Discriminant analysis application in spatiotemporal evaluation of water quality in South Florida

Mohammad Haji Gholizadeh, Assefa M. Melesse and Lakshmi Reddi

ABSTRACT

The discriminant analysis (DA) method was used to differentiate and classify the water quality of three major rivers in South Florida. In this study, DA was used to assess the water quality and evaluate the spatial and temporal variations in surface water quality in South Florida. DA, as an important data reduction method, was used to assess the water pollution status and analysis of its spatiotemporal variation. It was found by the stepwise DA that five variables (chl-a, dissolved oxygen (DO), total Kjeldahl nitrogen (TKN), total phosphorus (TP) and water temperature) are the most important discriminating water quality parameters responsible for temporal variations. Spatial variation in water quality was also evaluated and identified five variables (TKN, TP, ammonia-N, magnesium, and sodium) and seven variables (chl-a, DO, TKN, TP, ammonia-N, magnesium, and chloride) as the most significant discriminating variables in the wet and dry season, respectively of three selected rivers in South Florida. It is believed that the results of apportionment could be very useful to the local authorities for the control and management of pollution and better protection of important riverine water quality.

Key words | discriminant analysis, river water, South Florida, spatiotemporal analysis, water quality

Mohammad Haji Gholizadeh (corresponding author)
Department of Civil and Environmental Engineering,
Florida International University,
10555 W Flagler Street, EC3781,
Miami, FL 33174,
USA
E-mail: *mhaji002@fiu.edu*

Assefa M. Melesse
Department of Earth and Environment,
Florida International University,
AHC-5-390, 11200 SW 8th Street,
Miami, FL 33199,
USA

Lakshmi Reddi
Department of Civil Engineering,
Florida International University,
PC 230, 11200 SW 8th Street,
Miami, FL 33199,
USA

INTRODUCTION

Surface water quality has become a serious concern for urban planners and managers. The surface water quality in a region is a function of natural factors (precipitation, weather, basin physiography, soil erosion, etc.) and anthropogenic factors (urbanization, industrial and agricultural activities, etc.) (Carpenter *et al.* 1998; Jarvie *et al.* 1998). As a result of their impacts, nutrients, toxic substances, and petroleum products enter the rivers, estuaries, lakes and other waterbodies, reducing the quality of water. Anthropogenic factors, such as residential and industrial wastewater, are a constant polluting source in urban areas, whereas natural factors like rainfall, surface runoff, and groundwater level are seasonal phenomenon that are mainly affected by climate (Singh *et al.* 2004). Seasonal variations in precipitation, surface runoff, ground water flow, and water interception and abstraction have significant effects on river discharge, and subsequently, on the concentration of pollutants in river water (Vega 1998). Therefore, to better investigate and evaluate the water quality of watersheds, a study of temporal variations alongside the spatial variations of water quality seems to be inevitable.

To obtain reliable information about the inherent properties of water quality and to understand the spatial and temporal variations in the hydro-chemical and biological properties of water, continuous and regular monitoring programs are required (Singh *et al.* 2005). However, the generated databases are large and complex so that their analyses require robust analytical tools. Multivariate statistical techniques are widely used for the evaluation of both temporal and spatial variations and the interpretation of large and complex water quality data sets (Wunderlin *et al.* 2001; Simeonov *et al.* 2003; Singh *et al.* 2004, 2005; Kim *et al.* 2005; Kowalkowski *et al.* 2006; Shrestha & Kazama 2007; Schaefer & Einax 2010; Juahir *et al.* 2011; Mustapha & Aris 2012).

doi: 10.2166/hydro.2016.023

Discriminant analysis (DA) was conducted on the data set of water quality of three selected rivers in the study area to construct the discriminant functions (DFs) on two different standard and stepwise modes to identify the most significant variables that result in water quality spatial and temporal variation. In addition, this step was performed to optimize the monitoring program by decreasing the number of parameters monitored. Wunderlin *et al.* (2001) obtained a complex data matrix, which was treated using the pattern recognition techniques of cluster analysis (CA), factor analysis/principal component analysis, and DA. They found that the DA technique showed the best results for data reduction and pattern recognition during both temporal and spatial analysis. Singh *et al.* (2004, 2005) used multivariate statistical techniques to evaluate spatial and temporal variations in the water quality of the Gomti River (India). Zhou *et al.* (2007) used CA and DA to assess temporal and spatial variations in the water quality of the watercourses in the North Western New Territories of Hong Kong, over a period of five years (2000–2004), using 23 parameters at 23 different sites (31,740 observations). Their results showed that DA allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters that were responsible for most of the variations in water quality. Shrestha *et al.* (2008) used different multivariate statistical techniques for the evaluation of temporal/spatial variations and the interpretation of a large complex water quality data set of the Mekong River. They also concluded that DA showed the best results for data reduction and pattern recognition during both spatial and temporal analysis. Li *et al.* (2009) also used DA to evaluate temporal and spatial variations and to interpret a large and complex water quality data set collected from the Songhua River basin in China. The results of DA showed a reduction in the dimensionality of the large data sets, by delineating a few indicator parameters of the water quality. Zhang *et al.* (2011) applied different multivariate statistical techniques for the interpretation of a complex data matrix obtained between 2000 and 2007 from the watercourses in the Southwest New Territories and Kowloon, Hong Kong. DA provided better results both temporally and spatially. Chen *et al.* (2015) used the fuzzy comprehensive assessment, CA and DA to assess the water pollution status and analyze its spatiotemporal variation. Furthermore, long-term hydro-chemical data of shallow water bodies were evaluated using DA (Medina-Gomez & Herrera-Silveira 2003; Parinet *et al.* 2004; Solidoro *et al.* 2004).

Given the above considerations, a large data matrix obtained over a 15-year (2000–2014) monitoring period at 16 different sites for 12 water quality parameters, and in two wet and dry seasons (approximately 35,000 observations), was subjected to DA to identify the most significant water quality variables responsible for spatial and temporal variations in river water quality.

METHODS

Study area

Canals in South Florida form an extensive network to distribute water and to discharge seasonal excess flows into estuaries. They are biologically productive systems that support a variety of aquatic plants, animals, and microorganisms, many of which also thrive in ponds, sloughs, and marshes. In this study, three major rivers in South Florida, the Miami Canal, Kissimmee river and Caloosahatchee river, are investigated for their water quality by applying different multivariate analysis techniques. The average annual temperature range from 19.2 to 28.7 °C and the annual rainfall in the entire area of South Florida is generally approximately 1,400 mm. The major land uses in these watersheds include agricultural area, wetlands, cattle ranching and dairy farming, and urban areas. Figure 1 shows the location of the study area and the selected water quality monitoring sites on the three major rivers of South Florida.

Data set preparation

The hydrography network of the study area, generated using the 1:24,000 national hydrography data set obtained from the South Florida water management district's (SFWMD) geographic information systems (GIS) data catalog, was used to delineate the flow line of the three selected rivers. The most recent (2008–2009) land cover/land use (LCLU) map, provided by the SFWMD, was used in this study.

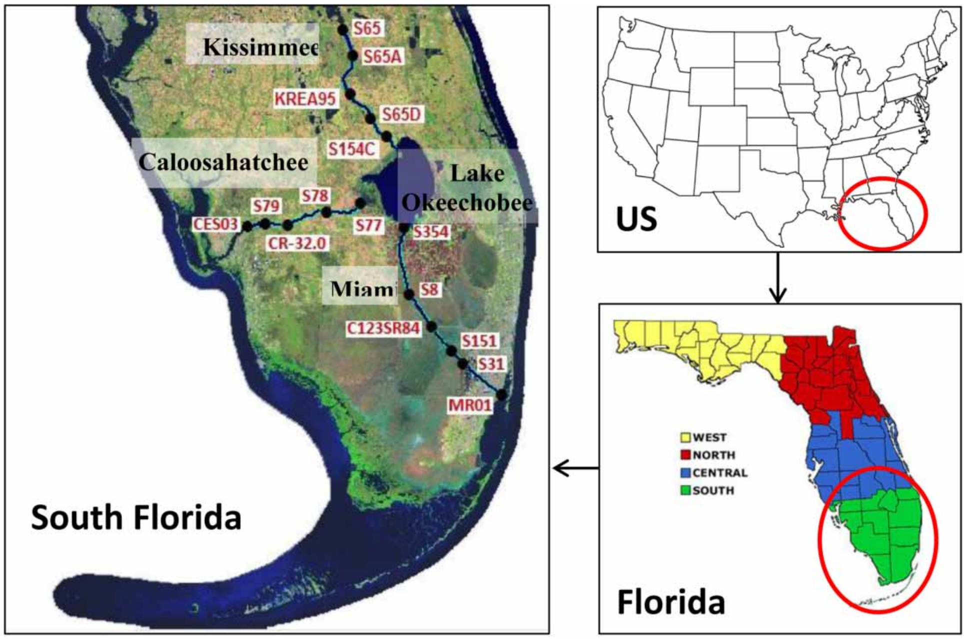

Figure 1 | The location of study area and water quality monitoring site.

These data were then clipped to fit our study area. The area of each type of land use within each watershed was calculated using an ESRI ArcGIS 10.0 platform. The monitoring stations, downloaded from the same source, were overlaid with the rivers' map in ArcGIS to design a network of sampling stations that include sufficient historical data to construct a robust statistical database of studied parameters, considering a suitable spatial distribution on the river. Then, the DBHYDRO (environmental database of SFWMD) was used to obtain continuous time series data for 12 selected water quality parameters from 2000 to 2014. This database was then divided into two dry and wet seasons (the wet season from May 15th to October 15th and the dry season from October 16th to May 14th). The selected water quality parameters for investigation in this study include chl-a, dissolved oxygen (DO), total Kjeldahl nitrogen (TKN), total phosphorus (TP), total phosphate, ammonia-N, water temperature (WT), total suspended solids (TSS), turbidity, magnesium, chloride, and sodium.

Multivariate statistical methods

Multivariate statistical methods for classification, modeling and interpretation of large data sets allow for the reduction of the dimensionality of the data and the extraction of information (Massart & Kaufman 1983). DA, also called the supervised pattern recognition technique, is a multivariate statistical analysis method that uses linear combinations of several variables to construct statistical classification of samples into categorical-dependent values. DA is usually performed with prior knowledge of the membership of objects to a particular group. This technique builds up a

Table 1 | Wilks' lambda and chi-square test for the temporal DA of water quality variations across two wet and dry seasons

Mode	DF	Canonical correlation (R)	Eigenvalue	Wilks' lambda	Chi-square	*p*-level (Sig.)
Standard mode	1	0.882	3.513	0.222	277.271	0.000
Stepwise mode	1	0.870	3.102	0.244	267.490	0.000

DF for each group, which operates on raw data (Johnson & Wichern 1992; Wunderlin *et al.* 2001; Singh *et al.* 2004). Generally, two different modes are used, standard and stepwise, to construct the DFs. In this study, DA was applied on the raw data matrix using both standard and stepwise modes in order to construct the DFs to differentiate and classify the water quality. The standard mode constructs discriminating functions containing all predictive variables, whereas in the stepwise mode, one variable that minimized the overall Wilks' lambda statistic was entered or removed at each step. Wilks' lambda is a statistical test used in multivariate analysis of variance to test the differences between the means of identified groups of subjects on a combination of dependent variables. It is a measure of how well each function separates cases into groups. Smaller values of Wilks' lambda indicate greater discriminatory ability of the function. The canonical DFs of the discriminating variables were used to discriminate among groups. The canonical DFs are defined as weighted linear combinations of the original variables, where each variable is weighted according to its ability to discriminate among groups. The first canonical DF defines the specific linear combination of variables that maximizes the ratio of among group to within group variance in any single dimension. It constructs a DF for each group as follows:

$$f(G_i) = K_i + \sum\nolimits_{j=1}^{n} W_{ij} \times P_{ij} \quad (1)$$

where i is the number of groups (G), k_i is a constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, and w_{ij} is the weight coefficient, assigned by DA to given parameters (P_{ij}). The weight coefficient maximizes the distance between the means of the dependent variable. In this study, 12 water quality parameters

Table 2 | Structure matrix for the temporal DA of water quality variations across two wet and dry seasons

Standard mode		Stepwise mode	
Parameters	**Function 1**	**Parameters**	**Function 1**
WT	0.940	WT	1.000
Dissolved oxygen	−0.321	Dissolved oxygen	0.613
TKN	0.195	TKN	−0.300
Chl-a	0.164	Chl-a	−0.337
Ammonia-N	0.091	Ammonia-N	−0.234
Total phosphate	0.092	TP	−0.179
TP	0.088	Total phosphate	−0.157
Magnesium	−0.047	Magnesium	0.094
Sodium	−0.037	Sodium	−0.016
Chloride	−0.019	Chloride	−0.017
Turbidity	0.014	Turbidity	0.016
TSS	0.012	TSS	−0.046

Table 3 | CFs coefficients for the temporal DA of water quality variations in wet and dry seasons

	Standard mode		Stepwise mode	
Parameter	**Dry coef.[a]**	**Wet coef.[a]**	**Dry coef.[a]**	**Wet coef.a**
Chl-a	−0.586	−0.532	−0.116	0.046
Dissolved oxygen	9.121	8.514	3.338	1.985
TKN	27.702	25.936	24.694	27.791
TP	41.686	78.262	5.835	11.915
Total phosphate	−126.870	−168.289		
Ammonia-N	43.744	45.375		
WT	7.452	9.299	5.541	7.319
TSS	−0.107	−0.180		
Turbidity	0.307	0.568		
Magnesium	−0.082	−0.141		
Chloride	0.017	0.020		
Sodium	0.039	0.048		
(Constant)	−123.968	−165.618	−60.360	−104.781

[a]Fisher's linear discriminant functions coefficients for wet and dry seasons correspond to w_{ij} as defined in Equation (1).

Table 4 | CMs for the temporal DA of water quality variations in wet and dry seasons

		Season assigned by DA	
Monitoring season	**% correct**	**Dry season**	**Wet season**
Standard mode			
Dry season	89.6	88	8
Wet season	97.9	1	95
Total	95.3	89	103
Stepwise mode			
Dry season	87.5	84	12
Wet season	99.0	1	95
Total	93.2	85	107

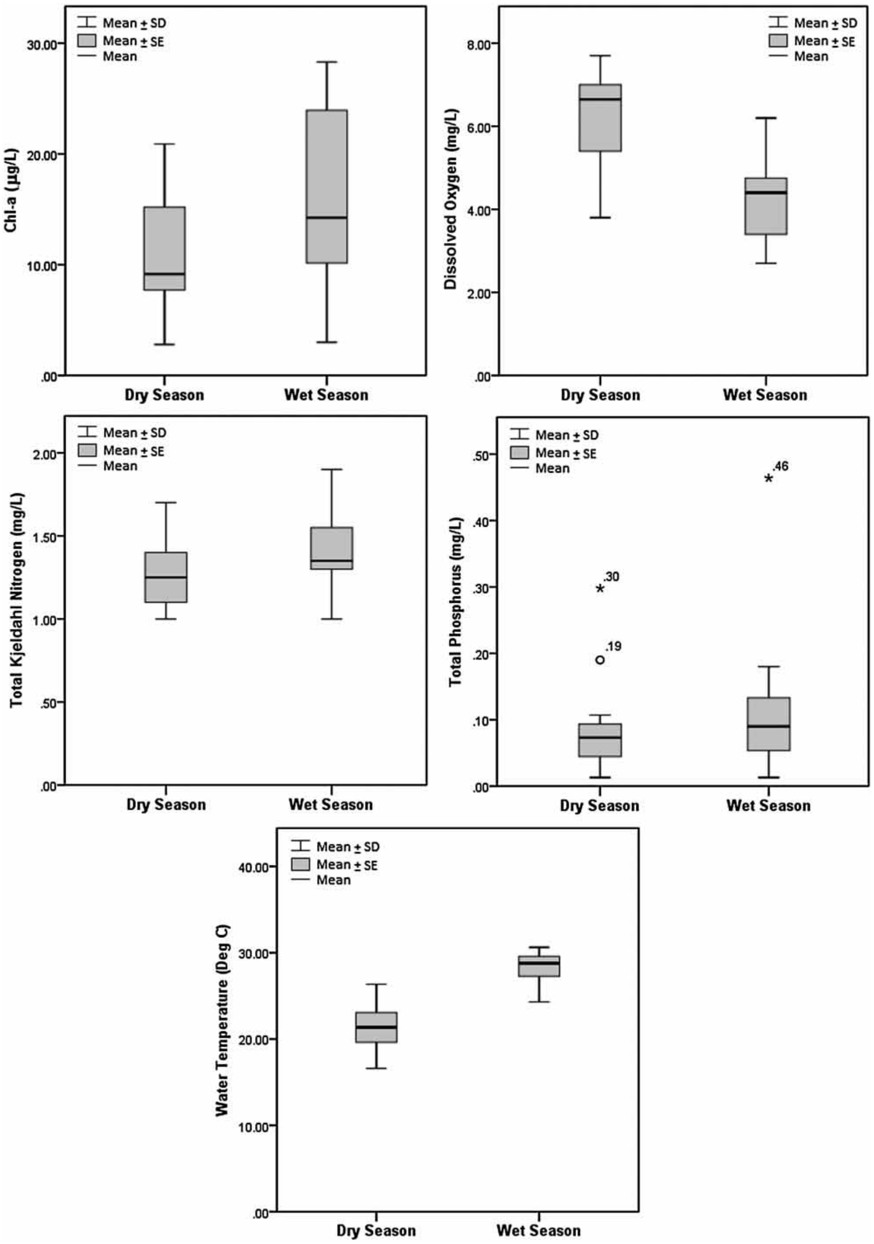

Figure 2 | Temporal variations in water quality of three major rivers of South Florida: chl-a; DO, TKN, TP, and WT.

Table 5 | Wilks' lambda and chi-square test for the spatial DA of water quality variations across three studied rivers in wet season

Mode	DFs	Canonical correlation (R)	Eigenvalue	Wilks' lambda	Chi-square	*p*-level (Sig.)
Standard mode	1	0.821	2.065	0.217	133.646	0.000
	2	0.578	0.503	0.665	35.635	0.000
Stepwise mode	1	0.787	1.625	0.306	107.666	0.000
	2	0.443	0.243	0.804	19.830	0.000

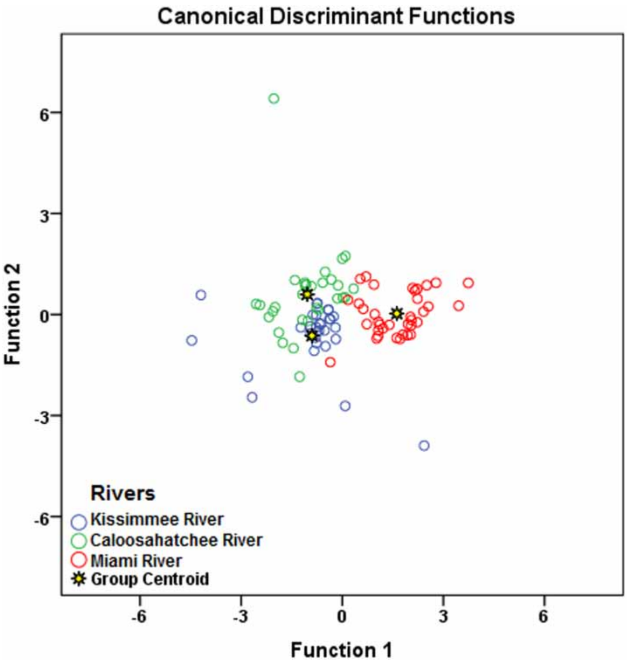

Figure 3 | Scatter plot for the spatial DA of water quality variations across three studied rivers in wet season (stepwise mode).

were considered for investigation including: chl-a, DO, TKN, TP, total phosphate, ammonia-N, WT, TSS, turbidity, magnesium, chloride, and sodium. Temporal DA was performed after dividing the whole data set into two groups of temporal (the wet season from May 15th to October 15th and the dry season from October 16th to May 14th). Furthermore, spatial DA was performed for each season data matrix based on three selected rivers of Kissimmee River, Caloosahatchee River, and Miami Canal, as spatial regions. The selected rivers (spatial) and the season (temporal) were the grouping (dependent) variables, whereas all the observed parameters constituted the independent variables. The SPSS 16.0 software package was employed for data treatment.

RESULTS AND DISCUSSION

Spatio-temporal variations in river water quality using DA

Temporal variations in water quality

Temporal variation in water quality was evaluated through DA. Both standard and stepwise modes were applied on the raw data after dividing the whole data set into two seasonal groups (wet and dry seasons). Season was the dependent variable, while all observed water quality parameters were independent variables. The values of Wilks' lambda and the chi-square statistic for each DF were obtained from each standard mode and stepwise mode. As shown in Table 1, the values varied from 0.222 to 0.244 and from 277.3 to 267.5

Table 6 | Structure matrix for the spatial DA of water quality variations across three studied rivers in wet season

Standard mode			Stepwise mode		
Parameters	**Function 1**	**Function 2**	**Parameters**	**Function 1**	**Function 2**
Total phosphate	−0.404	−0.070	Total phosphate	−0.451	−0.194
TP	−0.405	−0.038	TP	−0.453	−0.117
Sodium	−0.095	−0.206	Sodium	−0.098	−0.317
Chloride	−0.136	−0.125	Chloride	−0.200	−0.092
Magnesium	−0.081	0.070	Magnesium	−0.094	0.082
Ammonia-N	0.185	−0.038	Ammonia-N	0.210	−0.012
Dissolved oxygen	−0.139	0.333	Dissolved oxygen	−0.135	−0.009
Chl-a	−0.186	0.186	Chl-a	0.038	0.319
TKN	0.059	0.014	TKN	0.066	0.034
Turbidity	−0.227	−0.168	Turbidity	−0.125	0.119
TSS	−0.102	−0.142	TSS	0.133	0.070
WT	−0.094	0.256	WT	−0.051	0.218

Table 7 | CFs coefficients for the spatial DA of water quality variations across three studied rivers in wet season

	Standard mode			Stepwise mode		
Parameter	Kissimmee coef.[a]	Caloosahatchee coef.[a]	Miami coef.[a]	Kissimmee coef.[a]	Caloosahatchee coef.[a]	Miami coef.[a]
Chl-a	−1.401	−1.467	−1.549			
Dissolved oxygen	9.940	10.897	11.142			
TKN	58.595	65.589	75.088	41.668	44.773	51.713
TP	5.466	−21.418	−26.290	−18.42	−8.456	−72.102
Total phosphate	−158.888	−116.608	−186.954			
Ammonia-N	81.958	88.992	114.317	32.908	36.687	52.102
WT	17.873	18.044	18.041			
TSS	−0.773	−0.792	−1.006			
Turbidity	1.577	1.242	1.518			
Magnesium	−0.059	0.061	−0.049	0.182	0.26	0.197
Chloride	0.032	0.028	0.040			
Sodium	−0.033	−0.073	−0.047	−0.077	−0.108	−0.067
(Constant)	−297.426	−314.877	−323.796	−28.07	−33.348	−39.89

[a]Fisher's linear DFs coefficients for three groups of sites correspond to w_{ij} as defined in Equation (1).

for Wilks' lambda and the chi-square, respectively. Smaller values of Wilks' lambda indicate a greater discriminatory ability of the function. The associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. The small significance values (p-level <0.01) indicate that the DF does better than chance at separating the groups. Thus, the temporal DA was credible and effective. The first function in standard DA explained almost all ($R = 88.2\%$) of the total variance in the dependent variables. The stepwise DA had similar results, which indicated that 87% of the total group differences in the data set were explained by its first DF. Therefore, the first DF alone was sufficient to explain the difference of water quality among the two wet and dry seasons.

The stepwise DA identified five variables (chl-a, DO, TKN, TP and WT) as the most important discriminating variables. Table 2 shows that the first function in the stepwise DA was perfectly correlated with temperature (coefficient = 1.000), and then mostly correlated with DO (coefficient = 0.613). Classification functions (CFs) and the classification matrices (CMs) obtained from standard and stepwise modes of DA are shown in Tables 3 and 4. In the standard mode, all 12 variables were included to construct CFs which correctly classified 95.3% of the original grouped cases. In the stepwise mode, the DA correctly assigned 93.2% of the cases using only five discriminating variables. Therefore, the temporal DA results of the stepwise mode suggested that chl-a, DO, TKN, pH, and WT were the most significant parameters for discriminating differences between the wet season and dry season, and could be used

Table 8 | CMs for the spatial DA of water quality variations across three studied rivers in wet season

		Rivers assigned by DA		
Monitoring rivers	% correct	Kissimmee	Caloosahatchee	Miami
Standard mode				
Kissimmee	80.0	24	5	1
Caloosahatchee	76.7	5	23	2
Miami	83.3	2	4	30
Total	80.2	31	32	33
Stepwise mode				
Kissimmee	70.0	21	8	1
Caloosahatchee	70.0	7	21	2
Miami	94.4	1	1	34
Total	79.2	29	30	37

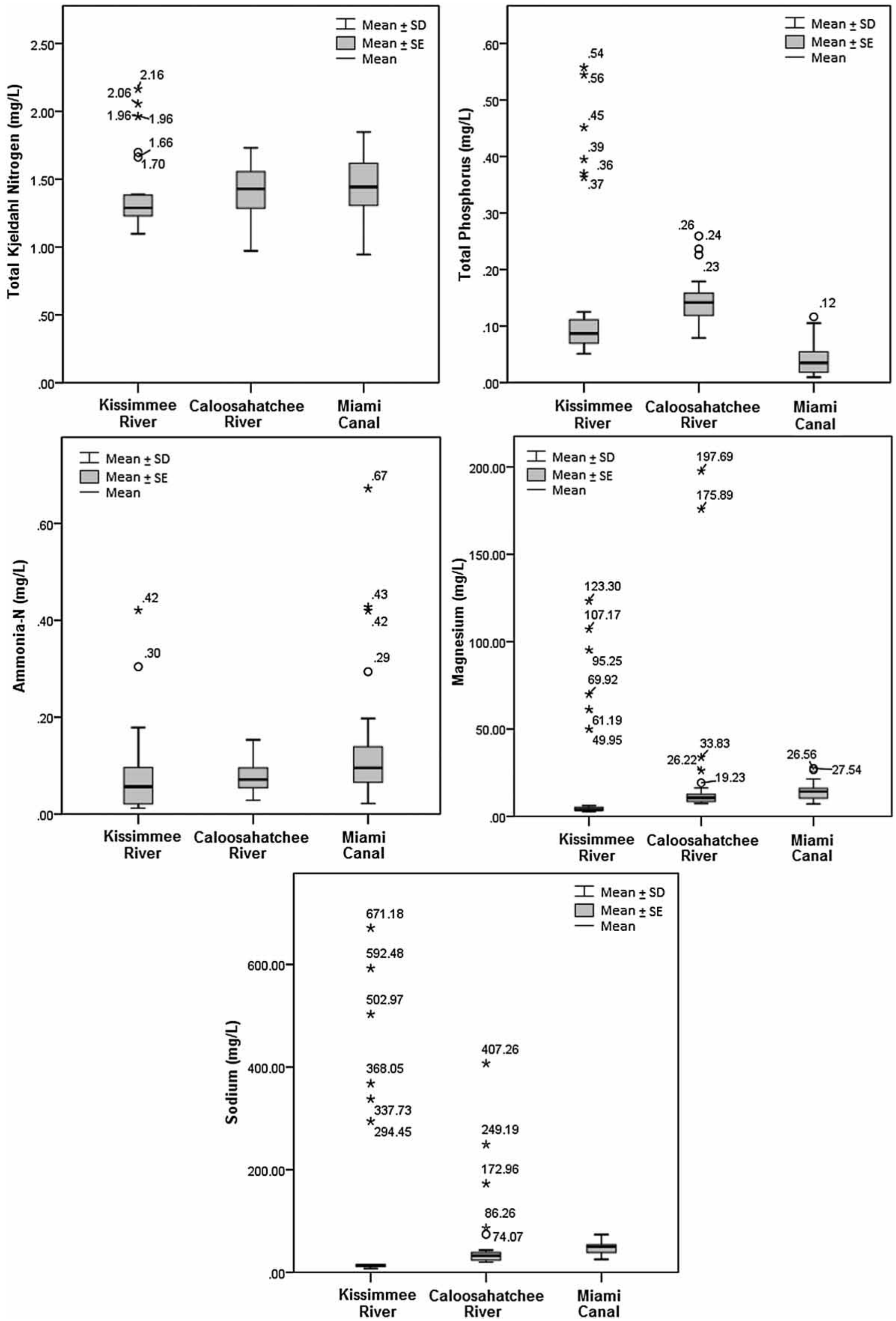

Figure 4 | Spatial variations in water quality of three studied rivers in wet season: TKN, TP, ammonia-N, magnesium, and sodium.

Table 9 | Wilks' lambda and chi-square test for the spatial DA of water quality variations across three studied rivers in dry season

Mode	DFs	Canonical correlation (R)	Eigenvalue	Wilks' lambda	Chi-square	*p*-level (Sig.)
Standard mode	1	0.837	2.348	0.168	156.314	0.000
	2	0.663	0.783	0.561	50.578	0.000
Stepwise mode	1	0.820	2.057	0.209	140.995	0.000
	2	0.601	0.567	0.638	40.419	0.000

to explain most of the expected temporal variations in water quality.

Box and whisker plots of discriminating parameters were constructed (stepwise mode) to evaluate different patterns associated with temporal variations in river water quality (Figure 2). The first pattern showed clear seasonal differences for chl-a, DO, and WT, in which chl-a and WT showed a clear inverse relationship with DO. This could be explained that as WT increases in the river, biological activity of aquatic organism strengthens, and therefore, the consumption of DO concentration increases. In addition, more oxygen dissolves in cooler water. The second pattern showed higher average concentrations of TKN and TP in the wet season. This could be due to the erosion of soil containing nutrients during the rain.

Spatial variations in water quality

Spatial variations in water quality between the studied rivers in wet season. Spatial variations in water quality between the studied rivers in wet and dry seasons were studied to evaluate the spatial patterns in the water quality of rivers. Three major rivers of South Florida were the grouping (dependent) variable, while the observed parameters in each season constituted the independent variables. Both standard and stepwise modes of DA were applied. First, the spatial variations in water quality between the studied rivers in the wet season were evaluated. As shown in Table 5, the values of Wilks' lambda and the chi-square for each DF varied from 0.217 to 0.804 and from 35.635 to 133.646. *P*-values were all less than 0.01, indicating that the spatial DA was credible and effective. In the stepwise DA, only five variables (TKN, TP, ammonia-N, magnesium, and sodium) were selected as the most important discriminating variables. The two DFs explained 78.7 and 44.3% of the group differences, respectively. The first DF separated Miami Canal from Kissimmee River and Caloosahatchee River (Figure 3), and was significantly and negatively correlated with total phosphate and TP (Table 6). The second DF established some separation between Kissimmee River and Caloosahatchee River, and was significantly correlated with chl-a. The CFs and CMs obtained from the two modes are shown in Tables 7 and 8. In the standard mode, all 12 variables were included and the constructed CFs produced 80.2% accuracy in assigning cases. However, in the stepwise mode, DA produced 79.2% correct assignment using only five discriminating variables.

Box and whisker plots of discriminating parameters identified by spatial DA (stepwise mode) were constructed to evaluate different patterns associated with variations in river water quality between three studied rivers in the wet season (Figure 4). The points are outliers. These are defined

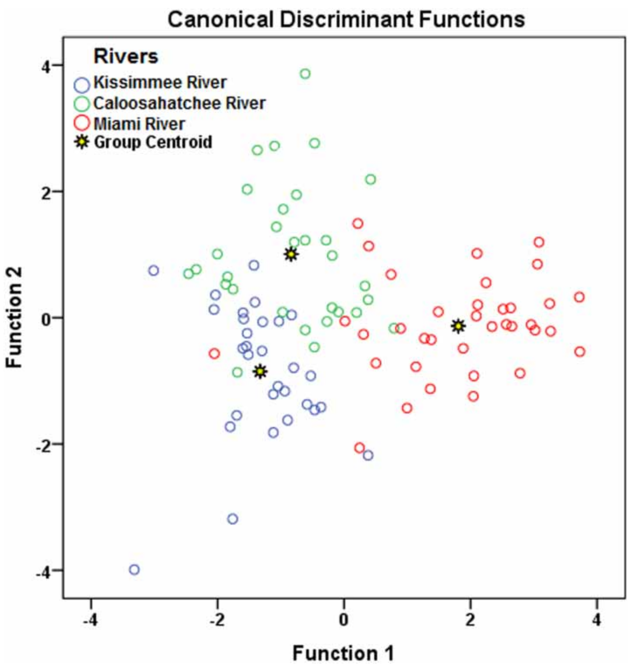

Figure 5 | Scatter plot for the spatial DA of water quality variations across three studied rivers in dry season (stepwise mode).

as values that do not fall in the inner fences. Outliers are extreme values. The asterisks or stars are extreme outliers. These represent cases that have values more than three times the height of the boxes.

The identified patterns of the most important discriminating variables show that, in general, the average concentrations of the water quality parameters in the Miami Canal are worse than the other two rivers. The Kissimmee River demonstrated even lower average values. Nonetheless, the outliers in Figure 4 are related to the average concentration of the represented variables in two highly polluted sites of S154C and CES03 in the Kissimmee River and the Caloosahatchee River, respectively. However, TP showed that a different pattern and average concentrations of this variable were much higher in the

Table 10 | Structure matrix for the spatial DA of water quality variations across three studied rivers in dry season

Standard mode			Stepwise mode		
Parameters	**Function 1**	**Function 2**	**Parameters**	**Function 1**	**Function 2**
Total phosphate	−0.297	−0.080	Total phosphate	−0.278	0.006
TP	−0.306	−0.033	TP	−0.327	−0.074
Sodium	−0.129	−0.052	Sodium	−0.211	0.004
Chloride	−0.158	−0.124	Chloride	−0.176	−0.129
Magnesium	−0.126	0.197	Magnesium	−0.135	0.226
Ammonia-N	0.210	−0.132	Ammonia-N	0.228	−0.157
Dissolved oxygen	−0.262	0.084	Dissolved oxygen	−0.285	0.107
Chl-a	−0.083	−0.250	Chl-a	−0.095	−0.274
TKN	0.450	0.000	TKN	0.493	−0.023
Turbidity	−0.147	0.008	Turbidity	−0.025	0.054
TSS	0.016	−0.003	TSS	−0.211	0.058
WT	0.134	0.311	WT	0.238	−0.001

Table 11 | CFs coefficients for the spatial DA of water quality variations across three studied rivers in dry season

	Standard mode			Stepwise mode		
Parameter	**Kissimmee coef.[a]**	**Caloosahatchee coef.[a]**	**Miami coef.[a]**	**Kissimmee coef.[a]**	**Caloosahatchee coef.[a]**	**Miami coef.[a]**
Chl-a	−0.722	−1.022	−0.738	−0.148	−0.368	−0.17
Dissolved oxygen	16.053	17.204	15.654	7.805	8.413	7.292
TKN	47.975	52.815	69.377	59.089	63.939	77.278
TP	−145.526	−194.743	−217.95	−76.051	−59.214	−114.523
Total phosphate	−45.610	28.722	−10.294			
Ammonia-N	124.317	140.425	165.566	113.794	128.452	156.259
WT	6.786	7.158	6.704			
TSS	0.194	0.191	0.338			
Turbidity	−0.298	−0.335	−0.620			
Magnesium	0.176	0.297	0.278	0.239	0.326	0.32
Chloride	0.033	0.018	0.021	−0.008	−0.023	−0.019
Sodium	0.007	−0.003	0.001			
(Constant)	−146.359	−166.618	−168.765	−61.155	−70.314	−80.84

[a]Fisher's linear DFs coefficients for three groups of sites correspond to w_{ij} as defined in Equation (1).

Caloosahatchee River and the Kissimmee River, respectively, in comparison to the Miami Canal. Besides the two mentioned highly polluted sites, TP was found higher at the Caloosahatchee River and the Kissimmee River than the Miami Canal sites. The percentage of agricultural and urbanized areas in the Miami Canal watershed was measured from the LULC map in GIS and was seen to be even more than the other two rivers. Therefore, this could be related to the effectiveness of eco-restoration projects implemented in its watershed and adjacent linked watersheds (the water conservation area-3, WCA-3) in order to decrease the amounts of nutrients.

Spatial variations in water quality between the studied rivers in dry season. As shown in Table 9, the values of Wilks' lambda and the chi-square for each DF varied from 0.168 to 0.638 and from 40.419 to 156.314. All *p*-values were less than 0.01, indicating that the spatial DA was credible and effective. In the stepwise DA, seven variables (chl-a, DO, TKN, TP, ammonia-N, magnesium, and chloride) were selected as the most important discriminating variables. The two DFs explained 82 and 60.1% of the group differences, respectively. Likewise, in the wet season, the first DF separated the Miami Canal from the Kissimmee River and the Caloosahatchee River (Figure 5), and was significantly correlated with TP, DO, and TKN (Table 10). The second DF established some separation between the Kissimmee River and the Caloosahatchee River, and was correlated with chl-a and magnesium. The CFs and CMs

Table 12 | CMs for the spatial DA of water quality variations across three studied rivers in dry season

		Rivers assigned by DA		
Monitoring rivers	**% correct**	**Kissimmee**	**Caloosahatchee**	**Miami**
Standard mode				
Kissimmee	83.3	25	4	1
Caloosahatchee	76.7	4	23	3
Miami	77.8	5	3	28
Total	79.2	34	30	32
Stepwise mode				
Kissimmee	80.0	24	5	1
Caloosahatchee	80.0	4	24	2
Miami	83.3	3	3	30
Total	81.3	31	32	33

obtained from the two modes are shown in Tables 11 and 12. In the standard mode, all 12 variables were included and the constructed CFs produced 79.2% accuracy in assigning cases. However, in the stepwise mode, DA produced 81.3% correct assignment using only seven discriminating variables.

Box and whisker plots of discriminating parameters identified by spatial DA (stepwise mode) were constructed to evaluate different patterns associated with variations in river water quality between the three studied rivers in the dry season (Figure 6). The points are outliers. These are defined as values that do not fall within the inner fences. Outliers are extreme values. The asterisks or stars are extreme outliers. These represent cases that have values more than three times the height of the boxes.

The first pattern showed clear spatial differences for chl-a and DO as a measure of life's vitality and the activity level of the plants and animals living in rivers. The higher average values of these two variables were found in the Kissimmee River and the Caloosahatchee River, which indicates the dynamism and strength of aquatic lives in this river. Besides, the Miami Canal had lower average concentrations of chl-a and DO, which indicated that organic pollution may play a significant role in the Miami canal, especially in urbanized areas which are under the influence of more domestic and industrial wastewater.

The second pattern showed higher average concentrations of TP in the Caloosahatchee River and the Kissimmee River, which consists of two highly polluted sites of CES03 and S154C, respectively. The asterisks or stars are extreme outliers observed in these two sites that represent cases that have values more than three times the height of the boxes. Previous analysis (Haji-Gholizadeh *et al.* 2016) indicated that these two highly polluted sites are extremely affected by urbanized areas, and also high-density environmental resource permits with more industrial effluent and domestic sewage.

The third identified pattern of the most important discriminating variables in the dry season showed that the average concentrations of TKN, magnesium, and chloride in the Miami Canal are worse than the other two rivers. The Kissimmee River demonstrated lower average values. Nonetheless, the outliers in Figure 6 are related to the average concentration of the represented variables in two highly

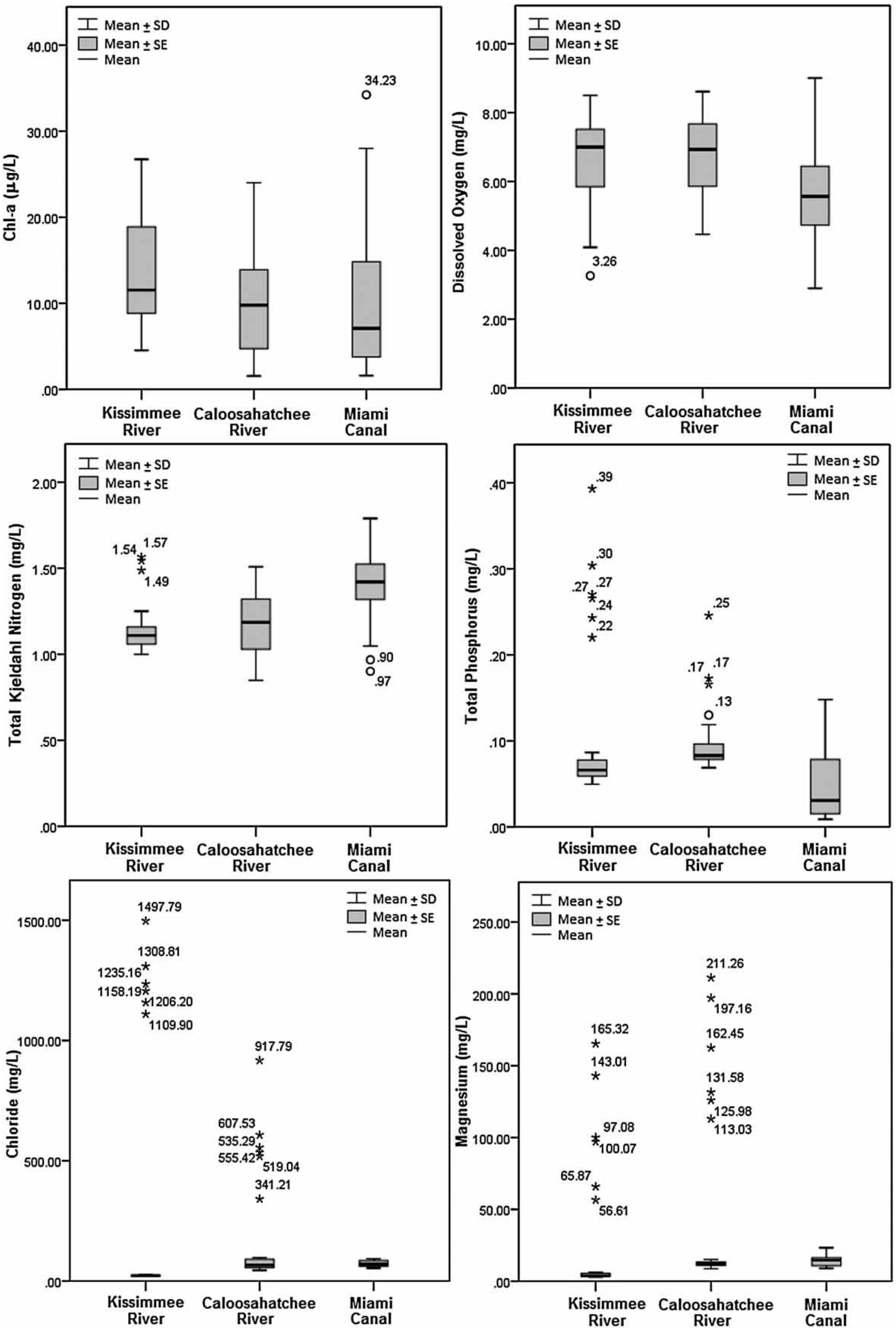

Figure 6 | Spatial variations in water quality of three studied rivers in dry season: chl-a, DO, TKN, TP, ammonia-N, magnesium, and chloride.

polluted sites of S154C and CES03 in the Kissimmee River and the Caloosahatchee River, respectively.

CONCLUSIONS

In this study, DA was applied to evaluate spatial and temporal variations in surface water quality of three major rivers of South Florida using 15 years (2000–2014) data sets of 12 water quality variables covering 16 sampling stations, with approximately 35,000 observations. DA, as an important data reduction method, was used to assess the water pollution status and analysis of its spatiotemporal variation. In temporal DA, 12 months of raw data were divided into two seasonal groups (wet and dry season) as the dependent variable, while all observed water quality parameters were independent variables. In the spatial DA, each river was separately considered as one spatial region to evaluate the patterns associated with spatial variations in each river's water quality. The three major rivers of South Florida were the grouping (dependent) variable, while the observed parameters in each season constituted the independent variables.

It was found by the stepwise DA that five variables (chl-a, DO, TKN, TP and WT) are the most important discriminating water quality parameters responsible for temporal variations. In spatial DA, the stepwise mode identified only five variables (TKN, TP, ammonia-N, magnesium, and sodium) and seven variables (hl-a, DO, TKN, TP, ammonia-N, magnesium, and chloride) as the most significant discriminating variables responsible for spatial variations in the wet and dry season, respectively. There were also different patterns associated with spatial variations that were identified, depending on the variables and considered season. However, it was found that the average concentrations of the most significant discriminating variables in both the wet and dry seasons in the Miami Canal was worse than the other two rivers, and the Kissimmee River demonstrated lower average values. Nonetheless, two highly polluted sites of S154C and CES03 in the Kissimmee River and the Caloosahatchee River require more attention and consideration.

This study showed the feasibility and reliability of DA in river water quality research. It is desirable that both state and local agencies pay more attention and consideration in order to improve and protect the vulnerable river quality. Additional studies will be required to assess precisely the unidentified sources of pollution and variation of further water quality parameters that were not analyzed in this study. Furthermore, the conclusion would be beneficial to water environment protection and water resources management in the future. The results of the spatial and temporal variations could be used to select the polluted areas and set the priority areas for the river water quality management in the study area.

ACKNOWLEDGEMENTS

The research was funded from Florida International University, Miami, USA. The observational data were obtained from South Florida Water Management District (SFWMD). We also thank the reviewers for providing insightful comments, as well as the conference officials.

REFERENCES

Carpenter, S., Caraco, N., Correll, D., Howarth, R. W., Sharpley, A. N. & Smith, V. H. 1998 Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* **8**, 559–568.

Chen, P., Li, L. & Zhang, H. 2015 Spatio-temporal variations and source apportionment of water pollution in Danjiangkou Reservoir Basin, Central China. *Water (Switzerland)* **7**, 2591–2611.

Haji-Gholizadeh, M., Melesse, M. A. & Reddi, L. 2016 Analysis of spatiotemporal trends of water quality parameters using cluster analysis in South Florida. In: *World Environmental and Water Resources Congress 2016*, pp. 519–528.

Jarvie, H. P., Whitton, B. A. & Neal, C. 1998 Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. *Sci. Total Environ.* **210–211**, 79–109.

Johnson, R. A. & Wichern, D. W. 1992 *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, NJ.

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E. & Mokhtar, M. 2011 Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ. Monit. Assess.* **173**, 625–641.

Kim, J.-H., Kim, R.-H., Lee, J., Cheong, T.-J., Yum, B.-W. & Chang, H.-W. 2005 Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrol. Process.* **19**, 1261–1276.

Kowalkowski, T., Zbytniewski, R., Szpejna, J. & Buszewski, B. 2006 Application of chemometrics in river water classification. *Water Res.* **40**, 744–752.

Li, Y., Xu, L. & Li, S. 2009 Water quality analysis of the Songhua River Basin using multivariate techniques. *J. Water Resour. Prot.* **1**, 110–121.

Massart, D. L. & Kaufman, L. 1983 *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis.* Wiley, New York.

Medina-Gomez, I. & Herrera-Silveira, J. A. 2003 Spatial characterization of water quality in a karstic coastal lagoon without anthropogenic disturbance: a multivariate approach. *Estuar. Coast. Shelf Sci.* **58**, 455–465.

Mustapha, A. & Aris, A. Z. 2012 Spatial aspects of surface water quality in the Jakara Basin, Nigeria using chemometric analysis. *J. Environ. Sci. Health. A. Tox. Hazard. Subst. Environ. Eng.* **47**, 1455–1465.

Parinet, B., Lhote, A. & Legube, B. 2004 Principal component analysis: an appropriate tool for water quality evaluation and management–application to a tropical lake system. *Ecol. Modell.* **178**, 295–311.

Schaefer, K. & Einax, J. W. 2010 Analytical and chemometric characterization of the Cruces River in South Chile. *Environ. Sci. Pollut. Res.* **17**, 115–123.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ. Model. Softw.* **22**, 464–475.

Shrestha, S., Kazama, F. & Nakamura, T. 2008 Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. *J. Hydroinform.* **10**, 43.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. 2003 Assessment of the surface water quality in Northern Greece. *Water Res.* **37**, 4119–4124.

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 Res.* **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. *Anal. Chim. Acta* **538**, 355–374.

Solidoro, C., Pastres, R., Cossarini, G. & Ciavatta, S. 2004 Seasonal and spatial variability of water quality parameters in the lagoon of Venice. *J. Mar. Syst.* **51**, 7–18.

Vega, M. 1998 Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* **32**, 3581–3592.

Wunderlin, A. D., Díaz, M. D. P., Amé, M. V., Pesce, F. S., Hued, A. C. & Bistoni, M. D. L. Á. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River basin (Córdoba-Argentina). *Water Res.* **35**, 2881–2894.

Zhang, X. A., Wang, Q. S., Liu, Y. F., Wu, J. & Yu, M. A. 2011 Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. *Environ. Monit. Assess.* **173**, 17–27.

Zhou, F., Liu, Y. & Guo, H. 2007 Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environ. Monit. Assess.* **132**, 1–13.

First received 26 February 2016; accepted in revised form 9 June 2016. Available online 29 July 2016