Poll Res. 41 (3) : 767-777 (2022)

ISSN 0257–8050
DOI No.: http://doi.org/10.53550/PR.2022.v41i03.001

RELATIVE DISTRIBUTION OF POLLUTANTS FROM URBAN CANAL AND AQUACULTURE FARM ONTO NATURAL WETLAND OF PHNOM PENH, CAMBODIA

R. CHAN[1, 2], R. CHAN[1, 2*], T. SOK[1, 2], S. BUN[1], V. KAING[1], M. MONG[1] AND C. OEURNG[1]

[1]*Faculty of Hydrology of Water Resources Engineering, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh 12156, Cambodia*
[2]*Research Innovation Center, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh 12156, Cambodia*

(Received 2 January, 2022; Accepted 11 February, 2022)

ABSTRACT

Under the rapid development of Phnom Penh, the pollutants control from different sources into the main natural wetland, Tamouk lake is consequentially required for its ecosystem as well as the sustainable development. Therefore, this work aims to investigate the distribution of the pollutants from the urban canal and aquaculture farm to Tamouk lake through the analysis of water quality in both dry and rainy seasons. Twelve sampling sites, i.e., two from urban canals, one from aquaculture farm, and nine from the investigated lake, were selected for their water quality assessment. As the results from both seasonal analyses, the maximum concentrations of total suspended solids (TSS), chemical oxygen demand (COD), and ammonia (NH_3^+) of the effluent from aquaculture farm and urban canal to the lake range of 107-134 mg l^{-1}, 76-184 mg l^{-1}, and 8.5-14.9 mg l^{-1} in the dry season and 105-263 mg l^{-1}, 75-140 mg l^{-1}, and 7.5-9.2 mg l^{-1} in the rainy season, respectively. In addition, TSS and COD in the lake were only 72-93 mg l^{-1} and 46-77 mg l^{-1} of both seasons, respectively, whereas NH_3^+ was estimated to be 7-14 times lower than that in urban canal and aquaculture farms. According to the result analysis, the lake's water pollution was mainly contributed by the discharge from the urban canal and aquaculture farm in both seasons. Higher release of pollutants from many sources during the rainy season could compensate for rainwater's dilution effect, resulting in the high remaining of organic and nitrogen in lake water. Hence, to reduce the risk of the lake water to humans and ecosystems, the effluent from aquaculture farms and urban water is highly suggested to have a proper treatment before discharging to the lake.

KEY WORDS: Physicochemical Properties, Urban Canal, Aquaculture Farm, PCA, Random Forest

INTRODUCTION

Lakes are essential multi-use features that serve as vital freshwater and livelihood sources for rural and urban communities. Lakes also play a crucial role in preserving biodiversity and collecting drainage or rainwater from metropolitan areas (Anand, 2014). Due to a lack of funds to invest in sanitation services, storm water and municipal wastewater are mixed in this application. Thus, rainwater shortens hydraulic retention in sewer lines after heavy rains, transporting mixed water down to the urban canals and finally to the Lake. As a result, the loss of lake water quality is inevitable and would pose a significant environmental challenge. Likewise, with the tremendous economic growth and increased population in urban areas, the lake has begun to suffer from various environmental stresses such as deteriorating its water quality with boosting nutrients and other chemical inputs (Ndungu *et al.*, 2015). However, excessive nutrient levels in the Lake have been attributed to harmful aquatic pollutions, i.e., eutrophication and hypoxia. Besides that, lake water quality pollution could also negatively affect human health through the food web. Thus, monitoring water quality in the urban Lake is often designed and collected from the points (Li and Heap, 2011).

The characteristics of the urban lakes are difficult to classify because they are based on a complex mix of anthropogenic, natural, and biogenic behavior, such as drainage pollutants from different sources in the urban environment, oxidation as a biological mechanism in the Lake, and the phenomenon of climate change (Sun *et al.,* 2019; Isiyaka *et al.,* 2019; Mir and Gani, 2019). One obvious disadvantage is that the obtained data resulted in a large and complex data matrix that is difficult to comprehend to make significant conclusions. To better understand the water quality investigation, multivariate statistical methods such as

Principal component analysis (PCR), cluster analysis (CA), and discriminant analysis (DA) were considered (Singh *et al.,* 2004; Li *et al.,* 2007; Kazi *et al.,* 2009). Among them, PCR has been widely used to identify the factors that influence water quality. This strategy effectively identifies underlying patterns in the original data (Ouyang *et al.,* 2005, Yang *et al.,* 2020). Likewise, an examination of the water quality index (WQI) might demonstrate the level of lake water contamination (Abdul *et al.,* 2010). Apart from that, the random forest (RF) model can handle many predictors and can be effectively trained even with a limited amount of data. Furthermore, the RF model measures the relative importance of descriptors, which may be used to identify the main variations that influence dependent variable variations (Genuer *et al.,* 2010). Furthermore, utilizing the electrical conductivity (EC) as a dependent variable and other water quality as an independent variable, the RF model could predict the occurrence and distribution of contaminants in the surface water environment during the dry and rainy season (Chan *et al.,* 2020). Moreover, geostatistical interpolation methods (GIT) to create a spatial-temporal water quality map has been extensively recognized as an alternative technique for mapping contaminants in lakes. Precisely, Ty *et al.* (2018) used the GIT method for water quality assessment in Tonle Sap Lake, while Todd *et al.*(2010) used a remote sensing approach for soil mapping and modelling of the spatial distribution of sediment oxygen demand at the watershed scale in a coastal area.

Numerous studies have extensively defined water quality monitoring techniques over the last decade with various alternative approaches; however, using each method in Cambodia for water quality monitoring remains limited. Likewise, the multivariate statistical analysis combined with the GIT technique to assess the quality of urban lakes polluted by anthrophonic pollutants has received less attention. Hence, the objective of the present research was to investigate the seasonal dynamics of pollutant distribution from the urban canal and aquaculture farm to Tamouk lake, a floodplain urbanized area in Phnom Penh, through the analysis of main water quality parameters. The correlation and sources of lake water pollution were identified using Pearson correlation and PCR analysis. Moreover, the occurrence and distribution of pollutants from each source to the lake were investigated using the RF model and GIT technique.

MATERIALS AND METHODS

Study area and sampling locations

Tamouk lake, also known as Kob Srov lake, was selected as the study area of this study. This lake is located in the northwestern part of Phnom Penh, between 11.61-11.66 North and 104.72-104.85 East. The lake covers the border of Phnom Penh and Kandal Province. It is roughly 24 km^2 and divided into two sections, with the big half-covering 16.6km^2 and the small part covering just 7.4km^2 (see Figure 1). The water depth is 3.0-4.5 m in the dry season and rises to 5-7.5 m in the rainy season. As seen in Figure 1, the lake is flanked by two inlets, two outlets, and three water gates. Pollutants in Tamouk lake may come from aquaculture farms around the lake, estimated at more than 50 farms. Moreover, another source may come from the urban wastewater, which was conveyed by two urban canals and discharged to the lake body by two pumping stations.

Figure 1 shows the water sampling locations, including the aquaculture farm, urban canals (C1–C2), and Tamouk lake (S1–S9). The effluent from the aquaculture farm is located near Tamouk lake and relies heavily on the lake water for the fish culture. The sampling sites C1 and C2 are the main urban canals that have been used to mitigate flash flooding from Phnom Penhcapital city. In addition, sampling sites S1–S9 are in Tamouk lake, while S1 and S3 are the points next to the effluent of urban canals, C1 and C2. Likewise, sampling sites S6 and S7 are close to the aquaculture farm effluent. All water samples were collected once per month in 2020 and 2021,

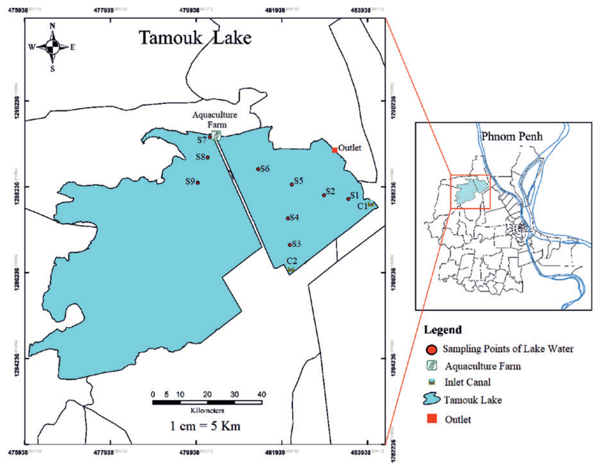

Fig. 1. Study area map and sampling points of Tamouk lake, Phnom Penh

corresponding to the rainy season (August-October 2020) and the dry season (November 2020-January 2021).

Water quality measurement

Effluents of the urban canal, aquaculture farm, and lake waters were collected and brought to analyze at Water and Environmental Laboratory, Institute of Technology of Cambodia (ITC) within 48 hrs. Water quality parameters were includedtotal suspended solids (TSS), chemical oxygen demand (COD), ammonia (NH_3^+), nitrate (NO_3^-), and phosphorus (PO_4^-) in triplicate following the Standard Method for Examination of Water and Wastewater (APHA, 2012), as shown in Table 1. Likewise, pH and electrical conductivity (EC) were measured by pH and EC meters (Hach HQ40D, US).

Water quality index (WQI) calculation

In this research, the WQI was used to estimate the value of physic-chemical characteristics of water quality in the urban canal, aquaculture farms, and lake in both dry and rainy seasons. Those parameters are included pH, EC, COD, NH_3^+, NO_3^-, and PO_4^-. The standard of drinking water from World Health Organization (WHO, 2011) and Cambodian Drinking Water Quality Standard (MIME, 2004) were used in this study. In the WQI calculation procedure, seven water quality parameters were assigned as weight based on their perceived effects on primary health. The assigned weight (AW_i) of each parameter was considered based on the collective expert opinions from previous studies of Kaswanto *et al.* (2012), Kangabam *et al.* (2017), and Wagh *et al.* (2016), as shown in Table 2. After defining the values of AW_i and the number of parameters (n), the relative weight (RW) of each parameter is calculated using Equation (1). The next step is to calculate a quality rating (q_i) for each parameter using Equation (2), where C_i is the concentration of each parameter in each water sample reported in mgl^{-1} and S_i is the WHO standard level of each parameter in each water sample presented in mgl^{-1}. Using Equations (3) and (4), the value of q_i was utilized to calculate the sub-index (SI_i) of each water quality parameter before computing WQI for overall water quality. The WQI calculation procedures employed in this study are based on previous worksof Pradhan *et al.* (2001), Yindana *et al.* (2010), and Bun *et al.* (2021). According to the existed studies of Mohanty *et al.* (2005), Ramakrishnaiah *et al.* (2009), Yadav *et al.* (2010), and Bun *et al.* (2021), the estimated WQI values were

Table 1. Analyzed water quality parameters and analytical methods

Parameters	Unit	Analytical Methods
pH	-	pH meter
Electrical conductivity (EC)	$\mu S\ cm^{-1}$	EC meter
Total suspended solid (TSS)	$mg\ l^{-1}$	2540D
Chemical oxygen demand (COD)	$mg\ l^{-1}$	5220D
Ammonia (NH_3^+)	$mg\ l^{-1}$	4500-NH_3
Nitrate (NO_3^-)	$mg\ l^{-1}$	4500-NO_3
Phosphorus (PO_4^-)	$mg\ l^{-1}$	4500-PO_4

Table 2. Drinking water quality standard of Cambodia and WHO (2011) with assigned weight (AW_i) and relative weight (W_i) of each parameter

Parameter	Cambodia Standard	WHO Standard (2011)	Assigned Weight (A.W.)	Relative Weight (R.W.)
EC ($\mu S\ cm^{-1}$)	1500	250	3.22	0.252
COD ($mg\ l^{-1}$)	-	10	2.00	0.156
NH_3^+ ($mg\ l^{-1}$)	1.5	1.5	3.00	0.235
NO_3^- ($mg\ l^{-1}$)	50	50	2.60	0.201
PO_4^- ($mg\ l^{-1}$)	-	2	2.00	0.156
		Total	**12.8**	**1.0**

divided into five levels, i.e., excellent, good, poor, extremely poor, and inappropriate water for drinking purposes as detailed in Table 3.

$$RW = \frac{AW_i}{\sum_{i=1}^{n} AW_i} \quad \text{Eq.1}$$

$$q_i = \frac{c_i}{s_i} \times 100 \quad \text{Eq.2}$$

$$SI_i = W_i \times q_i \quad \text{Eq.3}$$

$$WQI = \sum SI_i \quad \text{Eq.4}$$

Correlation assessment

Pearson correlation method (Benesty *et al.*, 2009) was applied to analyze the correlation matrix among water quality of urban canals (C1-C2), aquaculture farm effluent (S1-S9), and lakewater (S1-S9), including pH, E.C., COD, NH_3^+, NO_3^-, and PO_4^-. This method was used to evaluate the relationship between water quality variables. Moreover, this method could minimize the effect between-stations correlations and between-sampling campaigns relationships, while the correlation coefficient of each variable could identify the common source, uniform distribution, and similar behaviour of water. The higher absolute value of the correlation coefficient refers to the strong correlation among two variables, while 0.0-0.5 coefficient represents the medium and no correlation at a significant level of $p < 0.05$. Based on Triola 1999, Selected variables are important, whether those with correlation (r) in the module equal or higher than ±0.5 ($r \geq \pm 0.5$, $p \leq 0.05$).

Principle component analysis

Determining surface water quality status, identifying influencing elements of water quality, and improving surface water environment quality are complex and challenging tasks (Canobbio *et al.*, 2013). Thus, PCA was employed to monitor surface water quality in this scenario to provide a comprehensive view of all variables in the system. PCA was also used to provide information on the most meaningful parameters that define the whole data set interpretation and data reduction and summarize the statistical relationship among components in water, with the least amount of loss of original information. Furthermore, the PCA utilizes multi-indicators of original variables in a linear combination to assess correlation. Comprehensive indexes are used to build new variables that are as related to one another as possible and reflect the information of original variables. The preparation of the original data matrix among monitoring stations and water quality parameters, calculation of the Z-score standardization for eliminating the impact of dimension, calculation of the correlation coefficient matrix, calculation of the eigenvalues and eigenvectors of the correlation coefficient matrix, and observation of principal components were the five main operations in this study (Simeonov *et al.*, 2007; Liu *et al.*, 2003).

Random forest model analysis

Dilution lowers contaminant concentrations when wastewater from urban canals and aquaculture farms enters a river or stream. As a result, pollution concentrations in streams differ significantly between the dry and rainy seasons. After such transportation along the stream, the remaining pollutants might be determined using the WQI level. In this case, the dependent variable in the modelling procedure of this study was the WQI value from each targeted site, such as an aquaculture farm, urban canals, and sampling points around the lake. The dry and rainy seasons were also subjected to the same modelling approach. The modelling processes were carried out using the R program (version 3.6.1). The Random Forest (RF) model was used to investigate how WQI variation responded to environmental variables and identify the underlying water quality spatial variations patterns. The percentage decrease accuracy (%IncMSC) for each variable was obtained to measure variable importance in the model using the default settings of the "randomForest" package.

Table 3. Water quality scale from previous works

Water quality	Yadav *et al.* (2010)	Ramakrishnaianh *et al.* (2009)	Mohanty *et al.* (2004)	Bun *et al.* (2021)
Excellent	0-25	<50	<50	<50
Good	25-60	50-100	50-100	50-100
Poor	51-75	100-200	100-200	100-200
Very Poor	76-100	200-300	200-300	200-300
Unsuitable	Above 100	>300	>300	>300

From the model fitting, overall regression (R^2)was computed as an indicator of model fitness on training data. The built models were then applied to the test data, yielding predicted WQI in the test data for further model assessment. Finally, we plotted predicted WQI versus WQIcalculations to see if the RF models fits each season.

Geostatistical mapping techniques

Kriging assumes that at least some of the spatial variation observed in natural phenomena can be modelled by random processes with spatial autocorrelation and require that the spatial autocorrelation be explicitly modelled. This technique can describe and model spatial patterns, predict values at unmeasured locations, and assess the uncertainty associated with a predicted value at the unmeasured locations. In two methods of Kriging, i.e., Simple Kriging (SK) and Ordinary Kriging (OK) are expressed by formula as a general linear regression model among measured values:

$$Z(s) = \mu(s) + \varepsilon(s) \quad \text{Eq. 5}$$

Where Z(s) is the variable of interest, decomposed into a deterministic trend μ(s) and random, autocorrelated errors form [ε(s), and represent a general spatial location to vary continuously over some domain of interest.

A potential geostatistical method, namely Ordinary Kriging (OK), was applied to map water quality due to the outperformance among the other interpolation methods (Chum *et al.*, 2017). Therefore, this study selected OK as an interpolate technique for water quality mapping. OK techniques are generally based on classical statistics, which are affected by the distribution of the grade population underlying the data. Also, it is a spatial estimation method where the error variance is minimized.

$$\hat{Z}(S_0) = \sum_{i=1}^{n} Z(S_i) + \mu(S_0) \quad \text{Eq. 6}$$

Where: nis the number of measured sample points surrounding the opensdiction location used in the prediction.$\hat{Z}(S_0)$ is the value we are trying to predict for location S_0.$Z(S_i)$ is the observed value at the location $S_{0,}$ and μ is a known stationary mean.

RESULTS AND DISCUSSION

Water quality assessment

Figure 2 shows the water quality at all sampled points in the dry and rainy seasons. For aquaculture farms, the average pH value of effluent from the aquaculture pond was 6.9 in the dry season and 7.3 in the rainy season, with an average EC of 355 μScm^{-1} and 370 μScm^{-1}, respectively. In addition, TSS concentration varied from 134 mg l^{-1} to 262.5 mg l^{-1} among both seasons. Aside from that, the average COD and NH_3^+ concentrations in the dry season were 184 mg l^{-1} and 8.5 mg l^{-1} while they were observed of 140 mg l^{-1} and 7.5 mg l^{-1} in the rainy season. Likewise, concentrations of NO_3^- and PO_4^- of 0.7 mg l^{-1} and 1.7 mg l^{-1} in the dry season were slightly increased to 5.7 mg l^{-1} and 2.5 mg l^{-1} in the rainy season. However, fluctuations in COD and NH_3^+ concentrations in aquaculture farm effluent may be induced by the amount of fish feed or fertilizer used and the digesting capability of fish (Kolkovski, 2001). Another possibility is the function of aquatic organisms and microorganisms in the self-purification processes inside the pond; for example, high concentrations of NO_3^- during the rainy season may be caused by nitrifying bacteria oxidizing NH_3^+ to NO_3^- (Daims *et al.*, 2015).

The pH level of urban canal water was neutral (7.2-7.4) over both seasons, while the average EC varied from 305 μScm^{-1} to 401.7 μScm^{-1}, respectively. Furthermore, TSS and COD concentrations in the dry season were 66.7-106.7 mg l^{-1} and 64-76 mg l^{-1},

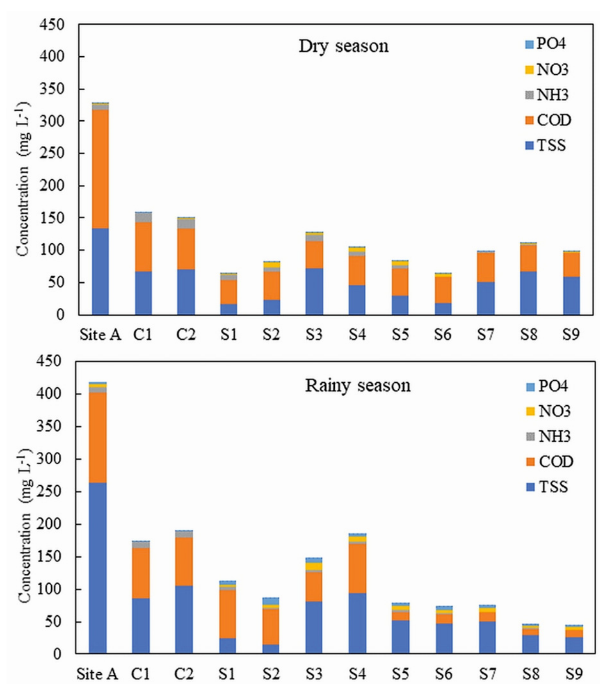

Fig. 2. The water quality at each sampling location on the aquaculture farm, canals, and Tamouk lakes. Site A, C1-C2 and S1-S9 on the horizontal axis refer to effluent from aquaculture farms, urban canals and Tamouk lake sampling sites.

respectively, but in the rainy season, their concentrations were 85-105 mg l^{-1} and 75-78 mg l^{-1}. Regarding nutritional components, NH_3^+ concentrations ranged from 14.3-14.9 mg l^{-1} in the dry season to 8.6-9.2 mg l^{-1} in the rainy season, whereas NO_3^- and PO_4^- concentrations were less than 1.2 mg l^{-1} in both seasons. For lake water quality, the average pH value increased from 8.0 in the dry season to 8.7 in the rainy season, while the average EC also slightly increased from 260 µScm^{-1} to 266 µScm^{-1}. In addition, TSS concentration was investigated at 15.7–71.8 mg l^{-1} in the dry season while in the rainy season was 14.5–93.3 mg l^{-1}. Even though the lake water quality varied at all sample stations, most sampling sites such as S1, S3, and S6 near the outlet of aquaculture farms and urban canals had high nutrient levels. Precisely, COD and NH_3^+ values at these sample locations were 46 mg l^{-1} and 9.7 mg l^{-1} in the dry season and 76.5 mg l^{-1} and 4.3 mg l^{-1} in the rainy season, respectively. But COD and NH_3^+ levels were only 10.4-37 mg l^{-1} and 0.5-0.8 mg l^{-1} over both seasons at other sample locations in the centre of the lake. Other nutrients, such as NO_3^- and PO_4^-, were 1.6 and 10 times higher in the rainy season than in the dry season, with maximum NO_3^- and PO_4^-concentrations of 10.5 mg l^{-1} and 10.9 mg l^{-1}.

WQI Value

Significant differences in water quality among sample locations in the lake suggest that the data are widely spread due to temporal variation caused by natural and anthropogenic polluting factors. Contaminants released from various pollution sources, such as aquaculture farms, urban canals, and other unknown areas, might pollute the lake. Regardless of the lower or higher contribution of contaminants in the lake, using lake water as a source of drinking water could pose a significant risk to human health. The water quality evaluation by WQI level was therefore estimated. WQI values of aquaculture farms, urban canals, and the lake determined in the dry season were 386, 320, and 141, as presented in Table 4. Within this range, according to the reports by Mohanty (2005), Ramakrishna *et al.* (2009), and Yandav *et al.* (2010), the water quality of the aquaculture farm and urban canalswere classified as an unsuitable level for drinking purposes while the lake water quality was also indicatedas poor and unsuitable conditions. There was no significant difference between the two seasons in the WQI values of effluent from the aquaculture farm and canals. However, the lake water might dilute with rainwater (an average of 362 mm per month) during the rainy season resulting in the decrease of certain WQI values. However, its water quality at all measured locations still varied from poor to unsuitable conditions. Therefore, pollutants from aquaculture farms and urban canals discharge to the lake throughout both seasons. Moreover, the high dilution effect of rainwater during the rainy season might enable pollutants in lake water to be flushed down. However, the high release of pollutants through direct discharge or runoff might compensate for this effect, so the lake water quality was determined to be similar to that of the dry season.

Pollutants Sources Estimation

After non-dimensionalizing the original monitoring data, the correlation coefficients matrix among all monitored sample sites in water characteristics were calculated in the R program (Version 3.6.1) using Pearson's correlation ($p \leq 0.05$), as shown in Table 5. In both seasons, pH was negatively correlated with all water parameters except NO_3^-, with correlation coefficient (r) values ranging from -0.58 to -0.94. Apart from pH, water EC was shown to have a

Table 4. Water quality indices and WQI values from aquaculture farm, urban canals and Lake throughout the dry and rainy season

Location	WQI Value	Water quality based on the scale suggested by		
		Yadav *et al.* (2010)	Ramakrishna *et al.* (2009)	Mohanty (2004)
Dry season				
- Aquaculture farm	386	Unsuitable	Unsuitable	Unsuitable
- Urban canal	320	Unsuitable	Unsuitable	Unsuitable
- Lake	141	Unsuitable	Poor	Poor
Rainy season				
- Aquaculture farm	392	Unsuitable	Unsuitable	Unsuitable
- Urban canal	322	Unsuitable	Unsuitable	Unsuitable
- Lake	135	Unsuitable	Poor	Poor

higher correlation with NH_3^+ (r=0.87) and PO_4^- (r=0.78) in the dry season, and TSS (r=0.84), COD (r=0.78), and NH_3^+ (r=0.8) in the rainy season. In addition, during the dry season, TSS also shown to have a positive correlation with COD (r = 0.76) and PO_4^- (r = 0.55), and negative correlation with NO_3^- (r=-0.5). For the rainy season, TSS was negatively correlated with PO_4^- (r = -0.5), but positively correlated with COD (r = 0.78) and NH_3^- (r = 0.61). Other compounds such as COD, NH_3^+, NO_3^-, and PO_4^- also positively correlatedin both seasons, showing the r-value of higher than 0.5.

A screen plot of the sorted eigen values from large to small was shown in Figure 3. The principal components (PCs) in the water dataset were identified in a previous study by Barakat *et al.* (2016) and Li *et al.* (2017), in which the eigen value derived from PCs is more significant than 1.0. Variables that have eigen values less than 1.0, on the other hand, were removed from the study due to their low significance. According to Table 6, two principal components (PC1 and PC2) were chosen in the dry and rainy seasons because their eigen values were 1.6-4.03. Total variance explained of 72.1 % was detected during the dry season. The first principal component (PC1) with a total variance of 49.3% had weak, moderate positive loading of PO_4^-, weak positive loading of pH, and weak negative loading of EC, TSS, COD, and NH_3^+. In addition, the total variation explained for the second principal component (PC2) was 22.8 %, indicating moderate positive loading of NO_3^-, weak positive loading of EC and weak negative loading of TSS and COD. The

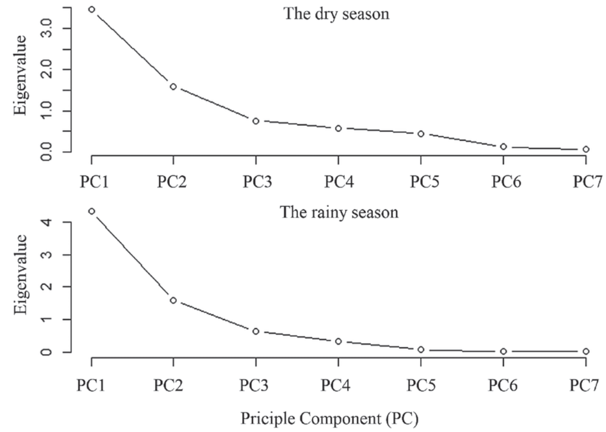

Fig. 3. Screen plot of components, quantifying by variance (eigen values) during the dry and rainy season

total variance explained during the rainy season, on the other hand, was 74.4 %, with 57.5 % found in the PC1 and 16.9 % in the PC2. The PC1 with eigenvalues of 4.03 revealed that pH had weak positive loading while EC, TSS, COD, and NH_3^+had weak negative loading. In contrast, PC2 showed a high positive loading of NO_3^-, buta weak positive loading of TSS and PO_4^-.

Meanwhile, there was a significant correlation between the physical compound (TSS) and chemical compounds (pH, EC, COD, NH_3^+, and PO_4^-), confirming that they have an origin, nevertheless, in the common part. However, the principal components of most variables were only detected at a low level during the dry and rainy seasons; PCs loading quantified the relationships between the

Table 5. Correlation coefficient matrix of water parameters (Pearson correlation coefficient(r)

Parameter	pH	EC	TSS	COD	NH_3	NO_3	PO_4
Dry season							
pH	1						
EC	**-0.83****	1					
TSS	-0.58*	0.17	1				
COD	-0.67*	0.34	0.76**	1			
NH_3	**-0.81****	**0.87****	0.42	0.31	1		
NO_3	0.22	0.17	-0.5*	-0.34	-0.11	1	
PO_4	**-0.94****	0.78*	0.55*	0.79*	0.69*	-0.2	1
Rainy season							
pH	1						
EC	-0.70*	1					
TSS	-0.74*	**0.84****	1				
COD	**-0.83****	0.78*	0.78*	1			
NH_3	**-0.87****	**0.8****	0.61*	**0.8****	1		
NO_3	0.29	-0.2	0.01	-0.18	-0.56*	1	
PO_4	-0.76*	-0.39	0.5*	-0.26	-0.54*	0.67*	1

Noted: * refers to medium correlation, ** refers to strong correlation

Table 6. Loading of seven variables among the dry and rainy seasons on the principal components for the whole datasets

PCs	pH	EC	TSS	COD	NH_3	NO_3	PO_4	Eigenvalues	% of variance
Dry season									
PC_1	0.33	-0.38	-0.35	-0.4	-0.43	0.17	0.50*	3.46	49.3
PC_2	-0.28	0.46	-0.43	-0.3	0.19	0.61*	0.05	1.60	22.8
Rainy season									
PC_1	0.42	-0.41	-0.39	-0.40	-0.46	0.18	0.27	4.03	57.5
PC_2	-0.004	0.24	0.35	0.19	-0.13	0.76*	0.41	1.19	16.9

Noted: **Strong > 0.75, *moderate: 0.5 ≥ factor loading ≥ 0.75, and weak: 0.3 ≥ factor loading ≥ 0

water quality variables of all monitored stations could be attributed to biogenic and anthrophonic pollution sources. Precisely, the COD is widely used as a metric for assessing waste concentration in rivers or streams (Kazi *et al.*, 2009). Otherwise, excessive concentrations of this compound in urban canals and lakes might be created by natural, residential sewage, agricultural, aquaculture, and industrial pollution discharge or leaching (Li *et al.*, 2017). Some nutrient compounds like NO_3^- and NH_3^+ are well-known waste products from domestic and aquaculture activities. PO_4^- also originates in agricultural areas where inorganic nitrogen and phosphorus fertilizers are often used (Li *et al.*, 2016). Other contaminants, such as TSS and pH, might come from runoff containing a high load of solids and waste from pollution sources such as farm fields and domestic areas (Gazzaz *et al.*, 2012). To sum up, the release of polluted water from various sources may affect lake water quality's spatial and temporal variation in both dry and rainy seasons. The discharging of pollutants from the urban canals and aquaculture farms through direct discharge or runoff might be a significant source of organic and nutrient pollution to Tamouk Lake, based on the topography of the study area.

Pollutants temporal analysis

RF model was used to determine the occurrence and distribution of contaminants in the urban canal and aquaculture wastewatersto Tamouk Lake. The RF model was calibrated using samples from all seven predictors (pH, EC, TSS, COD, NO_3^-, NH_3^+, and PO_4^-) and a single targeted variable (WQI). The percentage of variation explained by the RF model during the dry season was 86.75%. Based on %IncMSE, NH_3^+presented as the most significant in the model following by PO_4^-, COD, EC, TSS, pH, and NO_3^- (Figure 4). During the rainy season, the percentage of variation explained by the model was increased to 89.24%. In addition, COD was found as the most important variable in the model, showing the %IncMSE of 26%. Other variables also performed as significant factors. But their %IncMSE just varied from 7.7% to 18.9%, respectively.

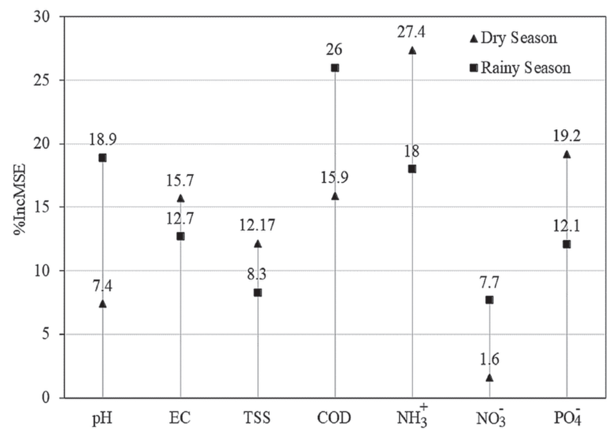

Fig. 4. Conditional ranking of the predator importance obtained by the RF model for urban canals, aquaculture farm, and Tamouk Lake during the dry and rainy seasons

Figure 5 illustrates the actual and predicted WQI values during the dry and rainy seasons. In the dry season, the coefficient of determination of the RF model was 0.96, but in the rainy season, the coefficient of determination was only 0.93. It can be notified that the model can estimate the WQI values with a discrepancyof about 20% for both seasons. Higher error can be observed at WQI between 50 and 150. This constructed model can be used to estimate WQI with acceptable performance.

Pollutants spatial analysis

Figure 6 shows the spatial and temporal variation of water quality of Tamouk Lake, quantified by the WQI value of all monitored stations. According to the spatial maps during both seasons, the poor lake water quality was highly detected at sampling sites surrounding the aquaculture farm and close to outlets of urban canals, while the water quality slightly became better when it moved toward the center of the lake. Based on RF models and spatial

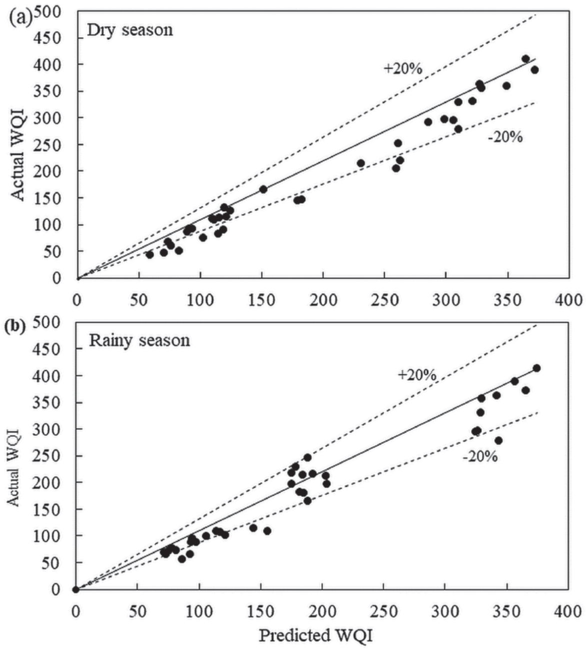

Fig. 5. Actual and predicted WQI values with 20% discrepancy: (a) dry season, (b) rainy season

maps, Tamouk Lake water pollution was highly affected by wastewater discharge and effluent from urban canals and aquaculture farms. Moreover, the distribution of pollutants from those sources to the river likely occurred during both seasons. From urban canals, the sources of pollutants in these canals may be essentially the polluted water from the capital city of Phnom Penh (Kelderman *et al.*, 2000). Those contaminants enter urban canals seasons due to pollutant sources being discharged from aquaculture farms and urban canals. Besidesthat, RF models in both seasons indicated that more frequently release of pollutants to the lake during the rainy season in order to prevent flooding in urban areas and overflow from fish pond culture could compensate for rainwater's dilution effect, resulting in the remaining high organic and nitrogen level in the lake water. In this case, to control the water quality in Tamouk Lake, urban canal water and effluent from aquaculture farms should be managed.

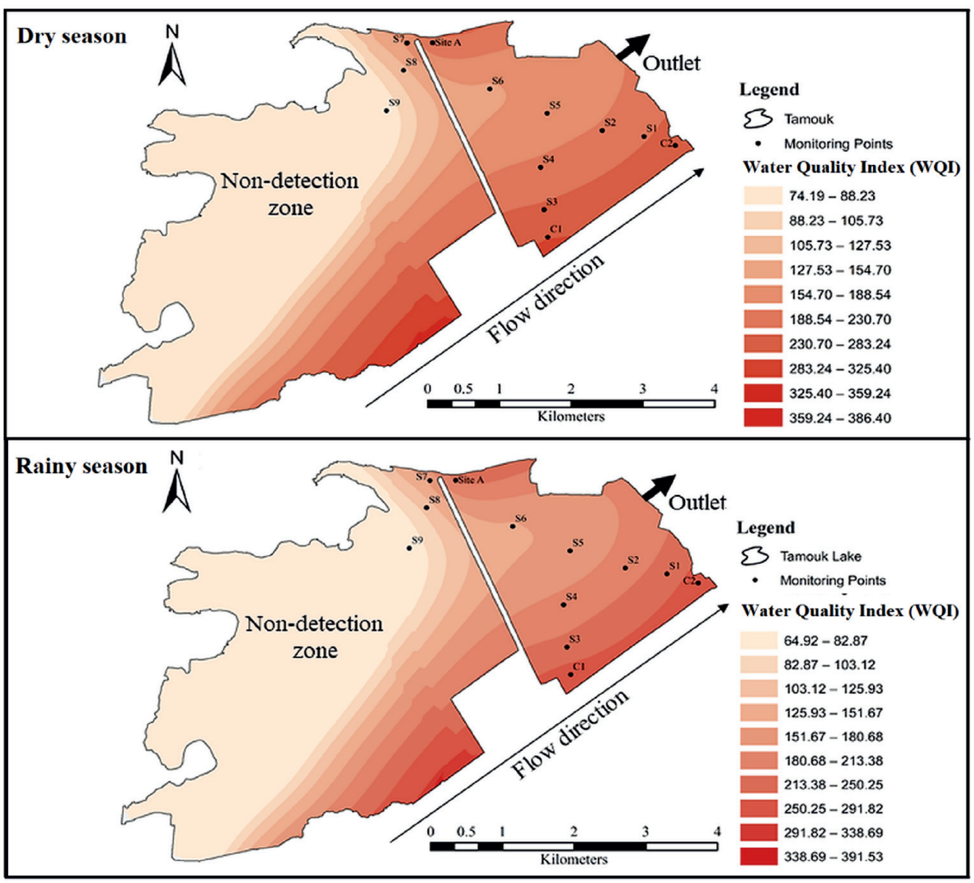

Fig. 6. Spatial distribution map of total pollutions from the aquaculture farm and canals to the Lake among dry and rainy seasons. Note that Site A and C1–C2 refer to all sampling locations of the aquaculture farm and urban canals, while S1–S9 refer to sampling locations of the Lake.

ACKNOWLEDGEMENT

The authors are grateful to the European Union (EU) and Agence Française de Développement (AFD) for financial support. The authors also expresstheir deep appreciation to AUN/SEED-Net Program of Japan International Cooperation Agency (JICA) through Alumni Support Program for Research No. ITC ASP-R 2101 for some financial support on water quality analysis.

Conflict of interest

The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCE

Abdul Hameed, M., Jawad, A., Haider, S, A. and Bahram, K. M. 2010. Application of water quality index for assessment of Dokan lake ecosystem, Kurdistan region, Iraq. *Journal of Water Resource and Protection*. 2 (9): 792-798.

American Public Health Association (APHA). 2012. Standard Methods for the Examination of Water and Wastewater, 22nd ed. American Public Health Association, Washington, D.C., USA. Pages 724.

Amini, M. A., Torkan, G., Eslamian, S., Zareian, M. J. and Adamowski, J. F. 2019. Analysis of deterministic and geostatistical interpolation techniques for mapping meteorological variables at large watershed scales. *Acta Geophysica*. 67(1): 191-203.

Anand, A. K. S. H. A. Y. 2014. *Linking urban lakes: assessment of water quality and its environmental impacts.* (Master's thesis, University of Twente).

Barakat, A., El Baghdadi, M., 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. *International Soil and Water Conservation Research*. 4(4): 284-292.

Benesty, J., Chen, J., Huang, Y. and Cohen, I. 2009. Pearson correlation coefficient. In: *Noise Reduction in Speech Processing. Springer*, Berlin, Heidelberg. Pages 1-4.

Boyd, C. E. 2003. Guidelines for aquaculture effluent management at the farm-level. *Aquaculture*. 226(1-4): 101-112.

Bun, S., Sek, S., Oeurng, C., Fujii, M., Ham, P. and Painmanakul, P. 2021. A Survey of Household Water Use and Groundwater Quality Index Assessment in a Rural Community of Cambodia. *Sustainability*. 13(18): 10071.

Canobbio, S., Azzellino, A., Cabrini, R. and Mezzanotte, V. 2013. A multivariate approach to assess habitat integrity in urban streams using benthic macroinvertebrate metrics. *Water Science and Technology*. 67(12): 2832-2837.

Chan, R., Wandee, S., Wang, M., Chiemchaisri, W., Chiemchaisri, C. and Yoshimura, C. 2020. Fate, transport and ecological risk of antibiotics from pig farms along the bang pakong River, Thailand. *Agriculture, Ecosystems & Environment*. 304(2020): 107123.

Chum, K., Ty, S., Chan. R, Heng, B., Sieve, S., Yang, H., Yushimora, C., Layheang, S. and Chantha, O. 2017. Assessment of Spatial Interpolation Methods to Map Water Quality in Tonle Sap Lake. In: *the 2nd International Symposium on Conservation and Management of Tropical Lakes*, Siem Reap, Cambodia.

Daims, H., Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M. and Wagner, M. 2015. Complete nitrification by Nitrospira bacteria'. *Nature*. 528(7583): 504-509.

Gazzaz, N. M., Yusoff, M. K., Ramli, M. F., Aris, A. Z. and Juahir, H. 2012. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Marine Pollution Bulletin*. 64(4): 688-698.

Genuer, R., Poggi, J. M. and Tuleau-Malot, C. 2010. Variable selection using random forests. *Pattern Recognition Letters*. 31(14): 2225-2236.

Isiyaka, H. A., Mustapha, A., Juahir, H. and Phil-Eze, P. 2019. Water quality modelling using artificial neural network and multivariate statistical techniques. *Modeling Earth Systems and Environment*. 5(2), 583-593.

Kangabam, R. D., Bhoominathan, S. D., Kanagaraj, S. and Govindaraju, M. 2017. Development of a water quality index (WQI) for the Loktak Lake in India. *Applied Water Science*. 7(6): 2907-2918.

Kaswanto, R. L., Arifin, H. S. and Nakagoshi, N. 2012. Water quality index as a simple indicator for sustainability management of rural landscape in West Java, Indonesia. *International Journal of Environmental Protection*. 2(12): 17-27.

Kazi, T.G., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, H.I., Sarfraz, R.A., Baig, J.A. and Shah, A.Q. 2009. Assessment of water quality of polluted Lake using multivariate statistical techniques: A case study. *Ecotox. Environ. Safe*. 72(2): 301-309.

Kelderman, P., Drossaert, W. M. E., Min, Z., Galione, L. S., Okonkwo, L. C. and Clarisse, I. A. 2000. Pollution assessment of the canal sediments in the city of Delft (the Netherlands). *Water Research*. 34(3): 936-944.

Kolkovski, S. 2001. Digestive enzymes in fish larvae and juveniles-Implications and applications to formulated diets. *Aquaculture*. 200(1-2): 181-201.

Kistemann, T., Claßen, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J. and Exner, M. 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Applied and Environmental Microbiology*. 68(5): 2188-2197.

Li, B., Yang, G., Wan, R., Hörmann, G., Huang, J., Fohrer, N. and Zhang, L. 2017. Combining multivariate statistical techniques and random forests model to assess and diagnose the trophic status of Poyang Lake in China. *Ecological Indicators*. 83: 74-83.

Li, R., Dong, M., Zhao, Y., Zhang, L., Cui, Q. and He, W. 2007. Assessment of water quality and identification of pollution sources of plateau lakes in Yunnan (China). *J. Environ. Qual*. 36(1): 291-297.

Li, B., Yang, G., Wan, R., Zhang, Y., Dai, X. and Chen, Y., 2016. Spatiotemporal variability in the water quality of Poyang Lake and its associated responses to hydrological conditions. *Water*. 8(7): 2-16.

Li, J. and Heap, A. D. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*. 6(3-4): 228-241.

MIME. 2004. Drinking Water Quality Standard. Ministry of Industry Mines and Energy. Phnom Penh, Cambodia

Mir, R. A. and Gani, K. M. 2019. Water quality evaluation of the upper stretch of the river Jhelum using multivariate statistical techniques. *Arabian Journal of Geosciences*. 12(14): 1-19.

Mohanty, S. K. 2005. Water Quality Index of four Religious Ponds and its Seasonal Variation in the Temple City, Bhubaneswar. *Fundamentals of Limnology*. Pages162.

Ndungu, J., Augustijn, D. C., Hulscher, S. J., Fulanda, B., Kitaka, N. and Mathooko, J. M. 2015. A multivariate analysis of water quality in Lake Naivasha, Kenya. *Marine and Freshwater Research*. 66(2): 177-186.

Ouyang, Y. 2005. Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*. 39(12): 2621-2635.

Pradhan, S. K., Patnaik, D. and Rout, S. P. 2001. Water Quality Index for the ground water around a phosphatic fertilizer plant. *Indian Journal of Environmental Protection*. 21(4): 355-358.

Ramakrishnaiah, C. R., Sadashivaiah, C. and Ranganna, G. 2009. Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *E-Journal of Chemistry*. 6(2): 523-530.

Singh, K.P., Malik, A., Mohan, D. and Sinha, S. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India): A case study. *Water Res*. 38(18): 3980-3992.

Sun, X., Zhang, H., Zhong, M., Wang, Z., Liang, X., Huang, T. and Huang, H. 2019. Analyses on the temporal and spatial characteristics of water quality in a seagoing river using multivariate statistical techniques: A case study in the Duliujian River, China. *International Journal of Environmental Research and Public Health*. 16(6): 1020.

Todd, M. J., Lowrance, R. R., Goovaerts, P., Vellidis, G. and Pringle, C. M. 2010. Geostatistical modeling of the spatial distribution of sediment oxygen demand within a Coastal Plain blackwater watershed. *Geoderma*. 159(1-2): 53-62.

Ty, S., Ilan, I., Heejun, Y., Chantha, O., Layheang, S., Sokly, S. and Chihiro, Y. 2018. Spatio-temporal variability of water quality in a large shallow lake in southeast Asia: Tonle Sap Lake, Cambodia. In :*17th World Lake Conference*. Pages 2-4.

Wagh, V., Panaskar, D., Mukate, S., Lolage, Y. and Muley, A. 2017. Groundwater quality evaluation by physico chemical characterization and water quality index for Nanded Tehsil, Maharashtra, India. In: *Proceedings of the 7th International Conference on Biology, Environment and Chemistry*, *IPCBEE* (Vol. 98).

WHO 2011. Guidelines for Drinking-Water Quality, the 4th Edition. World Health Organization,Geneva. Switzerland.

Yadav, A. K., Khan, P. and Sharma, S. K. 2010. Water Quality Index Assessment of Groundwater in Todaraisingh Tehsil of Rajasthan State, India-A Greener Approach. *E-Journal of Chemistry*. 7(S1): S428-S432.

Yang, W., Zhao, Y., Wang, D., Wu, H., Lin, A. and He, L. 2020. Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang river in Huangshan, China. *International Journal of Environmental Research and Public Health*. 17(8): 2942.

Yidana, S. M. and Yidana, A. 2010. Assessing water quality using water quality index and multivariate analysis. *Environmental Earth Sciences*. 59(7): 1461-1473.