

Assessment of Surface Water Quality by Using Multivariate Statistical Analysis Techniques: A Case Study of Nhue River, Vietnam

Nguyen Huu Hue and Nguyen Huu Thanh

Abstract—The aim of this study is to assess the spatial variability and to determine the main contamination sources in surface water quality of the Nhue River, Viet Nam by using multivariate statistical analysis techniques, including principal component analysis (PCA) and cluster analysis (CA). Eight water quality parameters were measured at 21 sites along the Nhue River and its tributaries during irrigated periods from 2016 to 2019. The spatial variability of water quality in the Nhue River and its tributaries was determined separately from cluster analysis. The result determined two tributaries, including Yen Xa Canal (NT9 monitoring site) and To Lich River (NT3 monitoring site) leading to severe pollution at To Bridge (N4 monitoring site) region in the Nhue River. The PCA determined a reduced number of two principal components that explained 47.75% of the total variation in the data. The first PC indicated that water temperature (WT) and pH are the dominant polluting factors which are attributed to craft villages, domestic sewage and industrial wastewater. Following is nitrate nitrogen NO_3^- in the second PC which is related to fertilizer application in the farms nearby. The results indicated that CA multivariate statistical analysis technique is useful for the assessment of the spatial water quality variability in a river which has a number of tributaries.

Index Terms—Water quality, principal component analysis, cluster analysis, multivariate statistical analysis.

I. INTRODUCTION

Rivers play an important role as a primary input for a huge array of human consumption and economic activities, including domestic and industrial water use, irrigation, agriculture, recreation and transport. Surface water quality is both influenced by natural processes and anthropogenic activities [1] through the discharge of industrial and domestic wastewater, craft villages, urban runoff as well as agricultural drainage to the rivers. At present, many river basins in Vietnam such as Cau, Nhue, Day and Dong Nai basins belong to the most polluted river group threatening health of the surrounding residents and native eco lives. Therefore, the government has been monitoring water quality of these rivers for years in order to manage water resources efficiently. However, long-term monitoring datasets are large, with

complex matrixes comprising numerous physico-chemical parameters. Consequently, it is often difficult for planners to extract meaningful information from these datasets, identify significant parameters, and apportion pollution sources [2]. The multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) have been widely accepted as a valuable tool in: (1) analyzing and interpreting the complex data matrices to better understand the water quality and ecological status of the studied systems; (2) verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality; and (3) identifying the possible factors/sources that influence water systems [3]-[5]. Due to the ability of multivariate statistical techniques to treat large volume of spatial and temporal data from a variety of monitoring sites, they have become popular for a better understanding of water quality and ecological status in recent year [6].

In the present study, a large data matrix of 8 water quality parameters were selected and collected from 2016 to 2019 at 21 monitoring sites in the Nhue River, Viet Nam. They were subjected to the PCA and CA multivariate techniques to extract information of water quality data. Firstly, similarities and dissimilarities among 10 sites along the Nhue River, and among 11 sites on its tributaries were classified by mean of CA. Then, the complex water quality datasets were analyzed to extract latent water quality factors using PCA. Finally, the possible pollution sources on water quality were identified.

II. MATERIALS AND METHODS

A. Study Area

The Nhue River takes its source from the Red River at the Lien Mac 1 Gate to the north west of Ha Noi (Fig. 1). The upstream part of the Nhue River with length of about 20 km mainly lies in the urban areas of Ha Noi city. The rest of the river is located in areas where agriculture is the predominant land use [7]. Like other rivers running in the Red River Delta of Viet Nam, the Nhue River flows south and southeast throughout its course without abrupt redirection or disruption. In order to prevent flood and to ensure irrigation during crop season, the discharge in the river is entirely regulated by dams/slucice gates such as Lien Mac 1, Nhat Tuu, Luong Co which were constructed from 1936 to 1941, during French colonization. The river basin is bordered by the Red River to the north and the east, the Day River to the west, the Chau Giang River to the south. To the west, the Day River almost runs parallel with the Nhue River at the distance of 10 km. To the south, the Nhue River joins the Day River at Phu Ly city,

Manuscript received January 28, 2020; revised August 12, 2020. This work was supported in part by the Office of National Science and Technology Research Programs, the Ministry of Science and Technology under Grant KC.08.27/16-20.

Nguyen Huu Hue is with the Faculty of Civil Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam (e-mail: nguyenuhuuhue@tlu.edu.vn).

Nguyen Huu Thanh is with the Institute of Civil Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam (e-mail: nguyenuhuuthanh@tlu.edu.vn).

74 km from its source. To the east, the Nhue basin is actually limited by the national highway number 1 and this highway also runs parallel with the river. The Nhue River has also several significant inflows such as Dam, Cau Nga, Xuan La, Phu Do, Trung Van, La Khe, Yen Xa, To Lich. Of which the To Lich River, a tributary of the Nhue River, is reportedly responsible for 77.5 km² portion of Ha Noi area. The Duy Tien and Van Dinh Canals were built and canalized for irrigation and drainage of the surrounding paddy fields.

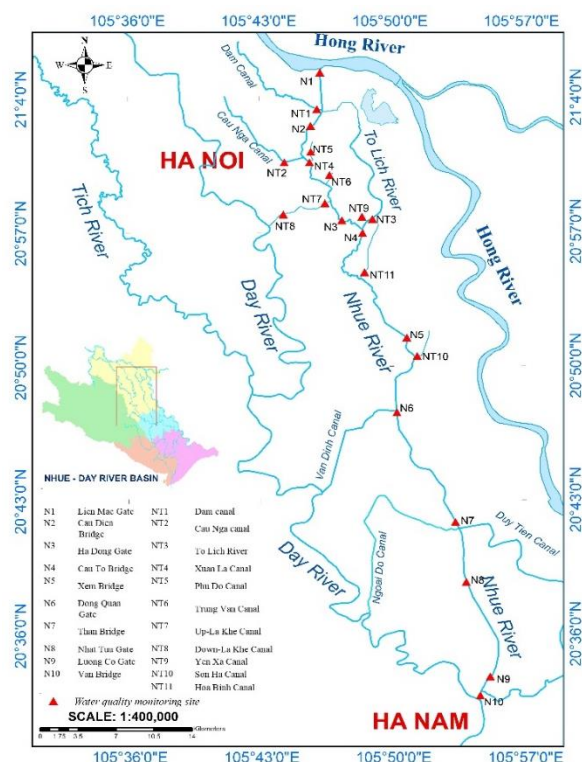


Fig. 1. Location map of study area and water quality monitoring sites in the Nhue River and its tributaries.

The Nhue River is considered as an inter-province hydro-agriculture system with different industrial and agricultural zones which are responsible for activate irrigation of 81,710 ha. All along its course, the river presents a netted canalization system in charge of irrigation for paddy fields. It is observed that irrigation canals spread all over the river basin and hook up to the Nhue every 2 or 3 km. At the moment, the Nhue River is seriously polluted by large amount of wastewater from domestic waste, hospital waste, industrial waste, agricultural waste, and craft village waste discharging directly into the river without being treated. To deal with this situation, the Directorate of Water Resources (DWR) under Ministry of Agriculture and Rural Development (MARD) has been monitoring water quality at 21 sites (which covers all surface water types, including the main river and tributaries) along the river to make decision and action plan for irrigation.

B. Monitoring Sites and Water Quality Parameters

In this study, the datasets of 21 water quality monitoring sites of the Nhue River during irrigated periods from 2016 to 2019. In which there are 10 monitoring sites along the main river and 11 sites on its tributaries. The water quality parameters, including water temperature (WT), pH, turbidity (Tur), dissolved oxygen (DO), total dissolved solid (TDS),

ammonia nitrogen (NH₄⁺), nitrate nitrogen (NO₃⁻), and conductivity (EC) were selected and analyzed according to the American Public Health Association standard (APHA, 1998). All the water quality parameters are expressed in mg/L, except temperature (°C), pH, and turbidity (NTU). The statistical summary of the water quality parameters monitored at 21 sites was shown in Table II.

TABLE I: THE DETAILS OF MONITORING SITES IN THE NHUE RIVER

Site	Name	Source of wastewater	Location (N E)
N1	Lien Mac Gate	Upstream	21°05'20"; 105°46'16"
N2	Cau Dien Bridge	Industrial, Domestic	21°02'35"; 105°45'43"
N3	Ha Dong Gate	Domestic, Craft village	21°57'40"; 105°47'24"
N4	Cau To Bridge	Domestic, Hospital	21°56'56"; 105°48'29"
N5	Xem Bridge	Domestic, Craft village	21°51'28"; 105°50'49"
N6	Dong Quan Gate	Urban areas of Hanoi, Agricultural drainage	21°47'35"; 105°50'48"
N7	Than Bridge	Urban areas of Hanoi, Agricultural drainage	21°41'53"; 105°53'20"
N8	Nhat Tuu Gate	Urban areas of Hanoi	21°38'46"; 105°53'52"
N9	Luong Co Gate	Urban areas of Hanoi	21°33'47"; 105°55'08"
N10	Van Bridge	Urban areas of Hanoi	21°32'49"; 105°54'36"
NT1	Dam canal	Domestic, Agricultural drainage	21°03'25"; 105°46'05"
NT2	Cau Nga canal	Domestic, Agricultural drainage	21°00'40"; 105°44'20"
NT3	To Lich River	Domestic	21°57'43"; 105°49'01"
NT4	Xuan La Canal	Domestic, Agricultural drainage	21°00'39"; 105°45'40"
NT5	Phu Do Canal	Domestic, Craft village	21°01'12"; 105°45'43"
NT6	Trung Van Canal	Domestic, Agricultural drainage	21°00'02"; 105°46'44"
NT7	Upstream of La Khe Canal	Domestic	21°58'33"; 105°46'30"
NT8	Downstream of La Khe Canal	Domestic, Craft village	21°57'58"; 105°44'17"
NT9	Yen Xa Canal	Domestic, Hospital, Craft village	21°57'50"; 105°48'29"
NT10	Son Ha Canal	Domestic, Agricultural drainage	21°50'31"; 105°51'22"
NT11	Hoa Binh Canal	Domestic, Agricultural drainage	21°54'54"; 105°48'36"

C. Data and Multivariate Statistical Methods

Descriptive statistics (mean) and multivariate statistical techniques, including cluster analysis (CA) and principal component analysis (PCA) were performed on the datasets using SPSS 25.0.

CA: CA is a notable method that assembles objects into aggregations based on their interdependent variables or characteristics [8]. Hierarchical agglomerative clustering is the most common approach, which starts with each case in a

separate cluster and joins the clusters together step by step until only one cluster remains and is typically illustrated by a dendrogram (tree diagram) [9]. The dendrogram provides a visual summary of the clustering process, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of original data. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between analytical values from samples.

In this study, cluster analysis was applied to detect spatial similarity for grouping of sites under the monitoring network. This work was performed on the standardized data by means of Ward’s method, with Euclidean distance as a measure of similarity.

TABLE II: MEAN OF WATER QUALITY PARAMETERS AT 21 MONITORING SITES IN THE NHUE RIVER DURING 2016-2019

Site	Water quality parameters							
	WT	pH	Tur	DO	TDS	NH ₄ ⁺	NO ₃ ⁻	EC
N1	20.50	7.51	30.73	1.87	0.35	5.95	1.06	0.48
N2	20.50	7.52	32.13	0.96	0.40	18.78	1.16	0.54
N3	22.13	7.41	31.04	0.14	0.50	21.63	1.89	0.66
N4	22.63	7.41	33.95	0.15	0.53	50.60	2.57	0.73
N5	23.25	7.53	32.69	0.17	0.51	23.72	2.60	0.68
N6	24.63	7.47	37.93	0.83	0.41	12.07	1.72	0.57
N7	26.33	7.57	31.37	0.56	0.42	22.91	1.14	0.57
N8	24.88	7.59	31.89	2.28	0.39	9.45	1.13	0.53
N9	24.50	6.64	19.58	3.90	0.35	3.61	1.90	0.48
N10	24.50	7.46	23.16	4.27	0.30	2.83	2.04	0.42
NT1	20.50	7.47	63.10	1.48	0.38	9.28	1.30	0.53
NT2	20.75	7.21	77.90	0.29	0.46	6.35	1.32	0.62
NT3	22.75	7.51	32.50	0.10	0.65	96.13	3.38	0.88
NT4	20.63	7.44	32.78	0.37	0.58	16.64	1.70	0.75
NT5	21.00	7.46	43.08	0.41	0.63	28.27	2.20	0.80
NT6	21.13	7.51	78.33	0.43	0.60	28.76	1.90	0.79
NT7	21.75	7.52	38.96	0.58	0.45	25.75	1.29	0.61
NT8	21.13	7.61	28.19	1.11	0.48	50.28	1.41	0.64
NT9	23.63	7.62	60.73	0.62	0.71	236.3	2.11	0.90
NT10	24.63	7.55	29.39	3.35	0.44	12.80	1.13	0.57
NT11	24.17	7.50	21.35	4.20	0.35	8.93	1.19	0.49

Water temperature (T) is in °C, electrical conductivity (EC) is in μS/cm, and other parameters are mg/l.

PCA: PCA is a helpful recognition technique applied for extracting information concisely by transforming many original, interrelated variables into fewer, uncorrelated variables named principal components (PCs). PCA focuses on the information from the most meaningful parameters, which minimizes the original data with the least loss of information. Reference [10] classified the factor loading as ‘strong’, ‘moderate’, and ‘weak’ corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively.

Before performing CA and PCA, Kaiser–Meyer–Olkin (KMO) and Bartlett’s test were performed to examine the suitability of the data. High KMO value (close to 1) generally indicates that PCA may be useful.

III. RESULTS AND DISCUSSION

A. Cluster Analysis

In this study, the spatial variability of water quality in the Nhue River and its tributaries was determined separately from cluster analysis. Two mean log values for each variable

from the datasets were calculated for every monitoring site with the dimension of 10 (sites) × 8 (parameters), and 11 (sites) × 8 (parameters) for the main river and tributaries respectively.

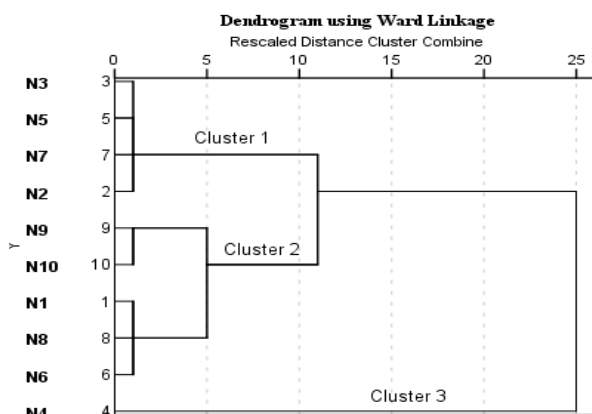


Fig. 2. Dendrogram showing clustering of sampling sites according to water quality characteristics of the Nhue River.

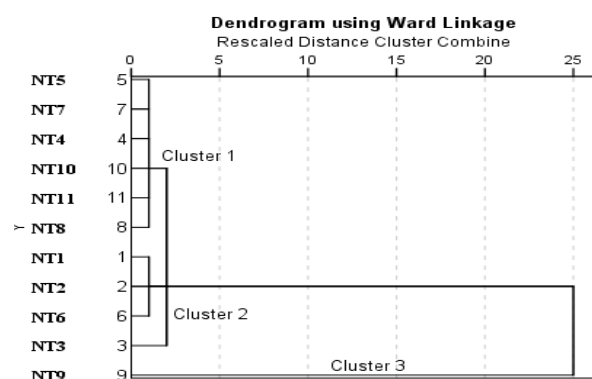


Fig. 3. Dendrogram showing clustering of sampling sites according to water quality characteristics of the tributaries.

For main river cluster analysis, the resulted dendrogram (Fig. 2) grouped all the ten sampling sites in the Nhue River into three statistically significant clusters. The Cluster 1 includes Sites N2, N3, N5 and N7 located in the middle region of the Nhue River, the Cluster 2 comprises Sites N1, N6, N8, N9 and N10 located in the upstream and downstream parts of the main river, and Cluster 3 consists of only Site N4. Site N4 at Cau To Bridge placed in Cluster 3 corresponds to relatively highly polluted (HP) region. At this section, the Nhue River receives wastewater from Yen Xa Canal and To Lich River through Thanh Liet dam. The Thanh Liet dam is located at the downstream of To Lich River where receives all domestic wastewater from Ha Noi city.

For tributaries, cluster analysis rendered a dendrogram (Fig. 3), where all the 11 sampling sites were grouped into three statistically significant clusters as cluster 1 (sites NT4, NT5, NT7, NT8, NT10, NT11), cluster 2a (sites NT1, NT2, NT6), cluster 2b (site NT3) and cluster 3 (site NT9). Sites NT3 (To Lich River) and NT9 (Yen Xa Canal) receive pollution mostly from domestic, hospital, craft village wastewater.

Cluster analysis applied to determine separately the spatial variability of water quality among monitoring sites in the main river, and among ones in tributaries brings an important result. Wastewater from the tributaries, including Yen Xa

Canal (NT9) and To Lich River (NT3) flowing into the Nhue River at To Bridge (N4) leads to severe pollution at this river region. Thus, Managers and Leaders should prioritize action plan controlling pollution in these tributaries to minimize pollution for the middle section of the Nhue River.

B. Principal Component Analysis

PCA based on factor analysis was used to explore the possible sources of pollution and to recognize considerable factors accounting for variation in water quality. Kaiser–Meyer–Olkin (KMO) and Bartlett’s tests were performed on the parameter correlation matrix to examine the validity of the PCA. The results of the KMO and Bartlett’s tests were 0.68 and 667.499, respectively indicating that PCA was useful for data reduction and that significant relationships were present among the variables. PCA was applied to a standardized dataset to identify the latent factors. The aim of this analysis was primarily to create an entirely new, smaller set of factors compared to the original dataset. The PCA revealed two PCs with eigenvalues >1 that explained about 45.75% of the total data variance in the water quality datasets where a correlation greater than 0.75 is considered “strong”; 0.75–0.50, “moderate”; and 0.50–0.30, as “weak” significant factor loading.

TABLE III: LOADING OF VARIABLES ON ROTATED PCs FOR COMPLETE DATA SET

Variables	PC1	PC2
WT	0.971	0.239
pH	0.935	0.269
Tur	-0.598	-0.086
DO	0.425	-0.147
TDS	0.234	0.166
NH ₄ ⁺	0.159	0.056
NO ₃ ⁻	0.248	0.968
EC	0.007	0.190
Eigenvalue	2.498	1.163
% Total variance	31.221	14.533
Cumulative % variance	31.221	45.754

Bold values indicate strong loadings (>0.70).

The first PC was responsible for the 31.22 % variance and was positively correlated (loading >0.70) with WT ($r=0.971$) and pH (0.935). The strong positive loading of WT observed was consistent with the fact that surface water in the Nhue River is influenced by anthropogenic activities such as domestic sewage and industrial wastewater from Ha Noi city. Moreover, the source of the Nhue River is regulated by Lien Mac Gate. Discharge of municipal and industrial wastewater adds heavy metals in the riverine water that also results in pH fluctuation [11]. The upstream and middle sections of the Nhue River comprised of craft villages, mechanical instruments, paint, chemical and electronic industries.

The PC2 in this study accounting for 14.53 % of the total variance was correlated with NO₃⁻ ($r=0.968$). The presence of NO₃⁻ originate from fertilizer application in the farms; agricultural land use strongly influences nitrate nitrogen. This is evident as farmers practice cereal plantation around the area.

Generally, it is evident from the analysis that surface water in the Nhue River is severely polluted by agricultural sources,

domestic sewage, craft villages, and industries.

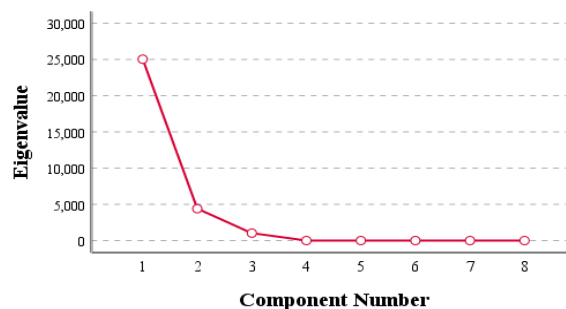


Fig. 4. Screen plot of the eigenvalues of PC

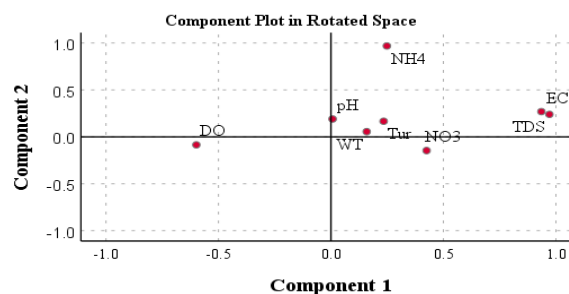


Fig. 5. Scores of the two principal components (PC1 and PC2) explained 45.75% of the total variance.

IV. CONCLUSION

In this case study, different multivariate statistical techniques were used to evaluate spatial variation and to distinguish sources of pollution in surface water quality of the Nhue River. Hierarchical cluster analysis applied to determine separately the spatial variability of water quality among 10 monitoring sites in the main river, and among 11 ones in tributaries brings useful information. The result indicated that two tributaries, including Yen Xa Canal (NT9) and To Lich River (NT3) lead to severe pollution at To Bridge (N4) in the Nhue River. PCA revealed that parameters related to craft village, industrial and agricultural activities, and domestic sewage discharge were the most important parameters contributing to water quality variation in the Nhue River. This study brings noticeable results which are useful to be applied for plan river management.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Mr. Nguyen Huu Hue is the corresponding author of this paper. He was mainly responsible for conducting the research as well as analyzing the data. Mr. Nguyen Huu Thanh was in charge of writing the paper. All authors had approved the final version.

REFERENCES

[1] B. Anna, J. Łukasz, and L. B. Anna, “The effect of anthropogenic and natural factors on the prevalence of physicochemical parameters of water and bacterial water quality indicators along the river Białka, southern Poland,” *Environmental Science and Pollution Research International*, vol. 25, no. 10, pp. 10102-10114, 2018.
 [2] Y. Wang, G. Zhu, and R. Yu, “Assessment of surface water quality using multivariate statistical techniques: A case study in China,”

- Irrigation & Drainage Systems Engineering*, vol. 7, no. 3, August 2018.
- [3] M. F. Chow *et al.*, "Evaluation of surface water quality using multivariate statistical techniques: A case study of Fei-Tsui Reservoir basin, Taiwan," *Environmental Earth Sciences*, vol. 75, no. 6, December 2015.
- [4] X. Wang, Q. Cai, L. Ye, and X. Qu, "Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: A case study of the Xiangxi River basin, China," *Quaternary International*, vol. 282, pp. 137-144, 2012.
- [5] L. T. T. Huyen, Z. Stephanie, L. Malte, and M. Günter, "Multivariate statistical assessment of a polluted river under nitrification inhibition in the tropics," *Environmental Science and Pollution Research International*, vol. 24, no. 15, pp. 13845-13862, 2017.
- [6] B. Ahmed, E. B. Mohamed, R. Jamila, A. Brahim, and S. Mohamed, "Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques," *International Soil and Water Conservation Research*, vol. 4, no. 4, pp. 284-292, 2016.
- [7] P. T. M. Hanh *et al.*, "Anthropogenic influence on surface water quality of the Nhue and Day sub-river systems in Vietnam," *Environmental Geochemistry and Health*, vol. 32, pp. 227-236, June 2010.
- [8] Y. Wang *et al.*, "Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China," *Journal of Hydro-Environment Research*, vol. 7, no. 1, pp. 30-40, 2013.
- [9] X. Zhang *et al.*, "Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong," *Environmental Monitoring and Assessment*, vol. 173, no. 1-4, pp. 17-27, 2011.
- [10] W. L. Chen, H. L. Kao, and M. K. Yi, "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan," *Science of The Total Environment*, vol. 313, no. 1-3, pp. 77-89, September 2003.
- [11] K. Firdaus *et al.*, "Evaluation of surface water quality on spatiotemporal gradient using multivariate statistical techniques: A case study of River Chenab, Pakistan," *Pol. J. Environ. Stud.*, vol. 28, no. 4, pp. 2645-2657, April 2019.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).



Nguyen Huu Hue is a professor and he received the Ph.D in hydraulic engineering from Moscow State University of Civil Engineering, Moscow, Russia in 2007. He has over 20 years of experience in doing research and teaching on hydraulic engineering and water resources in Vietnam. His current research involves hydraulic structure, soil, water self-purification, and climate change.



Nguyen Huu Thanh is a researcher at Institute of Civil Engineering (ICE), Thuyloi University (TLU) in Hanoi, Vietnam. He has experiences in environmental management system, natural resources management issues, and managing natural hazards. Currently, he is involving in some related projects on increasing water self-purification capacity and HEC-RAS, MIKE model.