Decision tree techniques to assess the role of daily DO variation in classifying shallow eutrophicated lakes in Hanoi, Vietnam

Thu-Huong T. Hoang, Viet D. Nguyen, Anh D. Van and Hien T. T. Nguyen

ABSTRACT

Eutrophication is a serious phenomenon in shallow lakes in Hanoi. The most important effect is the change of dissolved oxygen (DO) in the water column. The value of DO in water changed over time, and the variation of DO proved to be more important than the absolute value. The research aimed to assess the role of daily DO variation in classifying the eutrophication status of shallow lakes in Hanoi. Data were collected in eight shallow eutrophicated lakes in Hanoi during 2015–2017 in relation to water quality parameters. A large variation of DO was observed with the maximum value achieved at 15–17 h (180% saturation DO) and the minimum value at 4–6 h (20% saturation DO). The lowest DO value reached 1.5 mg/L, which was much lower than the threshold of survival of some aquatic animals. The daily DO fluctuations correlated with phytoplankton density, the Trophic State Index (TSI), chlorophyll-*a* concentration and density, and the proportion of DIN:PO_4, showing the close relationship between DO variation and eutrophication in the lakes. The decision trees were developed and also selected the ΔDO as the driving variable in classifying the eutrophication status in lakes. The daily DO variation could be an important indicator for eutrophication.

Key words | DO variation, eutrophication, oxygen depletion

Thu-Huong T. Hoang (corresponding author)
Viet D. Nguyen
Anh D. Van
Hien T. T. Nguyen
School for Environmental Science and Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
E-mail: *huong.hoangthithu@hust.edu.vn*

INTRODUCTION

Dissolved oxygen (DO) is the most important parameter used for assessing the condition of the freshwater environment. The sources of dissolved oxygen in lakes include: oxygen diffused through the water–air interface; photosynthesis of phytoplankton in the water; respiration of aquatic organisms; bacterial oxidation of organic matter, chemical oxidation and consumption of DO by other reduced inorganic substances; and oxygen uptake by benthic sediments (Zhang *et al.* 2015). Stratification of DO is highly dependent on the thermal stratification pattern in deep lakes (Rahman *et al.* 2005) and only slightly affected by phytoplankton photosynthesis (Zhang *et al.* 2015). In shallow lakes with a depth of less than 10 m, the thermal stratification pattern is unclear, stratification of DO by depth is unremarkable and other sources play a more important role.

When household and industrial wastes are discharged into water, organic matter and nutrients present in wastewater are taken up by bacteria and other biological species such as algae using DO in the interacting processes. As these processes are parts of a food chain in the waterbody, the level of DO decreases due to various interactive biochemical and biodegradation processes (Mason 2002). When a waterbody is eutrophicated, a much more complex interacting phenomenon takes place involving algae, macrophytes, detritus and especially dissolved oxygen.

Under normal conditions, an increase of DO in a waterbody is mainly due to atmospheric diffusion through

doi: 10.2166/wqrj.2019.105

the water surface and, to a lesser extent, to its production through photosynthesis by algae and emerging macrophytes. However, the oxygen from these processes does not result in an increase in the concentration of DO due to negative impacts of algae blooms (Dachs *et al.* 2000; Xu *et al.* 2001). After blooming, algae die and decomposition of detritus formed by dead algae consumes DO and thus reduces DO concentration in the water, especially at the bottom of lakes.

In practice, during the day, production of DO by photosynthesis is much more powerful than oxygen consumption by other processes. This results in a short-term increase of DO in the waterbody. The concentration of DO measured during this time can be higher than saturation. The importance of DO to good water quality and healthy ecosystems is widely acknowledged by lake managers and is reflected in the establishment of standards to maintain minimum DO concentrations (Foley *et al.* 2012; Müller *et al.* 2012). Thus, understanding temporal DO variations, the factors affecting them, and their response to eutrophication status is important for developing strategies to adaptively protect and manage water quality in reservoirs.

The eutrophication of lakes all over the world has increased rapidly during the last decade due to increased urbanization and, consequently, increased discharge of nutrients per capita. The production of fertilizers has grown drastically in this century and the concentration of phosphorus in many lakes reflects this. However, the most critical effect from an ecological point of view is the reduced oxygen content of the hypolimnion, caused by the decomposition of dead algae. During summer, eutrophic lakes sometimes show a high oxygen concentration at the surface, but a low oxygen concentration in the hypolimnion, which can lethally and sub-lethally affect aquatic biota (Foley *et al.* 2012). If DO concentrations are persistently below 3 mg/L, species richness of invertebrate and fish communities will remarkably reduce (Franklin 2014). Sub-lethal effects of hypoxia indicated by several metrics including refusal to eat, rapid gulping, and the immunosuppressive effects, seemed to begin at an oxygen concentration of 4.5 mg/L. Distributions for the invertebrate and fish taxa of varying tolerance indicated that a DO threshold could be near 2.5 mg/L (Justus *et al.* 2014). The DO concentration can be used as an 'early warning' of trophic state changes in a lake. This occurs much earlier than algal productivity increasing and other common trophic state indicators. (Carlson & Simpson 1996). Eutrophic shallow waterbodies often suffer from uncontrolled algal blooms and can lead to severe oxygen depletion killing fish and other sensitive organisms. The DO levels in these shallow eutrophic lakes in warm regions can fluctuate rapidly during the day (Rabalais *et al.* 2001; Yin *et al.* 2004).

In 2015, there were 119 lakes and ponds located in six main urban districts in Hanoi, which were mainly shallow lakes. The total surface area of these ponds and lakes was 6,959,305 m^2 (Hoang *et al.* 2017). This system of lakes has created a unique ecological landscape in Hanoi. However, in recent years, due to urban development, the lakes have received domestic wastewater beyond their self-cleaning ability. In addition, lake management and exploitation issues overlapped. Urban lakes were severely polluted and their area has gradually shrunk in size. The majority of the urban lakes do not ensure rainwater regulation function. This has resulted in higher biological productivity in the lakes causing 'blooming' of algae in the water. Algae and micro corpses of phytoplankton have settled in the sediment at the bottom of the lakes.

In the eutrophicated lakes, due to an increase in algae density, the process of strong photosynthesis and respiration causes a very large change in dissolved oxygen concentration during the day. This phenomenon has a great influence on the lives of aquatic plant and animal species and causes many problems and difficulties in the assessment of water quality.

DO is widely measured as a water quality parameter for aquatic bodies; however, most of the studies have only limited point-by-time measurements. Hypereutrophicated lakes often have DO saturation and even DO supersaturation during the daytime, resulting in an overall higher average DO when compared to healthier lakes (Matthews & Effler 2006; O'Boyle *et al.* 2013). This shows that a simple comparison in DO level between point-by-time measurements, which is commonly done in current water quality monitoring and evaluation practices, is not a good approach for determining water quality changes.

In this study, temporal changes in dissolved oxygen concentrations in the lakes of Hanoi were investigated. The decision trees were applied to test the DO variation as a driving variable for classifying eutrophication level of lakes. The following hypotheses were tested: (1) eutrophic lakes can have large daily DO variation, (2) daily DO variation is an important parameter for analyzing the eutrophication status of shallow lakes. The results can provide better insight into the effect of eutrophication of the urban shallow lakes of Hanoi and develop conservation and management programs to improve the ecological health of lakes in Hanoi city.

MATERIALS AND METHODS

Study areas and experiment monitoring

Six lakes (Hoan Kiem, Ba Mau, Bay Mau, Hai Ba Trung, Truc Bach, and Dam Tri) located in five different districts in urban Hanoi were selected for investigation. The site selection was based on variation in lake structure (with or without embankment), usage and pollution status (fish pond, recreation lakes, drainage lakes), and eutrophication conditions of the lakes (from mesotrophic to hypertrophic). All lakes in urban Hanoi are shallow, and the lakes selected have a depth less than 4 m.

Monitoring to assess the DO fluctuations and to measure in-field parameters over time and seasons took place during seven campaigns: October 2015, January 2016, March 2016, July 2016, September 2016, November 2016, and January 2017.

The climate in the region is humid-tropical, with long hot and rainy summers and short mild and dry winters. The long-term annual temperature in the area is approximately 20.8 °C, with monthly averages ranging from 17.2 °C in January to 29.8 °C in June. Long-term annual precipitation in the area is approximately 1,585 mm, ranging from 366 mm in July to 6 mm in January (Hanoi Statistical Yearbook 2015).

Water for nutrient and phytoplankton analyses was collected as surface samples from two or three mid-lake points depending on lake surface area during the days when DO measurement took place. The sampling procedure and analysis followed the ISO 17025 standard. The sampling procedure followed the TCVN 5994:1995 (ISO 5667-4:1987) – Water quality – Sampling: Guidance on sampling in natural and artificial lakes (ISO/IEC 17025 2017).

Field measurements included depth (m, ultrasonic device), water temperature (°C), pH (pH Meter DREL/2010), dissolved oxygen (mg/L, the percentage of saturated oxygen %) (WTW Oxy 330), conductivity (WTW 249 electrode) and Secchi depth (SD) (cm). Chemical analysis was performed for COD (SMWW 5220C/D, ISO 6060:1989), BOD_5 (SMWW 5210B, ISO 10707:1994), Kjeldahl nitrogen (4500-Norg C, ISO 5663:1984), total P (4500-P E), PO_4^{3-}-P, dissolved inorganic nitrogen ($DIN = NH_4^+$-N (SMWW 4500-NH_3 F) + NO_2^-N (SMWW 4500-NO_2^- B) + NO_3^-N (SMWW 4500-NO_3^- E)), Chl-*a* (SMWW 10200 H), and phytoplankton density (APHA/AWWA/WEF 1998).

Algae were collected according to the standard method. Samples were fixed in Lugol's iodine solution. Chlorophyll-*a* concentration was quantified by the acetone extraction and spectrophotometric method (SMWW 10200 H). To extract chlorophyll-*a*, the sample first went through sonication in an acetone solvent and was then centrifuged for 20 minutes. The resulting supernatant was used for measuring the absorbance at 750 and 665 mm in UV-Vis Lambda 35 – PerkinElmer.

The U-22XD set was used to measure multiple parameters along the lakes of the depth profile in the survey to select the best position for DO measurement in the lake.

The original 1-hour interval DO data were measured (WTW Oxy 330) within 24 hours at approximately 0.5 m below the water surface. DO was recorded by concentration (mg/L) and percentage of saturated oxygen (%) every 1 hour during the day to study daily variance. The measurement was performed on the day of water sampling to study the impact of water quality on diel DO (denoting or involving a period of 24 hours).

Data analysis

Complex trophic state index

A modified Carlson's Trophic Status Index (TSI) (Carlson & Simpson 1996) was used to evaluate the eutrophication level of the lakes. TSI is a frequently used biomass-related trophic index. It is a numeric index of lake trophic status ranging from 1 to 100. TSI is simple to use, requires a minimum of data to compute and is generally easy to understand.

Only dissolved forms of nitrogen and phosphorus are directly available for algal growth; therefore, in this study, TSI was calculated based on available nutritional factors (DIN and DIP) (Primpas & Karydis 2011). TSI is thus a calculation based on four parameters: PO_4^{3-}P, Chl-*a*, SD, and DIN ($NH_4^+ + NO_3^- + NO_2$):

$$TSI = \frac{TSI_{PO4P} + TSI_{Chl} + TSI_{SD} + TSI_{DIN}}{4}$$

For classification, TSI in the range of <30 represents an oligotrophic (O), 30–50 a mesotrophic (M), 50–70 a eutrophic (E), and >70 a hypertrophic (H) condition. Since the lakes of Hanoi are already eutrophic and hypertrophic, the classification is subdivided in the range of >70 and ≤80 for hypertrophic – a (Ha), >80 and ≤90 for hypertrophic – b (Hb) and >90 for hypertrophic – c (Hc).

Data driving methods

The applicability of classification trees (CTs) for assessing ecological quality was examined based on their performance in categorizing sampling points. CTs classify discrete dependent variables into a class from the values of a set of independent attributes (Quinlan 1986). These attributes can be continuous or discrete variables. The 'Top-down induction of decision trees' (TDIDT, also called the C4.5 algorithm) is one of the most widely used classification tree induction methods. The J48 algorithm is a Java re-implementation of C4.5 and is a part of the machine learning package WEKA (Witten & Frank 2005). In this study, the J48 algorithm with binary splits was applied to induce classification trees. Standard settings were used except for pruning confidence factors (PCF). Tree pruning was applied to optimize model performance by changing confidence factor values. The model training and validation were based on five-fold cross-validation.

The classification tree (CT) method was applied to discover the driving variables for eutrophication in the lakes of Hanoi and it focused on the evolution of the diel DO. The applicability of CTs for assessing ecological quality was based on their performance in categorizing sampling points. The CTs use several input variables to explain the variation of a single response variable. CTs have been widely used for supporting water management (Dakou et al. 2007; Hoang et al. 2013). The use of classification models in this study provided an insight into the effect of environmental variables in order to identify important parameters affecting the process of eutrophication. The model used eight independent variables to predict nutrient status based on TSI, which has proven to be an effective index for studying eutrophication in the lakes of Hanoi (Hoang et al. 2017). The daily variation of dissolved oxygen (ΔDO) was also included in input variables to study the impact of this on the classification.

RESULTS AND DISCUSSION

Water quality assessment

Water quality was assessed based on single parameters as well as complex water quality indices (Table 1). Nutrient concentrations mostly exceeded the Vietnamese National Technical Regulations on surface water quality

Table 1 | Water characteristics in the studied lakes during the sampling period (2015–2016)

No.	Parameter	Unit	Min	Max	Average	QCVN 08-MT:2015/ BTNMT (B1)
1	Depth	m	1.0	3.5	2.	–
2	Clarity (SD)	cm	30	70	48	–
3	Temperature	°C	16	33.6	26	–
4	pH	–	6.09	10.33	8.2	5.5–9
5	DO	mg/L	0.76	14.35	5.7	>4
6	TDS	mg/L	57	414	252	
7	Conductivity	μg/cm^2	93	646	391	
8	NH_4-N	mg/L	0.1	5.98	1.15	0.9
9	COD	mg/L	2.58	249.6	30.6	30
10	BOD_5	mg/L	0.95	47.1	6.3	15
11	DIN	mg/L	0.15	33.10	7.9	–
11	TN	mg/L	0.54	39.30	16.98	–
13	TP	mg/L	0.04	5.45	1.15	–
14	$PO_{4_}$P	mg/L	0.004	3.17	0.62	0.3
15	DIN:PO_4-P	–	0.3	124.2	17.2	–
16	Chlorophyll-*a*	μg/L	7.6	360.8	91.6	–

(QCVN 08-MT:2015/BTNMT). Most of the lakes were polluted by NH_4^+ and PO_4^{3-}, with maximum concentrations six and ten times higher than the allowable value, respectively. The highest concentration of PO_4^{3-} (in Ba Mau Lake) was ten times higher than the allowable value of 0.3 mg/L. The main cause may be the perennial bottom sediment and also the geological nature of the lakes, which store large amounts of phosphorus (Mason 2002).

Calculation of the ratio of DIN:PO_4-P for almost all the lakes was 17. The results could be explained by a natural source of phosphorus as well as an artificial load of nitrogen in the lakes. In fact, studies have shown that the status of DIN:PO_4-P >16 is conditionally conductive to the bloom of Cyanophyta and Chlorophyta (Ye & Cai 2010). The presence of these algae groups is an indicator of the level of eutrophication in lakes. These two main groups of algae appeared in very high density in the lakes studied. The Chlorophyta division was the most diverse with 17 genera, followed by Cyanophyta with 16 genera.

The results of eutrophicated lake classifications by TSI are presented in Figure 1. All of the lakes were hypereutrophic during some or all of the study period, and the complex index shows that eutrophication levels gradually increased. In 2016, September and November seemed to provide favorable conditions for eutrophication due to long sunny days with high solar intensity, and it was more serious at this time.

All lakes studied had full embankments to prevent encroachment of lakeshore and landscape areas. However, the embankments completely obstructed the flow of water and the nutrient exchange between waterbodies and the lakeshore corridors. Most of the narrow and shallow lakes, which include Hoan Kiem, Bay Mau, Ba Mau, and Truc Bach, had very high trophic values (TSI >70 at all times). Truc Bach receives output from a local domestic wastewater treatment plant and is the site of fish farming. The water quality changes regularly according to discharge regimes and the TSI was very high. Lotuses planted in Dam Tri Lake hinder the photosynthesis and growth of algae. The TSI was lowest in this lake in the summer (TSI < 70). However, in the winter, the lotuses wilt and water is pumped out of the lake around March each year. As a result, the eutrophication level of Dam Tri Lake increased appreciably (TSI > 80). Hai Ba Trung is the smallest of the studied lakes. It is located in an area of many large trees with a great many shadows and thus received relatively less solar intensity compared to other lakes.

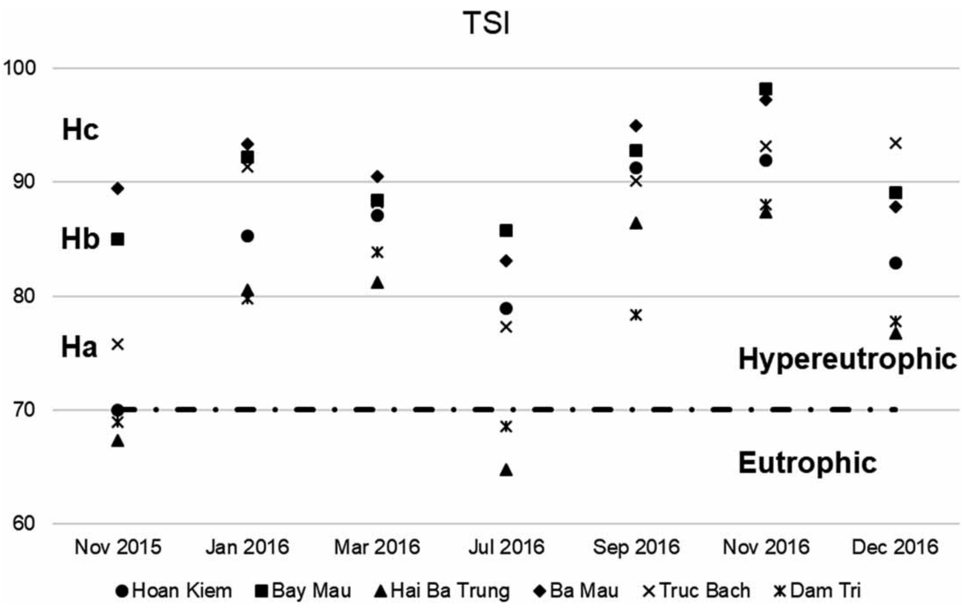

Figure 1 | Classification of study lakes based on TSI values.

Temporal variation of diel DO

The concentration of dissolved oxygen greatly depends on the gas exchange through the water–air interface and photosynthesis processes in the lakes. Thus, DO concentration varies according to depth due to changes in solar irradiation and thermal stratification.

Figure 2 shows DO concentrations over the depth of the lakes. Although thermal stratification is unclear, stratification of DO was observed in very shallow lakes. An increase in DO was observed in surface water and reached the maximum value at a depth of about 0.5 m. This aquifer also had the highest algae density in the lake. This effect can be explained by strong photosynthesis during the day, causing a significant increase in dissolved oxygen. The oxygen was retained, mostly in the form of tiny air bubbles in the algae. The oxygen was not prone to diffuse into the air as when it is in the surface water layer. Measurement of the DO was taken at the depth of around 0.5 m to get the highest value of the DO.

The DO concentration within a eutrophicated waterbody experienced large daily fluctuations. The minimum value of DO was often reached between 4 and 6 h (40% saturation DO) and DO peaks were reached around 15–17 h. The value reached 180% and even 200% DO saturation in some lakes. During the day, due to high radiation intensity, the process of photosynthesis of algae is strong and generates a high amount of oxygen as a by-product of photosynthesis. At night, the respiration of algae, because of their high density, consumes large amounts of dissolved oxygen, which makes the DO plummet (Figure 3).

DO variability differs by the hour of the day and by the time of year. In autumn, the DO variation in lakes was the highest. In autumn, Truc Bach, Bay Mau, and Ba Mau lakes had great variability with maximum DO saturation values reaching 180% and 200%, respectively. The higher variation in the cases of Truc Bach, Bay Mau, and Ba Mau lakes corresponded to the higher eutrophic level of these lakes.

Diel DO concentration and saturation began rising in the early morning (5:00 a.m. to 6:00 a.m.), reaching their peaks in the late afternoon (3:00 p.m. to 4:00 p.m.), and declining in the evening until reaching their troughs shortly before sunrise (Figure 4). The difference between the daily DO maximum and minimum during summer and autumn (10 mg/L) was more than three times as large as the difference during winter (2.8 mg/L). Also, no supersaturation was found in the diel DO trends during January and March 2016, while even 14 hours of DO supersaturation within a day were observed in September 2016 (from 9:00 a.m. to 22:00 p.m.) in some hypertrophic lakes (Figure 4).

The DO fluctuations were observed the highest in productive lakes. The results confirmed a hypothesis that

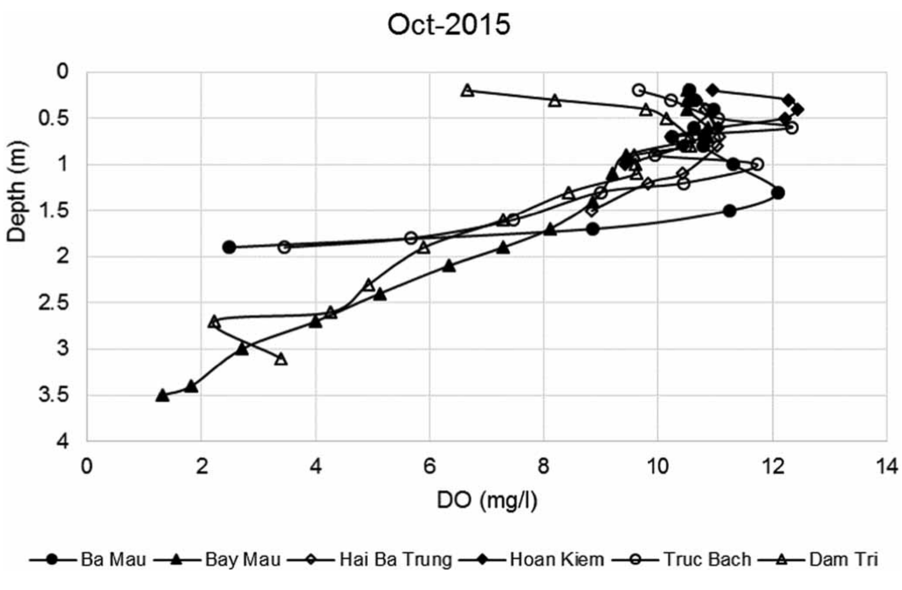

Figure 2 | Variation of DO across stratification of shallow lakes.

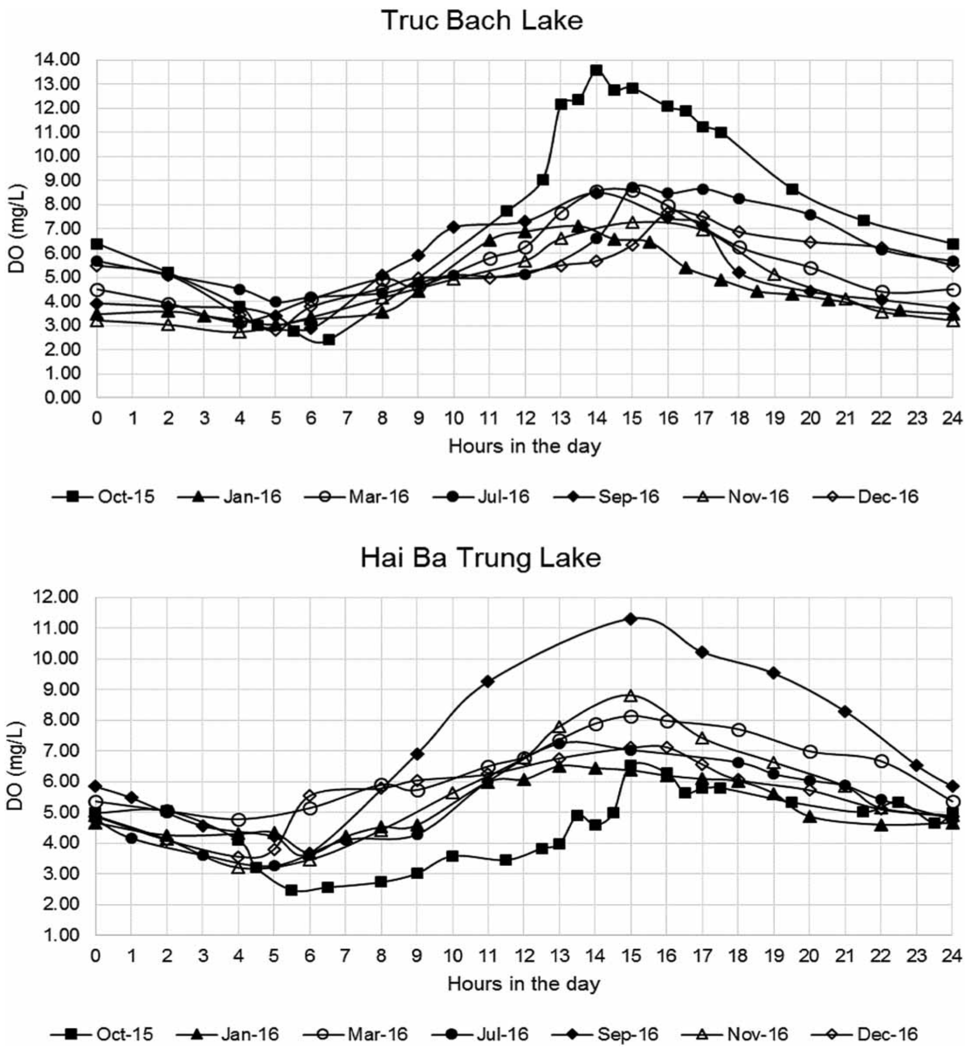

Figure 3 | DO fluctuation in Hai Ba Trung and Truc Bach lakes.

in these productive lakes, the DO concentration is usually lowest just before sunrise and highest just before sunset. In these lakes, oxygen dissolved in water is consumed by respiration at night much faster than it is received from oxygen diffusing from the atmosphere (Müller *et al.* 2012). When the DO value reached critical, lake organisms including fish, macroinvertebrate larvae may die from lack of oxygen or suffering oxygen shocks due to huge diel DO variation (Chapman 1986; Franklin 2014; Justus *et al.* 2014).

In addition, the minimum value of DO at night in some lakes went below the limit for avoidance (4 mg/L). This minimum level was deemed critical to the survival of the benthos that supplies both fish and upper trophic level organisms (Barton & Taylor 1996). A massive fish die-off is observed in several hypertrophic lakes in Hanoi every year, usually in summer and autumn, which have long sunny days with the highest solar intensity during the year. For example, on 9th June 2016, about 6 tonnes of spoilt fish were taken out of the Hoang Cau Lake inside Hanoi city. In October 2016, about 200 tonnes of spoilt fish were collected from the West Lake, the biggest lake in Hanoi (Vietnam Net 2016).

The level of DO variation also changed over time of year in Hanoi. In October 2015 (autumn) when the air temperature was recorded at about 35–37 °C, the DO variability in Truc Bach Lake was the largest at 12 mg/L, with a

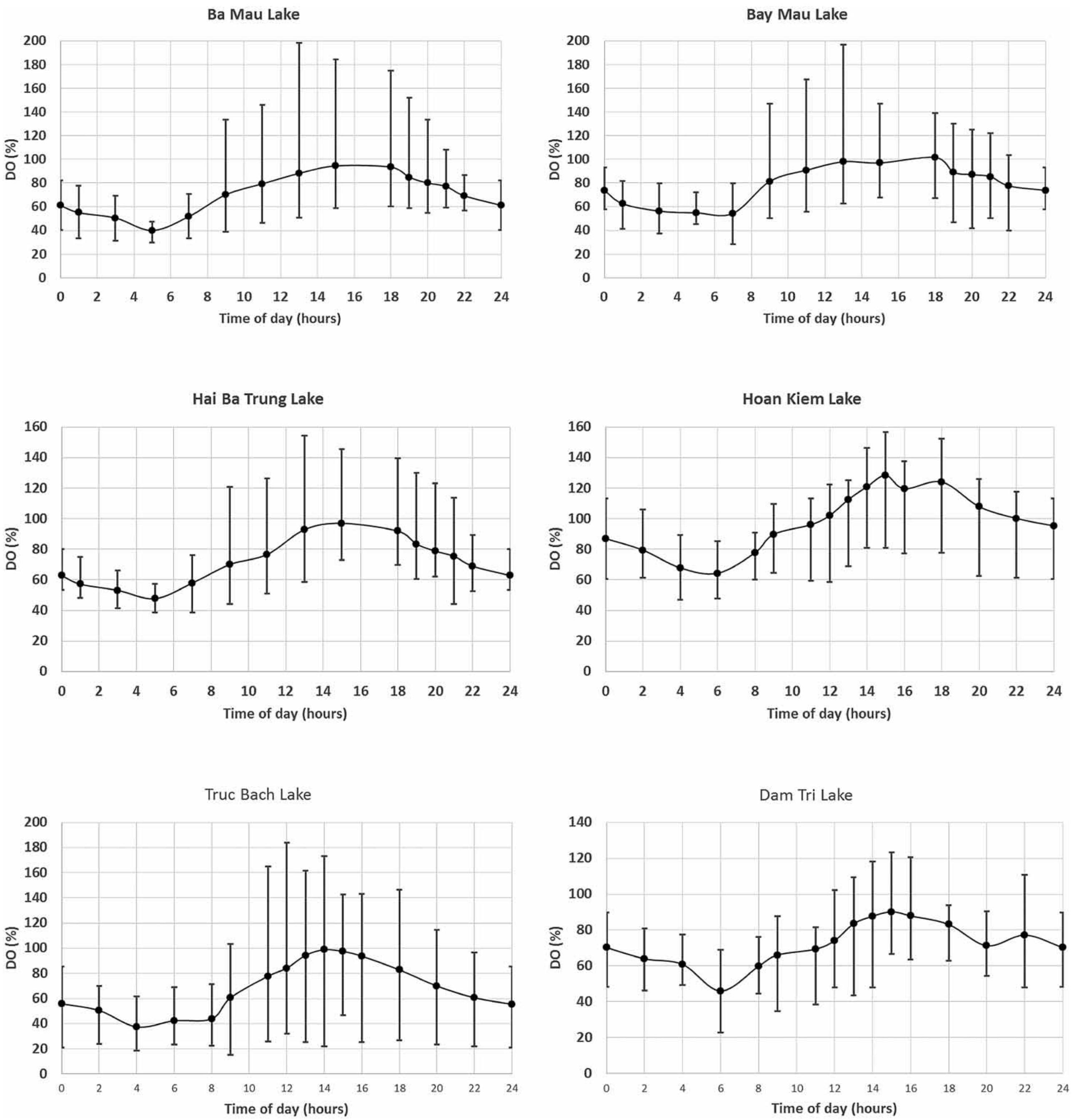

Figure 4 | Diel DO concentration in the lakes studied (n = 7).

maximum value of 14 mg/L and a minimum of 2 mg/L. In January 2016 (the coldest time of year), when the water temperature was about 15 °C, the weather was not suitable for growth of algae and the maximum and minimum levels of DO were only 6 mg/L and 3 mg/L, respectively. Similar seasonal variation in the DO was observed in the remaining lakes (Figure 4). The higher the daily temperature, the higher the variation of DO that was observed in the lakes.

After the incidence of massive fish dies in the West Lake in 2016, an automatic monitoring station was installed in summer 2017 to monitor several parameters of water quality in the West Lake including dissolved oxygen over time. The

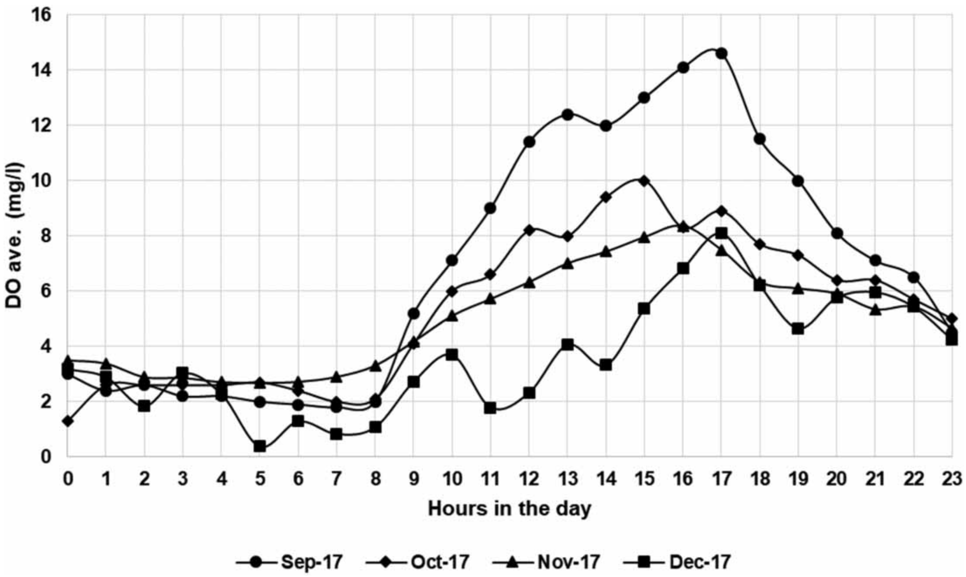

Figure 5 | The daily DO variation collected by automatic monitoring station in West Lake (Hanoi Environment & Natural Resources Department 2017).

average daily fluctuation (Figure 5) showed the DO during the night reached an extremely low value, which was considered dangerous for aquatic animals. In September, with the highest solar radiation during the year, the DO value reached the highest value reflecting intensive algae bloom in the lake. Both DO value and daily variation of DO in the winter time in December were much lower.

The role of DO variation in classification of eutrophication level

Daily variation of DO was large in the lakes studied. The correlation between the variation of DO and the parameters related to eutrophication is shown in Table 2.

The results showed that the level of DO variability was positively correlated with algae density, TSI, and Chl-*a*. This indicates that eutrophication is an important factor causing the large variation of DO concentration in the lakes. The variation of dissolved oxygen is mainly owing to algae photosynthesis, respiration, and death decomposition processes. This correlation confirmed results obtained in other research, which showed that the variance of DO in urban lakes was the result of eutrophication and algal bloom, rather than other reasons (Ma & Zaho 2011; Huang & Chen 2013).

In addition, the proportion of DIN:PO_4 and BOD_5 levels also correlated with the variation of diel DO ($r = 0.58$ and $r = 0.66$, respectively, $p < 0.05$). This also confirmed that the pollution source causing eutrophication in the lakes of Hanoi was nutrient and organic enrichment from domestic wastewater (Hoang *et al.* 2017).

The classification tree developed for TSI using six inputs with 120 instances showed the best classification results with CCI = 65% and Kappa = 0.32. Besides BOD_5 and the nutrient-related parameters (NH_4-N and TP), the tree selected ΔDO% as a driving variable for the tree structure (Figure 6). The results showed that dissolved oxygen concentration in the lakes can be an important indicator for eutrophication conditions.

Table 2 | Correlation between variations of % DO saturation and eutrophication parameters

	Log (algae density)	TSI	Chl-*a*	DIN:PO_4	BOD_5
ΔDO%	0.76	0.67	0.64	0.58	0.66
P	<0.05	<0.05	<0.05	<0.05	<0.05

DISCUSSION

Although a large diel fluctuation in DO concentration is known as one of the symptoms of eutrophication, DO is

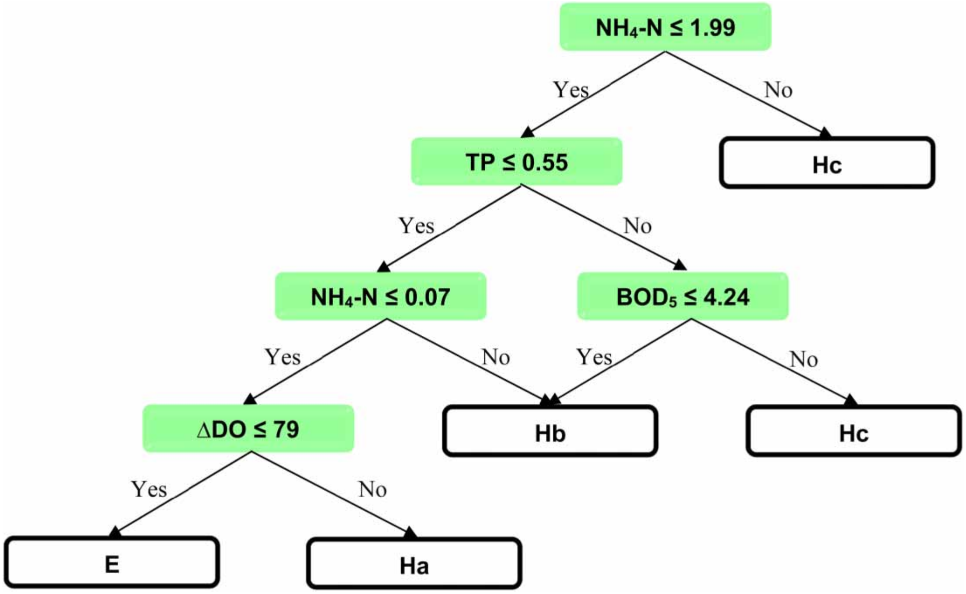

Figure 6 | Classification tree (standard setting and pruning confidence factor (PCF) = 0.05) predicting eutrophication classes based on Trophic Status Index (TSI), developed with 120 instances. Correctly classified instances (CCI) = 65%; Kappa = 0.32.

typically not used in trophic states' determination (Foley *et al.* 2012). One of the reasons for this may be due to the large daily and seasonal variation of DO.

The use of DO value in water quality assessment, as well as the calculation of the eutrophic index and WQI, should be carefully considered. In the calculation of the Vollenweider trophic index (TRIX), for example, oxygen as absolute [%] deviation from saturation was used (Vollenweider *et al.* 1998) and thus supersaturation or depletion of DO can be claimed as an indicator for poor water quality. However, the DO measure is strongly time-dependent.

Hypoxic conditions at the sediment–water interface may additionally promote phosphorus release from the sediments, particularly in shallow lake waters. Thus, oxygen depletion could positively feedback to eutrophication via the redistribution of phosphorus into more bioavailable forms (Andersen & Ring 1999).

The change of DO, and particularly the depletion of DO to the critical value, is the main reason for fish die-off and biodiversity declination in lake ecosystems. West Lake is the biggest and most biologically diverse lake in Hanoi. In the year 2000, the water quality was classified as good and in mesotrophic condition. The ecosystem of West Lake was rather diverse with 18 species of macrophyte, 141 species of phytoplankton, 43 species of zooplankton, 16 species of benthos, and about 39 species of fish. In addition, there were 214 species of plants and 58 species of birds around the lake (IEBR 2012).

Research in West Lake showed the relatively high density of phytoplankton at 65×10^6 cells/L. Cyanophyta presented at the highest density, accounting for 60% of total algae density in the lake, followed by Chlorophyta and Euglenophyta at the lowest density. The results showed signs of organic pollution in the lake (IEBR 2012). Cyanophyta is known as an indicator for eutrophic bodies. Studies showed an increased number of Cyanophyta species, proving that the water quality changed to become more eutrophic. From 2007 to the present, the number of Chlorophyta species in West Lake has drastically reduced from 70 to 23 species. While the number of species of Cyanophyta, Euglenophyta, and Bacillariophyta have drastically increased, diatoms seem to be reduced. This situation proved that water quality in lakes can change due to nutrient enrichment.

Wild fish species are decreasing in number and have been overwhelmed by cultured fish species. Some species, previously very popular and playing an important role in the yield, have now become rare and uncommon. Some species that were recorded previously may now have disappeared. The economic role of wild fish at West Lake decreased, together with the reduction of wild fish species in West Lake (IEBR 2012).

The decline in DO causing the death of aquatic animals led to a reduction in biodiversity in the eutrophicated lakes in Hanoi. The hypothesis was that serious eutrophication, together with weather conditions during the long sunny days, caused oxygen depletion in the lake during the night. The lack of oxygen was the main reason for the mass fish die-off. Eutrophication in shallow urban lakes in warm regions can often be accelerated by climatic and anthropogenic environments. DO levels in these lakes can drop from oversaturation to hypoxia within a few hours leading to possible oxygen shock for aquatic biota (Yin *et al.* 2004). The oversaturation of the DO in the daylight measured by the automatic monitoring station can be an early warning for the dangerous condition for fish and other aquatic biotas in the lakes during the night.

The decision trees use several input variables to explain the variation of a single response variable. Presently, the decision trees have been applied widely for practical applications in water management (Dakou *et al.* 2007; Hoang *et al.* 2013). The use of classification models allows us to gain insight into the effect of environmental variables to identify important parameters affecting the process of eutrophication. In this study, the diel DO variation was selected by the model as the classifying factor between eutrophicated and hypereutrophicated lakes. The results of developed models support measures selected for improving the ecological quality of lakes. Driving variables revealed by optimized models thus provide meaningful information for lake management. They enable management decisions to be made, such as the establishment of standards for organic and nutrient pollutants, and also actions to be taken within the catchment to achieve desired goals.

CONCLUSIONS

The diel dissolved oxygen variation is a very important parameter characterizing the eutrophication status, as well as a reason for biodiversity degradation, in the lakes of Hanoi.

Dissolved oxygen in eutrophicated lakes had large fluctuations within the same day, with the maximum value achieved at 15–16 h and the minimum value reached at 5–6 h. The DO variation significantly depended on the weather. In hot weather, the minimum value of DO concentrations was 1.5 mg/L, much lower than the limit of avoidance of aquatic animals (4 mg/L). The results of this study suggest that decision trees reveal the DO variation is an important water quality regulation key variable that affects algae bloom occurrence and is a promising tool for lake management.

The daily DO fluctuations correlated with trophic indices TSI, Chl-*a*, algae density, BOD_5, and the proportion of DIN:PO_4. These showed the close relationship between DO variation during the day and eutrophication and algae density in the lakes. The classification trees technique selected DO as an important variable to assess eutrophication levels of the lakes. Therefore, DO should be seriously reconsidered as the main parameter to assess the water quality of eutrophicated waterbodies. A practical methodology can be developed to use DO variation as an indicator of eutrophication.

ACKNOWLEDGEMENTS

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.99-2012.17.

REFERENCES

Andersen, F. & Ring, P. 1999 Comparison of phosphorus release from littoral and profundal sediments in a shallow eutrophic lake. *Hydrobiologia* **408/409**, 175–183.

APHA/AWWA/WEF 1998 *Standard Methods for the Examination of Water and Wastewater*, 20th edn. American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC, USA.

Barton, B. A. & Taylor, B. R. 1996 Oxygen requirements of fishes in northern Alberta Rivers with a general review of the adverse effects of low dissolved oxygen. *Water Quality Research Journal of Canada* **31** (2), 361–409.

Carlson, R. & Simpson, J. 1996 *A Coordinator's Guide to Volunteer Lake Monitoring Methods*. North American Lake Management Society, Madison, WI, USA.

Chapman, G. 1986 *Ambient Water Quality Criteria for Dissolved Oxygen*. United States Environmental Protection Agency, Washington, DC, USA.

Dachs, J., Eisenreich, S. J. & Hoff, R. M. 2000 Influence of eutrophication on air–water exchange, vertical fluxes and phytoplankton concentration of persistent organic pollutants. *Environmental Science and Technology* **34**, 1095–1102.

Dakou, E., D'heygere, T., Dedecker, A., Goethals, P., Lazaridou-Dimitriadou, M. & De Pauw, N. 2007 Decision tree models for prediction of Macroinvertebrate taxa in the river Axios (Northern Greece). *Aquatic Ecology* **41**, 399–411.

Foley, B., Jones, I. D., Maberly, S. C. & Rippey, B. 2012 Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication. *Freshwater Biology* **57** (2), 278–289.

Franklin, P. A. 2014 Dissolved oxygen criteria for freshwater fish in New Zealand: a revised approach. *New Zealand Journal of Marine and Freshwater Research* **48** (1), 112–126.

Hanoi Environment and Natural Resources Department 2017 *Water Monitoring Index*. https://chisoquantracnuoc.vn/ (accessed 22 December 2017).

Hanoi Statistical Office. 2015 *Hanoi Statistical Yearbook 2015*. Statistical Publishing House, Hanoi, Vietnam.

Hoang, T. H., Mouton, A., Lock, K. & Goethals, P. 2013 Integrating data-driven ecological models in an expert-based decision support system for water management in the Du River basin (Vietnam). *Environmental Monitoring and Assessment* **185**, 631–642.

Hoang, T. T. H., Van, D. A. & Nguyen, T. T. H. 2017 Driving variables for eutrophication in lakes of Hanoi by data-driven technique. *Water and Environment Journal* **31**, 176–183.

Huang, Y. & Chen, M. 2013 Variation of dissolved oxygen in the experiments of occurrence and disappearance for *Microcystis* bloom. *Procedia Environmental Sciences* **18**, 559–566.

Institute for Ecology and Biological Resources (IEBR) 2012 *Investigation and Assessment of Current Status of the Water Quality and Ecological System of the West Lake, Hanoi*. Project report, Hanoi People's Committee.

ISO/IEC 17025 2017 *2017: General Requirement for the Competence of Testing and Calibration Laboratories*, 3rd edn. International Organization for Standardization, Geneva, Switzerland.

Justus, B. G., Mize, S. V., Wallace, J. & Kroes, D. 2014 Invertebrate and fish assemblage relations to dissolved oxygen minima in lowland streams of southwestern Louisiana. *River Research and Application* **30** (1), 11–28.

Ma, C. & Zaho, D. 2011 The model of dissolved oxygen in ponds based on genetic algorithm and RBF networks. *China Rural Water and Hydropower* **2**, 14–17.

Mason, C. F. 2002 *Biology of Freshwater Pollution*, 4th edn. Prentice Hall, Inc., England.

Matthews, D. A. & Effler, S. W. 2006 Assessment of long-term trends in the oxygen resources of a recovering urban lake, Onondaga Lake, New York. *Lake and Reservoir Management* **22** (1), 19–32.

Müller, B., Bryant, L., Matzinger, A. & Wüest, A. 2012 Hypolimnetic oxygen depletion in eutrophic lakes. *Environmental Science and Technology* **46** (18), 9964–9971.

O'Boyle, S., McDermott, G., Noklegaard, T. & Wilkes, R. 2013 A simple index of trophic status in estuaries and coastal bays based on measurements of pH and dissolved oxygen. *Estuaries Coasts* **36** (1), 158–173.

Primpas, I. & Karydis, M. 2011 Scaling the Trophic Index (TRIX) in oligotrophic marine environments. *Environmental Monitoring and Assessment* **178**, 257–269.

Quinlan, J. R. 1986 Induction of decision trees. *Machine Learning* **1**, 81–106.

Rabalais, N. N., Turner, R. E. & Wiseman, W. J. 2001 Hypoxia in the Gulf of Mexico. *Journal of Environmental Quality* **30** (2), 320–329.

Rahman, A., Al Bakri, D., Ford, P. & Church, T. 2005 Limnological characteristics, eutrophication and cyanobacterial blooms in an inland reservoir, Australia. *Lakes & Reservoirs: Research and Management* **10** (4), 211–220.

Vietnamese National Standard TCVN 5994:1995: Water Quality – Sampling – Guidance on Sampling From Natural Lakes and man-Made Lakes 1995, Directorate for Standards, Metrology and Quality, Ministry of Science and Technology, Hanoi, Vietnam.

Vietnamese National Technical Regulation QCVN 08-MT:2015/BTNMT: Surface Water Quality 2015, Ministry of Natural Resources and Environment, Hanoi, Vietnam.

VietnamNet 2016 *Over 200 Tons of Fish die in West Lake: VIETNAMNET Bridge*. http://english.vietnamnet.vn/fms/environment/164790/over-200-tons-of-fish-die-in-west-lake.html (last updated 5th October 2016). (accessed 22 April 2017).

Vollenweider, R. A., Giovanardi, F., Montanari, G. & Rinaldi, A. 1998 Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: proposal of a trophic scale, turbidity and generalized water quality index. *Environmetrics* **9**, 329–357.

Witten, J. H. & Frank, E. 2005 *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. Elsevier, San Francisco, CA, USA.

Xu, F. L., Tao, S., Dawson, R. W., Li, P. G. & Cao, J. 2001 Lake ecosystem health assessment indicators and methods. *Water Research* **35** (13), 3157–3167.

Ye, L. & Cai, Q. 2010 Spring phytoplankton blooms in Xiangxi Bay of Three-Gorges Reservoir: spatiotemporal dynamics across sharp nutrient gradients. *Journal of Freshwater Ecology* **26** (1), 11–18.

Yin, K., Zhifeng, L. & Zhiyuan, K. 2004 Temporal and spatial distribution of dissolved oxygen in the Pearl River Estuary and adjacent coastal water. *Science Direct* **24**, 1935–1948.

Zhang, Y., Wu, Z., Liu, M., He, J., Shi, K., Zhou, Y., Wang, M. & Liu, X. 2015 Dissolved oxygen stratification and response to thermal structure and long-term climate change in a large and deep subtropical reservoir (Lake Qiandaohu, China). *Water Research* **75**, 249–258.

First received 27 January 2018; accepted in revised form 21 June 2019. Available online 18 July 2019