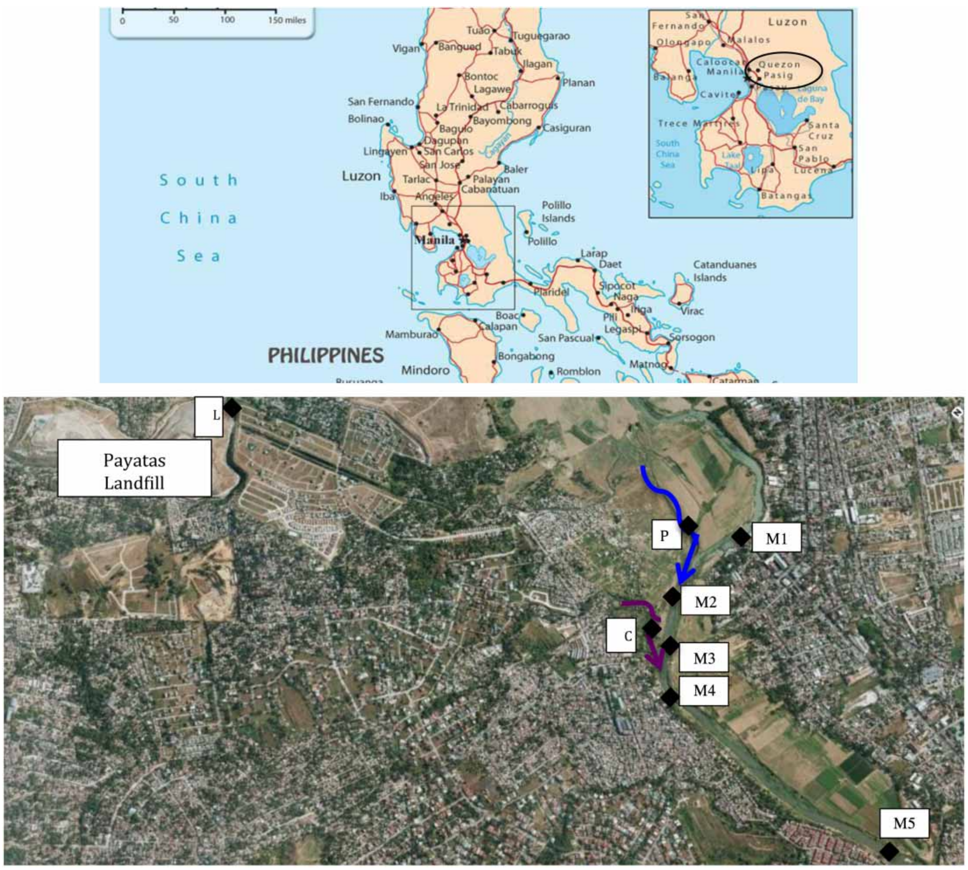

Figure 1 | Location of sampling sites and sampling network (source: Google Maps).

established to study the effects of Payatas landfill leachate and creeks discharging to the river. Station M1, located upstream of Payatas creek, has no influence from the landfill. Station M2 was approximately 0.7 km from M1 and downstream of the Payatas creek outfall. Station M3 was approximately 1.1 km downstream of M2 but upstream of where a second creek that carries municipal waste meets the river. M4 was approximately 0.9 km downstream of M3 and downstream of the second creek outfall. M5 was approximately 1.5 km downstream of M4. Stations P, C and L were located on Payatas creek, the creek carrying municipal waste and the landfill site, respectively (Figure 1). It was noted that three small creeks carry domestic waste in to the Marikina River between M4 and M5 and the area has a lot of agricultural activity.

Monitored parameters

Davis & Cornwell (2006) discuss the commonly used parameters for water quality characterization. The important physical parameters are temperature, pH, electrical conductivity (EC), total dissolved solids (TDS) and turbidity, whereas the dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), anions (Cl^-, NO_3^--N, SO_4^{2-}, PO_4^{3-}-P) and relevant heavy metals (Cd, Pb, etc.) are the common chemical parameters. The 12 physico-chemical parameters (pH, EC, TDS, total suspended solids (TSS), turbidity, DO, COD, BOD, Cl^-, NO_3^-, SO_4^{2-}, PO_4^{3-}) were chosen based on the literature review and general knowledge of the region and landfill. BOD and COD are representative of the total organic content of leachate and river water. Heavy metal contents of the Payatas leachate and the river are very low based on experimental studies (Chounlamany 2015).

Sampling was done during March to December 2013 and January and February 2014, except in August and November 2013 due to heavy rain. The samples were analyzed for the 12 parameters mentioned previously. All water quality parameters are expressed in mg L^{-1}, except pH, EC ($\mu S\ cm^{-1}$), temperature (°C) and turbidity (NTU). The experimental procedures and data quality for water quality testing were ensured by following *Standard Methods for the Examination of Water and Wastewater* (American Public Health Association 1985). The procedure includes blank samples of distilled water and three duplicate samples for each parameter measured. The averages of samples were used in the analysis. Details of the experimental methods and equipment used are given elsewhere (American Public Health Association 1985; Chounlamany 2015).

Multivariate statistical methods

Cluster analysis

CA is an unsupervised pattern recognition technique that uncovers the intrinsic structure or underlying behavior of a data set without making a priori assumption about the data, in order to classify the objects of the system into categories or clusters based on their nearness or similarity (Vega *et al.* 1998; Shaw 2003). Hierarchical clustering is the most common approach, in which clusters are grouped sequentially, by starting with the most similar pair of objects and forming higher clusters step by step, and is typically illustrated by a dendrogram. A dendrogram provides a visual summary of the clustering processes, presenting the map of groups with a dramatic reduction in dimensionality of the original data.

Principal component analysis

PCA is a multivariate statistical technique which is designed to transform the complexity of input variables with a large volume of information into new, uncorrelated variables, called principal components (Shaw 2003). It allows identification of hidden patterns in the data. In terms of statistical analysis, PCA mainly involves the following six major steps (Shaw 2003): (1) start by coding the variables x_1, x_2, $x_3, \ldots, x_p$ to have zero means and unit variance, and standardize the variables to make sure they have equal weight in the further analysis; (2) calculate the covariance matrix; (3) calculate the correlation matrix; (4) calculate the eigenvalues and the corresponding eigenvectors; (5) rank eigenvalues and corresponding eigenvectors by the order of numerical values and discard components interpreting a small part of total variance in the data set; (6) develop the variable loading matrix to infer the principal parameters.

RESULTS AND DISCUSSION

Descriptive measures of river water quality data

PCA and CA of water quality data sets were performed using XLSTAT software (version 2014.1), and all experimental data was normalized to zero mean and unit variance (Shaw 2003). The range, mean and coefficient of variation

Table 1 | Range, mean and CV of water quality parameters at five stations

Parameter		M1	M2	M3	M4	M5
Temp. (°C)	Range	23.6–27.9	23.4–28.2	23.7–29.2	23.8–29.1	24.1–31.3
	Mean (CV)	25.9 (0.055)	26.0 (0.061)	26.1 (0.064)	26.2 (0.06)	27.1 (0.097)
pH	Range	7.3–7.9	7.6–8.1	7.2–8.1	7.3–8.0	7.5–8.0
	Mean (CV)	7.7 (0.027)	7.9 (0.025)	7.8 (0.036)	7.7 (0.031)	7.7 (0.021)
EC (μS/cm)	Range	175–457	296–823	168–544	213–568	209–548
	Mean (CV)	355 (0.273)	607 (0.283)	387 (0.333)	436 (0.280)	415 (0.301)
Turb. (NTU)	Range	82–468	79–423	67–497	61–420	57–555
	Mean (CV)	217 (0.627)	205 (0.512)	239 (0.573)	199 (0.598)	199 (0.824)
DO (mg/L)	Range	1.74–6.73	0.68–5.20	1.09–6.36	1.06–6.38	0.19–6.33
	Mean (CV)	3.75 (0.464)	2.46 (0.646)	3.42 (0.544)	3.16 (0.576)	2.41 (0.938)
TDS (mg/L)	Range	109–289	194–526	108–348	138–363	133–351
	Mean (CV)	229 (0.275)	390 (0.279)	252 (0.313)	282 (0.273)	270 (0.304)
COD (mg/L)	Range	10–34	28–136	13–52	18–70	16–66
	Mean (CV)	21 (0.433)	77 (0.475)	28 (0.457)	37 (0.395)	35 (0.460)
BOD (mg/L)	Range	6–30	16–79	9–74	3–25	6–46
	Mean (CV)	18 (0.549)	34 (0.749)	26 (0.940)	11 (0.736)	23 (0.753)
Cl^- (mg/L)	Range	5.8–23	14.5–99	7.6–34	8.3–42	8.7–34
	Mean (CV)	17.2 (0.336)	43.3 (0.576)	21.6 (0.391)	26.0 (0.414)	23.1 (0.392)
SO_4^{2-} (mg/L)	Range	15–63	15–61	14–64	14–61	16–67
	Mean (CV)	30.98 (0.418)	31.8 (0.444)	30.6 (0.480)	32.8 (0.473)	30.29 (0.431)
NO_3^- (mg/L)	Range	0.65–1.65	0.09–1.23	0.55–1.88	0.54–2.47	0.43–1.84
	Mean (CV)	1.21 (0.347)	0.53 (0.774)	1.20 (0.392)	1.19 (0.479)	1.11 (0.405)
PO_4^{3-} (mg/L)	Range	0.22–0.98	0.23–1.58	0.25–1.11	0.26–1.02	0.26–1.39
	Mean (CV)	0.60 (0.700)	0.79 (0.544)	0.63 (0.397)	0.69 (0.362)	0.74 (0.446)

(CV) of the water quality data of Marikina River are given in Table 1. The mean values of temperature as well as pH at different stations do not vary much, and the CVs (given in parentheses) are small, implying that spatial and temporal effects are relatively minor. This is consistent with tropical environments like the Philippines (Flores & Balagot 1969). Singh *et al.* (2005) also observed similar behavior in their study of Gomti River in India.

EC, turbidity, DO, TDS, COD, BOD, Cl^- SO_4^{2-}, NO_3^- and PO_4^{3-} show high dispersion of data (high CV). The highest mean values of EC, TDS, COD, BOD and Cl^- are observed at M2, whereas the smallest mean values of DO are at M2 and M5. M1 has the highest average DO value and the lowest average COD. The highest variation in mean values between stations is observed in the case of DO, BOD and COD, which confirms the high organic content coming from pollution sources as well as decomposition. Singh *et al.* (2005) also observed similar variation between sampling stations in their study of a river with point and non-point sources of pollution. It is also observed that TDS and Cl^- have high mean values at M2 compared with other stations, which implies high inorganic loading from landfill (Kjeldsen *et al.* 2002; Yusof *et al.* 2009). Mean values of inorganic parameters SO_4^{2-}, PO_4^{3-} and NO_3^- (except M2) are not substantially different between stations, implying minor spatial effects. However, the CVs of these parameters at different stations are relatively high, confirming substantial temporal variation resulting from seasonal factors such as rain, flooding and agricultural activity.

Table 2 summarizes the results of a correlation analysis. Note that data from all river sampling stations were combined to calculate the correlation matrix. EC correlates well with TDS and Cl^-, while turbidity also shows good correlation with TSS. TDS and Cl^- represent inorganic matter, which naturally correlates well with EC (Davis & Cornwell 2006). Turbidity correlates well with TSS because it depends

Table 2 | Correlation matrix of physico-chemical parameters ($p < 0.01$ with one-tailed test)

Variables	Temp.	pH	EC	Turb.	DO	TDS	TSS	COD	Cl^-	NO_3^-	SO_4^{2-}	PO_4^{3-}
Temp.	**1**											
pH	−0.640	**1**										
EC	0.386	0.056	**1**									
Turb.	−0.342	−0.054	−0.460	**1**								
DO	−0.491	0.210	**−0.769**	0.359	**1**							
TDS	0.382	0.062	**0.999**	−0.465	**−0.771**	**1**						
TSS	−0.346	−0.018	−0.478	**0.873**	0.389	−0.488	**1**					
COD	0.104	0.194	**0.713**	−0.180	−0.452	**0.710**	−0.262	**1**				
Cl^-	0.360	−0.117	**0.799**	−0.168	−0.625	**0.797**	−0.205	**0.763**	**1**			
NO_3^-	0.421	−0.457	−0.103	−0.377	−0.158	−0.101	−0.296	−0.376	−0.207	**1**		
SO_4^{2-}	0.131	−0.442	−0.090	0.660	−0.065	−0.098	0.583	0.014	0.175	−0.159	**1**	
PO_4^{3-}	0.335	−0.297	0.487	0.115	−0.690	0.476	0.082	0.370	0.679	−0.080	0.372	**1**

on the suspended solids in the water. TDS also correlates well with Cl^-, whereas COD shows strong correlation with Cl^-. Cl^- and COD correlation could be due to the same source of origin from leachate (Kjeldsen *et al.* 2002). DO shows negative correlation with temperature, since the solubility of oxygen decreases with temperature (Li & Liao 2003). COD, Cl^-, TDS and PO_4^{3-} are also negatively correlated with DO, as organic matter is partially oxidized by DO, while nutrients cause eutrophication of water, which results in an increase of organic matter, leading to a further increase in oxygen demand (Li & Liao 2003).

Temporal similarity and period grouping

Figure 2 shows a dendrogram generated by temporal CA that confirms the 10-month monitoring period can be represented by three temporal clusters. Cluster 1 (the first period) includes January to May and July, generally corresponding to the dry season in the Philippines (traditionally December to April), which is the low flow period of the river.

Cluster 2 (second period) includes October and December, representing the transition period from wet to dry season, approximately corresponding to the typical mean flow period. Cluster 3 (third period) includes June and September, corresponding to the wet season (traditionally May to November) involving high flow rates. July was quite dry, while heavy rain was noted in August and November (no data collected) with dangerously high water levels and high flow rates.

Figure 2 confirms that river monitoring must be done at least three times, corresponding to the three clusters. It also shows that the temporal patterns of water quality data have to take into consideration the transition from wet to dry season as well as localized weather patterns within a year. A good example is the month of July, which would normally be considered a wet month but was relatively dry in 2013. Other researchers also found that temporal clusters do not necessarily align with traditional dry and wet seasons, and adjustments are needed to recognize the actual seasonal weather patterns (Wang *et al.* 2013).

Spatial similarity and station grouping

Figure 3 shows that the stations can be grouped into four clusters, which also implies pollution from different sources and quality of water. Group 1 consists of Station M1, the upstream station treated as the base station for pollution assessment. Station M3 belongs to Group 2; however, the dendrogram indicates that M1 and M3 can be clustered into a single group (see also Table 1). This behavior implies that the river still retains some self-purification capacity, as Payatas creek adds a substantial organic load between M1 and M3. Group 3 consists of downstream stations M4 and M5, where pollution is increased (see Table 1) due to agricultural and domestic waste disposal through several creeks. Group 4 consists of Station M2, which is directly affected by leachate loading from Payatas Creek. Table 1 shows that M2 is more polluted compared with the other four stations based on the average values of COD, BOD, DO, TDS and Cl^-. In summary, the station clustering observed here is similar to the three clusters noted by Singh *et al.* (2004) associated

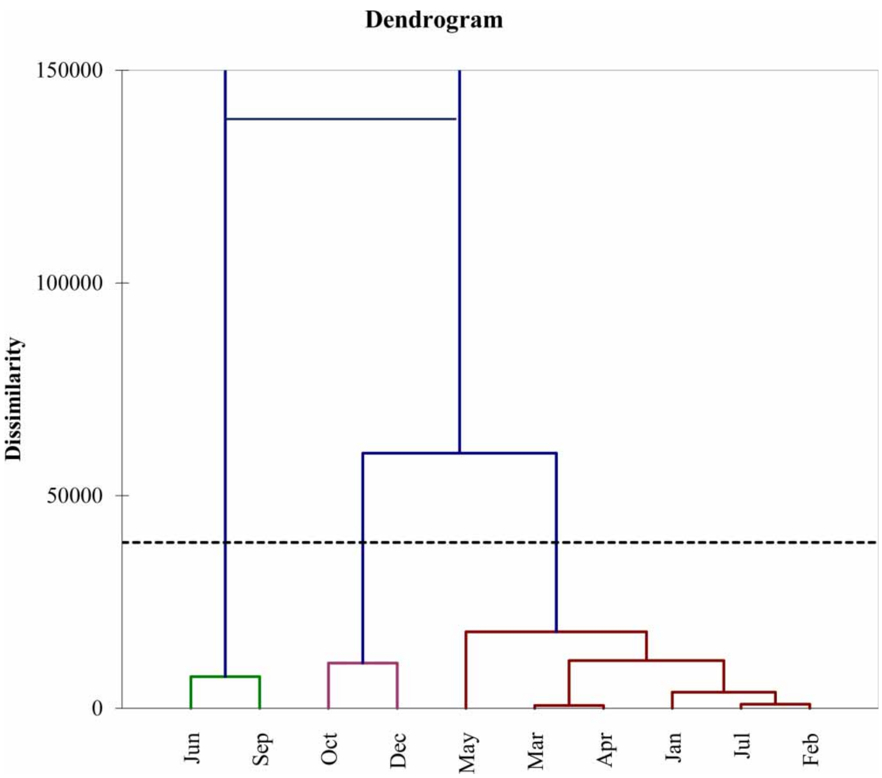

Figure 2 | Dendrogram showing clusters of monitoring periods.

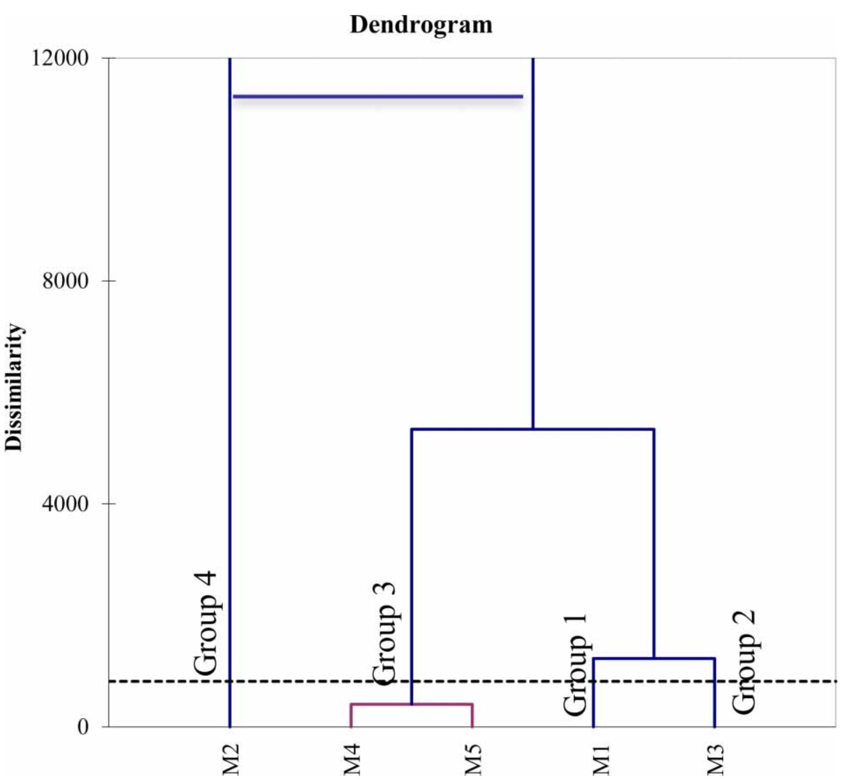

Figure 3 | Dendrogram showing clusters of monitoring stations.

with relatively low (M1 and M3), moderate (M4 and M5) and high pollution (M2) states.

The results from CA indicate that efficient and rapid assessment of water quality of Marikina River segment is possible using the representative sites from temporal and spatial groups. This would lead to the design of an optimal monitoring strategy with fewer sampling stations and monitoring times at reduced costs (Cui *et al.* 2010; Wang *et al.* 2013). According to the CA results, the frequency of monitoring can be decreased to three times a year (e.g. January, June and October) and only four (or three) stations are needed (M1 and/or M3; M2; M4 or M5).

Box plots of water quality parameters

The box plots of the individual water quality parameters showing the temporal variations corresponding to the three periods from CA are shown in Figure 4. These were prepared by combining data at all stations corresponding to each period as shown in Figure 2 for a given parameter. The median value, first (Q_1) and third (Q_3) quartile values, lowest value and highest values for a given period were determined for each parameter by analyzing the data for all stations for the specific period. The line across the box shows the median concentration and the bottom and top

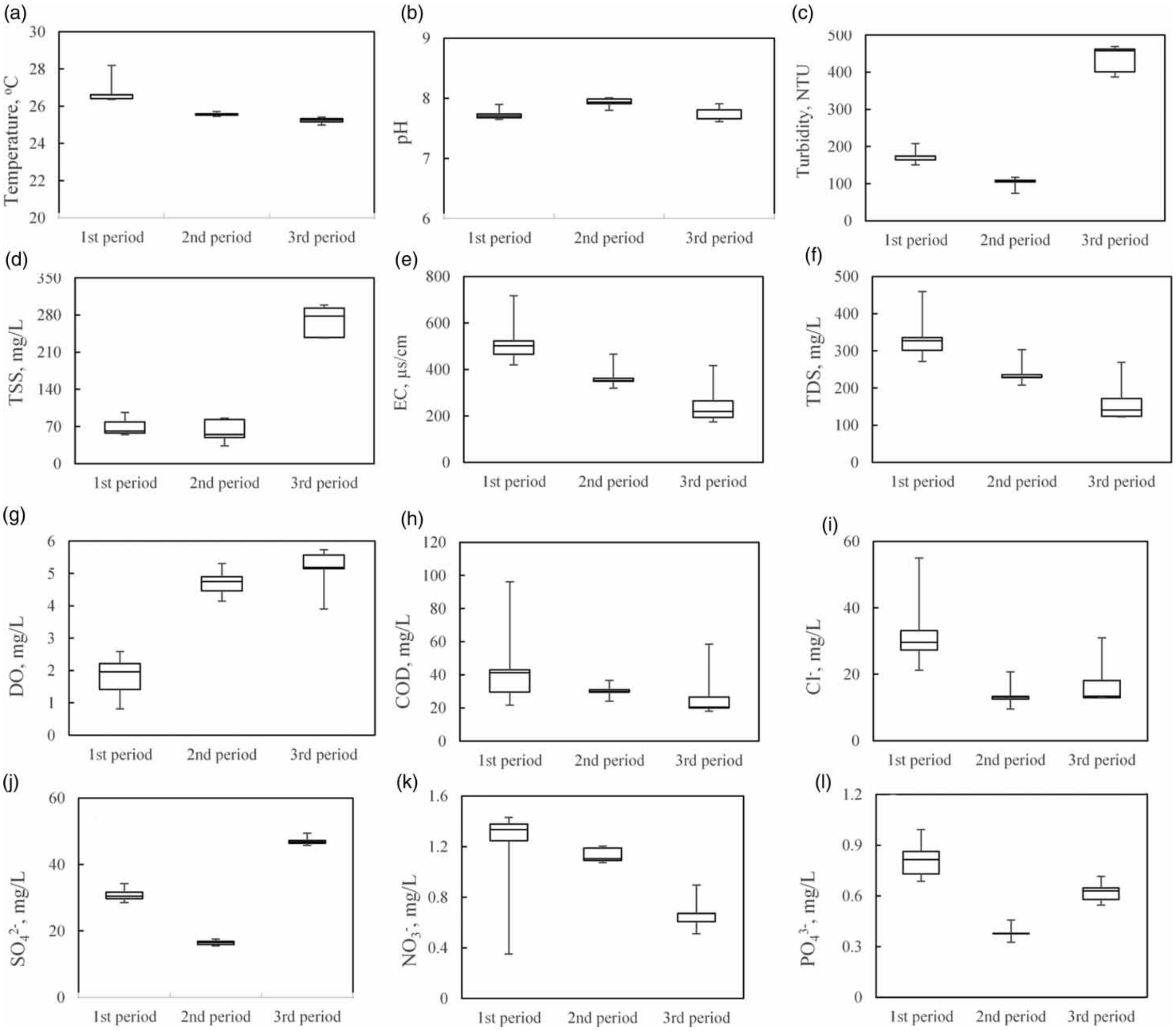

Figure 4 | Box plots of (a) temperature, (b) pH, (c) turbidity, (d) TSS, (e) EC, (f) TDS, (g) DO, (h) COD, (i) Cl^-, (j) SO_4^{2-}, (k) NO_3^- and (l) PO_4^{3-} for different temporal clusters.

of the box correspond to Q_1 and Q_3. The vertical lines that extend from the bottom and top of the box correspond to the lowest and highest observations.

The box plots of temperature and pH are similar, indicating minor influence of seasons and river flow rates as well as changes in pollution loading. However, turbidity, TSS, TDS, DO, EC, Cl^-, COD, SO_4^{2-}, NO_3^-, and PO_4^{3-} show substantial variations between the three periods and quite different individual patterns. DO shows (Figure 4(g)) a larger magnitude from the first period (dry season) to the third period (rainy period). This is due to increasing river flow rates that cause more aeration as well as aeration directly resulting from raining (Davis & Cornwell 2006). The distribution of DO data is quite balanced during the first and second periods, but during the third period the distribution is skewed towards the lower concentrations (longer whisker at the bottom).

Turbidity and TSS concentrations increase from period 1 to period 3 due to heavy rain bringing more sediment from upstream, nearby agricultural land as well as erosion of river banks (Wang *et al.* 2013). EC, TDS, COD, and NO_3^- show a decreasing magnitude from the first period to the third period. This is due to the fact that during dry seasons the river flow rate is low and pollution loading from landfill and non-point sources results in higher concentrations of these parameters. As the river flow rate increases substantially during the rainy months, the concentrations get reduced. Furthermore, the box plots of these variables for the first period show long whiskers, implying a large spread of magnitudes and distributions skewed towards either higher or lower concentrations. This also implies a higher effect of pollution sources during dry season. In general, the second period, which corresponds to transition from wet to dry season and stable river conditions, shows, the smallest box plots, implying relatively stable conditions along the river.

Cl^-, SO_4^{2-} and PO_4^{3-} show decreases from the first period to the second period, then an increase in the third period. As noted by Chounlamany (2015) through creek water sample testing during dry months, the concentrations of these water quality parameters are relatively high. In addition, their distribution is also larger except for SO_4^{2-}. The second period (transition from dry to rainy months) has moderately high river flow rates that cause dilution of water quality parameters. The concentrations increase during the rainy months (third period) as regional flooding and drainage increase the pollution loading coming into the river from domestic, agricultural and other sources.

It was noted during field visits that a substantial amount of domestic waste got carried away into the river through creeks and regional flooding during heavy rain. During rainy months, the concentration of some chemical constituents increases at the upstream station M1 due to contributions from sources upstream of M1, which also results in a general increase of concentrations at downstream stations. Based on the box plots shown in Figure 4, it can be concluded that water quality characteristics of the stretch of the Marikina River under study are generally better during the rainy season compared with the dry season. The lowest spatial variations of water quality parameters are observed during the second period (transition from wet to dry season), while both dry and wet seasons show substantial variations of water quality along the river.

A second set of box plots are shown in Figure 5 to examine the influence of the spatial clustering shown in Figure 3. The Group 1 and Group 2 box plots are similar with relatively small differences of mean values, as noted previously, through CA. pH and temperature show minor dependence on the grouping and smallest seasonal changes. Turbidity and TSS also show closer median values between the groups. However, as these box plots have a larger spread, the influence of seasonal changes is significant.

Group 4 box plots are the largest for EC, TDS, COD, Cl^- and PO_4^{3-} in Figure 5 and show the highest spread of data with larger bottom and/or top whiskers for these parameters. This means that seasonal variations as well as concentrations are highest at M2, confirming the significant role of Payatas landfill leachate. The DO concentration in Figure 5(g) shows a decreasing magnitude from Group 1 to Group 4. This confirms again that M1 is the least polluted amongst the five stations in terms of organic content and M2 has the highest organic pollution level. This observation is also consistent with the increasing magnitude observed in Figure 5(h) for COD for the different groups, as DO and COD are negatively correlated (Table 2) (Li & Liao 2003). The SO_4^{2-} concentration does not show substantial differences between the four groups, as the SO_4^{2-} loading from Payatas creek (also leachate) was relatively low (Chounlamany 2015). The concentration of NO_3^- shows not much variation from Group 1 to 3 but shows a reduction at M2, possibly due to consumption of nitrites by pollutants from Payatas creek (David & Cornwell 2003). The municipal creek also causes a noticeable increase in the pollution level of Marikina, as DO levels decrease and TDS, EC, COD, Cl^- and PO_4^{3-} levels increase for Group 3 (Stations M4 and M5) due to discharge from three small creeks carrying domestic waste and agricultural runoff. River pollution

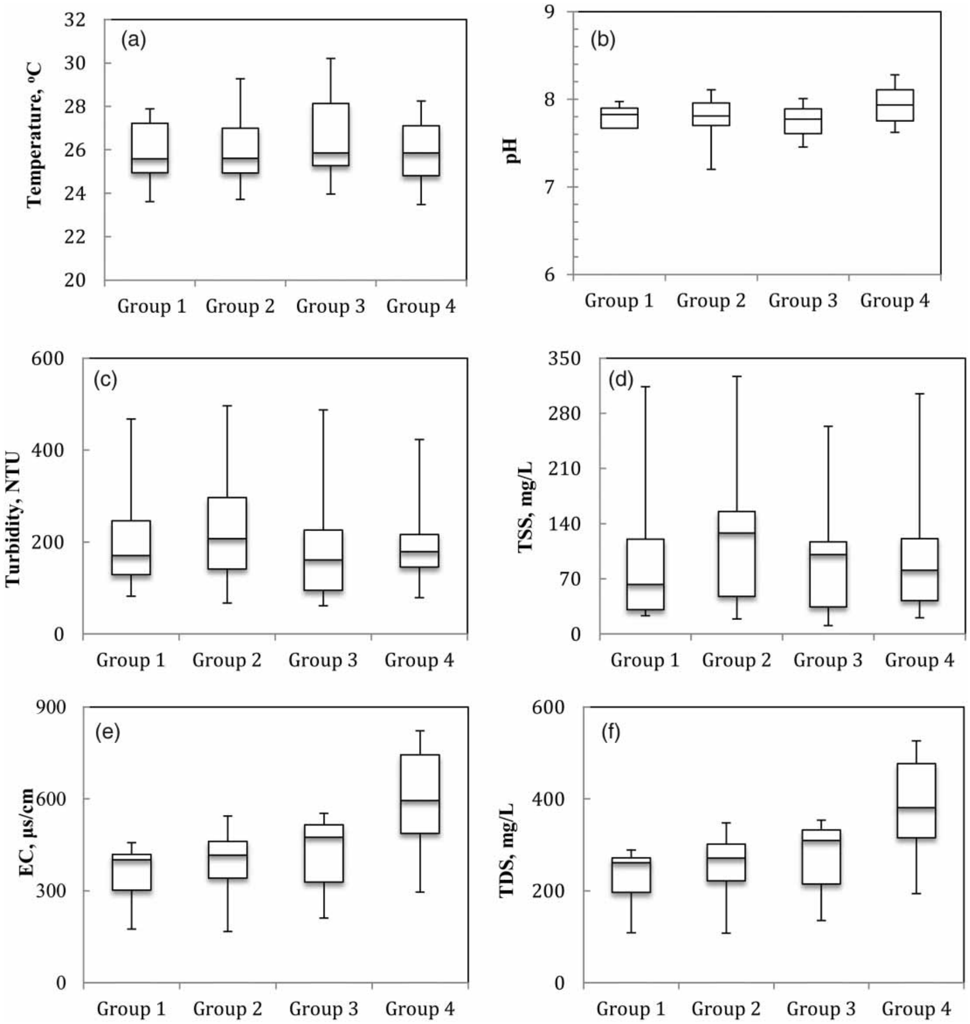

Figure 5 | Box plots of (a) temperature, (b) pH, (c) turbidity, (d) TSS, (e) EC, (f) TDS, (g) DO, (h) COD, (i) Cl^-, (j) SO_4^{2-}, (k) NO_3^- and (l) PO_4^{3-} for different spatial clusters. *(Continued.)*

characteristics observed in Figure 4 are quite similar to the observations of Singh *et al.* (2004).

PCA of river water quality data

PCA of the water quality data set covering the five stations was performed and corresponding eigenvalues obtained. Following Boyacioglu & Boyacioglu (2008), eigenvalues greater than 1 were retained. It is found that the first three eigenvalues together explain over 82% of the variance or information contained in the original data set. The percentage of variances corresponding to the first two (64%) and three (84%) eigenvalues are better than the variances of the first two and three eigenvalues reported in the previous studies focused on river pollution (e.g., Singh *et al.* 2004; Cui *et al.* 2010; Cai *et al.* 2012; Wang *et al.* 2013). This confirms the strong applicability of PCA to the current set of data. Based on the theory of PCA (Shaw 2003), the projections of water quality parameters on the axes of PCs are called loadings and represent the correlation coefficients between PCs and variables. Table 3 shows the loading of the retained PCs. According to Liu *et al.* (2003), factor loadings were classified as 'strong', 'moderate' and 'weak' for *absolute* loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively.

PC1 corresponds to 42% of the variance and is strongly (>0.7) contributed by EC, TDS, COD, BOD, Cl^- and DO (negative) and moderately (0.5–0.7) by PO_4^{3-} (nutrients from leachate and other pollutant sources), TSS (negative) and turbidity (negative). These water quality parameters also showed similar positive or negative correlation in Table 2. This PC represents organic pollution indicator parameters associated with anthropogenic pollution sources

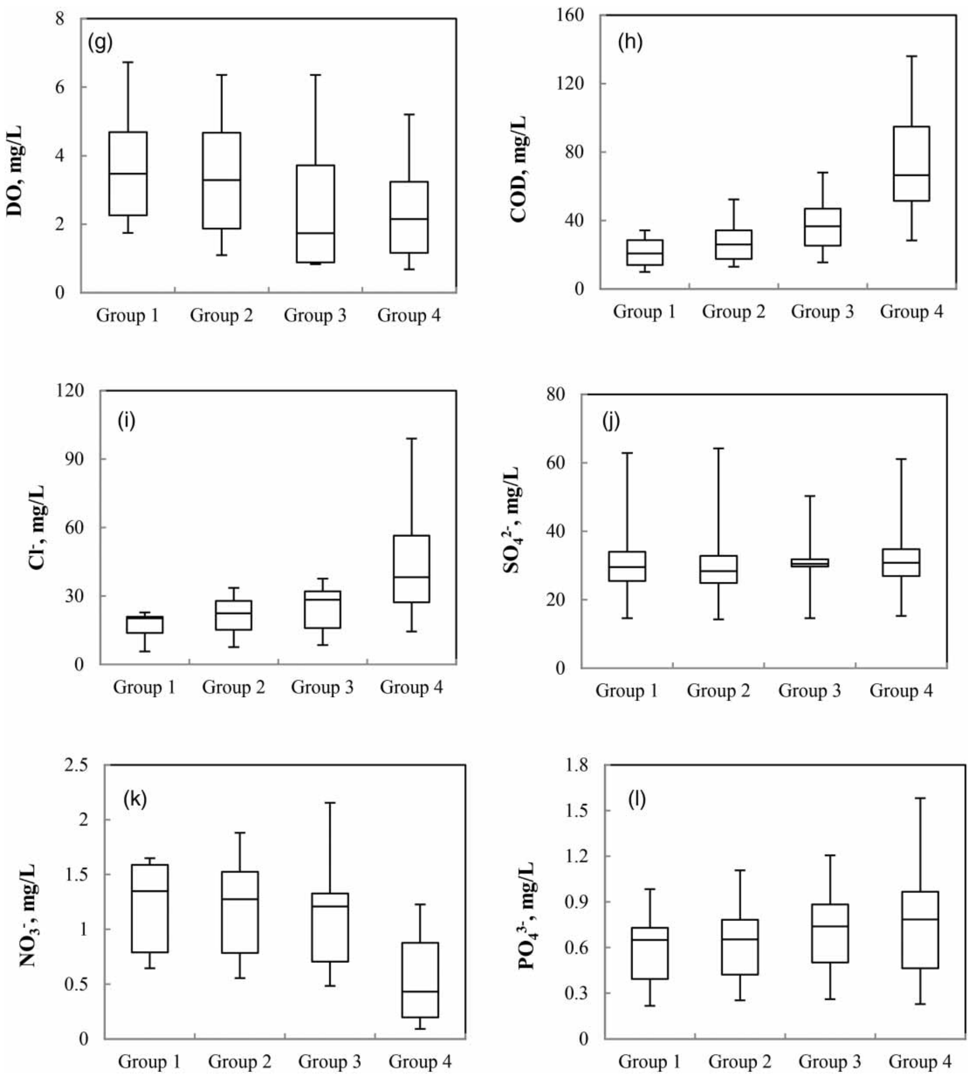

Figure 5 | Continued.

(Li & Liao 2003). Dissolved organic matter at higher concentrations consumes large amounts of oxygen and undergoes anaerobic fermentation processes leading to formation of ammonia and organic acids. Hydrolysis of these acidic materials could also cause a decrease of water pH values, as noted by the negative loading of pH in Table 2 (Wang et al. 2013). Furthermore, the negative loading of DO, TSS and turbidity agrees with past studies and negative correlation of these parameters with anthropogenic pollution sources. PC2 is strongly contributed by turbidity, TSS and SO_4^{2-} and moderately by PO_4^{3-} (increased flow of suspended solids and nutrients from leachate and other sources during rain) and NO_3^- (negative). This PC represents the effect of seasonal factors such as rain and agricultural activity. PC3 is strongly contributed by pH and NO_3^- (negative) and moderately by temperature (negative). PC3 is contributed by physiochemical sources of variability (Shrestha & Kazama 2007). The negative correlation between pH and temperature and pH and NO_3^- can also be seen from Table 2.

Figure 6 displays a plot of the water quality parameters in the first two principal components space. PC1 has strong to moderate positive loading on COD, TDS, Cl^-, BOD and PO_4^{3-} and is influenced by inorganic and organic pollution from Payatas leachate and domestic wastewater through several creeks. PC1 has strong negative loading on DO, similar to the observations by Zhao et al. (2011), and confirms the negative correlation of DO with organic pollution (Li & Liao 2003; Singh et al. 2004). PC2 has positive loading on TSS, turbidity and SO_4^{2-} (associated with leachate, domestic wastewater and agricultural

Table 3 | Factor loading of water quality parameters

	PC1	PC2	PC3
Temperature	0.549	−0.083	−0.673
pH	−0.135	−0.159	**0.892**
EC	**0.944**	0.031	0.199
Turbidity	−0.533	**0.801**	−0.014
DO	− **0.856**	−0.087	0.183
TDS	**0.944**	0.021	0.204
TSS	−0.592	**0.719**	0.030
COD	**0.704**	0.254	0.448
Cl^-	**0.840**	0.376	0.106
NO_3^-	0.037	−0.505	−0.701
SO_4^{2-}	−0.074	**0.820**	−0.374
PO_4^{3-}	0.604	0.558	−0.215
BOD	0.573	0.328	0.482
Eigenvalue	5.042	2.645	2.254
Variability (%)	42.020	22.042	18.784
Cumulative (%)	42.020	64.062	82.846

runoff), which are mostly increased by rain, flooding and soil erosion. These parameters negatively correlate with organic/inorganic pollution and have negative PC1 loading. Inorganic pollutants primarily affect EC, whereas TDS is affected by both organic and inorganic pollution. Moderate negative loading of PC2 on NO_3^- points to the use of fertilizer and agriculture activity, which often happens during dry season and at the end of rainy season. Both PC1 and PC2 have weak negative loading on pH, as it is known to negatively correlate with BOD, and rain increases the acidity of water (Zhao *et al.* 2011).

Figure 7 shows a score plot of the river water quality status at different stations and months in the first two principal components space. It is clear that the highest state of pollution is observed at station M2 from March to May 2013 (dry season). With respect to the stations M3 and M5, the worst pollution is observed in May 2013, whereas for station M4 it is in April 2013. The lowest inorganic and organic pollutants (high DO content) are found in September at M1, M2 and M3 and in December at M4 and M5. The highest seasonal effect was found in October at M4 and M5 and in June at M1, M2 and M3. Comparison of the water quality state of different stations from Figure 7 clearly shows that station M2 is significantly affected compared with other stations.

In general, it can be concluded that pH and NO_3^- do not influence the Marikina River stretch under consideration. It

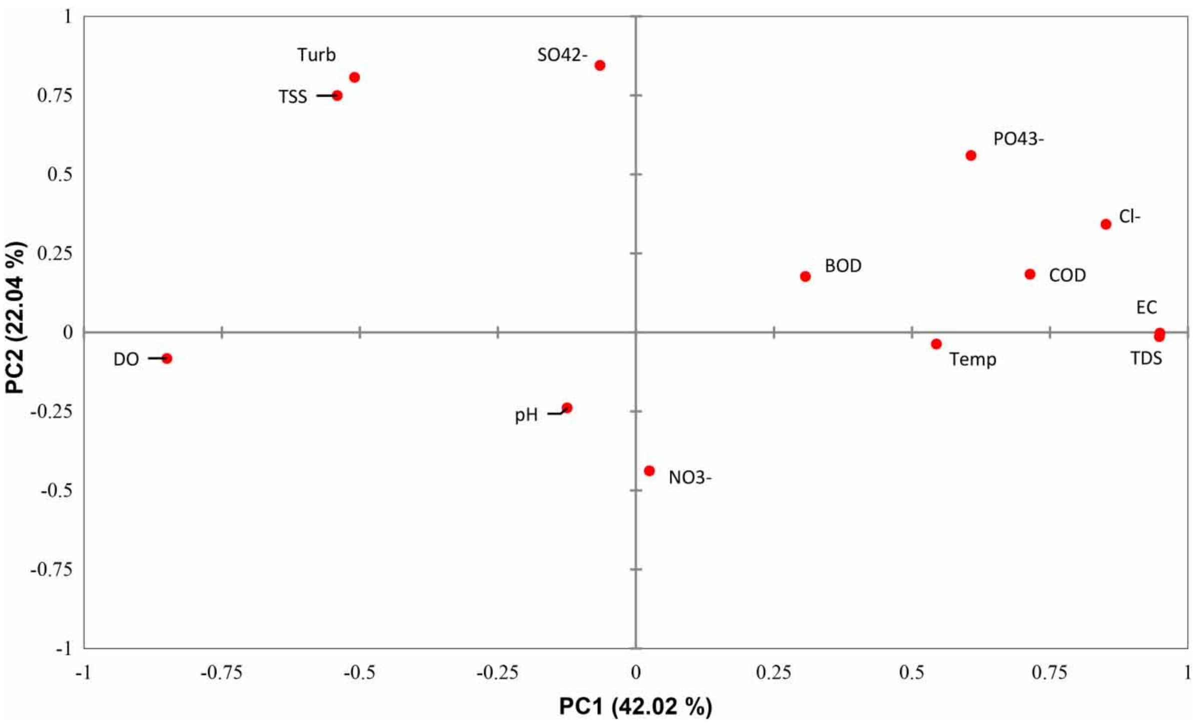

Figure 6 | PC1 and PC2 loading of water quality parameters.

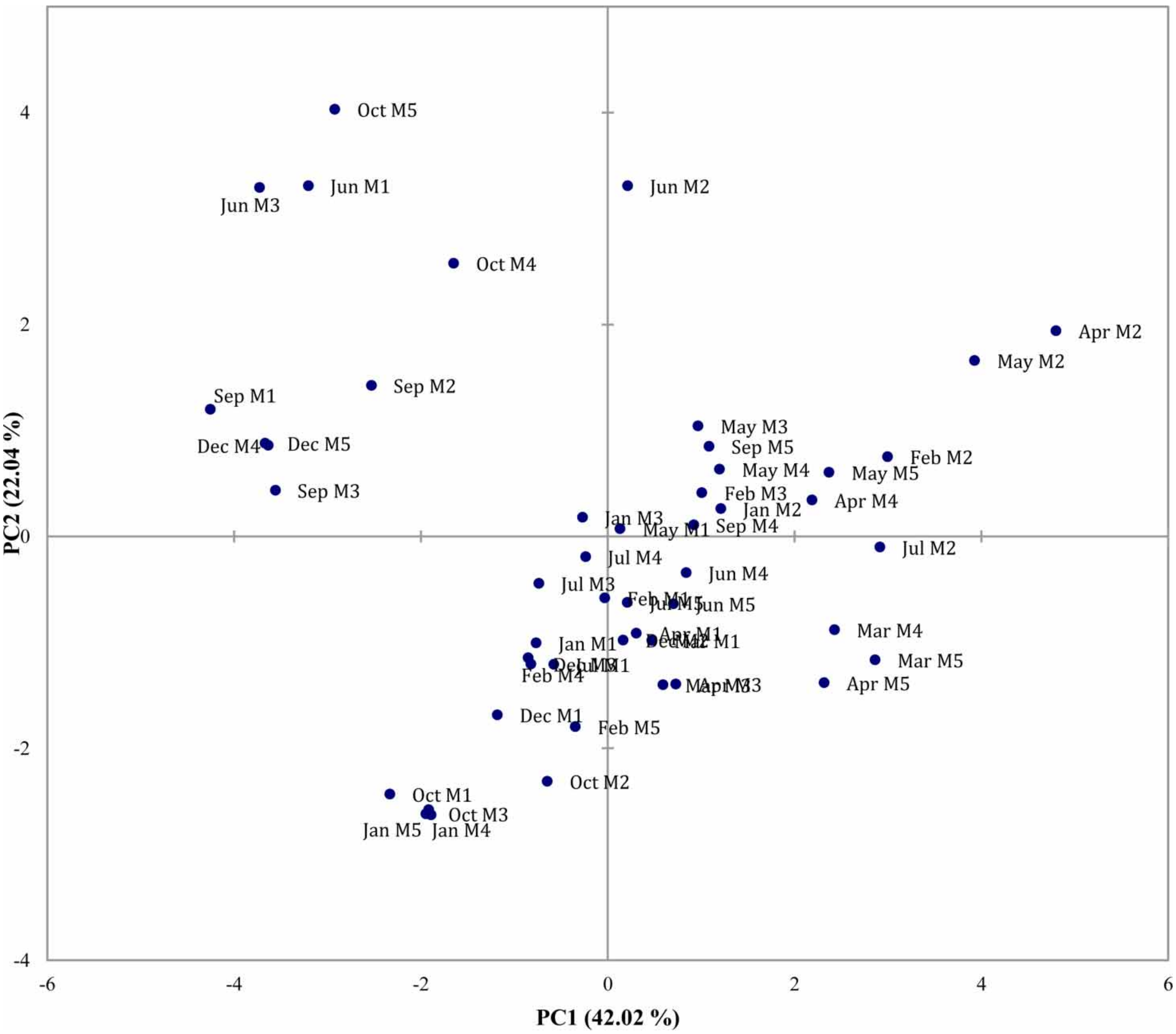

Figure 7 | Representation of temporal and spatial behavior of sampling network in the principal component space.

is therefore recommended that in the next sampling study, it is unnecessary to measure these parameters. Since turbidity, TSS and SO_4^{2-} have influence mostly during raining season, these parameters can be measured only during wet season, whereas temperature and pH require only one measurement per year. The most significant parameters for the Marikina River pollution are TDS, Cl^-, DO, COD, BOD and PO_4^{3-}.

CONCLUSIONS

CA shows that the monitoring period can be divided into three temporal clusters representing dry, wet and dry-wet transition periods. The five sampling stations represent four spatial groups, and the frequency of monitoring can be limited to three times per year using only four river stations, M1, M2, M3, and M4 or M5.

PCA confirms that three latent factors are responsible for the present water quality data set explaining 83% of total variance. TDS, Cl^-, DO, COD, BOD and PO_4^{3-} are the most significant water quality parameters for Marikina River, indicating that pollution is mainly from point sources of anthropogenic/eutrophication nature. TSS, turbidity and SO_4^{2-} are mostly influenced by seasonal factors and soil erosion. pH and NO_3^- have a minor influence on Marikina River. The highest state of pollution is observed at M2 from March to May 2013 and for M3 and M5 in May. These findings could guide design of a comprehensive river water quality monitoring program for Marikina River and regional water bodies.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by the AUN-SEED Net through the Japan International Cooperation Agency (JICA). Access to research facilities of the University of Philippines Diliman and the Toyohashi University of Technology is acknowledged.

REFERENCES

American Public Health Association 1985 *Standard Methods for the Examination of Water and Wastewater,* 16th edn. American Public Health Association, American Water Works Association and Water Pollution Control Federation, Washington, DC, USA.

Boyacioglu, H. & Boyacioglu, H. 2008 Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environment Geology* **54**, 275–282.

Cai, Q., Wang, X., Ye, L. & Qu, X. 2012 Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: a case study of the Xiangxi River basin, China. *Quaternary International* **282**, 137–144.

Cho, K. H., Park, Y., Kang, J.-H., Ki, S. J., Cha, S., Lee, S. W. & Kim, J. H. 2009 Interpretation of seasonal water quality variation in the Yeongsan Reservoir, Korea using multivariate statistical analyses. *Water Science and Technology* **59**, 2219–2226.

Chounlamany, V. 2015 *Effect of the Quezon City Controlled Dumping Facility Leachate on the Water Quality of a Segment of Marikina River*. PhD Dissertation, University of the Philippines, Diliman.

Cui, B., Fan, A., Zhao, H., Zhang, Z. & Zhang, H. 2010 Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Science* **2**, 1220–1234.

Davis, M. & Cornwell, D. 2006 *Introduction Environmental Engineering,* 4th edn. McGraw-Hill, New York.

DENR 1990 Administrative Order No. 34, Department of Environment and Resources-Environmental Management Bureau, http://emb.gov.ph/wp-content/uploads/2016/04/DAO-1990-34.pdf (accessed June 2012).

Feher, I.-C., Zaharie, M. & Oprean, I. 2016 Spatial and seasonal variation of organic pollutants in surface water using multivariate statistical techniques. *Water Science and Technology* **74**, 1726–1735.

Flores, J. F. & Balagot, V. F. 1969 Climate of the Philippines. In: *World Survey of Climatology*, Vol. 8 (K. Arakawa, ed.). Climates of Northern and Eastern Asia, Elsevier, Amsterdam, The Netherlands.

Kjeldsen, P., Barlaz, M. A., Rooker, A. P., Anders, B. A., Ledin, A. & Christensen, T. H. 2002 Present and long-term composition of MSW landfill leachate: a review. *Critical Reviews in Environmental Science and Technology* **32**, 297–336.

Li, J. X. & Liao, W. G. 2003 An analysis on the possibilities of eutrophication in the Three Gorges Reservoir. *Science and Technology Review* **9**, 49–52.

Liu, C. W., Lin, K. H. & Kuo, Y. M. 2003 Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *Science of the Total Environment* **313**, 77–89.

Mor, S., Ravindra, K., Dahiya, R. P. & Chandra, A. 2006 Leachate characterization and assessment of groundwater pollution near municipal solid waste landfill site. *Environmental Monitoring and Assessment* **118**, 435–456.

Shaw, P. J. A. 2003 *Multivariate Statistics for the Environmental Sciences.* Hodder-Arnold, UK.

Shrestha, S. & 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**, 464–475.

Sia Su, G. L. 2005 Water-borne illness from contaminated drinking water sources in close proximity to a dumpsite in Payatas, The Philippines. *Journal of Rural and Tropical Public Health* **4**, 43–48.

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.

Singh, K. P., Malik, A. & Sinha, S. 2005 Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques – a case study. *Analytica Chimica Acta* **538**, 355–374.

Vega, M., Pardo, R., Barrado, E. & Debán, L. 1998 Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research* **32**, 3581–3592.

Wang, P., Wang, Y., Bai, Y., Tian, Z., Li, J., Shao, X., Mustavich, L. F. & Li, B. L. 2013 Assessment of surface water quality via multivariate statistical techniques: a case study of the Songhua River Harbin region, China. *Journal of Hydro-Environment Research* **7**, 30–40.

Yusof, N., Haraguchi, A., Hassan, M. A., Othman, M. R., Wakisaka, M. & Shirai, Y. 2009 Measuring organic carbon, nutrients and heavy metals in rivers receiving leachate from controlled and uncontrolled municipal solid waste (MSW) landfills. *Waste Management* **29**, 2666–2680.

Zhao, J., Fu, G., Lei, K. & Li, Y. 2011 Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. *Journal of Environmental Sciences* **23**, 1460–1471.

First received 6 August 2016; accepted in revised form 2 May 2017. Available online 6 June 2017