Pollution 2022, 8(2): 463-477
DOI: 10.22059/POLL.2021.329252.1165

RESEARCH PAPER

Analysis of Surface Water Quality using Multivariate Statistical Approaches: A case study in Ca Mau Peninsula, Vietnam

Nguyen Thanh Giao*

Department of Environmental Management, College of Environment and Natural Resources, Can Tho University, Can Tho city 900000, Vietnam

Received: 21.08.2021, Revised: 21.12.2021, Accepted: 23.12.2021

ABSTRACT

The study was conducted to assess surface water quality in Ca Mau peninsula using multivariate statistical analysis. Fifty-one water samples with the parameters of pH, dissolved oxygen (DO), total suspended solids (TSS), biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), ammonium ($N\text{-}NH_4^+$), orthophosphate ($P\text{-}PO_4^{3-}$) and total coliform were used in the evaluation. Water quality is assessed using national standard and water quality index (WQI). The methods of cluster (CA), discriminant (DA), principal component analysis (PCA) were used to analyze the variation patterns of water quality. The surface water was contaminated with organic matters, suspended solids, nutrients, and microorganisms. DA revealed that DO, TSS, BOD_5 and pH contributed 76.91% to the seasonal variation of water quality. Water quality is classified from bad to heavily polluted. CA grouped water quality into 7 clusters and DO, TSS, BOD_5, COD and coliform of the clusters 1-3 were significantly higher than those of the clusters 4-7. PCA presented that PC1-PC3 was the main sources affecting water quality, explaining 85.45% of the variation in water quality. The sources of pollution can be human (domestic wastewater, waste from agriculture, fisheries, industry, landfills), natural (hydrological regime, rainwater overflow, river bank erosion). pH, DO, BOD_5, COD, TSS, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ and coliform have an impact on water quality and need to be continuously monitored. However, for the multivariate statistical method to be more effective, an initial data set with several water quality parameters sampling locations is needed. The current results provide scientific information and support local water quality monitoring activities.

Keywords: water quality, multivariate analysis, organic matters, nutrients, Ca Mau

INTRODUCTION

Water quality monitoring is an important task to manage and maintain water quality for socio-economic development. For surface water quality monitoring, physico-chemical and biological parameters can be selected (Cao et al., 2007; Wijeyaratne and Kalaotuwave, 2017; Giao, 2019; Giao and Nhien, 2020). The physical and chemical parameters include water temperature, pH, total suspended solids (TSS), turbidity, dissolved oxygen (DO), biological oxygen demand (BOD_5), chemical oxygen demand (COD), ammonium ($N\text{-}NH_4^+$), orthophosphate ($P\text{-}PO_4^{3-}$), heavy metals (Fe, Al, Mn, Cr, Cd), chloride (Cl^-), sulfate (SO_4^{2-}), pesticides, antibiotics or biological factors such as *Escherichia coli*, coliform (MPN/100mL) (Cho et al., 2009; MONRE, 2015; Chounlamany et al., 2017; Zeinalzadeh and Rezaei, 2017). In Vietnam, water quality monitoring is carried out under the national monitoring program and each province has set up a different water quality monitoring network. Most provinces

* Correspondence author Email: ntgiao@ctu.edu.vn

only assess water quality at the WQI calculation level based on the guidance on WQI calculation of the Vietnam Environment Administration (2019) or assessment of individual criteria according to national technical regulation on surface water quality (QCVN 08 MT: 2015/BTNMT) (MONRE, 2015). Currently, multivariate analysis methods are commonly used to analyze water quality data in river and lake systems (Cho et al., 2009; Chounlamany et al., 2017; Zeinalzadeh and Rezaei, 2017). By these analyzes, potential sources of pollutant generation, water quality classification, identification of criteria that make the difference in water quality between the sampling times are identified (Vega et al., 1998; Singh et al., 2005; Chounlamany et al., 2017; Zeinalzadeh and Rezaei, 2017; Giao, 2020). It can be seen that the multivariate statistical method is a useful tool to analyze water quality data to support the development of surface water quality monitoring (Chounlamany et al., 2017; Giao, 2020; Giao and Nhien, 2021; Giao and Minh, 2021, Giao et al., 2021).

Ca Mau province is the southernmost point of Vietnam, under the combined effects of climate change and human activities, leading to water quality deterioration. Besides, surface water in Ca Mau is mainly used for aquaculture and daily life; therefore, the assessment of surface water quality of the province is very necessary. However, the data has been relatively collected, but little consideration for how data will be used and those who collect data often lack the tools for analyzing large amounts of data effectively. This may lead to the monitoring program not exploiting the dataset effectively; therefore, when water quality monitoring programs are more beneficial if the data are comprehensively exploited for their potential impacts and variability. Through this, the study was conducted to apply multivariate statistical methods to analyze surface water quality in Ca Mau peninsula, a coastal province in the Vietnamese Mekong delta. The results can be useful for analysis, extracting important information from monitoring data sets to better serve water quality monitoring in the study area.

MATERIALS AND METHODS

Ca Mau is the province in the Mekong Delta region, consisting of two parts the mainland and the sovereign sea. The mainland has an area of 5,294.87 km^2. Sovereign waters are over 70,000 km^2 with a coastline of 254 kilometers. It borders Kien Giang province to the north, Bac Lieu province to the northeast, the East and South China Sea to the east, and the Gulf of Thailand to the west. Ca Mau is located on a peninsula, has a rather special geographical position, with three sides adjacent to the sea. Ca Mau has a system of rivers and canals interlaced, occupying 3.02% of the natural area, in which there are many large and deep rivers such as Cua Lon, Ganh Hao, Bay Hap, Song Doc, Dam Doi, Cai Tau, Trem. The total length of rivers is about 7,000 km. Due to the influence of two tidal regimes and many estuaries connecting to the sea, the entire land area of the province is saline and the tidal transmission regime is very complicated (People's Committee of Ca Mau province, 2020).

Ca Mau has been restructuring agriculture through the transformation of production models adapted to climate change, for examples, the model of production linkage along the shrimp-forest value chain; hi-tech industrial shrimp farming model; model of planting melaleuca, acacia hybrid intensive farming; rice-shrimp production model; converting extensive shrimp farming to improved extensive shrimp farming. Many small seafood processing establishments have not built wastewater treatment systems, leading to increasingly serious pollution levels. Moreover, in some concentrated shrimp and fish farming areas, the discharge of eutrophic organic substances, microbial toxins (including pathogens) and indiscriminate domestic waste causes environmental degradation, disease outbreaks and cause significant economic losses as well as ecological environmental conditions of water sources in rivers and canals (People's Committee of Ca Mau province, 2020). In addition, chemicals, pesticides, and waste containing pathogens during the improvement of ponds by shrimp farmers also pollute the river environment. Wastewater from specialized agricultural activities containing toxic components such as pesticides, chemical fertilizers has been causing the risk of environmental pollution of soil, underground water and surface water in neighboring areas (People's Committee of Ca Mau province, 2020). Currently, Ca Mau province has three industrial zones and one economic zone. However, there is no facility to build a centralized wastewater treatment system. Therefore, the monitoring and analysis of water quality data is one of the important tasks of the local environmental management agency.

A total of 51 surface water quality samples were collected in the water bodies of Ca Mau peninsula in March (dry season) and September (dry season) of 2020. Sampling locations are divided according to impact sources. Sources of impacts on surface water environment due to urban areas, residential areas, concentrated markets, tourist areas (Residential-market) were locations from NM1 to NM20; the impact sources because of industrial production activities (Industry) included locations from NM21 to NM28. The sources of impacts due to aquaculture and seafood processing (Aquaculture) were the locations from NM29 to NM39 and the impact sources due to agricultural production (Agriculture) included the locations from NM40 to NM42. Finally, the national park and the landfill impact on surface water quality included location NM43 – NM51 and 1 location NM52, respectively. At each sampling site, mixed water samples were collected in plastic bottles at a depth of 30 - 50 cm below the water surface. Details of sampling locations are shown in **Figure 1**. The water quality evaluation parameters include pH, dissolved oxygen (DO), total suspended solids (TSS), biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), ammonium ($N\text{-}NH_4^+$), orthophosphate ($P\text{-}PO_4^{3-}$) and coliforms. The pH and DO parameters were measured in the field using hand-held instruments. TSS, BOD_5, COD, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ and coliform parameters were analyzed according to standard analytical methods (APHA, 1998) **(Table 1)**. To ensure quality control of the analysis in the field and laboratory measurements for parameters, the study carried out repeated measurements for a number of random samples for the water quality parameters. The measurement results of the above replicate samples showed that the percentage of the standard deviation of the duplicate samples varied from 0.26 – 0.82 (field) and 0.00 – 8.00 (laboratory); which is less than 20%. Therefore, the monitoring results for these samples have satisfactory requirements according to the regulations of the Ministry of Natural Resources and Environment (2012).

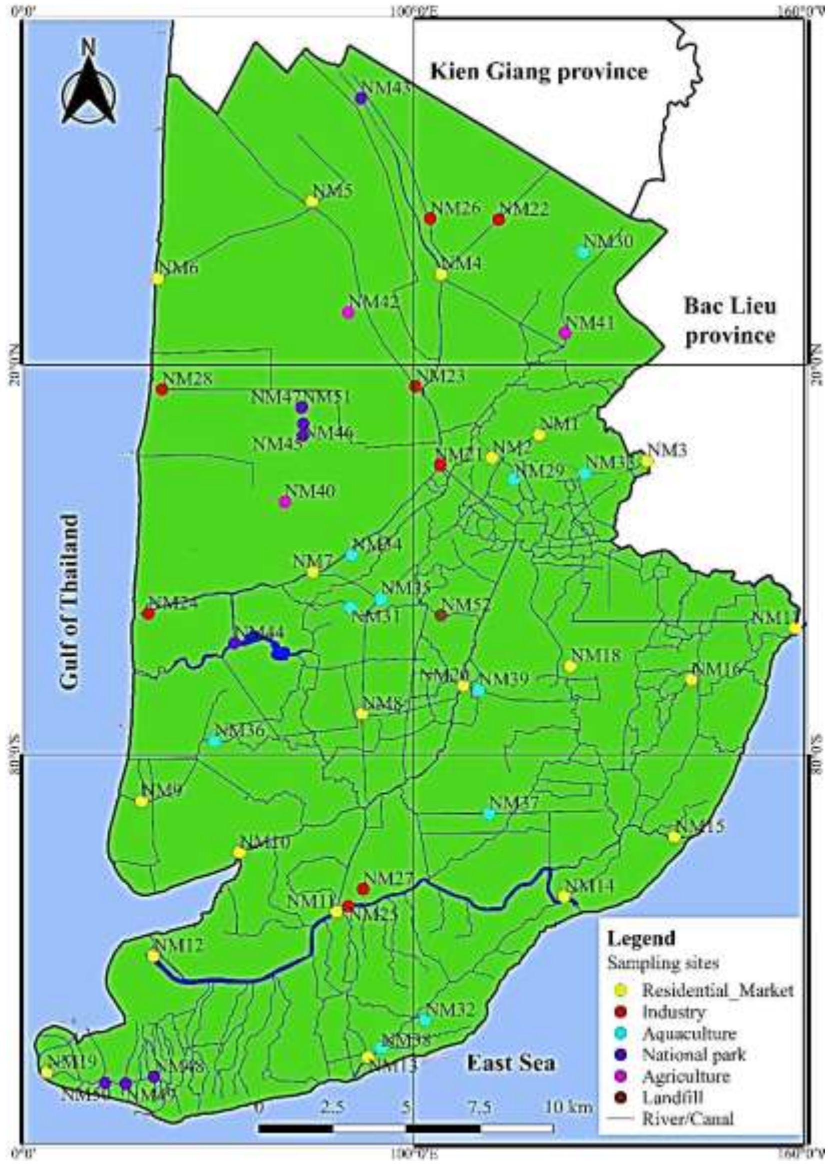

Fig 1. Map of the sampling locations

Surface water quality is assessed using each individual criterion, compared with the national technical regulation on surface water quality QCVN 08 MT: 2015/BTNMT, column B1 **(**MONRE, 2015) (Table 1). Besides that, water quality index (WQI) was also applied to give a unique status of water quality in each sampling site. Water quality assessment index calculated according to the guidance of Decision No. 1460/QĐ-TCMT on technical guidance for calculation of water environment quality index present in Vietnam (VEA, 2019). The differences of surface water quality in seasonal and socio-economic sectors were analyzed using Independent Sample T-Test and One-way ANOVA, respectively using SPSS Version 20.0 software (IBM Corp., Armonk, NY, USA). The relationship between water quality parameters was evaluated using Pearson Correlation analysis. The discriminant analysis (DA) method was used to determine the different water quality parameters between the dry season and the rainy season (Shrestha and Kazama, 2007). Cluster analysis (CA) was used to group water quality according to Ward's method (Salah et al., 2012). The important criteria affecting the water quality, and potential sources of water pollution, were determined by Principal Component Analysis (PCA) (Chounlamany et al., 2017; Varol, 2020). In which, the values of

KMO (Kaiser-Meyer-Olkin) was recorded about 0.7; therefore, the sample all is a whole is adequate for principal component analysis (Kaiser, 1974). DA, CA, and PCA were performed using Statgraphics Centurion version XVI software (Statgraphics Technologies Inc., Virginia state, USA).

Table 1. Analytical methods and limit values of water quality parameters

Variables	Unit	Analytical methods	Device	Limit values*
pH	-	TCVN 6492:2011	pH 3110i (WTW, Germany)	5.5-9
DO	mg/L	ASTM D 888 -12	HQ 30d (Hach, USA)	≥ 4
TSS	mg/L	TCVN 6625:2000	Cole-parmer (USA) Memmert Unb 500 (Germany)	50
BOD_5	mg/L	TCVN 6001-1:2008	Foc 225E (Velp scentifica, Italia)	15
COD	mg/L	SMEWW 5220C:2012	Eco 25 (Velp scentifica, Italia)	30
$N\text{-}NH_4^+$	mg/L	SMEWW 4500$NH_{3,}$ F:2012	Jasco V-530 Spectrophotometer (Japan)	0.9
$P\text{-}PO_4^{3-}$	mg/L	SMEWW 4500-P,E:2012	Jasco V-530 Spectrophotometer (Japan)	0.3
Coliform	MPN/100mL	TCVN 6187-2:2009	HL-342 (USA)	7,500

* Limit values in column B1 of National Technical Regulation on Surface Water Quality (QCVN 08 MT: 2015/BTNMT).

RESULTS AND DISCUSSION

The pH values in the water bodies of Ca Mau peninsula in the dry season and the rainy season were 7.07-7.77 and 6.43-7.57, respectively (Table 2). Fluctuations of pH according to the survey sites did not have a statistically significant difference. Former studies reported that pH in the Mekong River ranged from 6.7-7.1 (MRC, 2015), 6.3-8.0 in the Hau River (Lien et al., 2016), 6.7-7.4 in the Tien River (Giao and Minh, 2021). pH of the studied water bodies did not differ between the dry season and the rainy season. Previous studies have shown that pH in rivers and canals in the Mekong Delta and the tropics does not fluctuate much (Singh et al., 2005; Onlgley, 2009; MRC, 2015; Chounlamany et al., 2017; Giao et al., 2021; Giao and Minh, 2021). This pH value is still within the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (pH 5.5-9).

The dissolved oxygen in the dry season ranged from 3.51 to 5.41 mg/L and 2.46 to 3.95 mg/L in the rainy season (Table 2). DO in the dry season has no statistically significant difference while DO in the water bodies in the rainy season had differences between sampling locations. During the rainy season, DO in residential areas and markets was significantly higher than in the other locations. DO at the location affected by the landfill was the lowest in both the rainy and dry seasons. DO has a clear seasonal variation, in which DO in the dry season tended to be higher than in the rainy season. The depletion of DO in the rivers can be explained by the effects of the decomposition of organic matter in the water (Jerves-Cobo, et al., 2020). Besides that, this observation was reported in a previous study by Huang et al. (2004) on the abundance of phytoplankton in the rainy season was higher than that of the dry season. In the rainy season, DO at all locations was lower than the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (DO≥4 mg/L). Meanwhile, DO in the dry season was higher than the allowable limit except for the locations in agricultural production areas and landfill (Table 2). According to the previous studies have shown that DO in Hau rivers was varied 5.3-5.6 mg/L (Giao, 2020), in Tien River 5.5-6.0 mg/L (Giao and Minh, 2021), and in canals of Dong Thap province was 5.07-5.18 (Giao et al., 2021) and the values of DO was seasonal fluctuation (Lien et al., 2016; Ut et al., 2016; Ly and Giao, 2018; Giao and Minh, 2021). The DO in this study was lower than in that in previous studies, indicating that the water bodies in Ca Mau peninsula were heavily organically polluted (Kazi et al., 2009; Zeinalzadeh and Rezaei, 2017).

The total suspended solids content in the dry and wet season was in the range of 83.5-152 mg/L and 76-152.2 mg/L, respectively (Table 2). TSS at all locations and between the dry and wet seasons were not statistically different, except for the location in the national park. TSS concentrations at the national park sites during the rainy season were significantly higher than that in the dry season. Because the province has a rather special geographical position, with three sides adjacent to the sea. The rainy season is affected by the amount of rainwater, and the dry season is affected by saline intrusion in the estuaries. In addition, the main livelihood is aquaculture - regularly uses and discharges water into the river. Therefore, the flow is frequently disturbed, which may have resulted in TSS values not having a significant difference between the two seasons. TSS in all water bodies of Ca Mau peninsula has exceeded QCVN 08 MT: 2015/BTNMT, column B1 (50 mg/L) by 1.7 to 3 times. High TSS in water and its seasonal fluctuations are the common trends in the water bodies in the Mekong Delta region (MRC, 2015; Ly and Giao, 2018; Giao and Minh, 2021; Giao et al., 2021). BOD_5 concentrations in the studied water bodies in the dry and rainy seasons were 15-24 mg/L and 13.3-17.3 mg/L, respectively (Table 2). BOD_5 concentration in the dry season was the highest in the national park area and lowest in the industrial production area. In the rainy season, there was no statistically significant difference in BOD_5 between sampling locations. BOD_5 in the national park areas in the dry season was significantly higher than that in the rainy season. BOD_5 in water bodies affected by residential-market areas, aquaculture, and national parks all exceeded the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (15 mg/L). Previous studies showed that BOD_5 in canals in An Giang, Hau river, Tien river, and water bodies in Tien Giang were all lower than BOD_5 in this study (Lien et al., 2016; MRC, 2016; Ly and Giao, 2018; Giao, 2020; Giao and Minh, 2021; Giao et al., 2021).

COD concentrations in the dry season and the rainy season were 23.7-43.2 mg/L and 25.0-34.8 mg/L, respectively (Table 2). During the dry season, COD concentrations were the highest in the national park and the lowest in agricultural and industrial areas. Meanwhile, COD in the rainy season did not have a statistically significant difference between the surveyed water bodies. Among the surveyed water bodies, only COD in the national park in the dry season was significantly higher than that in the same location in the rainy season. In all other locations, there was no significant difference in COD between wet and dry seasons. COD in some locations such as national parks (both rainy and dry seasons), landfills (rainy season) exceeded the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (30 mg/L). Former studies reported that COD in Tien River was in the range of 10.75-15.5 mg/L (Giao et al., 2021); in canals in Hau Giang province averaged at 17.9±4.3 mg/L; in water bodies in Dong Thap province was 21.3-23.1 mg/L (Giao et al., 2021). Previous studies have shown that COD has seasonal variation in that the rainy season tends to be higher than the dry season (Giao and Minh, 2021; Giao et al., 2021). This trend is similar to the analysis in the present study (except for national parks); this can be explained by the high temperature, humidity and less waterlogging that facilitate the decomposition of organic matter. BOD_5 and COD indicators for organic pollution (Galal-Gorchev et al., 1993; Kazi et al., 2009).

The $N\text{-}NH_4^+$ in the dry season ranged from 0.49 to 0.76 mg/L while $N\text{-}NH_4^+$ in the rainy season was from 0.18 to 1.37 mg/L. There is no statistically significant difference between the sampling locations in the dry season; in contrast, $N\text{-}NH_4^+$ at the agricultural production site in the rainy season was significantly different from the rest of the sites (Table 2). $N\text{-}NH_4^+$

concentrations at aquaculture sites and agricultural production differred by seasons. $N\text{-}NH_4^+$ at agricultural areas in the rainy season exceeded the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (0.9 mg/L), while all survey sites in the dry season were within the allowable limit of Vietnamese standards. The content of $N\text{-}NH_4^+$ in the study area is similar to previous reports in similar water bodies (Tuan et al., 2019; Giao, 2020; Giao et al., 2021). In addition, $N\text{-}NH_4^+$ in water bodies often is seasonally fluctuated (Lien et al., 2016; MRC, 2015; Giao and Minh, 2021, Giao et al., 2021). There was no difference between sampling sites and between seasons for $P\text{-}PO_4^{3-}$. $P\text{-}PO_4^{3-}$ at the survey locations in the water bodies of Ca Mau peninsula was low and within the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (0.3 mg/L). The previous reports also show that the concentration of $P\text{-}PO_4^{3-}$ tended to fluctuate similarly in water bodies in the Mekong Delta (Ly and Giao, 2018; Tuan et al., 2019; Giao, 2020). $P\text{-}PO_4^{3-}$ is a water environment problem in water bodies of the Mekong Delta (Lien et al., 2016; Truc et al., 2019; Giao, 2020; Giao and Minh, 2021; Giao et al., 2021). Water-soluble phosphorus is often derived from fertilizers, detergents, cultivation, animal husbandry and industry (Barakat et al., 2016). The concentrations of $N\text{-}NH_4^+$ and $P\text{-}PO_4^{3-}$ in the study area pose potential risk of eutrophication.

The coliform density fluctuated greatly and there was no statistically significant difference by the survey locations and seasons. The coliform concentration in the study area exceeded the allowable limit of QCVN 08 MT: 2015/BTNMT, column B1 (7,500 MPN/100mL) by 1-12 times and 2.1-10.5 times in the dry and rainy seasons, respectively (Table 2). Previous studies also showed that coliform density in river networks in An Giang exceeded the national regulation by 2.14-7.04 times (Ly and Giao, 2018); the density of coliform in Tien and Hau rivers exceeded QCVN 08-MT:2015/BTNMT, column A1 from 1.1 to 6.5 times (Giao, 2020). In summary, the current results show that the surface water environment in Ca Mau peninsula is contaminated with organic matters (low DO, high BOD_5 and COD), suspended solids, nutrients, and microorganisms in which DO, BOD_5, COD, TSS, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ at some locations are subjected to seasonal fluctuations.

Table 2. Summary of surface water quality in the study area

Var.	Season	Resid.-Mark.	Industry	Aquaculture	Agriculture	National park	Landfill
pH	Dry	7.57 ± 0.19^{ns}	7.49 ± 0.41^{ns}	7.5 ± 0.24^{ns}	7.63 ± 0.17^{ns}	7.07 ± 0.69^{ns}	7.77
	Wet	7.57 ± 0.27^{ns}	7.37 ± 0.28^{ns}	7.45 ± 0.18^{ns}	7.5 ± 0.12^{ns}	6.43 ± 1.5^{ns}	7.56
DO	Dry	4.8 ± 1.64^{ns}	5.41 ± 2.19^{ns}	4.73 ± 1.7^{ns}	3.94 ± 0.31^{ns}	4.07 ± 1.84^{ns}	3.51
	Wet	3.95 ± 1.55^{a}	3.66 ± 1.35^{ab}	2.86 ± 1.43^{b}	2.76 ± 0.04^{ab}	3.16 ± 2.11^{ab}	2.46
TSS	Dry	93.3 ± 39.8^{ns}	92.3 ± 32.6^{ns}	83.5 ± 31.1^{ns}	85 ± 57.8^{ns}	$95.3\pm33.7^{ns(y)}$	152
	Wet	122.6 ± 51.6^{ns}	132.1 ± 67.2^{ns}	104.9 ± 36.3^{ns}	94 ± 35.1^{ns}	$152.2\pm105^{ns(x)}$	76
BOD_5	Dry	16.5 ± 5.3^{ab}	14.6 ± 2.0^{b}	17.3 ± 7.3^{ab}	15.0 ± 2.0^{b}	$24.0\pm12.8^{a(x)}$	15
	Wet	16.9 ± 7.1^{ns}	14.8 ± 2.4^{ns}	17.3 ± 8.6^{ns}	13.3 ± 3.2^{ns}	$17.3\pm5.7^{ns(y)}$	15
COD	Dry	29.2 ± 9.2^{ab}	26.5 ± 3.4^{b}	29.6 ± 14.2^{ab}	23.7 ± 4.1^{b}	$43.2\pm23.5^{a(x)}$	27
	Wet	29.9 ± 11.5^{ns}	27.1 ± 5.2^{ns}	30.3 ± 15.8^{ns}	25.0 ± 5.6^{ns}	$34.8\pm14.9^{ns(y)}$	32
$N\text{-}NH_4^+$	Dry	0.65 ± 0.73^{ns}	0.59 ± 0.37^{ns}	$0.49\pm0.25^{ns(x)}$	$0.69\pm0.17^{ns(y)}$	0.67 ± 0.32^{ns}	0.76
	Wet	0.68 ± 0.84^{ab}	0.81 ± 0.57^{ab}	$0.37\pm0.35^{b(y)}$	$1.37\pm1.03^{a(x)}$	0.59 ± 0.43^{b}	0.18
$P\text{-}PO_4^{3-}$	Dry	0.11 ± 0.15^{ns}	0.13 ± 0.07^{ns}	0.21 ± 0.36^{ns}	0.13 ± 0.12^{ns}	0.06 ± 0.04^{ns}	0.09
	Wet	0.13 ± 0.28^{ns}	0.11 ± 0.13^{ns}	0.23 ± 0.44^{ns}	0.18 ± 0.2^{ns}	0.08 ± 0.07^{ns}	0.10
Coliform	Dry	$37{,}415\pm61{,}373^{ns}$	$16{,}750\pm4{,}234^{ns}$	$89{,}754\pm21{,}923^{ns}$	$7{,}600\pm3939^{ns}$	$21{,}877\pm21{,}838^{ns}$	30,000
	Wet	$29{,}130\pm0^{ns}$	$19{,}750\pm0^{ns}$	$79{,}000\pm0^{ns}$	$11{,}200\pm0^{ns}$	$15{,}533.33\pm0^{ns}$	30,000

Notes: Letters a, b in the same row indicated significant different at significance level of 5% ($p < 0.05$). NS: There was no significant difference between the impact sources.

The correlation between seasonal water quality parameters is presented in Figure 2. In the dry season, pH was a positive and weak correlation with DO, a good negative correlation with BOD_5 and COD. DO concentration was negatively correlated, on average with BOD_5 and COD, and weakly negatively correlated with $N\text{-}NH_4^+$. BOD_5 was a very strong positive correlation with COD but has a weak correlation with $N\text{-}NH_4^+$ and coliform. BOD_5 has a weak positive correlation with $N\text{-}NH_4^+$ and coliform. The analysis results also showed that $P\text{-}PO_4^{3-}$ was strongly and positively correlated with coliform (Figure 2a). In the rainy season, pH has a positive correlation with DO, but there was a weak and negative correlation with BOD_5 and COD. In the rainy season, pH was also negatively correlated on average with TSS. DO was moderately negatively correlated with BOD_5 and COD, and negatively correlated with $N\text{-}NH_4^+$. Besides, DO was negatively correlated to a weak level with $P\text{-}PO_4^{3-}$ and coliform. Unlike the dry season, TSS in the rainy season had a weak positive correlation with BOD_5 and COD. BOD_5 had a very good positive correlation with COD, a good positive correlation with $P\text{-}PO_4^{3-}$, coliform and a weak positive correlation with $N\text{-}NH_4^+$. COD during the rainy season was moderately correlated with $P\text{-}PO_4^{3-}$ and coliform while having a weak positive correlation with a $N\text{-}NH_4^+$. $N\text{-}NH_4^+$ has a weak positive correlation with $P\text{-}PO_4^{3-}$ while $P\text{-}PO_4^{3-}$ had a good, positive correlation with coliform (Table 2b). The water quality variables had a change in its correlation characteristics between the rainy season and the dry season, showing the seasonal change of the water quality parameters in the study area.

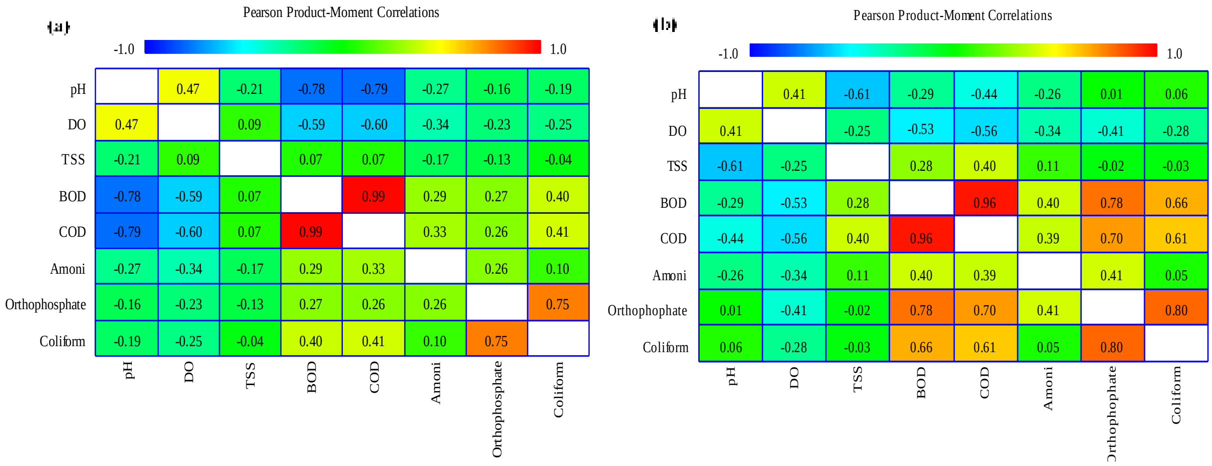

Fig 2. Pearson correlation for water parameters in (a) dry and (b) rainy season

The change over time of water quality in Ca Mau peninsula was evaluated through discriminant analysis (DA) by dividing the data set into 2 periods of dry season and rainy season. The discriminant functions (DF) were built using standard mode and stepwise mode, the obtained the classification matrices were presented in Table 3. As can be seen that the standard mode was characterized by Eigenvalues of 0.389, Wilks Lambda constant of 0.720, and the p-value of lower than 0.05. Similarly, for stepwise mode, it was found that the Eigenvalues, Wilks Lambda constant, and the p-value were 0.382, 0.723, ($p<0.05$), respectively. Therefore, both standard mode and stepwise mode discriminant functions were successfully applied in the calculations. The standard mode method accurately explained the variation of water quality in the dry season and the rainy season by 84.62% and 69.23%, respectively (average 76.92%) using 8 water quality parameters (Table 3). However, the stepwise mode method could explain the change of water quality at an average of 76.92% using only 4 water quality parameters (DO, TSS, BOD_5 and pH) (Table 3). Previous study also showed that DO, TSS, and BOD_5 were indicators that cause seasonal variations in surface water quality (Giao and Nhien, 2021). Thus, the stepwise method can be used to determine the criteria that cause the seasonal difference in water quality in the scope of this study.

Table 3. Classification function coefficients (a) and classification matrix (b) for the temporal discriminant analysis of water quality

(a)

Parameter	Standard mode		Stepwise mode	
	Dry	Wet	Dry	Wet
pH	49.06	49.51	38.01	38.52
DO	1.47	0.72	0.62	-0.13
TSS	0.21	0.23	0.23	0.25
BOD_5	-3.30	-3.46	1.59	1.47
COD	3.21	3.22	-	-
$N\text{-}NH_4^+$	6.29	6.22	-	-
$P\text{-}PO_4^{3-}$	-10.67	-9.74	-	-
Coliform	0.00	0.00	-	-
Constant	-217.75	-217.47	-168.38	-168.67

(b)

Season	% Correct	Season assigned by DA	
		Dry	Wet
Standard mode			
Dry	84.62	44	8
Wet	69.23	16	36
Mean	76.92	60	44
Stepwise mode			
Dry	84.62	44	8
Wet	69.23	16	36
Mean	76.92	60	44

According to the instructions of VEA (2019) regarding the WQI ranges, water quality is divided into 6 levels, including very good (WQI = 91-100) which is good for domestic water use; good (WQI = 76-90) that is used for water supply but need suitable treatment measures; medium (WQI = 51-75) which is used for irrigation and other equivalent purposes; bad (WQI = 26-50) which is used for transportation and other equivalent purposes; heavily polluted (WQI = 10-25) which is severely polluted water requiring future treatment measures; very heavily polluted (WQI < 10) that is contaminated water, required remedial measures. The results of the WQI calculation are shown in Figure 3. Surface water quality in Ca Mau peninsula is classified from good to very polluted. In which, good water quality was found at only one location (1.9%) at NM14, medium at 5 locations (9.6%), bad at 25 locations (48.1%), and heavily polluted at 19 locations (36.5%) and very heavily polluted at two locations (3.8%). Thus, the water quality in the study area is mainly from bad to heavily polluted. Upstream of the Mekong River is classified from bad to good based on the WQI index (15-71) (Giao and Nhien, 2021). This shows that water pollution is one of the problems that can hinder the development of social-economic sectors related to water uses.

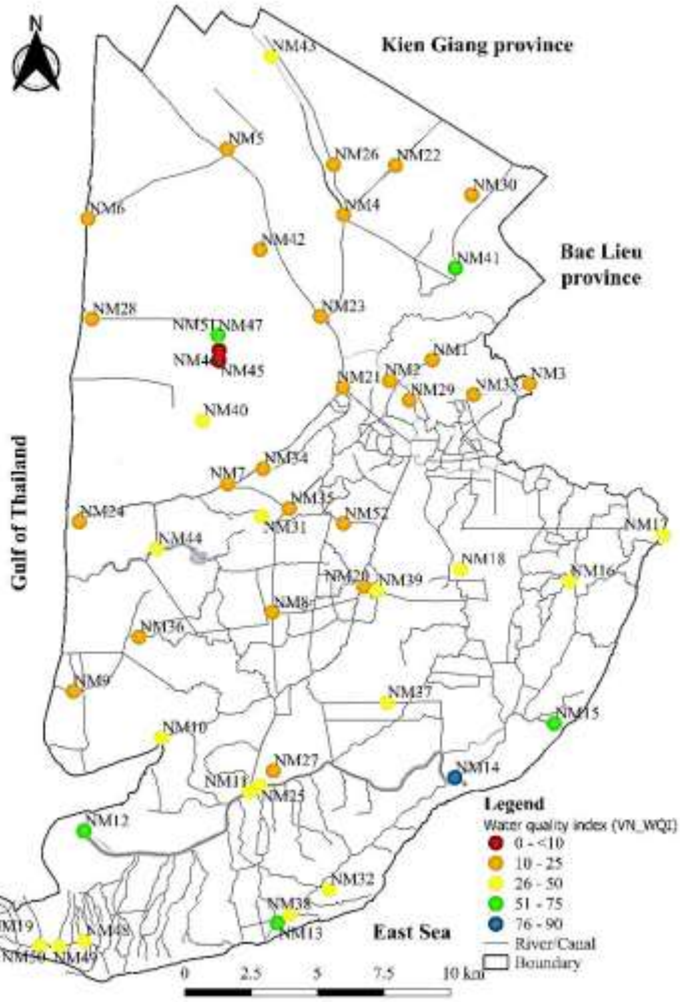

Fig 3. Spatial distribution of WQI in water bodies in Ca Mau province

Spatial variation of surface water quality in Ca Mau peninsula

The results of water quality classification are shown in Figure 4. At Euclidean distance of 40, water quality in the area was classified into 7 clusters. Cluster 1 includes 2 locations (NM1, NM2), cluster 2 includes 1 location (NM29), cluster 3 includes 4 locations (NM45, NM46, NM47, NM51), cluster 4 includes 12 locations (NM3, NM8, NM11). , NM17, NM18, NM21, NM25, NM28, NM33, NM36, NM38, NM52), cluster 5 includes 8 locations (NM9, NM13, NM14, NM15, NM16, NM24, NM27, NM32), cluster 6 includes 15 locations (NM4, NM7, NM10, NM12, NM19, NM20, NM31, NM34, NM37, NM39, NM40, NM44, NM48, NM49, NM50), cluster 7 includes 10 locations (NM5, NM6, NM22, NM23, NM26, NM30) , NM35, NM41, NM42, NM43). The results showed that the water quality in clusters 1-3 was similar with the average WQI of 20.1 indicating heavily polluted water, and water quality in clusters 4-7 was identical with average WQI of 30.8 signaling bad water quality.

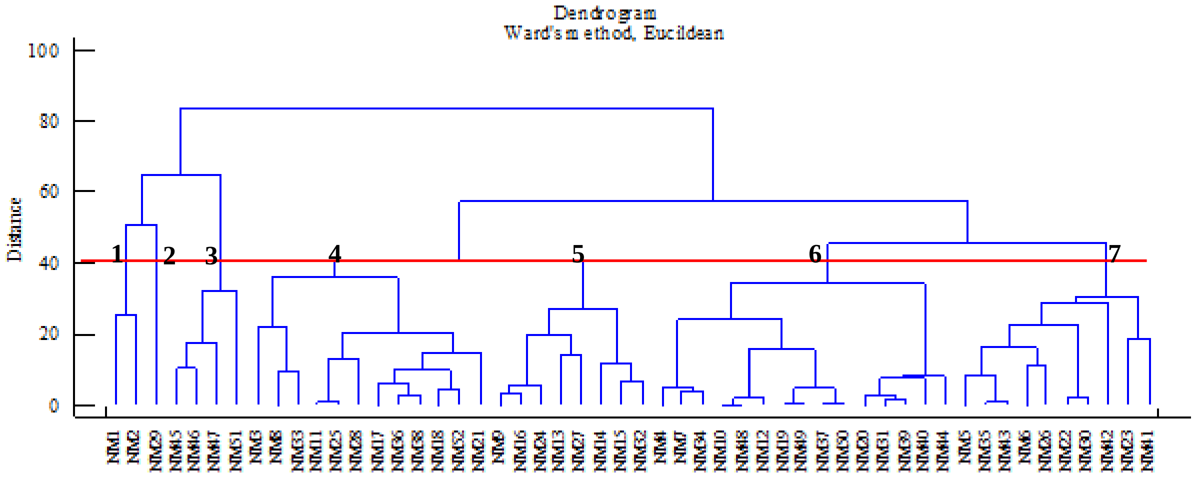

Fig 4. Clustering surface water quality in Ca Mau peninsula

Representative values of water quality which is typical for each cluster were presented in Table 4. Water quality in cluster 1-3 was characterized by low DO and high TSS, BOD_5, COD, $N\text{-}NH_4^+$, and coliform. In contrast to the water quality characterization through the WQI index, the similarity of NM45, NM46 and NM51 positions in cluster 3 was noted due to the significant similarity of pH and TSS compared with the remaining clusters. In cluster 4, the pollution characteristic of this cluster was similar to those of cluster 1-3, but the concentration of water pollutants was lower. Cluster 5-6 was mainly polluted by TSS and coliform while cluster 7 was polluted by low DO and TSS, $N\text{-}NH_4^+$, high coliform. The results confirmed that surface water quality is subjected to spatial variation.

Table 4. Characteristics of surface water quality in the identified clusters

Cluster	pH	DO	TSS	BOD_5	COD	$N\text{-}NH_4^+$	$P\text{-}PO_4^{3-}$	Coliform
1	7.3	**2.1**	**86.8**	**28.8**	**50.3**	**2.6**	0.7	**116,000**
2	7.2	**1.5**	**105.0**	**40.0**	**73.0**	0.5	**1.4**	**675,000**
3	5.7	**1.8**	**182.6**	**29.9**	**58.3**	**0.9**	0.1	**8,275**
4	7.5	**3.5**	**122.0**	**17.6**	**31.1**	0.3	0.1	**29,221**
5	7.6	5.7	**151.7**	13.6	23.7	0.5	0.0	**12,569**
6	7.6	5.0	**79.0**	13.3	23.6	0.4	0.1	**24,660**
7	7.5	**3.3**	**72.5**	14.8	26.6	**1.0**	0.1	**20,495**
Limits	**5.5-9**	**≥4**	**50.0**	**15.0**	**30.0**	**0.9**	**0.3**	**7,500**

The results of principal component analysis showed that 100% of the water quality variation in Ca Mau peninsula was explained by 8 PCs, of which PC1-PC3 accounted for 85.45% of the water quality variation (Table 5). Therefore, there are at least three main potential sources affecting the surface water quality in the study area. According to Shrestha and Kazama (2007), the Eigenvalue coefficient greater than 1 of PCs makes PCs the main component. PC4-PC8 explained 14.55% of the variation in surface water quality in the Ca Mau peninsula so as sub-components. PC1 had a weak correlation with all survey water parameters. PC2 has a weak correlation with pH, $P\text{-}PO_4^{3-}$, coliform and had a moderate correlation with TSS. PC3 had weak correlation with DO, TSS, coliform and high correlation with $N\text{-}NH_4^+$. PC4 had moderate correlation with DO and $N\text{-}NH_4^+$. PC5 has a weak correlation with pH and moderate with DO and TSS. PC6 had a weak correlation with COD, coliform and moderate correlation with pH and BOD_5. PC7 had a moderate correlation with coliform, which correlated well with $P\text{-}PO_4^{3-}$. PC8 was moderately correlated with BOD_5 and COD.

pH was affected by PC6. In this study, the cause of the pH change may be due to the mixing of fresh, brackish, sea water and or seasonal variation. In addition, pH changes can be attributed to the presence of algae, nitrification (Phu and Ut, 2006); DO was found to be affected by the PC4-PC5. DO in rivers depends on diffusion, the presence of phytoplankton and organic matter from nature or by human activities (MRC, 2015; Giao et al., 2021); TSS was affected by PC2 and PC5. The main causes of high TSS can be rainwater runoff, river bank erosion, algae growth, river bed dredging, water exchange between rivers and aquaculture ponds (Phu and Ut, 2006; MRC, 2015; MONRE, 2015; People's Committees of Ca Mau province, 2020); BOD_5 was governed by PC6 and PC8 while COD was affected by PC8. The origin of BOD_5 is mainly due to agricultural activities (cultivation, livestock), domestic wastewater, wastewater from landfills, industrial and service activities (Kazi et al., 2009; MRC, 2015; Chea et al., 2016); $N\text{-}NH_4^+$ was found to be influenced by PC3-PC4. $N\text{-}NH_4^+$ is present in fertilizers used in agriculture and aquaculture, in the decomposition of organic matter arising from human activities (MRC, 2015; Zeinalzadeh and Rezaei, 2017; Chounlamany et al., 2017; People's Committees of Ca Mau province, 2020); $P\text{-}PO_4^{3-}$ was influenced by PC7. The origin of phosphorus can be due to the use of washing powder, human waste in agricultural, fishery and industrial production, wastewater from unsanitary landfills (Barakat et al., 2016; Zeinalzadeh and Rezaei, 2017; Chounlamany et al., 2017; People's Committees of Ca Mau province, 2020); and coliform was affected by PC7. Previous studies have shown that coliforms are present in areas where animal carcasses, human faeces and poultry manure are present (Bolstad and Swank, 1997; UNICEF, 2008; WHO, 2008). The current results showed that the sources of impact on water quality indicators is very complex and the water quality indicators of pH, DO, BOD_5, COD, TSS, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ and coliform all have an impact on surface water quality. Previous studies also confirmed that pH, TSS, DO, BOD_5, COD, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ and coliform criteria and a number of other criteria all have major influences on surface water quality in water bodies in Vietnamese Delta River (Giao, 2020; Giao et al., 2021; Giao and Nhien, 2021; Giao and Minh, 2021).

Table 5. Loading of water quality parameters for PCs

Parameter	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
pH	0.32	-0.48	0.06	0.07	0.48	**0.64**	0.10	0.08
DO	0.37	-0.08	0.31	**0.67**	**-0.56**	0.08	-0.03	-0.01
TSS	-0.15	**0.54**	0.40	0.44	**0.57**	-0.02	-0.03	0.01
BOD_5	-0.48	0.05	0.10	-0.02	-0.20	**0.50**	-0.02	**0.69**
COD	-0.47	0.10	0.09	-0.02	-0.27	0.40	0.08	**0.72**
$N\text{-}NH_4^+$	-0.25	-0.08	**-0.73**	**0.57**	0.10	-0.04	0.24	-0.01
$P\text{-}PO_4^{3-}$	-0.35	-0.49	0.09	0.18	0.11	-0.20	**-0.73**	0.06
Coliform	-0.32	-0.46	0.42	0.04	0.05	-0.36	**0.62**	-0.05
Eigenvalue	4.03	1.78	1.03	0.55	0.34	0.21	0.05	0.01
% Variance	50.43	22.20	12.82	6.83	4.29	2.62	0.67	0.15
% Cumulative	50.43	72.63	85.45	92.27	96.56	99.18	99.85	100.00

CONCLUSIONS

The study results showed that the surface water environment in Ca Mau peninsula was contaminated with organic matters (low DO, high BOD_5 and COD), total suspended solids, nutrients, and microorganisms. DO, BOD_5, COD, TSS, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ were subjected to seasonal fluctuations. In the dry season, pH had a good negative correlation with organic matters, coliform had a good positive correlation with $P\text{-}PO_4^{3-}$. DO had a weak negative correlation with organic matters. In addition, $N\text{-}NH_4^+$ and coliform also had a weak positive correlation with organic matters. In the rainy season, in addition to the correlation with organic matters, pH also had a good negative correlation with TSS, DO had a weak negative correlation with nutrients and coliform. The nutrient indicators, coliform in the rainy season were correlated from weak to good with organic matters. DA results that 4 water quality parameters including DO, TSS, BOD_5 and pH contributed to the seasonal variation, explaining exactly 76.91% of the seasonal variation in water quality. CA divided water quality in Ca Mau peninsula into 7 clusters in which the clusters 1-3 were more heavily polluted than that in the clusters 4-7 mainly due to low concentration of DO and high concentrations of TSS, BOD_5, COD and coliform. PCA results revealed that 100% of water quality variation in Ca Mau peninsula was governed by 8 PCs, of which PC1-PC3 explained 85.45% of water quality variation and was considered as the main potential sources of water quality degradation. Meanwhile, sub-sources of PC4-PC8 only explained 14.55% of surface water quality variation. PCA also showed that pH, DO, BOD_5, COD, TSS, $N\text{-}NH_4^+$, $P\text{-}PO_4^{3-}$ and coliform all have an impact on water quality. The current findings show that the multivariate statistical methods can be useful in analyzing surface water quality monitoring data.

ACKNOWLEDGEMENTS

The authors would like to express our sincere attitude toward the Department of Natural Resources and Environment Ca Mau province for data provision. The scientific and personal views presented in this paper do not necessarily reflect the views of the data provider.

GRANT SUPPORT DETAILS

The present research did not receive any financial support

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research

REFERENCES

American Public Health Association, American Water Works Association, Water Environment

Federation. (1998). Standard Methods for the Examination of Water and Wastewater, 20th Ed. American Public Health Association: Washington, D.C.

Barakat, A., Baghdadi, M. E., Rais, J., Aghezzaf, B. and Slassi, M. (2016). Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. Int. Soil Water Conserv. Res., 4(4); 284-292.

Bolstad, P. V. and Swank, W. T. (1997). Cumulative impacts of landuse on water quality in a southern Appalachian watershed. J. Am. Water Resour. Assoc., 33(3); 519–533.

Cao, L., Guisen, D., Bingbin, H., Qingyi, M., Huimin, L., Zijian, W. and Fu, S. (2007). Biodiversity and water quality variations in constructed wetland of Yongding River system. Acta Ecol. Sin., 27(9); 3670-3677.

Chea R., Grenouillet G. and Lek, S. (2016). Evidence of Water quality degradation in Lower Mekong Basin Revealed by Self-Organizing Map. PLOS ONE/DOL: 10.1371/journal.pone.0145527.

Cho, K. H., Park, Y., Kang, J-H, Ki, S. J., Cha, S., Lee, S. W., and Kim, J. H. (2009). Interpretation of sea-sonal water quality variation in the Yeongsan Reser-voir, Korea using multivariate statistical analyses. Water Sci. Technol., 59(11); 2219-2226.

Chounlamany, V., Tanchuling, M. A. and Inoue, T. (2017). Spatial and temporal variation of water qual-ity of a segment of Marikina river using multivariate statistical methods. Water Sci. Technol., 66(6); 1510-1522.

Galal-Gorchev, H., Ozolins, G. and Bonnefoy, X. (1993). Revision of the WHO guidelines for drinking water quality. Ann. Ist. Super. Sanità, 29; 335-45.

Giao, N. T and Nhien, H. T. H. (2020). Phytoplankton-Water quality relationship in water bodies in the Mekong Delta, Viet Nam. Appl. Environ. Res., 42(2); 1-12.

Giao, N. T and Nhien, H. T. H. (2021). Application of GIS and multicriteria statistical techniques in assessing water quality in the coastal province of Vietnamese Mekong delta. Appl. Environ. Res., 43(3); 17-33.

Giao, N. T. (2019). The use of zoobenthos for the assessment of water quality in canals influenced by landfilling and agricultural activity. J. Viet. Env., 11(2); 33-42.

Giao, N. T. (2020). Evaluating current water quality monitoring system on Hau River, Mekong delta, Vietnam using multivariate statistical technique. Appl. Environ. Res., 2020, 42(1):14-25.

Giao, N. T. and Minh, V. Q. (2021). Evaluating surface water quality and water monitoring variables in Tien River, Vietnamese Mekong Delta. J. Teknol., 83(3); 29-36.

Giao, N. T., Nhien, H. T. H., Anh, P. K. and Ni, D. V. (2021). Classification of water quality in low lying area in Vietnamese Mekong Delta Using set pair analysis method and Vietnamese water quality index. Environ. Monit. Assess., 193(6); 1-16.

Huang, L., Jian, W., Song, X., Huang, X., Liu, S., Quian, P., Yin, K. and Wu, M. (2004). Species diversity and distribution for phytoplankton of the Pearl River estuary during rainy and dry seasons. Mari. Pollut. Bull., 49(7-8); 588-596.

Jerves-Cobo, R., Forio, M. A. E, Lock, K., Butsel, J. A., Pauta, G., Cisneros, F., Nopens, I. and Goethals, P. L. M. (2020). Biological water quality in tropical rivers during dry and rainy seasons: A model-based analysis. Ecologi. Indica, 108; 105769.

Kaiser, H.F. (1974). An index of factorial simplicity. Psychometrika, 39(1); 31–36.

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A. and Shah, A. Q. (2009). Assessment of water quality of polluted reservoir using multivariate statistical techniques: A case study. Ecotoxicol Environ Saf, 72(20); 301-9.

Lien, N. T. K., Huy, L. Q., Oanh, D. T. H., Phu, T. Q. and Ut, V. N. (2016). Water quality in mainstream and tributaries of Hau River. CTUJS, 43; 68-79. (In Vietnamese).

Ly, N. H. T. and Giao, N. T. (2018). Surface water quality in canals in An Giang province, Viet Nam, from 2009 to 2016, J. Viet. Env., 10(2); 113-119.

Mekong River Commission (MRC). (2015). Lower Mekong regional water quality monitoring report. ISSN: 1683-1489. MRC Technical Paper No.51.

Ministry of Environment and Natural Resources (MONRE). 2012. Circular No. 21/2012/TT-BTNMT dated December 19, 2012 providing for quality assurance and quality control in environmental monitoring; Hanoi, Vietnam: MONRE; 2012.

Ministry of Environment and Natural Resources (MONRE). 2015. National technical regulation on surface water quality (QCVN 08-MT: 2015/BTNMT); Hanoi, Vietnam: Ministry of Natural Resources and Environment (MONRE).

Ministry of Natural Resources and Environment (MONRE). (2015). National state of environment-surface water quality.

Ministry of Natural Resources and Environment (MONRE). (2015). National technical regulation on surface water quality (QCVN 08-MT: 2015/BTNMT). Hanoi, Vietnam: MONRE.

Ongley, E. D. (2009). Chapter 12: Water Quality of the Lower Mekong River. In: Campbell, I.C. (ed.): The Mekong: Biophysical Environment of an International. USA: Academic Press.

People's Committee of Ca Mau province. (2020). Report on the state of environment in five years (2015 -2020) of Ca Mau province, 211pages.

Phu, T. Q, Ut, V. N. (2006). Water quality for pond aquaculture. Can Tho University: College of Aquaculture and Fisheries (in Vietnamese).

Salah, E. A. M., Turki, A. M. and Othman, E. M. A. (2006). Assessment of water quality of Euphrates River using cluster analysis. J. Environ. Prot., 3; 1629–1633.

Shrestha, S. and Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin. Environ Model Softw., 22; 464-75.

Singh, K. P., Malik, A. and 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.

Truc, D.T., Phat, P. H., Nam, N. D. G., Toan, P. V. and Tri, V. P. D. (2019). Surface water quality of Tien River flowing through Tan Chau area, An Giang province. CTUJS, 55; 12-22.

Tuan, D. D. A., Thu, B. A. and Trung, N. H. (2019). Assessing quality of surface water for urban water supply source for Soc Trang City. CTUJS, 4A; 61-70. (In Vietnamese).

United Nations Children's Fund (UNICEF). (2008). UNICEF Handbook on Water Quality. New York: United Nations Children's Fund.

Ut, V.N., Loan, N. B., Giang, H. T., Oanh, D. T. H., Lien, N. T. K., Quoc, N. B., Ngoan, N. V., Hoa, A. V. and Tu, P. T. C. (2013). Study on measures to limit the growth of *Limnoperna fortunei* living on *Cipangopaludina lecithoides* in Cho Lach district, Ben Tre. Final report on ministerial-level projects, 100 pages.

Varol, M. (2020). Spatio-temporal changes in surface water quality and sediment phosphorus content of a large reservoir in Turkey. Environ. Pollut., 259; 113-860.

Vega, M., Pardo, R., Barrado, E. and Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res., 32; 3581-92.

Vietnam Environment Administration. (2018). Water quality - Sampling Part 6: Guidance on sampling of rivers and streams (TCVN 6663-6:2018). Hanoi, Vietnam: Ministry of Science and Technology

Vietnam Environment Administration. (2019). Decision 1460/QD-TCMT dated November 12, 2019 regarding the promulgation of Technical Guidelines for calculation and publication of Vietnam's water quality index.

Wijeyaratne, W. M. D. N. and Kalaotuwave, K. M. B. P. P. (2017). Evaluation of the water and sediment quality of a lotic water-body in the western coastal region of Sri Lanka using Rapid Bioassessment Protocol II (RBP II) of benthic macroinverterbrates. Sri Lanka J. Aquat. Sci. 22(2); 85-97.

World Health Organization (WHO). (2008). Guidelines for drinking-water quality - Volume 1: Recommendations Third edition, incorporating first and second addenda. Geneva: World Health Organization.

Zeinalzadeh, K. and Rezaei, E. (2017). Determining spatial and temporal changes of surface water quality using principal component analysis. J. Hydrol. Reg. Stud., 13; 1-10.