



Research article

Long-term prediction of algal chlorophyll based on empirical models and the machine learning approach in relation to trophic variation in Juam Reservoir, Korea

Sang-Hyeon Jin^a, Namsrai Jargal^a, Thet Thet Khaing^a, Min Jae Cho^a, Hyeji Choi^a, Bilguun Ariunbold^a, Mnyagatwa Geoffrey Donat^a, Haechan Yoo^a, Md Mamun^{a,b}, Kwang-Guk An^{a,*}

^a Department of Bioscience and Biotechnology, Chungnam National University, Daejeon, 34134, Republic of Korea

^b Department of Earth Sciences, Southern Methodist University, Dallas, TX, 75205, USA

ARTICLE INFO

Keywords:

Algal chlorophyll
Empirical analysis
Nutrient regimes
Machine learning
Summer monsoon
Temperate reservoir

ABSTRACT

This study analyzed spatiotemporal variation and long-term trends in water quality indicators and trophic state conditions in an Asian temperate reservoir, Juam Reservoir (JR), and developed models that forecast algal chlorophyll (CHL-a) over a period of 30 years, 1993–2022. The analysis revealed that there were longitudinal gradients in water quality indicators along the reservoir, with notable influences from tributaries and seasonal variations in nutrient regimes and suspended solids. The empirical model showed phosphorus was found to be the key determinant of algal biomass, while suspended solids played a significant role in regulating water transparency. The trophic state indices indicated varying levels of trophic status, ranging from mesotrophic to eutrophic. Eutrophic states were particularly observed in zones after the summer monsoons, indicating a heightened risk of algal blooms, which were more prevalent in flood years. The analysis of trophic state index deviation suggested that phosphorus availability strongly influences the reservoir trophic status, with several episodes of non-algal turbidity at each site during Mon. Increases in non-algal turbidity were more prevalent during the monsoon in flood years. This study also highlighted overall long-term trends in certain water quality parameters, albeit with indications of shifting pollution sources towards non-biodegradable organic matter. According to the machine learning tests, a random forest (RF) model strongly predicted CHL-a ($R^2 = 0.72$, $p < 0.01$), except for algal biomass peaks ($>60 \mu\text{g/L}$), compared to all other models. Overall, our research suggests that CHL-a and trophic variation are primarily regulated by the monsoon intensity and predicted well by the machine learning RF model.

1. Introduction

Reservoirs are crucial freshwater sources because they serve multiple functions, including drinking water supply, hydropower generation, flood control, and recreational activities [1]. However, human activities such as industrialization, urbanization, and intensive agriculture in catchment areas have led to disturbances impacting the water quality of lentic ecosystems like lakes and

* Corresponding author.

E-mail address: kgan@cnu.ac.kr (K.-G. An).

<https://doi.org/10.1016/j.heliyon.2024.e31643>

Received 13 August 2023; Received in revised form 20 May 2024; Accepted 20 May 2024

Available online 28 May 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

reservoirs [2]. Among the challenges these systems face, nutrient enrichment stands out, resulting in eutrophication and diminished water quality [3]. Both anthropogenic discharges and internal processes contribute to this issue, fostering excessive algal growth [4,5]. Hence, monitoring water quality in artificial reservoirs becomes imperative to comprehend nutrient dynamics and accurately predict algal biomass.

Eutrophication management primarily relies on empirical models that establish the relationship between nutrients (nitrogen and phosphorus (P)) and algal chlorophyll (CHL-a) levels [6,7]. While some models focus on individual lentic systems [8–10], others encompass broader datasets covering numerous lakes at national and regional levels [7,11,12]. The positive relation between nutrients and CHL-a underscores the significance of P-loading, mainly associated with anthropogenic pollution sources, to mitigate excessive algal biomass and manage eutrophication [4,6,13]. However, empirical relationships between P and CHL-a can vary due to geographical location, land use, reservoir morphology, and climatic region [14–18], as well as local moderating factors such as light availability, grazing pressure, water temperature (WT), and water transparency [19–23].

Trophic state classification is fundamental for assessing reservoir health and eutrophication impact on water quality. Various concentration criteria and trophic state indices are employed for this purpose at regional and global scales [24,25]. The Trophic State Index (TSI) by Carlson [24] stands out as one of the widely used methods, utilizing parameters like total P (TP), CHL-a, and Secchi depth (SD) to categorize reservoirs into oligotrophic, mesotrophic, eutrophic, and hypereutrophic states. These classifications are pivotal for developing predictive eutrophication models to curb algal blooms and identify causative factors. Also, a two-dimensional graphical approach based on the trophic state index deviation (TSID) can provide important insight into the spatial and seasonal changes in nutrient levels, light availability, and zooplankton grazing that potentially influence algal growth seasonally and spatially [19].

In regions experiencing seasonal monsoons, rainfall intensity significantly influences water quality and algal growth, primarily through variations in inflow rates and hydraulic retention time (HRT) [8,26,27]. South Korea experiences summer monsoons, mainly from July to August [14]. It encounters substantial changes in nutrient levels, suspended solids (SS), light availability, and algal CHL-a in reservoirs due to pollutant-rich runoff and sediment resuspension [11,28,29]. These seasonal fluctuations, including rapid flushing and dilution, can alter algal growth, challenging the predictive accuracy of statistical models [23,26,30].

The complexity of environmental factors can diminish the predictive accuracy of algal biomass models for individual systems based on a single factor, such as key nutrients [30,31]. Machine-learning approaches are increasingly used to predict algal growth in lentic systems because they can provide insights into the importance of input variables and their interactions for predicting algal CHL-a [30,32,33]. Random forest (RF), support vector machine (SVM), and artificial neural network (ANN) models have been applied to predict CHL-a levels, blue-green algal blooms, and water transparency in lacustrine systems globally [32–36]. However, each machine learning algorithm has its own advantages and disadvantages. Therefore, selecting the most effective algorithm or combination of

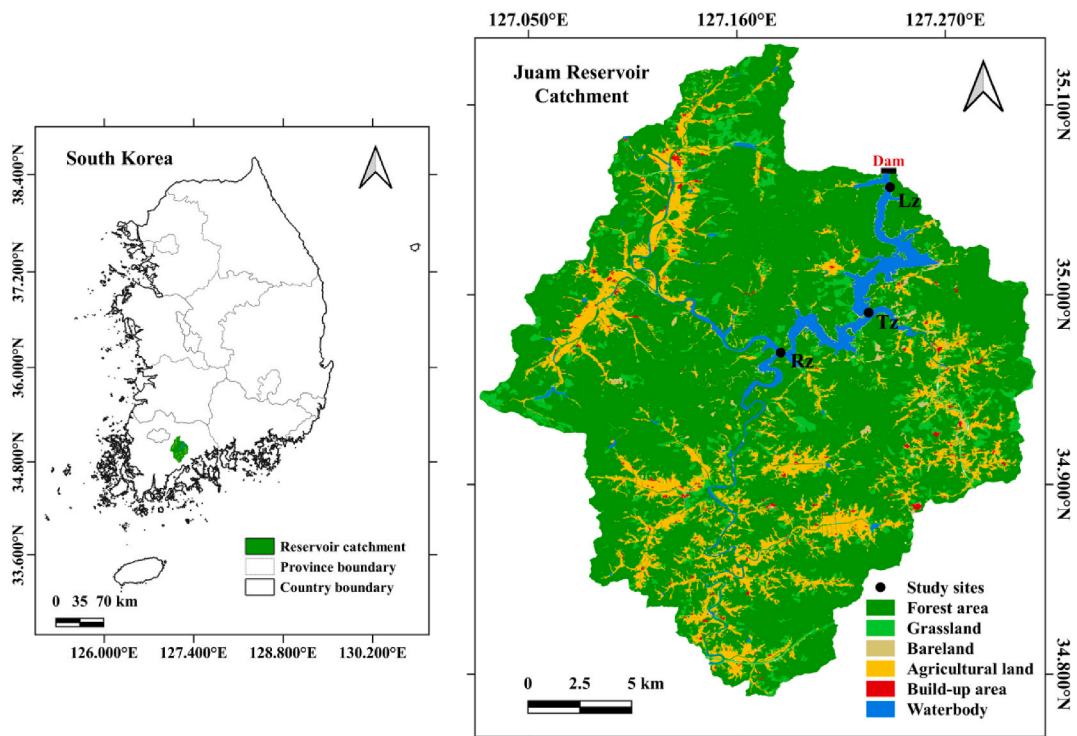


Fig. 1. Location of Juam Reservoir and the study sites, including lacustrine (Lz), transition (Tz), and riverine (Rz) zones, with land cover and land use patterns.

algorithms for predicting algal biomass in lentic systems may be challenging. Huang et al. [37] predicted algal CHL-a from environmental factors in Chinese lentic systems using RF and a generalized linear model. They found TP was the main factor influencing CHL-a. Meanwhile, Mamun et al. [30] utilized multiple linear regression (MLR), SVM, and ANN to predict algal CHL-a and SD in a monsoon-region reservoir and found that SVM performed better than MLR and ANN.

The Juam Reservoir (JR), located within a monsoonal climate region, is a pivotal water resource catering to various societal needs, including water supply, irrigation, and recreational activities. Despite the absence of substantial water quality problems, previous research endeavors have shed light on the hydrological conditions, water quality dynamics, and nutrient loading patterns within the reservoir's watershed. Han et al. [38] identified phosphorus as a critical factor driving eutrophication in JR, while Yi et al. [39] assessed hazardous sub-watersheds contributing to non-point source pollution, providing valuable insights for targeted management strategies. The study of [40] comprehensively monitored water quantity and quality in an upstream tributary of JR, highlighting the significant influence of land use practices on nutrient loading. Furthermore, Chung et al. [41] elucidated the relationship between harmful cyanobacteria and environmental factors in JR, underscoring the impact of runoff-induced phosphorus inputs on algal biomass dynamics during rainfall events.

This study, based on 30 years of monthly data for JR, comprehensively analyzed the longitudinal-seasonal dynamics of physicochemical indicators of water quality, CHL-a, and trophic state conditions to identify critical factors determining algal biomass and water transparency with the help of multivariate analysis and empirical linear models. We also compared the differences in water quality indicators, TSI, and TSID between drought and flood year regimes, depending on the intensity of the summer monsoon. Furthermore, we assessed overall long-term trends of water physicochemical indicators and CHL-a using the Mann-Kendall test and innovative trend analysis and tested the machine learning technique for long-term CHL-a and SD prediction models. Overall, this study would provide important insights into the factors influencing water quality and CHL-a dynamics in the JR and help future decision-making for sustainable water supplies and ecological health.

2. Materials and methods

2.1. Study system

The JR is an artificial lake that was initiated in 1992 through the construction of a multi-purpose dam on the Seomjin River in South Korea (Fig. 1). It is a warm monomictic reservoir that is about 40 km long and has a surface area of 33 square kilometers, with an average depth of 14 m, maximum depth of 47 m, and basin area of 1010 square kilometers. It is a critical freshwater source for the region, serving the water needs of Gwangju, Naju, and Mokpo cities by supplying 25 million cubic meters of drinking water daily [32]. The reservoir catchment area mainly consists of forested mountains (70.5 %), with agriculture (13.2 %) and built-up areas (1.5 %) as the primary land uses (Fig. 1). Non-point pollution sources are key managing issues for JR due to high nutrient loadings released from agricultural lands during summer monsoons [38,39]. In this study, we used long-term data from three study sites representing the longitudinal zones of the reservoir: Rz (riverine zone), Tz (transition zone), and Lz (lacustrine zone) (Fig. 1).

2.2. Data sources and water quality analysis

The physicochemical data used in this study were obtained from the Korean Water Environment Information System (<https://water.nier.go.kr/>), managed by the Korea Ministry of Environment (MOE). The analysis of the monsoon season was conducted by taking samples during three distinct periods - the premonsoon period (Pre), which spans from May to June; the monsoon period (Mon), which occurs during July and August; and the postmonsoon period (Post), which lasts from September to October. Thirteen physicochemical variables were studied: WT, electrical conductivity (EC), SS, biological oxygen demand (BOD), chemical oxygen demand (COD), total organic carbon (TOC), total nitrogen (TN), ammonium-nitrogen (NH₄-N), nitrate-nitrogen (NO₃-N), TP, phosphate-phosphorus (PO₄-P), algal CHL-a, and SD. The study period was 1993–2022, and monthly data were collected from the water surface layer (at 0.5 m depth) at each site. Monthly data on water level and precipitation were obtained from the Water Information Portal (<https://www.water.or.kr>), managed by the Korea Water Resources Corporation (K-water).

WT and EC were measured on-site using a YSI Sonde 6600 multi-parameter sensor (Environmental Monitoring Systems, Yellow Springs, OH, USA). SD was measured on-site using a standard 20-cm Secchi disk as water transparency indicator. Determination of the levels of organic matter (BOD, COD, and TOC), SS, nutrients, and CHL-a was followed by a standard procedure of MOE [42]. SS was determined in samples filtered through a 1.2- μ m pore size glass microfiber filter (grade GF/C; Whatman) and dried for 1 h at 105 °C [43]. COD was determined by oxidizing the water sample with potassium permanganate at 100 °C for 30 min after acidifying it with sulfuric acid [42,44]. A second derivative procedure was applied for TN after the digestion of persulfate [45], while ascorbic acid was used after oxidation to determine TP [46]. To measure CHL-a concentrations, the water sample was passed through a GF/C filter and extracted in ethanol before being assessed spectrophotometrically (DU-530; Beckman Coulter, Brea, CA, USA).

2.3. Statistical analysis

2.3.1. Empirical regression models and principal component analysis

The data were first log₁₀-transformed to address normality and heteroscedasticity concerns. The transformed spatial-seasonal averages of the designated variables were then used for empirical linear models (EMs), multiple linear regression (MLR), and principal component analysis (PCA). The regression analyses were performed using SigmaPlot v14.5 (Systat, Santa Clara, CA, USA), and

PCA was conducted using PAST v4.12 (University of Oslo, Norway).

2.3.2. Long-term trend analysis

We conducted a long-term trend analysis to detect significant trends in yearly mean values of water quality indicators in the reservoir from 1993 to 2022. The Mann–Kendall (MK) test and innovative trend analysis (ITA) were applied for this purpose.

The MK test is a non-parametric statistical method [47,48]. It checks the null hypothesis that there is no trend in the data against the alternative hypothesis that there is a trend. The following steps are carried out to perform the MK test. The formula for the MK test statistic S is:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & (\text{if } x_j - x_i > 0) \\ 0 & (\text{if } x_j - x_i = 0) \\ -1 & (\text{if } x_j - x_i < 0) \end{cases}$$

The variance is computed as:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^m t_i(t_i-1)(2t_i+5)}{18}$$

Then, the standard normal test statistic is calculated as:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & (\text{if } S > 0) \\ 0 & (\text{if } S = 0) \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & (\text{if } S < 0) \end{cases}$$

Sen's slope is a non-parametric method used to estimate the magnitude and direction of the trend in time series data [49] and is defined as:

$$\text{Sen's slope} = \text{Median} \left\{ \frac{x_j - x_i}{j - i} : i < j \right\}$$

for the set of pairs (i, x_i) where x_i is time series data. The slope represents the rate of change of the data over time and can be used to estimate the magnitude and direction of the trend. These methods have been widely used to detect trends in hydrological time series data because they can extract useful information on the possibility of future changes in water quality variables. Moreover, no assumptions about the distribution of the data must be met [50–52]. We performed these tests using the “*mannKen*” function in the R package “*wql*”, a statistical programming language used for data analysis.

The ITA is used to identify deterministic trends in observed time series [53]. It can be used without assumptions, distinguishing it from classical approaches like the MK trend test and Spearman's rho, which typically impose certain assumptions. The method begins by dividing the time series into two equal parts, which are then sorted in ascending order. The x-axis represents the first half of the time series, while the y-axis represents the second half. When the data points align along the ideal 1:1 line, it indicates the absence of a trend in the time series. However, if the data points accumulate in the region above the ideal line, it suggests the presence of an increasing trend in the time series. Conversely, if the data points are below the 1:1 line, it indicates a decreasing trend in the time series [53,54]. We implemented ITA using the “*innvtrend*” function in the R package “*trendchange*” and the 10 % confidence level was applied to ensure accurate analysis.

2.3.3. TSI and TSID

TSI values were calculated individually from CHL-a, TP, and SD according to the following formulas [24]:

$$\text{TSI (TP, } \mu\text{g/L)} = 10 \times [6 - \ln(48/\text{TP}) / \ln 2]$$

$$\text{TSI (CHL - a, } \mu\text{g/L)} = 10 \times [6 - (2.04 - 0.68 \ln(\text{CHL} - \text{a})) / \ln 2]$$

$$\text{TSI (SD, m)} = 10 \times [6 - \ln(\text{SD}) / \ln 2]$$

The TSI (CHL-a) values were then compared with the TSI (TP) and TSI (SD) values, and the resulting deviations were determined. Two-dimensional plots were then constructed based on these deviations. The degree of P limitation can be inferred from the TSI (CHL-a) – TSI (TP) relationship deviations, while deviations in TSI (CHL-a) and TSI (SD) represent the level of underwater light availability relative to sestonic particles [19]. The two-dimensional plots of TSID were constructed using SigmaPlot v14.5.

2.3.4. Machine learning predictions of algal biomass and water transparency based on RF

In addition to EMs and MLR models, we employed the random forest (RF) machine-learning technique to predict the long-term variations in algal CHL-a and SD levels by assessing by assessing and simulating environmental variables. The RF algorithm is a non-parametric machine learning model that randomly selects subsets of weak predictors and builds shallow regression trees from the training set [55]. The algorithm then aggregates the predictions of all the trees to generate the final output. It is a powerful tool with self-adaptability, self-organization, and error tolerance, and can minimize risk and the upper limit of generalization while also enhancing the generalization ability [37,56,57].

3. Results and discussion

3.1. Longitudinal variation in water quality and the effect of monsoon season

The data monitoring conducted over a long-term period revealed a longitudinal gradient in water quality indicators along the reservoir (Table 1). The averages of various indicators such as WT, EC, phosphorus contents, SS, and CHL-a increased from the Lz to the Rz, while TN:TP ratio and SD showed decreasing patterns. Studies have found that reservoir systems exhibit a noticeable change in water quality parameters as one moves from the headwaters to the dam zone [10,58–60]. Such changes include variations in suspended solids, nutrients, and primary productivity, which could be due to the morpho-hydrological conditions of the reservoirs. As conditions shift from river-type to lake-type over the longitudinal dimension of the reservoirs, changes in flow velocity, hydraulic retention time, and water depth could cause these variations [8,61]. In addition, external inputs from tributaries could also impact water quality indicators, especially nutrient regimes, inorganic suspended solids, and water transparency. As tributaries carry pollutants, mainly from intensive non-point pollution sources [38–40], into the riverine zone, the water quality can be significantly impacted, leading to higher levels of algal biomass, as measured by CHL-a, in the zones compared to other zones (Table 1). Overall, the temperate reservoir showed low organic matter and a low to moderate state of phosphorus contents and CHL-a, according to Ref. [25].

The longitudinal variations in most of the indicators further showed seasonality (Fig. 2a–u). Their means and outliers increased during the Mon in each zone, particularly in terms of suspended solids (Fig. 2d–f) and phosphorus contents (Fig. 2p–r). The elevated levels of SS and TP observed during the Mon generally accorded with increased precipitation and inflow volumes. These findings suggest that there was a greater amount of runoff and phosphorus loading during high rainfall periods due to monsoon [8,38,39]. However, CHL-a levels increased at each site in the Post (Fig. 2s–u). Monsoon-induced nutrient availability and hydrologically stable conditions after the monsoon period could be attributable to the increase in CHL-a level [26,27,31].

3.2. Seasonal variation in water quality parameters in drought and flood years

Eleven water quality indicators in drought (1996, 2016, and 2019) and flood (2002, 2009, and 2020) years were analyzed to determine the effects of monsoon rainfall on water quality (Figs. 3 and 4). Our analysis indicated that the summer monsoon's intensity

Table 1

The measured water quality parameters along the longitudinal zones (Lz – Lacustrine zone, Tz – Transition zone, Rz – Riverine zone) from 1993 to 2022 (WT: water temperature, EC: electrical conductivity, TN: total nitrogen, NH₄-N: ammonium-nitrogen, NO₃-N: nitrate nitrogen, TP: total phosphorus, PO₄-P: Orthophosphate, TN:TP: TN, TP ratio, COD: chemical oxygen demand, BOD: biological oxygen demand, CHL-a: chlorophyll-a, SS: suspended solids, SD: Secchi depth).

Water Quality Parameter		Lz	Tz	Rz	Overall
		Mean ± SD (Min-Max)	Mean ± SD (Min-Max)	Mean ± SD (Min-Max)	Mean ± SD (Min-Max)
WT/ionic indicators	WT (°C)	12.07 ± 4.86 (2–25)	13.78 ± 6.11 (2–26)	15.48 ± 7.55 (0–29)	13.77 ± 6.42 (0–29)
	EC (µS/cm)	75.53 ± 38.03 (35–220)	77.37 ± 26.03 (37–223)	82.24 ± 29.99 (36–265)	78.38 ± 31.87 (35–265)
Organic matter pollution indicators	BOD (mg/L)	0.99 ± 0.40 (0.4–3.8)	1.27 ± 0.43 (0.4–2.7)	1.45 ± 0.46 (0.4–4)	1.24 ± 0.47 (0.4–4)
	COD (mg/L)	2.86 ± 0.53 (1.5–6.7)	2.63 ± 0.38 (1.3–4.3)	2.78 ± 0.48 (1.4–6.0)	2.76 ± 0.48 (1.3–6.7)
	TOC (mg/L)	2.05 ± 0.42 (0.9–3.3)	1.71 ± 0.22 (1.2–2.3)	1.85 ± 0.3 (1.2–2.8)	1.89 ± 0.37 (0.9–3.3)
Nutrient pollution indicators	TN (mg/L)	0.78 ± 0.18 (0.32–1.91)	0.89 ± 0.28 (0.29–2.04)	0.96 ± 0.39 (0.29–2.75)	0.87 ± 0.31 (0.29–2.75)
	NH ₄ -N (mg/L)	0.05 ± 0.05 (0–0.28)	0.04 ± 0.05 (0–0.36)	0.05 ± 0.07 (0–0.68)	0.05 ± 0.06 (0–0.68)
	NO ₃ -N (mg/L)	0.48 ± 0.14 (0.00–1.24)	0.53 ± 0.18 (0–1.25)	0.56 ± 0.24 (0–1.48)	0.52 ± 0.19 (0–1.48)
	TP (µg/L)	14.25 ± 7.10 (0–54)	17.48 ± 8.62 (4–61)	22.17 ± 11.89 (6–98)	17.96 ± 9.96 (0–98)
	PO ₄ -P (µg/L)	2.41 ± 2.09 (0–11)	3.71 ± 3.11 (0–21)	5.07 ± 5.34 (0–57)	3.75 ± 3.93 (0–57)
	TN:TP	69.8 ± 61.1 (14.8–212.2)	59.9 ± 32.8 (8.7–277)	52.1 ± 32.8 (6.3–270.1)	60.61 ± 44.9 (6.3–277)
Algal bloom and transparency indicators	CHL-a (µg/L)	4.29 ± 10.42 (0.10–60.2)	6.12 ± 7.34 (0.10–64.5)	9.13 ± 10.62 (0.3–77.6)	6.52 ± 9.78 (0.1–77.6)
	SS (mg/L)	2.11 ± 1.06 (0.5–11.7)	2.62 ± 1.72 (0.3–11.3)	3.57 ± 3.34 (0.3–31.4)	2.76 ± 2.33 (0.3–31.4)
	SD (m)	3.35 ± 0.91 (0.9–6.1)	2.81 ± 64.5 (1–5.5)	2.38 ± 0.87 (0.5–5.9)	2.84 ± 0.96 (0.5–6.1)

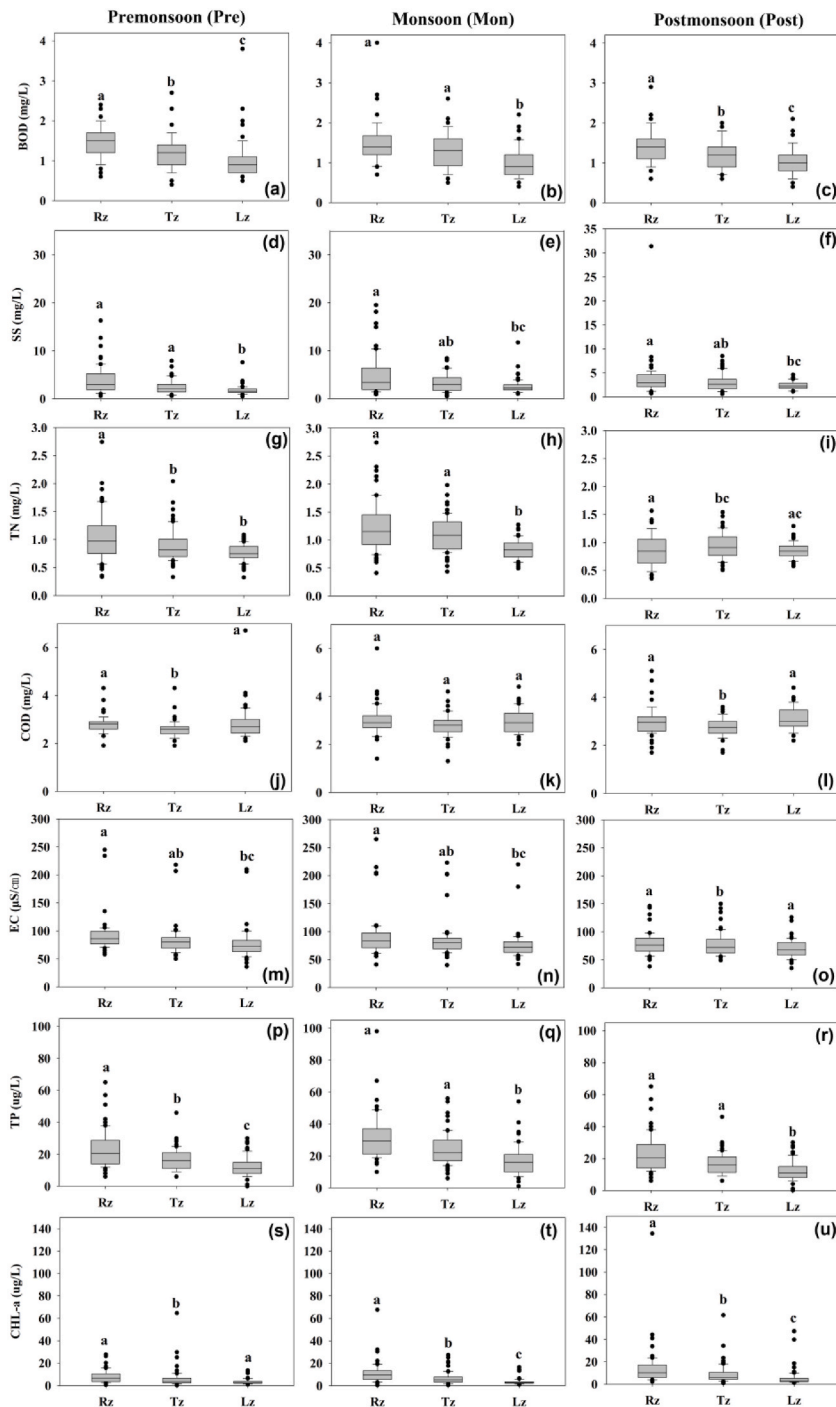


Fig. 2. Seasonal variations of water quality and Tukey test at each site (Rz: Riverine zone, Tz: Transition zone, Lz: Lacustrine zone), BOD: biological oxygen demand (a–c), SS: suspended solids (d–f), TN: total nitrogen (g–i), COD: chemical oxygen demand (j–l), EC: electrical conductivity (m–o), TP: total phosphorus (p–r), CHL-a: chlorophyll-a (s–u), the first highest mean receives the letter “a”, and the second and third highest mean receives the letter “b”, and “c”, respectively, and means with no significant difference receive the same letter).

significantly influenced the reservoir’s seasonal nutrient regimes, suspended solids, water transparency, and algal biomass. By comparing the patterns between years of drought and flood, we observed that the factors mentioned above were affected differently during each type of year.

There was no difference in WT between drought (=3–26) and flood (=4–26.7) years. The difference in EC between the Pre and Mon periods was small. The organic matter pollution indicators (COD and TOC) increased during the summer Mon, consistent with previous

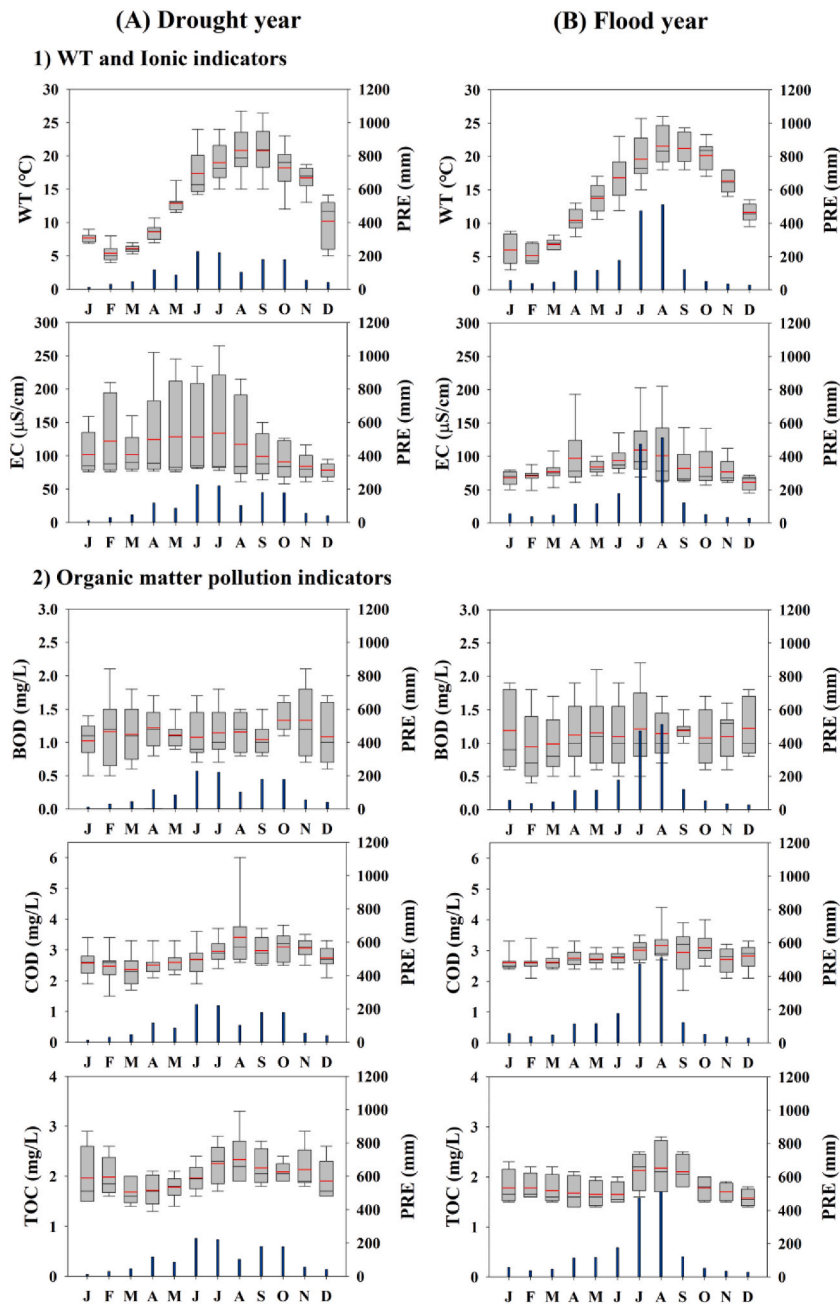


Fig. 3. Monthly dynamics in water temperature, the ionic indicator, and organic matter differed between drought (A) and flood years (B) in the reservoir (PRE-precipitation).

studies of Korean reservoirs [10,62,63]. Little difference in organic matter was observed between drought and flood years (Fig. 3A, B). The differences in BOD and COD concentrations were almost insignificant in the reservoir. These findings suggest low non-biodegradable organic matter in the reservoir. TN did not differ between the drought and flood years, and there was no significant N dilution during the Mon period, unlike other lakes in Korea [58]. TP increased significantly more (maximum = 91 μg/L) in flood years than in drought years (maximum = 56 μg/L) (Fig. 4A, B). The TN:TP ratio was higher in flood years during the Pre than Post periods, primarily attributed to increased TN (Fig. 4A, B). CHL-a increased during the Post in flood years, which is attributed to washout during the Mon [14]. The maximum value (61.5 μg/L) was observed in September, indicating the occurrence of an algal bloom. SS had the highest average value in September during a flood year, which was also attributed to an algal bloom (Fig. 4A, B). By contrast, SD decreased in July and August and increased in September during flood years, which is opposite to drought years. This suggests that inorganic SS dominates during the Mon [29], while organic matter algal blooms occur during the Post [23,31].

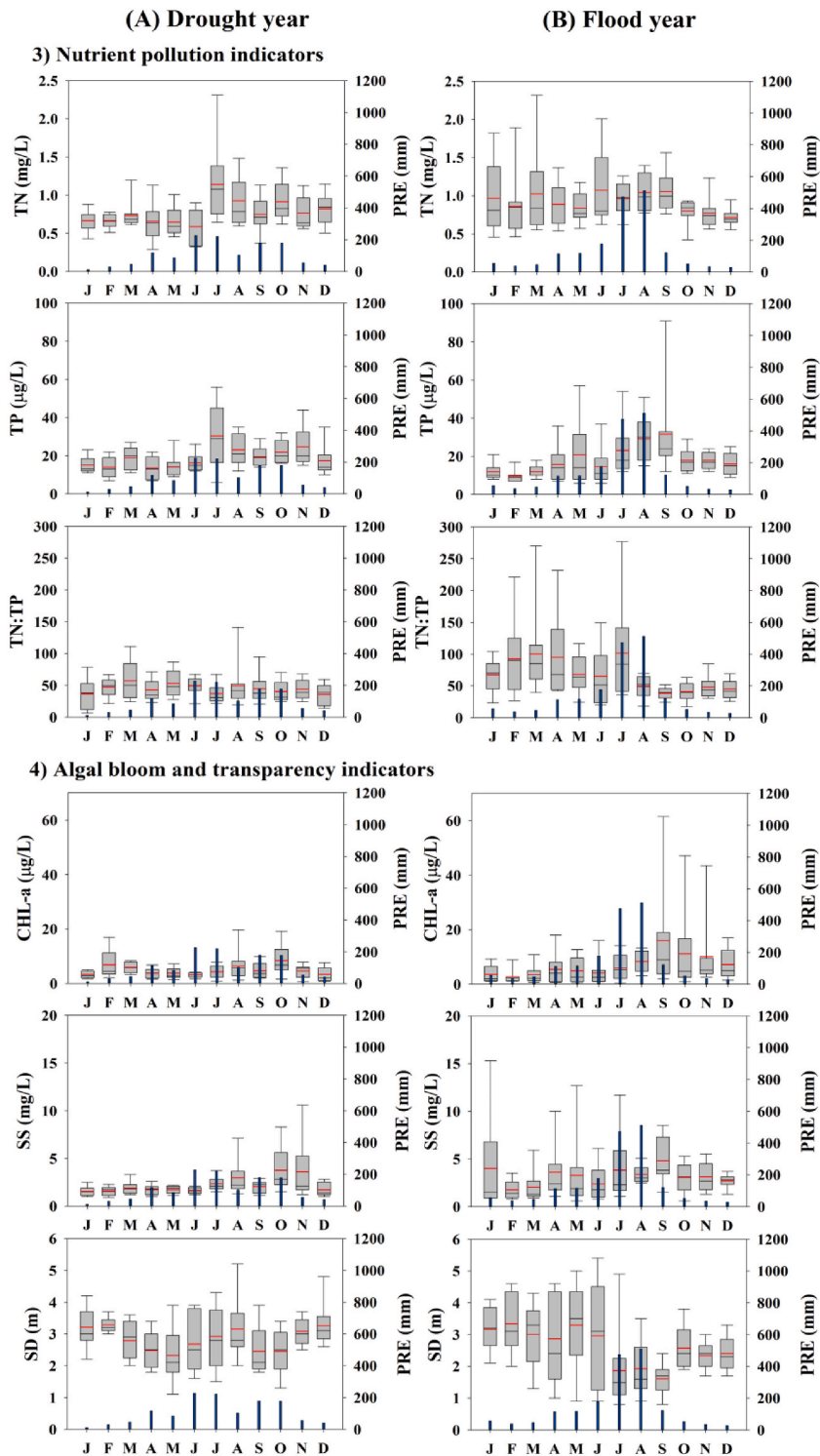


Fig. 4. Monthly dynamics in nutrients, TN:TP ratio, algal bloom, and transparency indicators differed between drought (A) and flood years (B) (PRE: precipitation).

Similar results were observed for TN:TP, EC, CHL-a, SD, and CHL-a:TP in drought years, although no significant differences were found among the Pre, Mon, and Post (Table 2). However, among the three seasons, COD, TOC, TN, TP, WT, and SS differed significantly in drought years ($p < 0.05$). Post hoc analysis revealed that COD was significantly lower in the Pre than Mon ($F = 5.201, p = 0.009$),

and TOC differed significantly between the Pre and Post ($F = 4.832$, $p = 0.014$), with higher values observed in the Post. TN and TP were significantly lower in the Pre than in both the Mon and Post ($p < 0.01$, Table 2). Similarly, significant differences were found in flood years for TOC, TP, TN:TP ratio, WT, CHL-a, and SD. Post hoc analysis revealed that TOC was significantly lower in the Pre than Post ($F = 4.211$, $p < 0.05$), and TP was significantly lower in the Pre than Post ($F = 4.52$, $p < 0.05$). The TN:TP ratio was significantly lower in Post than Mon ($F = 4.011$, $p < 0.05$), and WT was significantly higher in the Mon than Pre ($F = 22.41$, $p < 0.001$). Unlike drought years, CHL-a differed between the Pre and Post in flood years ($F = 4.898$, $p < 0.05$), with a unimodal peak in algal biomass in the Post. This finding is supported by previous studies of Korean reservoirs [17,31]. SD was significantly higher in the Pre than Mon ($F = 4.23$, $p = 0.02$). The CHL-a:TP ratio had a higher mean square value (0.4848) in flood years, indicating that CHL-a is formed in proportion to the amount of P supplied [29].

3.3. Relationship among water quality variables

The multivariate PCA showed the monsoon seasonal and longitudinal fluctuations in water quality indicators with their correlations (Fig. 5). The analysis method simplifies the handling of large datasets with multiple parameters by converting the original variables into smaller, independent ones. It can also identify the main sources of pollution that affect water quality [59,64]. This study used the first two principal components (PC I and PC II), which were selected based on criteria of an eigenvalue greater than 1, to illustrate seasonal and longitudinal patterns. Seasonal PCs accounted for 40.65 % of the total variance, with PC I explaining 23.19 % and PC II explaining 17.46 % (Fig. 5a). PC I had strong positive loadings for WT and TP and a moderate positive loading for SS. However, it also had a moderate negative loading of TN:TP. The variation based on these loadings was statistically acceptable between the Pre and Post according to 95 % confidence seasonal ellipses. The outcome suggests that seasonal increases in suspended solids and phosphorus contents were in the system due to summer monsoon, which led to fluctuations in nutrient ratio [12,14]. Additionally, PC II had strong positive loadings of CHL-a and CHL-a:TP, and a moderate negative loading of SD, indicating specifically that increases in algal biomass negatively correlated with water transparency.

In addition to seasonal monsoon patterns, spatial PCA indicated the longitudinal variations in water quality indicators along the reservoir, accounting for 51.32 % of the total variance (Fig. 5b). PC I explained 33.22 % of the variance. It had strong positive WT and CHL-a and moderate positive loadings of SS, BOD, pH, TP, and $PO_4\text{-P}$, suggesting longitudinal increases of the parameters from Lz to Rz. By contrast, SD exhibited strong negative loading. PC II accounted for 18.10 % of the variance and had moderate positive EC, TP, and $NH_4\text{-N}$ loadings. It also had TN:TP had strong negative loading of TN:TP and moderate negative loadings of CHL-a:TP and $NO_3\text{-N}$. The multivariate statistic results supported our recent findings on the longitudinal analysis and were consistent with previous studies for Korean reservoirs [10,65].

Table 2

One-way ANOVA test for drought and flood years (a: premonsoon (Pre), b: monsoon (Mon), c: postmonsoon (Post), Dr: Drought years, and Fl: Flood years).

Year	Parameters	Seasonal averages			Variance test					Post-hoc Comparison	
		Pre (a)	Mon (b)	Post (c)	SS	ν	MS	F-value	p-value	Tukey test	
Dr	BOD	1.09	1.32	1.19	0.027	2	0.013	0.763	0.472	a = b = c	
	COD	2.64	3.18	3.04	0.057	2	0.029	5.201	0.009^b	a < b (b = c)	
	TOC	1.88	2.29	2.13	0.043	2	0.022	4.832	0.014^c	a < c (a = b, c = b)	
	TN	0.62	1.03	0.83	0.43	2	0.215	9.72	0.0003^a	a < b (b = c)	
	TP	15.2	26.7	20.7	0.407	2	0.203	6.673	0.003^b	a < b (b = c)	
	TN:TP	43.1	47.1	42.1	0.11	2	0.055	1.666	0.199	a = b = c	
	WT	15.1	19.9	19.5	0.166	2	0.083	12.38	< 0.001^a	a < b (b = c)	
	EC	128	126	95	0.102	2	0.051	1.431	0.248	a = b = c	
	CHL-a	3.46	5.40	6.54	0.399	2	0.19	2.586	0.085	a = b = c	
	SS	1.70	2.66	2.90	0.353	2	0.177	5.717	0.006^b	a < b (b = c)	
	SD	2.50	3.04	2.46	0.097	2	0.049	2.535	0.089	a = b = c	
	CHL-a:TP	0.243	0.236	0.327	0.204	2	0.102	1.32	0.276	a = b = c	
	Fl	BOD	1.13	1.18	1.13	0.008	2	0.004	0.153	0.858	a = b = c
		COD	2.74	3.09	3.02	0.023	2	0.011	2.784	0.071	a = b = c
TOC		1.65	2.15	1.94	0.049	2	0.025	4.211	0.029^c	a < c (a = b, c = b)	
TN		0.96	1.00	0.93	0.016	2	0.008	0.515	0.601	a = b = c	
TP		17.7	26.1	24.8	0.52	2	0.265	4.52	0.016^c	a < c (a = b, c = b)	
TN:TP		83.4	43.9	44.3	0.446	2	0.223	4.011	0.024^c	c < b (a = b, a = c)	
WT		15.3	20.6	20.7	0.222	2	0.111	22.41	< 0.001^a	a < b (b = c)	
EC		89	105	83	0.071	2	0.036	1.89	0.162	a = b = c	
CHL-a		4.46	7.10	20.2	2.332	2	1.166	4.898	0.011^c	a < c (a = b, c = b)	
SS		2.83	3.61	5.41	0.654	2	0.327	3.018	0.058	a = b = c	
SD		3.13	1.89	2.09	0.391	2	0.196	4.23	0.02^c	b < a (a = c, b = c)	
CHL-a:TP		0.221	0.305	0.718	0.97	2	0.485	3.062	0.056	a < c (a = b, c = b)	

SS: sum of squares; ν : degrees of freedom; MS – mean square; Significance level.

^a Very strong.

^b Strong.

^c Moderate.

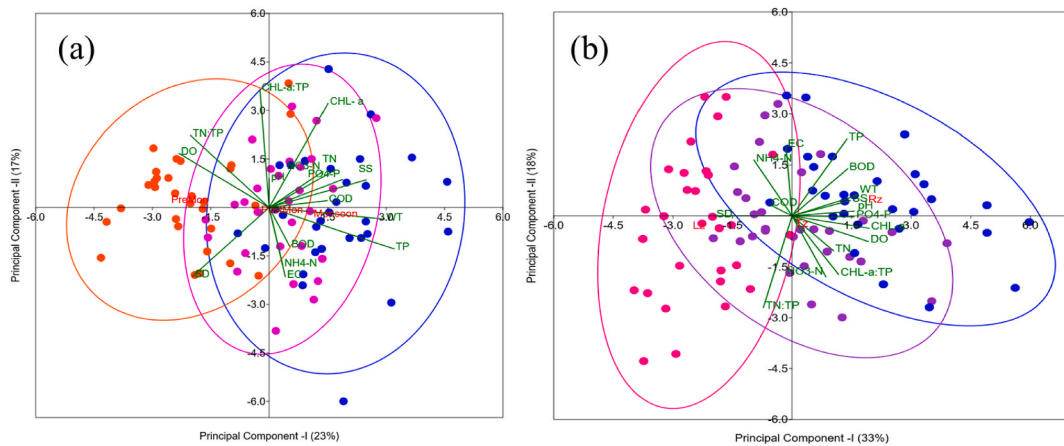


Fig. 5. Principal component analysis (PCA) of the reservoir studied during 1993–2022; (a) seasonal biplot of PCA (red = premonsoon, blue = monsoon, and pink = postmonsoon), (b) longitudinal zonal biplot of PCA (pink = Lz, purple = Tz, and blue = Rz).

3.4. Empirical models of reservoir algal biomass and water transparency

Empirical models based on simple linear regression showed that phosphorus was the most significant determinant for the variation in longitudinal-monthly algal CHL-a averages ($n = 36$) for 30 years, closely followed by TN:TP ratio (Fig. 6a–d). TP and TN:TP explained approximately 64 % and 57 % of the variation ($p < 0.001$) according to the coefficient determination (R^2). TN:TP showed a strong negative correlation with TP ($R^2 = 0.72$, $p < 0.001$), but it had a weak negative correlation with TN ($R^2 = 0.25$, $p < 0.01$). These correlations highlighted the importance of phosphorus availability relative to nitrogen in shaping algal growth dynamics, suggesting a phosphorus limitation within the system. This inference aligns with earlier empirical studies on Korean multipurpose and deep reservoirs [10,11,65]. Additionally, TN and WT significantly and positively correlated with CHL-a ($p < 0.001$), with respective R^2 values of 0.43 and 0.21 (Fig. 6a–d), indicating their influence on algal biomass. Therefore, we developed multiple linear regression models (MLR) using WT, TN, and TN:TP added on TP (Table S1). The best MLR was based on the interplay between TP, TN:TP, and WT and accounted for 67 % of the CHL-a variation, indicating a slight improvement in prediction accuracy as measured by R^2 .

Moreover, when SD was empirically modeled based on nutrients (TN and TP), CHL-a, and SS (Fig. 6e–h), SS accounted for a substantial part (81 %) of the variation in SD (Fig. 6h). This suggests that SS was the dominant regulator of water transparency and underwater light availability, due to its strong association with increased nutrients (TN and TP) and CHL-a levels (Fig. 6i–l). Notably, SS showed a particularly robust correlation with TP ($R^2 = 0.86$, $p < 0.001$). As previously discussed by Refs. [23,66], suspended solids tend to bind efficiently and transport phosphorus particles within the water column during the summer monsoon, especially in high concentrations, influencing SD variations. Consequently, the fluctuations in SD may be more attributable to inorganic suspended matter from monsoon runoffs rather than algal biomass. It aligns with the findings of [30], who observed a greater predictive power of SS over CHL-a in determining SD using machine learning techniques. Hence, employing SD as a metric for assessing trophic states in Korean reservoirs may be misleading, particularly considering the influence of monsoon-specific conditions. Similar observations have been reported in other regions, such as the Eastern Chinese Plains [18] and the Mediterranean [67], where reductions in SD were primarily attributed to inorganic turbidity and matter rather than CHL-a concentrations. Future studies should address the challenge of SD in trophic state assessment.

However, the interactive correlations among nutrients, suspended solids, SD, and CHL-a levels can underscore the dynamics of nutrient loading and its implications for algal growth and water transparency in the studied reservoir system. Elevated phosphorus levels, facilitated by suspended solids, can contribute to heightened CHL-a levels. As algal biomass increases, it can further diminish water transparency, impacting light penetration and potentially establishing a feedback loop where heightened algal biomass perpetuates decreased water transparency [61], exacerbating the situation.

3.5. Monsoon-seasonal analyses of TSI and TSID along the system with the influence of drought and flood years regimes

Assessing reservoir trophic status is crucial for its ecological health [11]. In this regard, a longitudinal and monsoon-seasonal analysis was conducted based on the TSIs of TP, SD, and CHL-a following Carlson's method [24]. The results revealed that the reservoir trophic status ranged from mesotrophic to eutrophic overall and varied between drought and flood years (Fig. 7).

TSI (TP) averages decreased longitudinally from Rz to Lz, with seasonal increases being evident in Mon at each zone (Fig. 7a). The overall TSI (SD) pattern was similar to that of TSI (TP) (Fig. 7b), and the seasonal increases in both indices were primarily attributed to heightened inorganic suspended solids carried with phosphorus due to monsoon runoff into the reservoir [23,29]. Also, a noticeable longitudinal decrease was observed in the TSI (CHL-a) averages from Rz to Lz (Fig. 7c). Seasonal increases were evident in both Mon and Post at Rz and Lz. Particularly, eutrophic states were observed in Rz during Mon and Post, indicating a heightened risk of algal blooms. Moreover, TSI (CHL-a) increased at Lz during the Post, with 75 % of its distribution pointing towards a eutrophic state

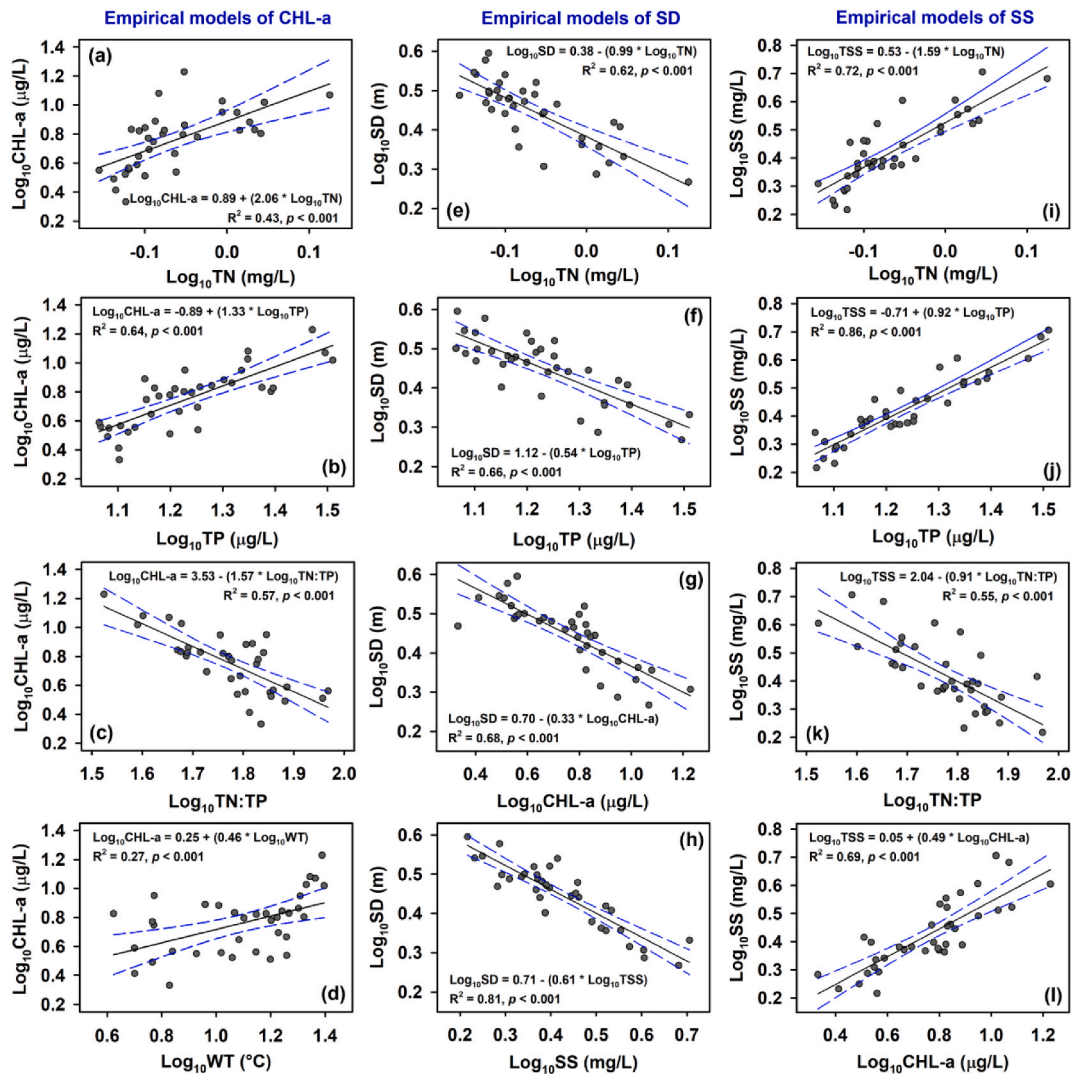


Fig. 6. Empirical log-linear models of algal CHL-a (a–d), water transparency as measured by SD (e–h), and suspended solids (SS; i–l) based on total nitrogen (TN), total phosphorus (TP), TN:TP ratio, and water temperature (WT).

(Fig. 7c).

Our analysis also revealed a crucial aspect regarding the impact of monsoon intensity on trophic state conditions, which varied significantly between flood and drought years (Fig. 7d–i), particularly for TSI (SD) and TSI (CHL-a). The variability in these indices for each season was considerably low in drought years (Fig. 7e and f). However, this variability increased during flood years (Fig. 7h and i). The TSI (SD) averages were notably higher in Mon with the eutrophic state during flood years (Fig. 7e) than in drought years (Fig. 7f). Furthermore, TSI (CHL-a) averages demonstrated an observable increase in eutrophication risk in Post after intense monsoon periods (Fig. 7i), particularly in Lz. In drought years, there were longitudinal decreases in TSI (CHL-a) averages for each season (Fig. 7f). However, during flood years, the spatial variation was interrupted, and all zones experienced eutrophic states in Post. The findings supported that intense rainfall events would trigger algal blooms in subsequent months, particularly in the lacustrine site [31].

TSID analyses, furthermore, suggested that phosphorus availability strongly influences the reservoir's trophic status, with several episodic impacts of non-algal turbidity at each site during Mon (Fig. S2), according to Ref. [19]. The findings were aligned with the insights obtained by empirical models of CHL-a and SD, supporting a P-limitation scenario in the temperate deep system along with the seasonal conditions of the summer monsoon [68]. Also, increases in non-algal turbidity influence (P-III) were evident during Mon in flood years compared to drought years, as indicated by mean values of TSID (Fig. 8). TSID also showed algal bloom occurrences (P-I) were more prevalent during Post in flood years compared to drought years.

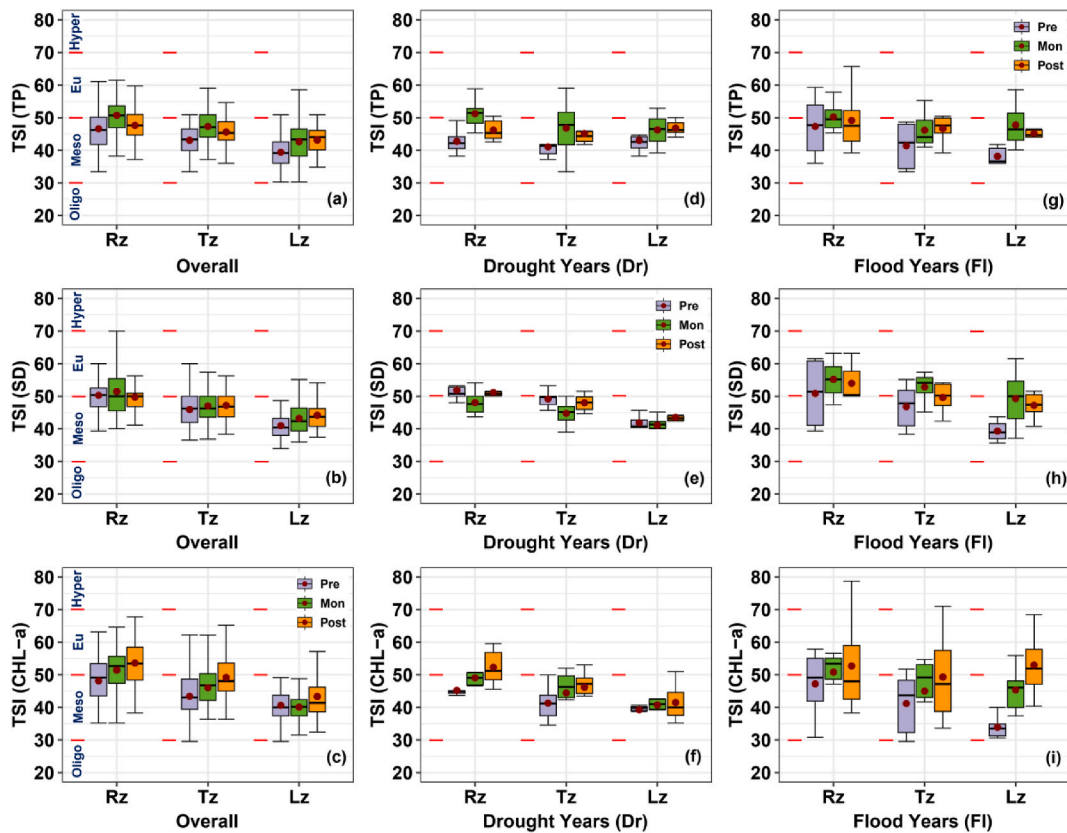


Fig. 7. Longitudinal and monsoon-seasonal variations in trophic state indices based on total phosphorus (TSI (TP)), Secchi depth (TSI (SD)), and CHL-a (TSI (CHL-a)) for overall (a–c), drought, and flood years (d–i). Seasons were categorized as premonsoon (Pre), monsoon (Mon), and postmonsoon (Post). The trophic status was classified as oligotrophic (Oligo), mesotrophic (Meso), eutrophic (Eu), and hypereutrophic (Hyper).

3.6. Long-term trend analysis

Our extensive dataset, covering a 30-year span, provided a platform for investigating the long-term trends in water quality indicators using the MK test and ITA. The analysis of yearly reservoir means revealed notable overall trends: BOD, TN, TP, TSS, and SD showed decreasing trends, whereas COD showed an increasing trend according to the MK test (Table 3). Conversely, TOC, EC, and CHL-a did not exhibit significant trends over the studied period. The ITA corroborated the findings, except for BOD and SD (Fig. 9a–i).

The contrasting trends observed between BOD and COD suggest a potential shift in the pollutants, possibly indicating an increase in the proportion of non-biodegradable organic matter [65]. While the decline in nutrients and suspended solids over the long-term period might suggest an improvement in reservoir water quality, it is important to consider other factors, such as increased water volume and storage capacity, which could also contribute to these changes. Early and recent studies [38–40] underscore the persistent influence of intense non-point pollution sources on reservoir water quality, potentially fostering algal blooms. Interannual-seasonal variations in hydrological conditions and nutrient availability can regulate the long-term dynamics of algal biomass during growing seasons, mainly triggered by the summer monsoon [10,59]. Monsoon-specific conditions might lead to no significant trend in algal CHL-a levels in the system studied. Also, diffuse nutrient losses worldwide are considered the main drivers of eutrophication [69]. This highlights the ongoing relevance and significance of monitoring and controlling the impacts of non-pollution sources on the system and non-biodegradable pollutants, particularly by considering the summer monsoon intensity.

3.7. Predictions of algal CHL-a and SD based on machine learning approach

The study also analyzed the modeling accuracies of the random forests (RF) machine learning, MLR, and EM models in predicting CHL-a and SD levels in the Juam Reservoir. The time series plots clearly illustrated the observed and predicted levels of both CHL-a and SD, generated using RF, MLR, and Ems (Fig. 10a and b). CHL-a values predicted by the RF model closely aligned with the observed values, indicating superior performance compared to predictions generated by both the best MLR and EM models. To validate this qualitative assessment, the study conducted a rigorous analysis utilizing the coefficient of determination (R^2) to compare the modeling accuracies of the three models. The RF model demonstrated a significantly higher R^2 value of 0.72, surpassing the performance of both the MLR and EM models. When employing the RF model, this elevated R^2 value indicates a robust correlation and low error between

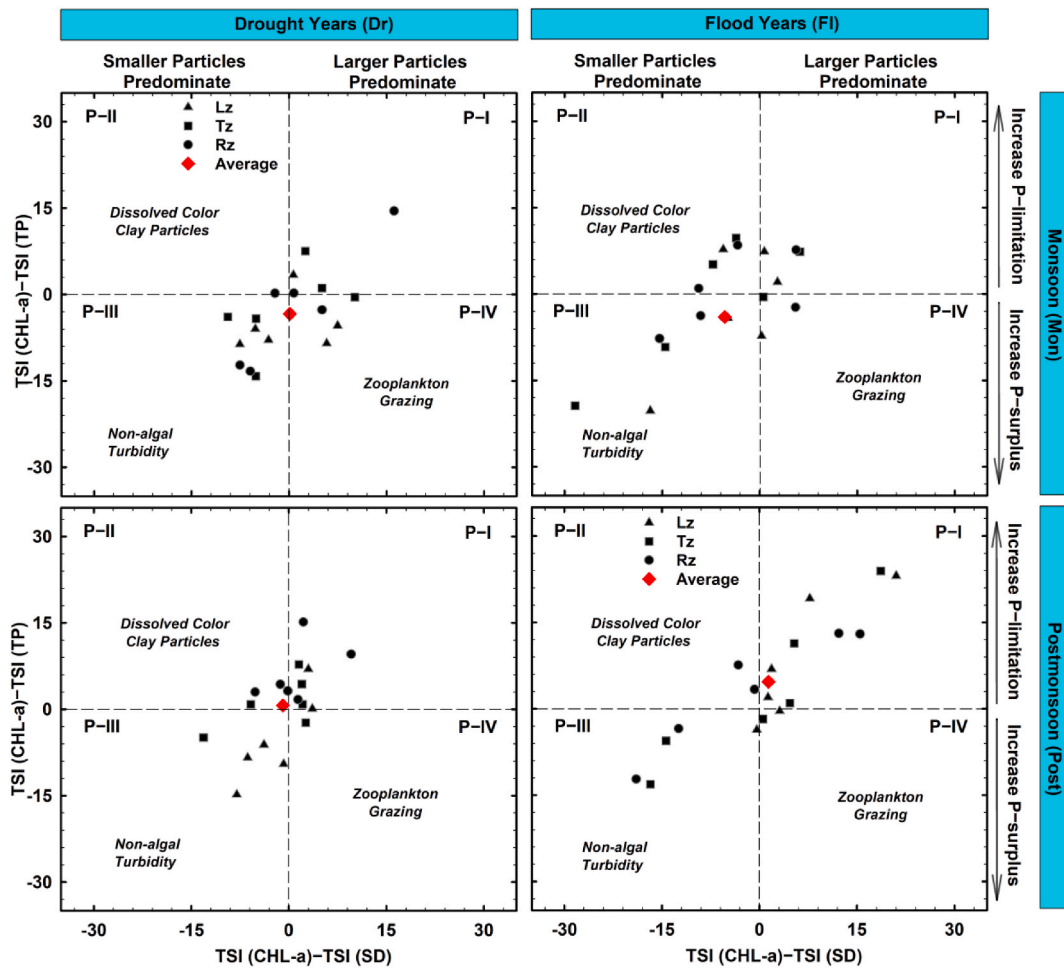


Fig. 8. Variations in trophic state index deviations during monsoon and postmonsoon seasons between drought and flood years.

Table 3

Long-term trend analysis of water quality indicators in the reservoir over the 30-year period (1993–2022).

Indicators	S	τ	Slope	Trend
BOD	-176	-0.405	-0.014 ^b	Decreasing
COD	240	0.552	0.019 ^a	Increasing
TOC	39	-0.371	0.019	No trend
TN	-145	-0.333	-0.007 ^c	Decreasing
TP	-156	-0.359	-0.194 ^b	Decreasing
EC	35	0.08	0.137	No trend
CHL-a	27	0.062	0.024	No trend
TSS	-117	-0.269	-0.038 ^c	Decreasing
SD	-66	-0.152	-0.010 ^b	Decreasing

^a $p < 0.001$.

^b $p < 0.01$.

^c $p < 0.05$.

the predicted and observed CHL-a levels, reaffirming its efficacy in modeling CHL-a dynamics within the Juam Reservoir. Based on these findings, the study recommends using the RF model to predict algal biomass and water transparency in multipurpose reservoirs.

4. Conclusions

This study provides important insights into the longitudinal and monsoon seasonal variation in water chemistry in the Asian temperate reservoir, highlighting its impact on algal biomass as indicated by CHL-a and long-term trends over three decades. The study

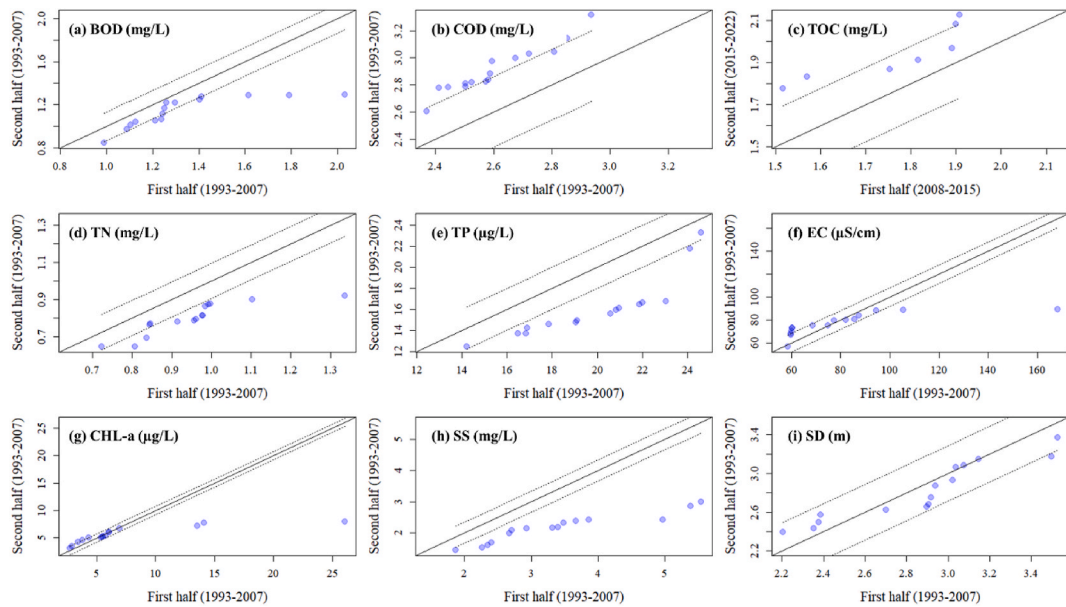


Fig. 9. Innovative trend analysis (ITA) of water quality indicators over the study period, including biological oxygen demand (BOD), chemical oxygen demand (COD), total organic carbon (TOC), total nitrogen (TN), total phosphorus (TP), electric conductivity (EC), algal CHL-a, suspended solids (SS), and Secchi depth (SD).

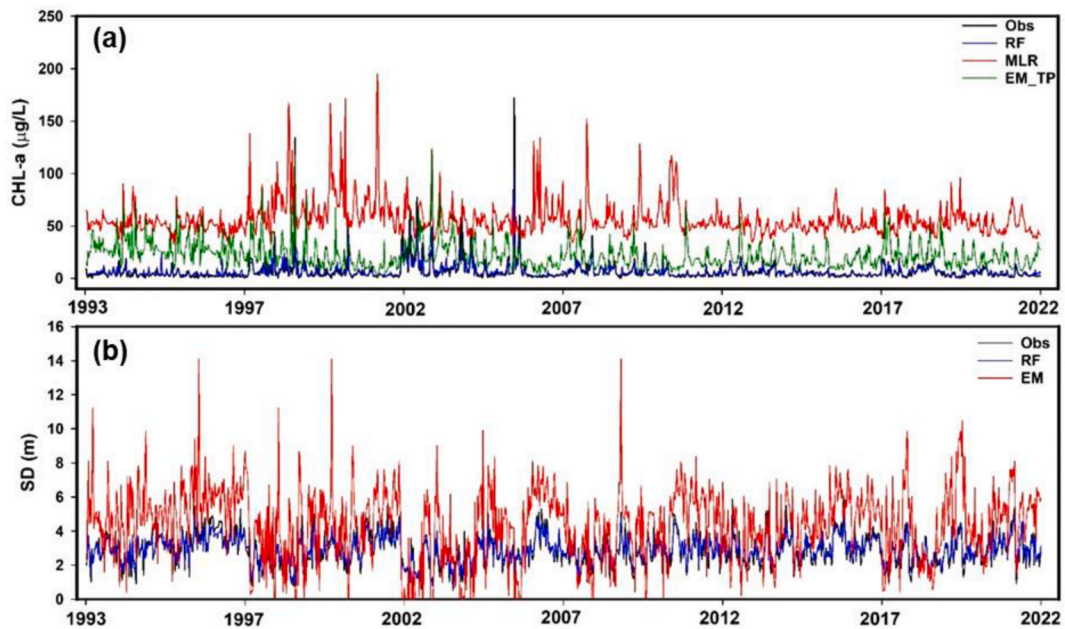


Fig. 10. The comparison of (a) predicted and observed CHL-a values based on the best multiple linear regression (MLR), empirical model (EM_TP), and random forest (RF), along with (b) SD predictions by RF and EM.

used various analytical methods, including multivariate analysis, empirical models, trophic state analysis, and machine learning approaches. The findings revealed longitudinal heterogeneity in nutrients, organic matter, CHL-a, and SS along the Rz, Tz, and Lz. Seasonal nutrient and suspended solid fluctuations and decreased water transparency were observed during the summer monsoon (Mon), influenced by external inputs and varying dynamics during flood and drought years. Furthermore, we found a monomodal increase in CHL-a levels after the summer monsoons, particularly during intense rainfall, indicating successful algal growth during the post-monsoon period.

According to empirical models, phosphorus was found to be the key determinant of algal biomass, while suspended solids played a

significant role in regulating water transparency. TSID analyses also suggested that phosphorus availability strongly influences the reservoir's trophic status, with episodes of non-algal turbidity at each site during the monsoons. The findings supported a P-limitation scenario in the temperate deep system along with the seasonal conditions of the summer monsoon.

Our study also highlighted overall trends in water quality parameters over the long period, albeit with the signaling of the pollutants towards non-biodegradable organic matter. Moreover, while declining nutrient and suspended solid levels could indicate improved water quality, the persistent influence of non-point pollution sources underscores the need for continued monitoring and control, particularly considering seasonal variations and the impact of the summer monsoon on algal biomass dynamics and nutrient loadings. The predictive capabilities demonstrated by random forests machine learning suggested the potential utility of such an approach in forecasting algal biomass and water transparency.

Data availability statement

Data will be made available on reasonable request.

CRedit authorship contribution statement

Sang-Hyeon Jin: Writing – original draft, Formal analysis. **Namsrai Jargal:** Writing – review & editing, Writing – original draft, Formal analysis. **Thet Thet Khaing:** Writing – original draft, Formal analysis. **Min Jae Cho:** Writing – original draft, Formal analysis, Data curation. **Hyeji Choi:** Writing – original draft, Formal analysis. **Bilguun Ariunbold:** Writing – original draft, Formal analysis. **Mnyagatwa Geoffrey Donat:** Writing – original draft, Formal analysis. **Haechan Yoo:** Writing – original draft, Formal analysis. **Md Mamun:** Validation, Formal analysis. **Kwang-Guk An:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition, Data curation.

Declaration of competing interest

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

Acknowledgment

This research is funded by the Daejeon Green Environment Center under the Research Development Program (Yr 2022) and the authors would like to acknowledge the assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e31643>.

References

- [1] B. Lehner, G. Grill, Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, *Hydrol. Process.* 27 (15) (2013) 2171–2186, <https://doi.org/10.1002/hyp.9740>.
- [2] J.R. Jones, C. Pollard, D.V. Obrecht, Factors influencing phosphorus in midcontinent impoundments (USA) and challenges for detecting abatement, *Inland Waters* 12 (1) (2022) 154–162, <https://doi.org/10.1080/20442041.2020.1816802>.
- [3] M.F. Chislock, E. Doster, R.A. Zitomer, A.E. Wilson, Eutrophication: causes, consequences, and controls in aquatic ecosystems, *Nature Education Knowledge* 4 (4) (2013) 10.
- [4] D.W. Schindler, S.R. Carpenter, S.C. Chapra, R.E. Hecky, D.M. Orihel, Reducing phosphorus to curb lake eutrophication is a success, *Environ. Sci. Technol.* 50 (17) (2016) 8923–8929, <https://doi.org/10.1021/acs.est.6b02204>.
- [5] W. Dodds, M. Whiles, *Freshwater Ecology: Concepts and Environmental Applications of Limnology*, third ed., Elsevier Academic Press, London, UK, 2019.
- [6] K.E. Havens, G.K. Nürnberg, The phosphorus-chlorophyll relationship in lakes: potential influences of color and mixing regime, *Lake Reservoir Manag.* 20 (3) (2004) 188–196, <https://doi.org/10.1080/07438140409354243>.
- [7] G. Phillips, O.P. Pietiläinen, L. Carvalho, A. Solimini, A. Lyche Solheim, A.C. Cardoso, Chlorophyll–nutrient relationships of different lake types using a large European dataset, *Aquat. Ecol.* 42 (2008) 213–226, <https://doi.org/10.1007/s10452-008-9180-0>.
- [8] K.G. An, J.R. Jones, Reservoir response to the Asian monsoon with an emphasis on longitudinal gradients, *J. Freshw. Ecol.* 17 (1) (2002) 151–160, <https://doi.org/10.1080/02705060.2002.9663878>.
- [9] D.G.F. Cunha, M.D.C. Calijuri, Limiting factors for phytoplankton growth in subtropical reservoirs: the effect of light and nutrient availability in different longitudinal compartments, *Lake Reservoir Manag.* 27 (2) (2011) 162–172, <https://doi.org/10.1080/07438141.2011.574974>.
- [10] N. Jargal, U. Atique, M. Mamun, K. An, Seasonal and long-term connections between trophic status, sestonic chlorophyll, nutrients, organic matter, and monsoon rainfall in a multipurpose reservoir, *Water* 13 (13) (2021) 1720, <https://doi.org/10.3390/w13131720>.
- [11] J.R. Jones, M.F. Knowlton, K.G. An, Trophic state, seasonal patterns and empirical models in South Korean reservoirs, *Lake Reservoir Manag.* 19 (1) (2003) 64–78, <https://doi.org/10.1080/07438140309353991>.
- [12] M. Mamun, K. An, Major nutrients and chlorophyll dynamics in Korean agricultural reservoirs along with an analysis of trophic state index deviation, *J. Asia Pac. Bus.* 10 (2) (2017) 183–191, <https://doi.org/10.1016/j.japb.2017.04.001>.
- [13] E. Jeppesen, M. Søndergaard, J.P. Jensen, K.E. Havens, O. Anneville, L. Carvalho, M.F. Coveney, R. Deneké, M.T. Dokulil, B. Foy, D. Gerdeaux, S.E. Hampton, S. Hilt, K. Kangur, J. Köhler, E.H.H.R. Lammens, T.L. Lauridsen, M. Manca, M.R. Miracle, B. Moss, P. Nöges, G. Persson, G. Phillips, R. Portielje, S. Romo, C.

- L. Schelske, D. Straile, I. Tatrai, E. Willén, M. Winder, Lake responses to reduced nutrient loading—an analysis of contemporary long-term data from 35 case studies, *Freshw. Biol.* 50 (10) (2005) 1747–1771, <https://doi.org/10.1111/j.1365-2427.2005.01415.x>.
- [14] K.G. An, S.S. Park, Indirect influence of the summer monsoon on chlorophyll-total phosphorus models in reservoirs: a case study, *Ecol. Model.* 152 (2–3) (2002) 191–203, [https://doi.org/10.1016/S0304-3800\(02\)00020-0](https://doi.org/10.1016/S0304-3800(02)00020-0).
- [15] J.M. Abell, D. Özkundakci, D.P. Hamilton, J.R. Jones, Latitudinal variation in nutrient stoichiometry and chlorophyll-nutrient relationships in lakes: a global study, *Fundamental Applied Limnology* 181 (2012) 1–14, <https://doi.org/10.1127/1863-9135/2012/0272>.
- [16] Q. Wang, R. Liu, C. Men, L. Guo, Application of genetic algorithm to land use optimization for non-point source pollution control based on CLUE-S and SWAT, *J. Hydrol.* 560 (2018) 86–96, <https://doi.org/10.1016/j.jhydrol.2018.03.022>.
- [17] U. Aitúque, K. An, Landscape heterogeneity impacts water chemistry, nutrient regime, organic matter and chlorophyll dynamics in agricultural reservoirs, *Ecol. Indic.* 110 (2020) 105813, <https://doi.org/10.1016/j.ecolind.2019.105813>.
- [18] W. Zou, G. Zhu, Y. Cai, A. Vilmi, H. Xu, M. Zhu, Z. Gong, Y. Zhang, B. Qin, Relationships between nutrient, chlorophyll a and Secchi depth in lakes of the Chinese Eastern Plains ecoregion: implications for eutrophication management, *J. Environ. Manag.* 260 (2020) 109923, <https://doi.org/10.1016/j.jenvman.2019.109923>.
- [19] R.E. Carlson, K.E. Havens, Simple graphical methods for the interpretation of relationships between trophic state variables, *Lake Reservoir Manag.* 21 (1) (2005) 107–118, <https://doi.org/10.1080/07438140509354418>.
- [20] J.R. Jones, M.F. Knowlton, Chlorophyll response to nutrients and non-algal seston in Missouri reservoirs and oxbow lakes, *Lake Reservoir Manag.* 21 (3) (2005) 361–371, <https://doi.org/10.1080/07438140509354441>.
- [21] W. Liu, Q. Zhang, G. Liu, Lake eutrophication associated with geographic location, lake morphology and climate in China, *Hydrobiologia* 644 (2010) 289–299, <https://doi.org/10.1007/s10750-010-0151-9>.
- [22] R. Quinlan, A. Filazzola, O. Mahdiyan, A. Shuvo, K. Blagrave, C. Ewins, L. Moslenko, D.K. Gray, C.M. O'Reilly, S. Sharma, Relationships of total phosphorus and chlorophyll in lakes worldwide, *Limnol. Oceanogr.* 66 (2) (2021) 392–404, <https://doi.org/10.1002/lno.11611>.
- [23] N. Jargal, E. Lee, K. An, Monsoon-induced response of algal chlorophyll to trophic state, light availability, and morphometry in 293 temperate reservoirs, *J. Environ. Manag.* 337 (2023) 117737, <https://doi.org/10.1016/j.jenvman.2023.117737>.
- [24] R.E. Carlson, A trophic state index for lakes 1, *Limnol. Oceanogr.* 22 (2) (1977) 361–369, <https://doi.org/10.4319/lo.1977.22.2.0361>.
- [25] G.K. Nürnberg, Trophic state of clear and colored, soft-and hardwater lakes with special consideration of nutrients, anoxia, phytoplankton and fish, *Lake Reservoir Manag.* 12 (4) (1996) 432–447, <https://doi.org/10.1080/07438149609354283>.
- [26] J.R. Jones, P. McEachern, D. Seo, Empirical evidence of monsoon influences on Asian lakes, *Aquat. Ecosys. Health Manag.* 12 (2) (2009) 129–137, <https://doi.org/10.1080/14634980902907342>.
- [27] S. Jung, M. Shin, J. Kim, J. Eum, Y. Lee, J. Lee, B. Kim, The effects of Asian summer monsoons on algal blooms in reservoirs, *Inland Waters* 6 (3) (2016) 406–413, <https://doi.org/10.1080/iw-6.3.967>.
- [28] S.J. Lee, K.G. An, Variations of limnological functions in a man-made reservoir ecosystem during high-flow year vs. Low-flow year, *Korean Journal of Ecology and Environment* 42 (4) (2009) 487–494.
- [29] M. Mamun, S. Kwon, J. Kim, K. An, Evaluation of algal chlorophyll and nutrient relations and the N:P ratios along with trophic status and light regime in 60 Korea reservoirs, *Sci. Total Environ.* 741 (2020) 140451, <https://doi.org/10.1016/j.scitotenv.2020.140451>.
- [30] M. Mamun, J. Kim, M.A. Alam, K. An, Prediction of algal chlorophyll-a and water clarity in monsoon-region reservoir using machine learning approaches, *Water* 12 (1) (2019) 30, <https://doi.org/10.3390/w12010030>.
- [31] N. Jargal, K. An, Seasonal and interannual responses of blue-green algal taxa and chlorophyll to a monsoon climate, flow regimes, and N:P ratios in a temperate drinking-water reservoir, *Sci. Total Environ.* 896 (2023) 165306, <https://doi.org/10.1016/j.scitotenv.2023.165306>.
- [32] Y. Park, K.H. Cho, J. Park, S.M. Cha, J.H. Kim, Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea, *Sci. Total Environ.* 502 (2015) 31–41, <https://doi.org/10.1016/j.scitotenv.2014.09.005>.
- [33] Z. Liao, N. Zang, X. Wang, C. Li, Q. Liu, Machine learning-based prediction of chlorophyll-a variations in receiving reservoir of world's largest water transfer project—a case study in the miyun reservoir, north China, *Water* 13 (17) (2021) 2406, <https://doi.org/10.3390/w13172406>.
- [34] M. Dou, X. Ma, Y. Zhang, Y. Zhang, Y. Shi, Modeling the interaction of light and nutrients as factors driving lake eutrophication, *Ecol. Model.* 400 (2019) 41–52, <https://doi.org/10.1016/j.ecolmodel.2019.03.015>.
- [35] N. Mellios, S.J. Moe, C. Lapidou, Machine learning approaches for predicting health risk of cyanobacterial blooms in northern European lakes, *Water* 12 (4) (2020) 1191, <https://doi.org/10.3390/w12041191>.
- [36] S. Lin, D.C. Pierson, J.P. Mesman, Prediction of algal blooms via data-driven machine learning models: an evaluation using data from a well-monitored mesotrophic lake, *Geosci. Model Dev.* (GMD) 16 (1) (2023) 35–46, <https://doi.org/10.5194/gmd-16-35-2023>.
- [37] H. Huang, W. Wang, J. Lv, Q. Liu, X. Liu, S. Xie, F. Wang, J. Feng, Relationship between chlorophyll a and environmental factors in lakes based on the random forest algorithm, *Water* 14 (19) (2022) 3128, <https://doi.org/10.3390/w14193128>.
- [38] K.H. Han, K.S. Yoon, J.W. Jung, S.G. Yoon, Y.J. Kim, Pollutant load characteristics of a rural watershed of Juam lake, *Journal of the Korean Society of Agricultural Engineers* 47 (3) (2005) 77–86, <https://doi.org/10.5389/KSAE.2005.47.3.077>.
- [39] H.S. Yi, S. Chong, S.J. Lee, Assessment of apprehensive area of non-point source pollution using watershed model application in Juam Dam watershed, *Journal of Korean Society of Environmental Engineers* 37 (10) (2015) 551–557, <https://doi.org/10.4491/KSEE.2015.37.10.551>.
- [40] T. Lee, Assessment of sediment and total phosphorus loads using SWAT in Oenam watershed, Hwasun, Jeollanam-do, *Journal of the Korean Association of Regional Geographers* 22 (1) (2016) 240–250.
- [41] H. Chung, M. Son, H.-S. Ryu, C.H. Park, R. Lee, M. Cho, C. Lim, J. Park, K. Kim, Variation of cyanobacteria occurrence pattern and environmental factors in Lake Juam, *Korean Journal of Environmental Biology* 37 (4) (2019) 640–651, <https://doi.org/10.11626/KJEB.2019.37.4.640>.
- [42] MOE, Korean Standard Methods of Water Pollution, seventh ed., Ministry of Environment (MOE), Gwacheon, Korea, 2000 (in Korean).
- [43] APHA, Standard Methods for the Examination of Water and Wastewater, twenty-first fifty tenth ed., American Public Health Association (APHA), New York, NY, USA, 2005.
- [44] T. Korenaga, Flow injection analysis using potassium permanganate: an approach for measuring chemical oxygen demand in organic wastes and waters, *Anal. Lett.* 13 (11) (1980) 1001–1011, <https://doi.org/10.1080/00032718008078020>.
- [45] W.G. Crumpton, T.M. Isenhardt, P.D. Mitchell, Nitrate and organic N analyses with second-derivative spectroscopy, *Limnol. Oceanogr.* 37 (4) (1992) 907–913, <https://doi.org/10.4319/lo.1992.37.4.0907>.
- [46] E.E. Prepas, F.H. Rigler, Improvements in quantifying the phosphorus concentration in lake water, *Can. J. Fish. Aquat. Sci.* 39 (6) (1982) 822–829, <https://doi.org/10.1139/f82-112>.
- [47] H.B. Mann, Nonparametric tests against trend, *Econometrica: J. Econom. Soc.* 13 (3) (1945) 245–259, <https://doi.org/10.2307/1907187>.
- [48] M.G. Kendall, *Rank Correlation Methods*, Griffin, London, 1975.
- [49] P.K. Sen, Estimates of the regression coefficient based on kendall's tau, *J. Am. Stat. Assoc.* 63 (324) (1968) 1379–1389, <https://doi.org/10.1080/01621459.1968.10480934>.
- [50] Y.S. Yu, S. Zou, D. Whittemore, Non-parametric trend analysis of water quality data of rivers in Kansas, *J. Hydrol.* 150 (1) (1993) 61–80, [https://doi.org/10.1016/0022-1694\(93\)90156-4](https://doi.org/10.1016/0022-1694(93)90156-4).
- [51] A. Mustapha, Detecting surface water quality trends using Mann-Kendall tests and Sen's slope estimates, *International Journal of Agriculture Innovations and Research* 1 (2013) 108–114.
- [52] Z. Hu, S. Liu, G. Zhong, H. Lin, Z. Zhou, Modified Mann-Kendall trend test for hydrological time series under the scaling hypothesis and its application, *Hydrol. Sci. J.* 65 (14) (2020) 2419–2438, <https://doi.org/10.1080/02626667.2020.1810253>.
- [53] Z. Şen, Innovative trend analysis methodology, *J. Hydrol. Eng.* 17 (9) (2012) 1042–1046, [https://doi.org/10.1061/\(asce\)he.1943-5584.0000556](https://doi.org/10.1061/(asce)he.1943-5584.0000556).

- [54] T. Caloiero, R. Coscarelli, E. Ferrari, Application of the innovative trend analysis method for the trend analysis of rainfall anomalies in southern Italy, *Water Resour. Manag.* 32 (2018) 4971–4983, <https://doi.org/10.1007/s11269-018-2117-z>.
- [55] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [56] J. Xu, Z. Xu, J. Kuang, C. Lin, L. Xiao, X. Huang, Y. Zhang, An alternative to laboratory testing: random forest-based water quality prediction framework for inland and nearshore water bodies, *Water* 13 (22) (2020) 3262, <https://doi.org/10.3390/w13223262>.
- [57] P.K. Jena, S.M. Rahaman, P.K. Das Mohapatra, D.P. Barik, D.S. Patra, Surface water quality assessment by Random Forest, *Water Pract. Technol.* 18 (1) (2023) 201–214, <https://doi.org/10.2166/wpt.2022.156>.
- [58] J.H. Han, J.Y. Lee, K.G. An, Interannual and seasonal variations of water quality in terms of size dimension on multi-purpose Korean Dam Reservoirs along with the characteristics of longitudinal gradients, *Korean Journal of Ecology and Environment* 43 (2) (2010) 319–337.
- [59] U. Atique, K.G. An, Reservoir water quality assessment based on chemical parameters and the chlorophyll dynamics in relation to nutrient regime, *Pol. J. Environ. Stud.* 28 (3) (2019) 1–19, <https://doi.org/10.15244/pjoes/85675>.
- [60] M. Mamun, K.G. An, Development of reservoir water quality index (WQI) based on long-term physicochemical parameters and their spatio-temporal variations, *Asian J. Water Environ. Pollut.* 17 (2) (2020) 55–63, <https://doi.org/10.3233/AJW200021>.
- [61] J. Kalf, *Limnology*, vol. 608, Prentice-Hall, Inc., 113 Sylvan Ave, Englewood Cliffs NJ 07632 USA, 2002, pp. 136–153.
- [62] A. Reza, J. Eum, S. Jung, Y. Choi, J.S. Owen, B. Kim, Export of non-point source suspended sediment, nitrogen, and phosphorus from sloping highland agricultural fields in the East Asian monsoon region, *Environ. Monit. Assess.* 188 (2016) 1–15, <https://doi.org/10.1007/s10661-016-5681-9>.
- [63] J. Im, C. Kang, K. Gil, Applicability on wet-land for management of NPS in organic matter and nutrients from agriculture and livestock farm area, *Journal of Wetlands Research* 23 (4) (2021) 372–380, <https://doi.org/10.17663/JWR.2021.23.4.372>.
- [64] Y. Ouyang, P. Nkedi-Kizza, Q. Wu, D. Shinde, C. Huang, Assessment of seasonal variations in surface water quality, *Water Res.* 40 (20) (2006) 3800–3810, <https://doi.org/10.1016/j.watres.2006.08.030>.
- [65] M. Mamun, J.Y. Kim, K.G. An, Multivariate statistical analysis of water quality and trophic state in an artificial dam reservoir, *Water* 13 (2) (2021) 186, <https://doi.org/10.3390/w13020186>.
- [66] M. Mamun, N. Jargal, K. An, Spatio-temporal characterization of nutrient and organic pollution along with nutrient-chlorophyll-a dynamics in the Geum River, *J. King Saud Univ. Sci.* 34 (7) (2022) 102270, <https://doi.org/10.1016/j.jksus.2022.102270>.
- [67] J. Armengol, L. Caputo, M. Comerma, C. Feijóo, J.C. García, R. Marcé, E. Navarro, J. Ordoñez, Sau reservoir's light climate: relationships between Secchi depth and light extinction coefficient, *Limnética* 22 (1–2) (2003) 195–210.
- [68] K.G. An, S.S. Park, K.H. Ahn, C.G. Urchin, Dynamics of nitrogen, phosphorus, algal biomass, and suspended solids in an artificial lentic ecosystem and significant implications of regional hydrology on trophic status, *J. Environ. Biol.* 24 (1) (2003) 29–38.
- [69] M. Le Moal, C. Gascuel-Oudou, A. Ménesguen, Y. Souchon, C. Étrillard, A. Levain, F. Moatar, A. Pannard, P. Souchu, A. Lefebvre, G. Pinay, Eutrophication: a new wine in an old bottle? *Sci. Total Environ.* 651 (2019) 1–11, <https://doi.org/10.1016/j.scitotenv.2018.09.139>.