

Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam

Dung Phung · Cunrui Huang · Shannon Rutherford · Febi Dwirahmadi ·
Cordia Chu · Xiaoming Wang · Minh Nguyen · Nga Huy Nguyen ·
Cuong Manh Do · Trung Hieu Nguyen · Tuan Anh Diep Dinh

Received: 8 September 2014 / Accepted: 24 March 2015 / Published online: 7 April 2015
© Springer International Publishing Switzerland 2015

Abstract The present study is an evaluation of temporal/spatial variations of surface water quality using multivariate statistical techniques, comprising cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA). Eleven water quality parameters were monitored at 38 different sites in Can Tho City, a Mekong Delta area of Vietnam from 2008 to 2012. Hierarchical cluster analysis grouped the 38 sampling sites into three clusters, representing mixed urban-rural areas, agricultural areas and industrial zone. FA/PCA resulted in three latent factors for the entire research location, three for cluster 1, four for cluster 2, and four for cluster 3 explaining 60, 60.2, 80.9, and 70 % of the total variance in the respective water quality. The varifactors from FA indicated that the parameters responsible for water

quality variations are related to erosion from disturbed land or inflow of effluent from sewage plants and industry, discharges from wastewater treatment plants and domestic wastewater, agricultural activities and industrial effluents, and contamination by sewage waste with faecal coliform bacteria through sewer and septic systems. Discriminant analysis (DA) revealed that nephelometric turbidity units (NTU), chemical oxygen demand (COD) and NH_3 are the discriminating parameters in space, affording 67 % correct assignment in spatial analysis; pH and NO_2 are the discriminating parameters according to season, assigning approximately 60 % of cases correctly. The findings suggest a possible revised sampling strategy that can reduce the number of sampling sites and the indicator parameters responsible for large variations in water quality. This study demonstrates the usefulness of multivariate statistical techniques for evaluation of temporal/spatial variations in water quality assessment and management.

D. Phung (✉) · C. Huang · S. Rutherford · F. Dwirahmadi ·
C. Chu
Centre for Environment and Population Health (CEPH),
Griffith University, 179 Kessels Road, Nathan, Brisbane,
QLD 4111, Australia
e-mail: d.phung@griffith.edu.au

X. Wang · M. Nguyen
Commonwealth Scientific and Industrial Research
Organisation (CSIRO), Dickson, ACT 2602, Australia

N. H. Nguyen · C. M. Do
Health Environment Management Agency, Ministry of
Health, Ha Noi, Ha Tay, Vietnam

T. H. Nguyen · T. A. D. Dinh
Department of Environmental and Natural Resources
Management, Can Tho University, Can Tho, Vietnam

Keywords Water quality · Temporal assessment · Spatial assessment · Mekong Delta · Vietnam

Introduction

The quality of water in a river is highly affected by both natural and anthropogenic sources since a river system, comprising both the main course and the tributaries, carries the one-way flow of a significant load of matter in dissolved and particulate phases. This load reflects the

major influences on the system: the lithology of the basin, atmospheric inputs, climatic conditions and anthropogenic inputs (Bricker and Jones 1995). Human activities are the major factors determining the quality of the river surface water through effluent discharges, use of agricultural chemicals, erosion of soils and land use (Niemi et al. 1990), whereas rainfall which causes surface water runoff is a seasonal phenomenon, largely affected by climate within the basin (Karbasi et al. 2008; Najafpour et al. 2008). Since the system of natural environment and anthropogenic sources is multivariate and complex, river quality assessment and management require a fundamental understanding of spatial and temporal variations in the water characteristics, including hydro-morphological, chemical and biological parameters. For this purpose, a monitoring program which can provide reliable estimation of the parameters of surface water quality is necessary. However, the program requires a large number of parameters to be measured, analysed and interpreted using multivariate methods (Shrestha and Kazama 2007; Simeonov et al. 2004).

The different multivariate statistical methods, comprising cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), have been reported to be effective methods to facilitate complex data matrices for characterizing and evaluating surface water quality, and they are useful in interpreting temporal and spatial variations of water quality parameters caused by natural and anthropogenic factors (Wunderlin et al. 2001; Helena et al. 2000; Lee et al. 2001; Adam et al. 2001; Simeonov et al. 2004; Singh et al. 2004; Reghunath et al. 2002; Papatheodorou et al. 2007). Thus, these methods have been considered as valuable tools for reliable assessment and management of water resources in order to help develop rapid solutions to pollution problems. Some recent studies applying multivariate statistical approaches to evaluate river water quality have been reported. Singh et al. (2004, 2005) used CA, PCA/FA and DA to evaluate spatial and temporal variations in water quality of the Gomti River (India). Similarly, Shrestha and Kazama (2007) also used multivariate statistical techniques to evaluate the surface water quality of the Fuji River Basin in Japan. Qadir et al. (2008) applied multivariate statistical techniques to assess spatio-temporal variations in water quality of the river Chenab, Pakistan, and Pejman et al. (2009) also evaluated spatial and seasonal variation in surface water quality using multivariate statistical techniques.

This study aims to assess the spatial and temporal variations of surface water quality in 38 sampling sites of the Mekong River throughout Can Tho City, in the central area of the Mekong Delta Basin in Vietnam. The data set obtained, comprising 11 parameters of water quality, which were monitored quarterly at key points throughout the city during the period of 2008–2011, was analysed using different multivariate statistical techniques to (i) examine the similarities or dissimilarities between sampling sites and season, (ii) identify water quality variables responsible for spatial and temporal variations in river water quality and (iii) evaluate the influence of possible sources on the water quality parameters in the research location.

Methods

Study area and monitoring sites

The study was conducted in Can Tho City, a typical city in Mekong Delta in Vietnam. Can Tho is the fourth most populous city in Vietnam with a total area of 1,411.49 km², a population of 1,188,390 people and a population density of 842 people/km² (Huyen 2012). In the heart of the Mekong Delta, where the river system is intertwined, Can Tho residential life is strongly affected by hydrological factors. Can Tho City has a tropical monsoon climate with two main seasons annually which comprise dry season (December–April) and wet season (May–November).

In the heavily urbanized areas of Can Tho City, there are reticulated water supplies but there is no centralized wastewater treatment, with sewers discharging into local waterways. In areas outside of the urban districts, there is inadequate access to water and sanitation services. Many households in rural and peri-urban areas depend on a range of water sources including rainwater, groundwater, piped water and river surface water. Similar to other lower Mekong Delta Basin localities, the main sources of surface water pollution in Can Tho comprise urban waste, industrial waste water and agricultural activities.

Surface water quality in Can Tho was quarterly monitored by the Can Tho Centre for Natural Resources and Environment, and the sampling method was conducted according to the guidelines of the National Technical Regulation on Surface Water Quality in Vietnam (VNMONRE 2008) (Table 1).

Table 1 Mean (standard deviation) values of water quality parameters at different sampling sites of the Mekong River in Can Tho City, Mekong Delta, Vietnam (2008–2012)

Sampling site	pH	BOD ₅ 20 °C (mg/l)	COD (mg/l)	Turbidity (NTU)	DO (mg/l)	SS (mg/l)	Fe _{ce} (mg/l)	NH ₃ -N (mg/l)	NO ₃ ⁻ -N (mg/l)	NO ₂ ⁻ -N (mg/l)	Total coliform (MPN/100 ml)
1	6.9 (0.45)	10.6 (4.4)	14.1 (5.3)	45.1 (32.8)	4.1 (1.2)	51.2 (37.6)	0.37 (0.34)	1.12 (1.4)	1.88 (3.3)	0.19 (0.35)	3.5 × 10 ⁴ (5.3 × 10 ⁴)
2	6.7 (0.36)	26.2 (20.6)	38.7 (29.3)	54.9 (31.5)	2.7 (0.95)	61.9 (36.5)	0.84 (0.75)	2.7 (1.8)	1.5 (0.92)	0.09 (0.12)	3.9 × 10 ⁵ (1.2 × 10 ⁵)
3	6.8 (0.33)	13.2 (5.1)	17.1 (6.2)	42.5 (24.9)	3.9 (1.3)	50.3 (31.4)	0.54 (0.54)	0.96 (0.79)	1.05 (0.5)	0.06 (0.06)	4.8 × 10 ⁴ (9.6 × 10 ⁴)
4	6.9 (0.37)	11.5 (8.2)	15.2 (8.9)	36.7 (20.3)	4 (1.2)	46.4 (21.6)	0.59 (0.49)	0.53 (0.67)	0.65 (0.41)	0.04 (0.04)	1.5 × 10 ⁴ (2.3 × 10 ⁴)
5	6.9 (0.35)	12.9 (6.9)	17.5 (9.8)	32.9 (21.1)	4.3 (1.1)	33.1 (18.4)	0.46 (0.44)	0.29 (0.34)	0.77 (0.46)	0.05 (0.06)	1.2 × 10 ⁴ (2.2 × 10 ⁴)
6	6.9 (0.4)	11.3 (6.5)	16.5 (8.9)	32.5 (22.7)	4.1 (1.1)	31.7 (20.9)	0.46 (0.41)	0.24 (0.33)	2.3 (0.85)	0.05 (0.06)	1.3 × 10 ⁴ (2.1 × 10 ⁴)
7	6.9 (0.4)	10.5 (4.3)	15.5 (6.2)	36.3 (16.5)	4.3 (0.9)	40.4 (21.2)	0.81 (1)	0.32 (0.47)	0.76 (0.36)	0.05 (0.06)	8.2 × 10 ³ (2 × 10 ⁴)
8	6.9 (0.41)	10.9 (4.4)	15.5 (5.5)	42.3 (25.3)	4.1 (1.03)	41.2 (20.9)	0.82 (0.68)	0.39 (0.52)	0.8 (0.45)	0.04 (0.04)	1.9 × 10 ⁴ (4.6 × 10 ⁴)
9	6.9 (0.33)	14.1 (4.58)	19 (5.4)	43.5 (20.7)	4 (0.87)	41.5 (20.4)	0.5 (0.4)	1 (0.77)	1.04 (0.47)	0.08 (0.09)	1.1 × 10 ⁴ (1.7 × 10 ⁴)
10	6.9 (0.37)	10.3 (5.2)	14.4 (6.2)	39.4 (24.5)	4.4 (1.4)	41.1 (21.9)	0.37 (0.44)	0.54 (0.73)	1.05 (0.67)	0.05 (0.07)	1.8 × 10 ⁴ (5.2 × 10 ⁴)
11	6.7 (0.39)	10.1 (4.9)	14 (5.4)	37 (19.9)	4.57 (1.09)	39.8 (17)	0.23 (0.3)	0.29 (0.32)	1.7 (2.2)	0.2 (0.41)	1.2 × 10 ⁴ (1.6 × 10 ⁴)
12	6.9 (0.36)	10.3 (2.9)	14.6 (3.5)	43.1 (21.8)	4.2 (0.89)	45.1 (17.9)	0.36 (0.39)	0.46 (0.42)	1.08 (0.52)	0.09 (0.16)	1.2 × 10 ⁵ (5.3 × 10 ⁵)
13	6.9 (0.35)	10.4 (4.3)	14.3 (5.5)	43.4 (25.7)	4.1 (1.2)	42.7 (21.2)	0.42 (0.45)	0.43 (0.49)	1.03 (0.63)	0.06 (0.13)	9 × 10 ³ (8.3 × 10 ³)
14	7 (0.38)	10.5 (5)	14.7 (14.7)	42.5 (24.8)	4.3 (1.03)	42.4 (21.4)	0.41 (0.36)	0.41 (0.47)	1.09 (0.77)	0.113 (0.24)	1.5 × 10 ⁴ (2.5 × 10 ⁴)
15	7 (0.39)	10.4 (5.1)	14 (6.2)	43.2 (28.5)	4.5 (1.08)	40.9 (23.2)	0.42 (0.44)	0.46 (0.54)	1.35 (1.1)	0.06 (0.08)	8 × 10 ³ (1.6 × 10 ⁴)
16	6.9 (0.4)	17.1 (5.6)	23 (6.9)	48 (26.8)	3.8 (0.88)	45.2 (24.9)	0.56 (0.42)	0.68 (0.6)	1.34 (1.1)	0.08 (0.14)	2.4 × 10 ⁴ (4 × 10 ³)
17	6.9 (0.35)	12.1 (4.68)	16.7 (5.8)	41.5 (20.7)	4.2 (0.9)	41.1 (19.1)	0.39 (0.36)	0.79 (0.73)	1.1 (0.7)	0.06 (0.06)	8 × 10 ³ (1 × 10 ⁴)
18	6.9 (0.3)	15.6 (6.2)	21.3 (7.8)	46.1 (22.3)	3.5 (0.6)	45.1 (20.1)	0.66 (0.59)	0.99 (0.7)	1.1 (0.95)	0.06 (0.08)	1.2 × 10 ⁴ (2.5 × 10 ⁴)
19	6.8 (0.35)	17.1 (7.4)	22.9 (9.1)	66.5 (50.4)	4.5 (4.2)	56.5 (35.1)	0.69 (0.64)	0.9 (0.63)	1.5 (1.1)	0.03 (0.03)	3.2 × 10 ⁴ (1 × 10 ⁵)
20	6.8 (0.39)	14.6 (5.9)	19.7 (6.8)	59.5 (35.6)	3.9 (0.8)	54.2 (32.3)	0.69 (0.5)	0.79 (0.59)	1.2 (0.96)	0.06 (0.15)	1.6 × 10 ⁴ (3.2 × 10 ⁴)
21	7 (0.37)	9.2 (3.38)	12.4 (3.7)	47 (28.6)	4.4 (1.3)	41.8 (23.7)	0.38 (0.36)	0.42 (0.49)	0.89 (0.46)	0.05 (0.06)	2 × 10 ⁴ (5.5 × 10 ⁴)
22	6.9 (0.57)	11.2 (4.1)	16.6 (6)	46.8 (26.7)	4.1 (1.2)	45.9 (25.2)	0.44 (0.33)	0.65 (0.56)	0.92 (0.48)	0.05 (0.07)	2.5 × 10 ⁵ (1 × 10 ⁶)
23	6.8 (0.34)	11.1 (3.25)	16 (4.95)	55.4 (32.8)	4.1 (1.2)	53.1 (42.4)	0.44 (0.49)	0.6 (0.49)	1.1 (0.66)	0.05 (0.1)	2.4 × 10 ⁴ (3.3 × 10 ⁴)
24	6.8 (0.38)	12 (3.3)	17.4 (4.5)	46.9 (19.8)	4 (1.1)	43.7 (21.9)	0.39 (0.29)	0.65 (0.55)	1.1 (0.58)	0.07 (0.16)	4.5 × 10 ⁴ (8.2 × 10 ⁴)
25	6.8 (0.32)	10.1 (3.2)	13.5 (3.4)	42.3 (19.7)	4.5 (1.2)	38.9 (22.9)	0.38 (0.38)	0.66 (0.92)	0.9 (0.59)	0.04 (0.05)	3.2 × 10 ⁴ (5.7 × 10 ⁴)
26	6.8 (0.3)	8.8 (2.94)	11.9 (4.6)	42.1 (20.6)	4.38 (1.1)	37.3 (18.4)	0.49 (0.56)	0.48 (0.39)	1.04 (0.75)	0.04 (0.05)	2.5 × 10 ⁴ (5.4 × 10 ⁴)
27	6.8 (0.4)	9.4 (3.9)	12.9 (4.8)	56.1 (38.1)	4 (1.3)	49.1 (32.1)	0.42 (0.36)	0.33 (0.27)	1.2 (0.94)	0.06 (0.1)	7.6 × 10 ⁴ (2.4 × 10 ⁵)
28	6.8 (0.42)	9.5 (2.8)	13.2 (3.1)	41.5 (14.9)	4.1 (1.1)	38.7 (15.7)	0.42 (0.3)	0.62 (0.58)	1.35 (0.9)	0.08 (0.13)	1.2 × 10 ⁴ (1.8 × 10 ⁴)
29	6.9 (0.41)	8.8 (3.4)	12.5 (3.9)	36.2 (12.1)	5.3 (4.1)	37.2 (10.9)	0.48 (0.41)	0.26 (0.48)	0.9 (0.6)	0.06 (0.09)	3.5 × 10 ⁴ (1 × 10 ⁵)
30	6.9 (0.42)	11.2 (3.6)	15.8 (4.8)	46.6 (26.7)	4 (0.86)	53.4 (25)	0.49 (0.49)	0.31 (0.44)	0.88 (0.41)	0.05 (0.05)	2.4 × 10 ⁴ (5.3 × 10 ⁴)

Table 1 (continued)

	pH	BOD ₅ 20 °C (mg/l)	COD (mg/l)	Turbidity (NTU)	DO (mg/l)	SS (mg/l)	Fe _c (mg/l)	NH ₃ -N (mg/l)	NO ₃ ⁻ -N (mg/l)	NO ₂ ⁻ -N (mg/l)	Total coliform (MPN/100 ml)
31	6.9 (0.39)	9.9 (2.9)	14.5 (4.3)	43.2 (12.4)	4.3 (1.04)	46.1 (12.7)	0.56 (0.39)	0.33 (0.35)	1.1 (0.42)	0.04 (0.04)	1.3 × 10 ⁴ (1.2 × 10 ⁴)
32	6.9 (0.38)	10.5 (3.57)	14.8 (4.4)	44.6 (28.9)	4.1 (1.05)	47.9 (26.9)	0.72 (0.75)	0.46 (0.48)	0.9 (0.39)	0.06 (0.07)	5 × 10 ⁴ (1.1 × 10 ⁵)
33	6.8 (0.35)	11.5 (2.9)	15.8 (3.4)	46.7 (25.1)	3.9 (1)	47.9 (24.5)	0.67 (0.5)	0.39 (0.27)	0.79 (0.35)	0.08 (0.07)	4 × 10 ⁴ (1 × 10 ⁵)
34	6.8 (0.34)	11.9 (3.35)	16.9 (3.9)	57.2 (37.7)	3.9 (0.84)	63.9 (38.4)	0.66 (0.44)	0.46 (0.53)	0.9 (0.5)	0.16 (0.27)	1 × 10 ⁴ (1.3 × 10 ⁴)
35	6.9 (0.35)	11.7 (3.73)	16.1 (3.9)	65.6 (62.9)	3.9 (0.97)	69.6 (70.9)	0.95 (0.86)	0.5 (0.6)	0.89 (0.5)	0.07 (0.08)	3.4 × 10 ⁴ (5.6 × 10 ⁴)
36	6.9 (0.33)	11.5 (3.1)	16.3 (3.47)	64.2 (59.5)	4.78 (5.47)	74.8 (54.1)	0.8 (0.7)	0.39 (0.57)	0.9 (0.4)	0.07 (0.08)	4.6 × 10 ⁴ (7 × 10 ⁴)
37	6.9 (0.44)	10.7 (3.6)	14.4 (4.7)	74.7 (89.1)	3.9 (0.82)	70.4 (81.9)	0.79 (0.81)	0.6 (1.64)	1.2 (0.9)	0.05 (0.05)	5.6 × 10 ⁴ (2 × 10 ⁵)
38	7 (0.44)	10.5 (3.38)	14.5 (4.58)	66.9 (76.7)	4.1 (0.68)	69.5 (86.1)	0.69 (0.84)	0.35 (0.42)	1.3 (0.93)	0.04 (0.04)	1.3 × 10 ⁴ (2.2 × 10 ⁴)
National guidelines											
A1	-	4	10	-	≥6	20	0.5	0.1	2	0.01	2500
A2	-	6	15	-	≥5	30	1.0	0.2	5	0.02	5000
B1	-	15	30	-	≥4	50	1.5	0.5	10	0.04	7500
B2	-	25	50	-	≥2	100	2.0	1	15	0.05	10,000

Note: **A1** For residential use and other purpose; **A2**: For residential use with proper treatment, preservation of aquatic plants or other purposes; **B1**: For irrigation or other purposes requiring the similar quality of water or other purposes; **B2**: For water transport and other purposes requiring low quality

BOD biochemical oxygen demand, *COD* chemical oxygen demand, *NTU* nephelometric turbidity units, *DO* dissolved oxygen, *SS* suspended solids

Water samples from the 38 sites (Fig. 1), which were located at key points throughout the city, were collected quarterly from January 2008 to December 2012 (Fig. 1). The data sets of the 38 water quality monitoring stations comprise 11 water quality parameters: pH, dissolved oxygen (DO) (i.e. in situ DO readings given in milligrams per litre and auto-compensated or corrected for variations), biochemical oxygen demand (BOD), chemical oxygen demand (COD), nephelometric turbidity units (NTU), suspended solids (SS), nitrate (NO₃) and nitrite-nitrogen (NO₂), ammonical nitrogen (NH₃), total Fe and total coliform.

Statistical analysis

Multivariate analysis of the river water quality data was conducted using four statistical techniques: cluster analysis (CA), factor analysis (FA), principal component analysis (PCA) and discriminant analysis (DA).

Cluster analysis (CA)

Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess (Shrestha and Kazama 2007). The similarity between two samples is usually evaluated by Euclidean distance which can be represented by the difference between analytical values derived from the samples (Otto 1998). This study performed hierarchical agglomerative CA on the normalized log transformation of water quality parameters using Ward's method, in which squared Euclidean distance was used as the measure of similarity (Singh et al. 2004). The spatial variability of water quality in the river basin was identified from CA, using linkage distance, reported as D_{link}/D_{max} representing the quotient between linkage distances to the maximal linkage. The quotient was then multiplied by 100 as an approach to standardize the linkage distance represented on the y-axis (Simeonov et al. 2004; Wunderlin et al. 2001). The previous studies indicated that CA reliably classifies surface water

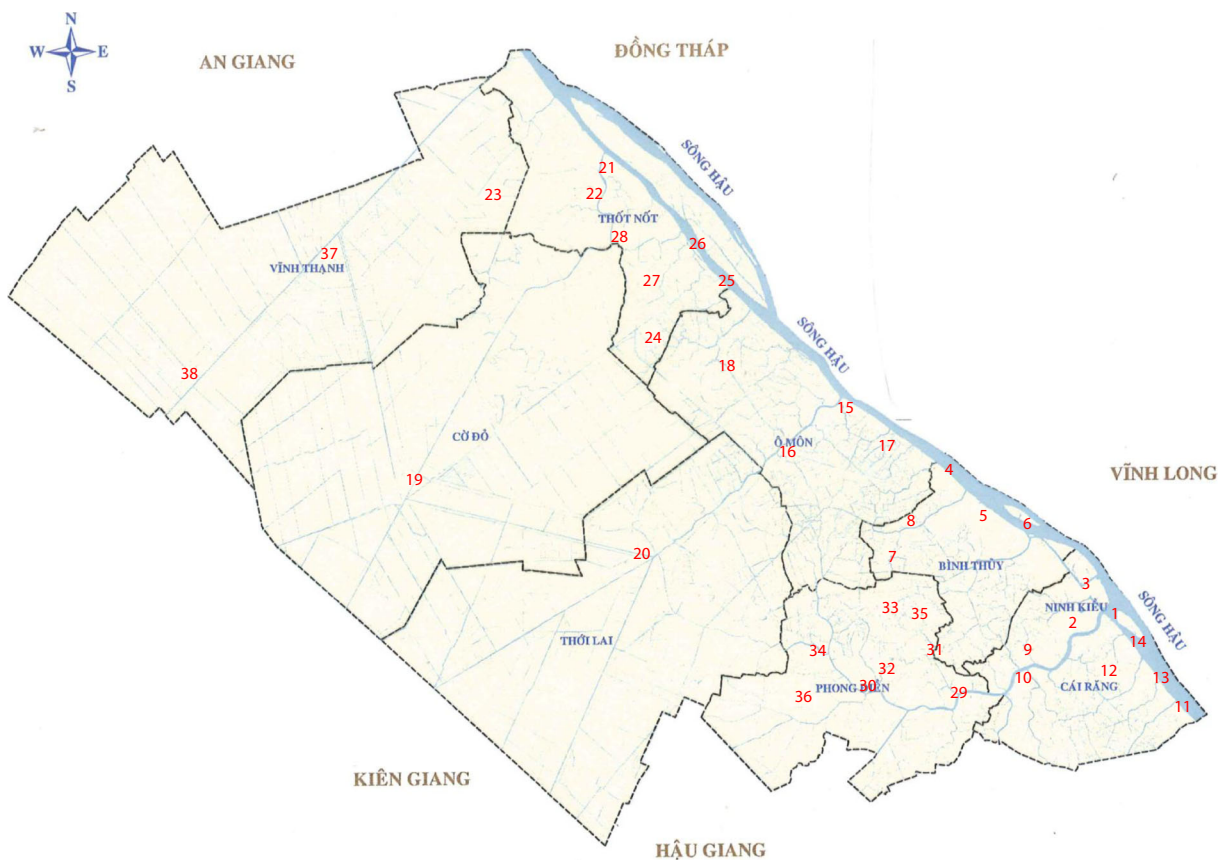


Fig. 1 Sampling sites for monitored surface water in Can Tho City, Mekong Delta, Vietnam

quality and can guide future sampling strategies (Simeonov et al. 2004; Singh et al. 2004; Wunderlin et al. 2001).

Factor analysis/principal component analysis

Factor analysis/principal component analysis (PCA/FA) was used to identify the important quality parameters of surface water and to investigate the possible sources of different pollutants, and this technique was performed on all research locations and each cluster identified from the previous CA analysis. The PCA technique (Wunderlin et al. 2001; Singh et al. 2004) starts with a covariance matrix presenting the dispersion of the original variables (water quality parameters), and then extracting the eigenvalues and eigenvectors. Principal components (PCs) are obtained by multiplying an eigenvector, which is a list of coefficients (loading or weighting), by the original correlated variables. A principal component is the product of the original data and an eigenvector is the result of the data projected on to a new axis. The number of PCs is the same as the number of the original variables, but the results provide information about the most meaningful parameters (Vega et al. 1998). In the following step of PCA/FA, FA gives factors meaning by the transformation process using Varimax rotation, in which FA reduces the contribution of less significant variables obtained from PCA and the new groups of variables, so-called varimax factors (VFs), are generated by rotating the axis defined by PCA. The VFs can include unobservable, hypothetical, and latent variables (Vega et al. 1998; Wunderlin et al. 2001; Singh et al. 2004; Helena et al. 2000).

Discriminant analysis

DA was used to determine the surface water quality parameters which best discriminate between groups identified by CA. The principle of DA is to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict membership (Qadir et al. 2008; Singh et al. 2004). DA was applied to the water quality data set without any standardization to define spatial and temporal variations in these parameters. DA was performed to construct discriminant functions (DFs) using standard, forward stepwise and backward stepwise modes. DFs were calculated using the equation below:

$$f(G_i) = ki + \sum_{j=1}^n w_{ij}P_{ij}$$

where i is the number of groups(G), ki the constant inherent to each group, n the number of parameters used to classify a set of data into a given group, w_i the weight coefficient, assigned by DA to a given parameter (p_i).

In DA, the sites which were identified by CA were taken as spatial grouping variables and seasons (wet and dry) were considered as temporal grouping variables. These grouping variables (spatial and temporal) were used in the analysis as dependent variables, and the water quality parameters were considered independent variables. In forward stepwise mode, DA variables were simultaneously added until no significant change occurs, while in backward stepwise mode, variables were removed starting from the least significant until significant change occurred.

Results and discussion

Overall water quality evaluation

A total of 608 water samples were collected from 38 sampling sites during the 4 years (2008–2012) and analysed for 11 parameters (pH, NTU, SS, Fe, DO, COD, BOD, NH₃, NO₃, NO₂, coliform). The descriptive statistics such as mean and standard deviation of water quality parameters are summarized in Table 1. The results shows that NO₃⁻ is within the permissible levels, while all other parameters—BOD (mean range, 8.8–26.2 mg/l), COD (mean range, 12.5–38.7 mg/l), DO (mean range, 2.7–5.3 mg/l), SS (mean range, 31.7–61.9 mg/l), Fe (mean range, 0.37–0.84 mg/l), NH₃⁻ (mean range, 0.24–2.7), NO₂⁻ (mean range, 0.04–0.19 mg/l) and total coliform (mean range, 8×10^3 – 3.9×10^5 MPN/100 ml)—exceed on average the levels set by national guidelines for residential use and other purposes. It is noteworthy that the BOD levels are quite high, with a minimum of 8.8 mg/l. However, the dissolved oxygen content of the water is relatively less with a minimum of 2.7 mg/l.

Spatial similarity of monitoring sites (CA)

Cluster analysis (CA) rendered a dendrogram which grouped the 38 sites into three clusters at (D_{link}/D_{max})

<60 (Fig. 2). The clustering procedure produced three groups of sites in a persuasive way, since the sites in these groups have similar features and natural background source types. Cluster 1 (C1) comprises 18 monitoring sites in five districts (Thot Not, Co Do, Thoi Lai, Phong Dien), which are mixed urban and rural areas. Cluster 2 (C2) is a small cluster comprising only two monitoring sites at Vinh Thanh District where agricultural land use is predominant. However, since this cluster is a predominantly agricultural district, the number of monitoring sites appears reasonable, with any increase in pollutants depending on the agricultural seasons and crops. Therefore, local authorities, experts, and field technicians will decide the number and location of additional sampling sites based on the real field situation. Cluster 3 (C3) comprises 18 monitoring sites in four districts (Cai Rang, Ninh Kieu, Binh Thuy) where urban and industrial zones are predominant. This implies that for rapid assessment of water quality, a reduced number of monitoring sites in each cluster may serve for spatial

assessment of the water quality of the whole network. The evidence also offers an opportunity to design a future spatial sampling strategy in an optimal manner, leading to a more cost-effective water monitoring program in this Mekong Delta area. The results of this current study correspond to the successful application of this approach in water quality programs reported from previous studies (Wunderlin et al. 2001; Simeonov et al. 2004).

Identification of potential sources (PCA/FA)

Waters are generally low in dissolved oxygen (means, 2.7–4.57 mg/l), reflecting organic loads, as indicated by BOD₅ and COD levels. The mean values of these parameters appear to be co-related. Elevated nutrient measurements in the surface are limited to inorganic forms of nitrogen. Clearly, there are abundant non-point and point sources of nitrogen (and phosphorus) nutrients in a mixed urban and agricultural environment. Ubiquitous

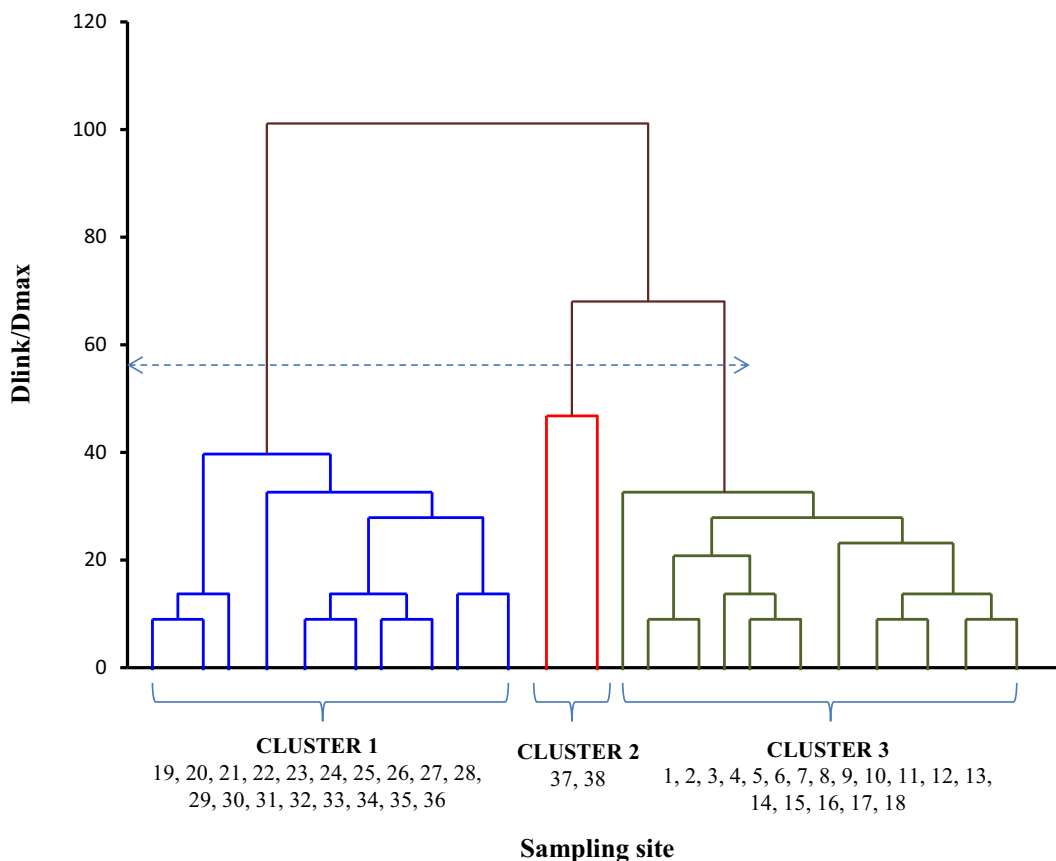


Fig. 2 Dendrogram showing clustering of sampling sites according to surface water quality characteristics of the Mekong Delta River in Can Tho City

microbiological contamination is indicated by coliform counts, presumably related to faecal sources.

Principal component analysis (PCA/FA) was conducted on the normalized log transformation of water quality parameters (11 variables) separately for the entire research location and for individual clusters, as delineated by CA techniques, to compare the compositional pattern between analysed water samples and to identify the factors influencing each one. The input data matrices (variables*cases) for PCA/FA were [11*760] for the whole research location, [11*460] for C1, [11*40] for C2 and [11*260] for C3.

PCA generated three PCs for the entire research location, three PCs for cluster 1, four PCs for cluster 2, and four PCs for cluster 3 with eigenvalues >1 explaining 60, 60.2, 80.9, and 70 % of the total variance in the respective water quality data sets. An example of the Scree plot representing eigenvalues for the entire research location is shown in Fig. 3. An eigenvalue representing a measure of the significance of the factor is considered significant if it is greater than 1.0 (Kim and Muller 1987). The highest eigenvalues are the most significant. Equal numbers of varifactors (VFs) were then identified for all cases through FA performed on the PCs. Corresponding VFs, variable loadings and

explained variance are presented in Table 2. Liu et al. (2003) classified the factor loadings as 'strong', 'moderate' and 'weak' corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively.

For both the entire research location and individual clusters, the VF1(s), accounting for 26.4, 25.9, 34.5, and 25.3 % of the total variance, have strong positive loading (>0.7) on nephelometric turbidity units (NTU) and suspended solids (SS), whereas these have a moderate loading (>0.6) on Fe and nitrates. These factors represent a result of soil erosion from disturbed land or inflow of effluent from sewage plants and industry, and a high loading of turbidity is also attributed to the presence of algae and plankton. The VF2(s), accounting for 23.5, 23.9, 24.4, and 24.5 % of total variance, have a strong positive loading on chemical and biochemical oxygen demand (BOD and COD), whereas they have only a moderate loading on ammonia. These factors represent influences from point sources, such as discharges from wastewater treatment plants, domestic wastewater, agricultural activities and industrial effluents. The VF3(s) of cluster 1, explaining 10.4 % of total variance, has a strong negative loading on pH. This factor can be explained by the anaerobic conditions in the river from the high loading of dissolved organic matter, which leads to

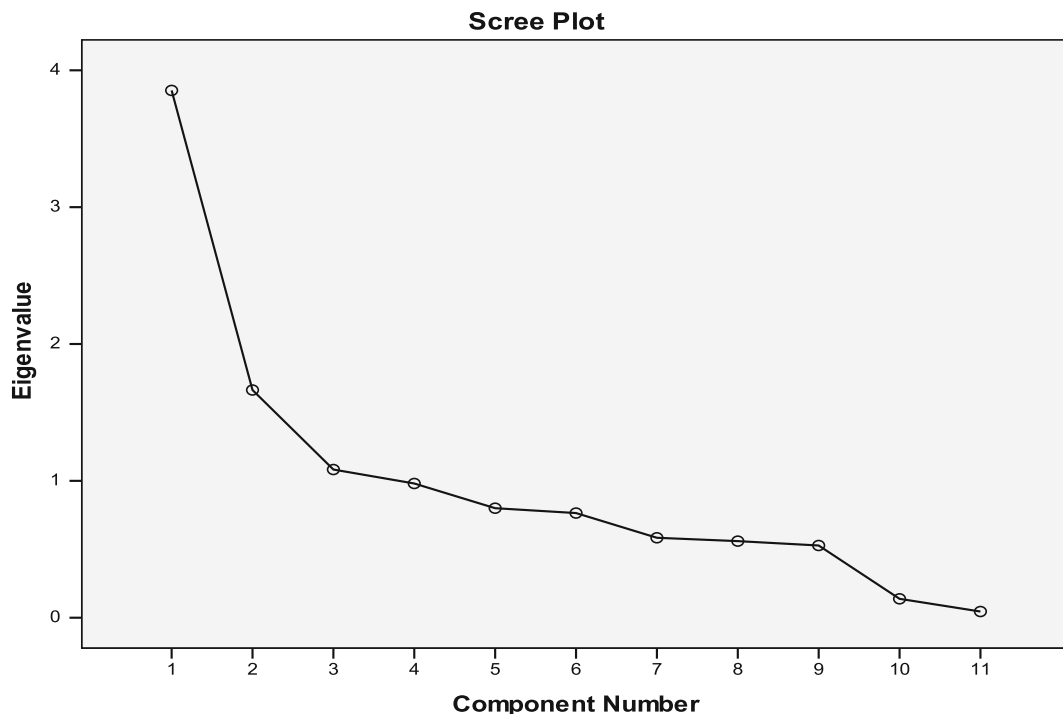


Fig. 3 The Scree plot represents eigenvalues for entire research location, Can Tho City, Mekong Delta, Vietnam

Table 2 Loading of 11 experimental variables on factor analysis parameters for entire research location and three clusters

Parameters	Entire location			Spatial cluster 1			Spatial cluster 2				Spatial cluster 3			
	VF1	VF2	VF3	VF1	VF2	VF3	VF1	VF2	VF3	VF4	VF1	VF2	VF3	VF4
pH	0.199	0.025	-0.810	0.249	0.010	-0.815	0.144	-0.004	0.893	-0.088	0.000	0.005	-0.036	0.969
NTU	0.919	0.020	0.061	0.888	0.080	0.101	0.924	-0.060	0.191	0.103	0.919	0.052	0.131	0.050
SS	0.847	-0.001	0.040	0.785	0.079	0.065	0.783	0.069	0.252	0.145	0.898	0.004	0.111	0.040
Fe	0.648	0.275	0.095	0.636	0.271	0.101	0.832	0.201	-0.207	0.096	0.694	0.305	-0.073	-0.072
DO	-0.429	-0.320	0.046	-0.565	-0.079	0.087	-0.400	0.556	0.556	0.405	-0.266	-0.485	-0.434	-0.248
COD	0.038	0.930	-0.610	-0.004	0.921	-0.115	-0.041	0.933	0.158	-0.031	0.073	0.934	0.063	-0.004
BOD	0.134	0.941	-0.049	0.134	0.927	-0.134	0.237	0.904	-0.033	-0.075	0.115	0.945	0.045	0.009
NH ₃	0.333	0.618	0.214	0.369	0.583	0.284	0.071	0.764	-0.364	-0.207	0.361	0.624	0.077	0.003
NO ₃ ⁻	0.613	0.325	-0.065	0.585	0.418	-0.013	-0.153	-0.153	0.294	-0.139	0.532	0.307	0.014	0.071
NO ₂ ⁻	0.391	0.412	0.247	0.329	0.552	0.232	0.127	0.127	-0.410	-0.052	0.385	0.335	0.035	-0.111
Coliform	0.202	0.043	0.562	0.249	0.000	0.526	0.165	-0.160	-0.087	0.873	0.035	0.066	0.951	-0.079
Eigenvalue	3.85	1.66	1.08	3.88	1.64	1.1	3.82	2.74	1.32	1.02	3.94	1.65	1.04	1.01
% Total variance	26.4	23.5	10	25.9	23.9	10.4	34.5	24.4	12.4	9.5	25.3	24.5	10.3	9.4
Cumulative % variance	26.4	49.9	60	25.9	49.8	60.2	34.5	59.0	71.4	80.9	25.3	49.8	60.2	69.6

BOD biochemical oxygen demand, *COD* chemical oxygen demand, *NTU* nephelometric turbidity units, *DO* dissolved oxygen, *SS* suspended solids, *VF* varifactor

formation of ammonia and organic acids resulting in a decrease in pH. In contrast, the VF3(s) of cluster 2 and cluster 3, accounting for 12.4 and 9.4 % respectively, have a strong positive loading on pH and represent the physiochemical source of variability. The VF4(s) of cluster 2 and cluster 3, accounting for the lowest total variance (9.5 and 9.4 %), have a strong positive loading on coliform bacteria. These factors indicate the contamination by sewage waste with faecal coliform bacteria through sewer and septic systems, feedlot and dairy runoff, runoff from broad acre farming, stormwater, and livestock and humans defecating directly into the water.

Spatial and temporal variation in river water quality (DA)

Spatial DA was conducted with the same raw data set comprising 11 parameters after grouping into three clusters as obtained through the CA technique. The clusters were the dependent variable, while all the measured parameters constituted the independent variables. Discriminant functions (DFs) and classification matrices (CMs), obtained from the standard, forward stepwise and backward stepwise modes of DA, are presented in

Tables 3 and 4. The results of standard DA and forward and backward stepwise modes constructed for the 11 parameters (Table 3) indicate that the coliform bacteria group coefficients are zero. The classification matrices assign more than 67 % of cases correctly for the standard and forward stepwise using 11 discriminant parameters, and the CM gave the similar result (66.8 % of cases) for backward stepwise mode using three parameters (Tables 3 and 4). Backward stepwise DA indicates that NTU, COD and NH₃ are the discriminating parameters in space.

Figure 4a–c shows box and whisker plots of discriminating parameters identified by spatial DA (backward stepwise mode) which was used to examine different patterns associated with spatial variations in river surface water quality in Can Tho City. The average COD is highest in cluster 3 as it is these urbanized and industrial zones which harbour domestic wastewater, wastewater treatment plants and industrial effluents. Ammonia levels are higher in clusters 3 and 1, which, combined with the high COD, suggests a high load of dissolved organic matter in these sites contributed by domestic wastewater, wastewater treatment plants and industrial effluents located upstream from the monitoring sites. This results in anaerobic conditions in the river, leading

Table 3 Classification function for discriminant analysis of spatial variation in Mekong River in Can Tho City, Mekong Delta, Vietnam

Parameters	Standard mode			Forward stepwise mode			Backward stepwise mode		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
pH	70.302	71.224	70.333	70.302	71.224	70.333			
BOD	-0.623	-0.674	-0.624	-0.623	-0.674	-0.624			
COD	0.545	0.563	0.607	0.545	0.563	0.607	0.225	0.212	0.287
NTU	-0.066	-0.053	-0.075	-0.066	-0.053	-0.075	0.038	0.062	0.035
DO	4.839	4.828	4.766	4.839	4.828	4.766			
SS	0.013	0.023	0.017	0.013	0.023	0.017			
Fe	5.701	5.889	5.958	5.701	5.889	5.958			
NH ₃	2.979	2.918	3.290	2.979	2.918	3.290	-0.271	-0.452	0.065
NO ₃ ⁻	2.590	2.742	2.581	2.590	2.742	2.581			
NO ₂ ⁻	-2.683	-5.362	-2.886	-2.683	-5.362	-2.886			
Coliform	0.000	0.000	0.000	0.000	0.000	0.000			
Constant	-256.035	-263.559	-257.114	-256.035	-263.559	-257.114	-3.598	-4.747	-4.777

BOD biochemical oxygen demand, *COD* chemical oxygen demand, *NTU* nephelometric turbidity units, *DO* dissolved oxygen, *SS* suspended solids

to formation of ammonia and organic acids. NTU levels are also higher in cluster 1 and 3 compared with cluster 2. This suggests a result of soil erosion from disturbed land or inflow of effluent from sewage plants and

industry, contributing to the presence of algae and plankton. Although turbidity, COD and ammonia are mostly attributed to urbanization and industrial activities in this study, the question of why there is less loading of these parameters in the predominantly agricultural area compared with urbanized and industrial zones needs to be addressed by further studies.

Table 4 Classification matrix for discriminant analysis of spatial variations in Mekong River, Can Tho City, Mekong Delta, Vietnam

Cluster	% correct	Clusters assigned by DA		
		Cluster 1	Cluster 2	Cluster 3
<i>Standard DA mode</i>				
Cluster 1	94.8	419	4	19
Cluster 2	15.4	31	6	2
Cluster 3	25.8	183	1	64
Total	67.1	623	11	85
<i>Forward stepwise DA mode</i>				
Cluster 1	94.8	419	4	19
Cluster 2	15.4	31	6	2
Cluster 3	25.8	183	1	64
Total	67.1	623	11	85
<i>Backward stepwise DA mode</i>				
Cluster 1	95.8	432	2	17
Cluster 2	15	32	6	2
Cluster 3	23.4	194	1	61
Total	66.8	658	9	80

Temporal DAs were also performed on the same raw data set comprising 11 parameters after grouping into seasonal groups (dry and wet season). Discriminant functions (DFs) and classification matrices (CMs) obtained from the standard, forward stepwise and backward stepwise modes of DA are shown in Tables 5 and 6. As with spatial DA, the standard DA- and forward stepwise DA mode-constructed DFs comprising 11 parameters, the coliform bacteria group coefficients are again zero. Both the standard and forward stepwise mode DFs using 11 parameters yielded the corresponding CMs assigning approximately 60 % cases correctly, whereas the backward stepwise mode DA yielded CMs with 56 % correct assignment using only two discriminant parameters (Table 5). Backward stepwise DA shows that pH and NO₂ are the discriminating parameters according to season. Box and whisker plots of discriminating parameters identified by the seasonal backward stepwise DA are presented in Fig. 5a, b. The pH reveals a subtle elevation during the dry compared with the wet season, whereas there are significantly

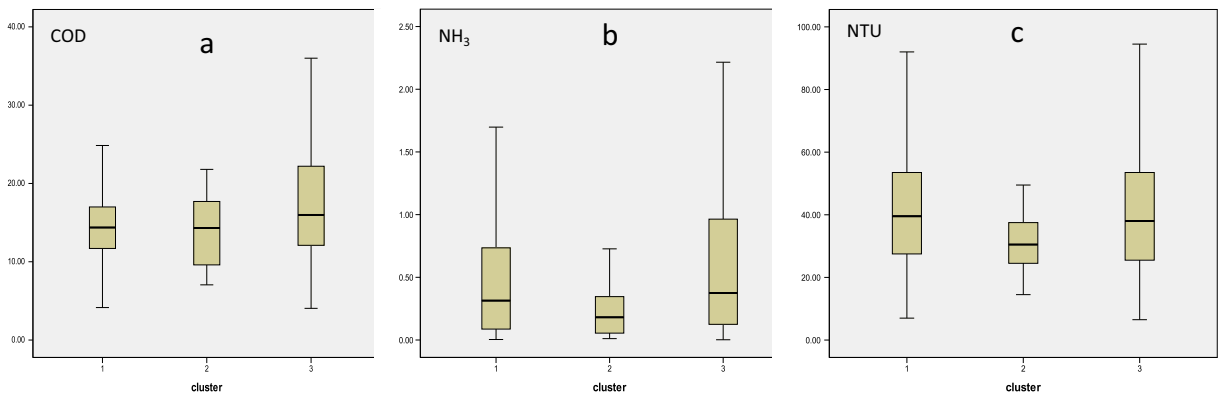


Fig. 4 Spatial variations: **a** COD, **b** NH₃ and **c** NTU in surface water quality of the Mekong River in Can Tho City, Mekong Delta, Vietnam

higher levels of NO₂ during the wet. This can be attributed to the runoff of nitrogenous fertilizers from agricultural activities during the wet season in the research location. In addition, nitrite-nitrogen might also come from other sources such as anaerobic wastewaters, low dissolved oxygen levels and urban runoff containing poorly degraded organic and ammonia-nitrogen waste during the wet season. The seasonal variation of pH values observed in this study is in agreement with the results of a previous similar study (Ekeh and Sikoki 2003; Ansa 2005) where the values of pH in dry season were higher than those in the late rainy season. This may be due to the influx and decay of debris in the area as well as imbalance in the level of H⁺ ions from surface

runoffs during the rains. In addition, the pH monitoring likely indicates the effects of acid sulphate soils, which is an issue in the Mekong Delta area (Neumann et al. 2013).

Conclusion

This study used various multivariate statistical techniques to evaluate temporal and spatial variations in surface water quality of the Mekong Delta River in Can Tho City, Vietnam. Cluster analysis grouped a large number of sampling sites (38 sites) into three clusters of similar water quality features which

Table 5 Classification function for discriminant analysis of temporal variation in Mekong River in Can Tho City, Mekong Delta, Vietnam

Parameters	Standard mode		Forward stepwise mode		Backward stepwise mode	
	Dry	Wet	Dry	Wet	Dry	Wet
pH	71.069	70.413	71.069	70.413	67.161	66.529
BOD	-0.589	-0.604	-0.589	-0.604		
COD	0.522	0.546	0.522	0.546		
NTU	-0.079	-0.074	-0.079	-0.074		
DO	4.691	4.791	4.691	4.791		
SS	0.014	0.012	0.014	0.012		
Fe	5.372	5.620	5.372	5.620		
NH ₃	3.425	3.176	3.425	3.176		
NO ₃ ⁻	2.661	2.581	2.661	2.581		
NO ₂ ⁻	-5.184	-2.905	-5.184	-2.905	1.087	2.836
Coliform	0.000	0.000	0.000	0.000		
Constant	-258.990	-255.315	-258.990	-255.315	-232.810	-228.590

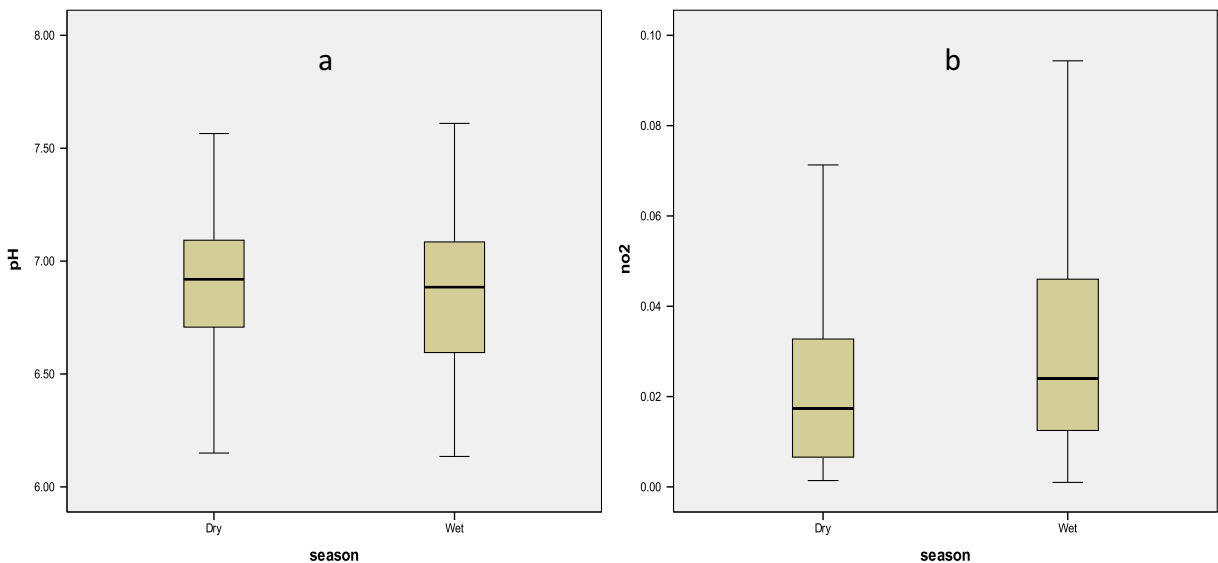
BOD biochemical oxygen demand, COD chemical oxygen demand, NTU nephelometric turbidity units, DO dissolved oxygen, SS suspended solids

Table 6 Classification matrix for discriminant analysis of temporal variations in Mekong River, Can Tho City, Mekong Delta, Vietnam

Cluster	% correct	Clusters assigned by DA	
		Dry	Wet
<i>Standard DA mode</i>			
Dry	53.8	190	163
Wet	62.5	141	235
Total	58.3	331	398
<i>Forward stepwise DA mode</i>			
Dry	53.8	190	163
Wet	62.5	141	235
Total	58.3	331	398
<i>Backward stepwise DA mode</i>			
Dry	53.9	202	173
Wet	58.4	158	222
Total	56.2	360	395

divided the river basin into three zones comprising mixed urban-rural area, agricultural area and mixed urban-industrial area. The results indicate the low quality of river $_{2\text{surface}}$ water in the Mekong Delta in Vietnam, which is proved by that fact that most of the quality parameters exceeded the national standard guidelines. The finding also suggests a possible revised sampling strategy that can reduce

the number of sampling sites and affiliated recurring costs. However, further study is needed to assess how sensitive the recommended sampling clusters are to the changes of environmental and seasonal factors as well as how they respond to climate change impacts such as increasing temperature, flooding and longer dry periods. The principal component analysis and factor analysis were performed to extract and recognize the factors or origins responsible for water quality variation over the entire location and in each cluster. The results indicated that the nephelometric turbidity units (NTU), suspended solids (SS), Fe and nitrate are the most significant parameters contributing to water quality variations for the entire location and for each cluster, which represent a result of soil erosion from disturbed land or inflow of effluent from sewage plants and industry. The high loading of turbidity is also attributed to the presence of algae and plankton. DA revealed that three parameters (NTU, COD and NH_3) discriminate clusters spatially, and two parameters (pH and NO_2^-) discriminate clusters by season (dry and wet season). The results of this study demonstrate the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, identification of pollution sources/factors, and evaluation of temporal/spatial variations in water quality assessment and management.

**Fig. 5** Temporal variations: **a** pH, **b** NO_2^- -N (mg/l) in surface water quality of the Mekong River in Can Tho City, Mekong Delta, Vietnam

Acknowledgments We offer sincere thanks to the staff of the Health Environment Management Agency (Ministry of Health), Can Tho Health Department, Can Tho Preventive Medicine Centre, and Can Tho University for providing kind support and assistance to our data collection.

Conflict of interest No conflict of interest

References

- Adam, S., Titus, R., Pietesen, K., Tredoux, G., & Harris, C. (2001). Hydrochemical characteristic of aquifer near Suntherland in the Western Karoo. *South Africa Journal of Hydrology*, *241*, 91–103.
- Ansa, E. J. (2005). Studies of the benthic macrofauna of the Andoni flats in the Niger Delta Area of Nigeria. *PhD Thesis University of Port Harcourt, Port Harcourt, Nigeria*, pp. 242.
- Bricker, O. P., & Jones, B. F. (1995). Main factors affecting the composition of natural waters. In B. Salbu & E. Steinnes (Eds.), *Trace Elements in Natural Waters* (pp. 1–5). Boca Raton: CRC Press.
- Ekeh, I. B., & Sikoki, F. D. (2003). The state and seasonal variability of some physio-chemical parameters in the New Calabar River, Nigeria. *Supplement Ad Acta Hydrobiologia*, *5*, 45–60.
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernadez, J. M., & Fernadez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Research*, *34*, 807–816.
- Huyen, N. C. (2012). *Atlas of Vietnamese City: Hanoi - Ho Chi Minh - Hai Phong - Da Nang - Can Tho*. Hanoi: Vietnam Publishing House of Natural Resources, Environment and Cartography.
- Karbasi, A. R., Nouri, J., Mehrdadi, N., & Ayaz, G. O. (2008). Flocculation of heavy metals during mixing of freshwater with Caspian Seawater. *Environmental Geology*, *53*, 1811–1816.
- Kim, J. O., & Muller, C. W. (1987). *Introduction to factor analysis: what is and how to do it. Quantitative Application in the Social Sciences Series*. Newbury Park: Sage University Press.
- Lee, J. Y., Cheon, J. Y., Lee, K., Lee, S. Y., & Lee, M. H. (2001). Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. *Journal of Environmental Quality*, *30*, 1548–1563.
- Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in Blackfoot disease area in Taiwan. *Science in the Total Environment*, *313*, 77–89.
- Najafpour, S., Alkarkhi, A. F. M., Kadir, M. O. A., & Najafpour, G. D. (2008). Evaluation of spatial and temporal variation in river water quality. *International Journal of Environmental Research*, *2*, 349–358.
- Neumann, L., Nguyen, M., Moglia, M., Cook, S., & Lipkin, F. (2013). Urban Water Systems in Can Tho, Vietnam: Understanding the current context for climate change adaptation. From: www.csiro.au.
- Niemi, G. J., Devore, P., Detenbeck, N., Taylor, D., & Lima, A. (1990). Overview of cases studies on recovery of aquatic systems from disturbance. *Environmental Management*, *14*, 571–587.
- Otto, M. (1998). Multivariate methods. In R. Keller, J. M. Mermet, M. Otto, & H. M. Wildmer (Eds.), *Analytical Chemistry*. Weinheim: Wiley-VCH.
- Papatheodorou, G., Lambrakis, N., & Panagopoulos, G. (2007). Application of multivariate statistical procedures to the hydrochemical study of coastal aquifer: an example from Crete, Greece. *Hydrocarbon Processing*, *21*, 1482–1495.
- Pejman, A. H., Nabi Bidhendi, G. R., Karbassi, A. R., Mehrdadi, N., & Esmaeili, B. (2009). Evaluation of spatial and seasonal variation in surface water quality using multivariate statistical techniques. *International Journal of Environmental Science and Technology*, *6*(3), 467–476.
- Qadir, A., Malik, R. N., & Husain, S. Z. (2008). Spatio-temporal variations in water quality of Nullah Aiktributary of the river Chenab, Pakistan. *Environmental Monitoring and Assessment*, *140*, 43–59.
- Reghunath, R., Murthy, T. R. S., & Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research*, *36*, 2437–2442.
- 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.
- Simeonov, V., Tsakovski, S., Lavric, T., Simeonova, P., & Puxbaum, H. (2004). Multivariate statistical assessment of air quality: a case study. *Microchimica Acta*, *148*, 293–298.
- 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.
- Vega, M., Pardo, R., Barrado, E., & Deban, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by explanatory data analysis. *Water Research*, *32*, 3581–3592.
- VNMONRE (2008). QCVN 08: 2008/BTNMT. National technical regulation on surface water quality.
- Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C., & Bistoni, M. A. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: suquia river basin (Cordoba, Argentina). *Water Research*, *35*, 2881–2894.