

## Research article

# Key drivers of microcystin-producing cyanobacteria in South Korean eutrophic waters determined with data-driven models

Jayun Kim <sup>a,b,\*</sup> , Jiyoung Lee <sup>a,c</sup>, Jason W. Marion <sup>d</sup> , Joonhong Park <sup>b</sup>

<sup>a</sup> Division of Environmental Health Sciences, College of Public Health, The Ohio State University, Columbus, OH 43210, USA

<sup>b</sup> Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Republic of Korea

<sup>c</sup> Department of Food Science & Technology, The Ohio State University, Columbus, OH 43210, USA

<sup>d</sup> Environmental Health Science & Sustainability Program, College of Health Sciences, Eastern Kentucky University, Richmond, KY 40475, USA

## ARTICLE INFO

Handling editor: Lixiao Zhang

### Keywords:

Algal bloom  
Cyanotoxin  
Machine learning  
MC-LR  
Microcystis  
Eutrophication

## ABSTRACT

The rise in cyanobacterial harmful algal blooms (CHABs), driven by eutrophication and climate change, necessitates understanding cyanotoxin conditions to mitigate risks. However, limited studies have explored the influencing factors of cyanobacterial community and cyanotoxins in freshwaters using large datasets and data-driven models (DDMs). This study examines cyanobacterial composition across regions and relates it to microcystin-LR, -RR, -LA, -YR, -LF, and -LY as well as to the taste and odor compounds geosmin and 2-methylisoborneol. It also assesses the impact of environmental factors using interpretable machine learning. In the Nakdong River (NR) and Daecheong Lake (DL), differences in cyanobacterial community emerged regardless of nitrogen to phosphorus (N/P) levels. Warmer NR areas exhibited prevalent *Microcystis* during CHAB events. 1300–1700 *Microcystis* cells/mL produced 0.03 µg microcystin-LR/L (model accuracy 95 %) and DDMs predicted geosmin and 2-methylisoborneol less accurately. Random forest models using environmental factors predicted *Microcystis* dominance with 52 % accuracy at NR and DL, highlighting water temperature (positively) and nutrient levels (total N/P or NH<sub>3</sub>-N/PO<sub>4</sub>-P, negatively) as the primary factors influencing *Microcystis* dominance. Higher N/P contributed to higher *Microcystis* dominance only at DL when the water temperature was high. Phosphorus input was significantly affected by rainfall in NR but not DL. This study demonstrates toxicity predictions using environmental variables through DDMs and underscores the need to manage nutrient pollution sources to prevent *Microcystis* proliferation and MC-LR in surface waters.

## 1. Introduction

Over the past decades, algal blooms have become increasingly common in surface waters worldwide, primarily because of climate change and eutrophication (Heisler et al., 2008; Tararu et al., 2012; Hou et al., 2022). Cyanotoxins and taste and odor (T&O) compounds, generated during cyanobacterial harmful algal blooms (CHABs), pose the greatest concern. The most problematic T&O compounds are geosmin and 2-methylisoborneal (MIB). Although these compounds do not directly impact human health, even a very small amount of their presence in water makes it difficult to drink (Kehoe et al., 2015).

Consequently, the costs of removing T&O compounds in water treatment plants are tremendous (Srinivasan and Sorial, 2011). On the other hand, cyanotoxins directly affect human health. Humans can be exposed to cyanotoxins via contaminated water consumption, dermal contact (Carmichael and Boyer, 2016; Cheung et al., 2013), and aerosol inhalation (Facciponte et al., 2018). They may even be ingested via foods grown with contaminated water and soil (Ai et al., 2020; Lee et al. 2017, 2021). Microcystins (MCs) are most common among cyanotoxins that can damage the liver, brain, and organs of animals and humans (Metcalf and Codd, 2012; Mrdjen et al., 2022; Zhang et al., 2021). Among the more than 300 identified MC congeners, MC-leucine-arginine (MC-LR) is

**Abbreviations:** ANOVA, Analysis of Variance; BOD, biochemical oxygen demand; CHAB, cyanobacterial harmful algal bloom; Chl-a, chlorophyll a; COD, chemical oxygen demand; DDM, data-driven model; DL, Daecheong Lake; DO, dissolved oxygen; DT, decision tree; DTN, dissolved total nitrogen; DTP, dissolved total phosphorus; EC, electrical conductivity; MC, microcystin; MC-LR, microcystin-leucine-arginine; MIB, 2-methylisoborneal; MLR, multiple linear regression; NR, Nakdong River; NSE, Nash-Sutcliffe Efficiency; RF, random forest; SHAP, SHapley Additive exPlanations; SS, suspended solids; SVM, support vector machine; T&O, taste and odor; TN, total nitrogen; TP, total phosphorus; WT, water temperature; XAI, explainable artificial intelligence.

\* Corresponding author. Division of Environmental Health Sciences, College of Public Health, The Ohio State University, Columbus, OH 43210, USA.

E-mail address: [nature9973@gmail.com](mailto:nature9973@gmail.com) (J. Kim).

<https://doi.org/10.1016/j.jenvman.2025.126616>

Received 1 May 2025; Received in revised form 7 July 2025; Accepted 15 July 2025

Available online 29 July 2025

0301-4797/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

considered the most toxic (Gupta et al., 2003; Yang et al., 2020). The International Agency for Research on Cancer (IARC) classifies MC-LR as a potential human carcinogen (IARC, 2010). Therefore, predicting occurrences of CHAB and elucidating the factors that influence their MC-LR content are crucial steps in proactively minimizing their adverse impacts.

Decades ago, studies presented the effects of environmental conditions such as temperature, sunlight, nutrients, carbon dioxide, and water movement on the growth of phytoplankton (Canale and Vogel, 1974; Capblancq and Catalan, 1994). Later research also examined the importance of nutrients for chlorophyll *a* (Chl-a) (Perkins and Underwood, 2000; Lau and Lane, 2002), findings that are consistent with those of recent studies (Park et al., 2015; Rousso et al., 2020). It is well known that nutrient reduction, especially phosphorus, can mitigate overall cyanobacteria abundance and bloom size; however, due to the environment becoming more nitrogen-rich relative to phosphorus, the environment can shift towards being more favorable for toxicogenic bloom formers which produce nitrogen-rich compounds like MCs (Hellweger et al., 2022). Common MC-producing cyanobacteria are *Microcystis*, *Planktothrix*, *Dolichospermum* (formerly *Anabaena*), and *Aphanizomenon* (Du et al., 2019; Svirčev et al., 2019; Zhang et al., 2021). Given the complexity of these dynamic aquatic ecosystems, it is important to characterize environmental conditions favorable to these various cyanobacteria by exploiting consistently collected long-term data, using standard methods, across various water bodies.

In 2012, the Republic of Korea enhanced regulations on water quality for the effluents from wastewater treatment facilities to mitigate eutrophication in surface waters. Approximately 301 million US dollars were allocated from 2010 to 2011 to install total phosphorus (TP) treatment facilities at wastewater treatment plants to comply with new guidelines (Shin, 2010). These guidelines stipulated a TP limit of 0.2 mg/L (for treatment facilities processing more than 500 m<sup>3</sup>/day), a significant reduction from the initial standard of 2 mg/L. Nevertheless, the nation continues to grapple with the challenges posed by CHABs in major water bodies that are used for drinking water sources and various purposes, especially in the Nakdong River (Park et al., 2021) and Daecheong Lake (Kim et al., 2021). Despite reducing phosphorus to lower levels from point source pollution in surface waters in Korea, the question of what factors promote CHABs and their toxicity remains unresolved.

Machine learning algorithms, leveraging large datasets that include numerous water quality and environmental variables, can be used to develop data-driven models (DDMs) capable of predicting ecological responses. These models have shown the potential to outperform conventional process-based models in predicting CHABs (Kim et al., 2023). In addition to prediction, DDMs can aid in elucidating the influential factors for these responses (Rousso et al., 2020). Meta-analyses have demonstrated nutrient levels being linked to phytoplankton (Faithfull et al., 2011) and cyanotoxin production (Brandenburg et al., 2020). Nutrient level variations are associated with algal blooms in China's hypereutrophic Lake Taihu (Wilhelm et al., 2011; Xu et al., 2017), and recent research using water quality data across 34 sampling points with interpretable machine learning aided in understanding the nutrients and pollutants affecting the relative abundance and cell density of two cyanobacterial genera (*Microcystis* and *Cyanobium*) in China's Chaobai River (Wang et al., 2024). Continued efforts are being made to improve the accuracy of algal bloom predictions through the application of novel DDM approaches (Busari et al., 2025; Nguyen et al., 2025). Understanding and predicting harmful substances associated with cyanobacteria, rather than overall algae, remains crucial. Nevertheless, the relationships between environmental factors, cyanobacterial communities, geosmin, MIB, and various congeners of MCs in various drinking water sources have not yet been thoroughly examined or understood. For addressing the concern and uncertainties associated with CHABs, as well as the temporal and spatial distribution of the toxin MCs, the factors associated with their increases need to be reliably determined.

To improve understanding of key drivers for problematic cyanobacteria and MC-toxicity risks in eutrophic source waters, this study focused on developing a predictive model for MC-LR, the most toxic MC congener, to aid in devising a robust and health-relevant CHAB-related risk mitigation and prevention plan for the study region. Accordingly, this study aimed to (i) characterize the spatial distribution of cyanobacteria genera and their relationship with geosmin, MIB, and MCs; and (ii) identify the major environmental factors influencing MC-LR levels in waters severely affected by CHABs, using DDMs. To achieve these aims, data on cyanobacteria community composition, T&O compounds, distribution of MC congeners, and environmental factors were collected from multiple sites within a eutrophic lake and a river. Explainable artificial intelligence (XAI) models were then employed to analyze the data.

## 2. Materials and methods

### 2.1. Study areas and data acquisition

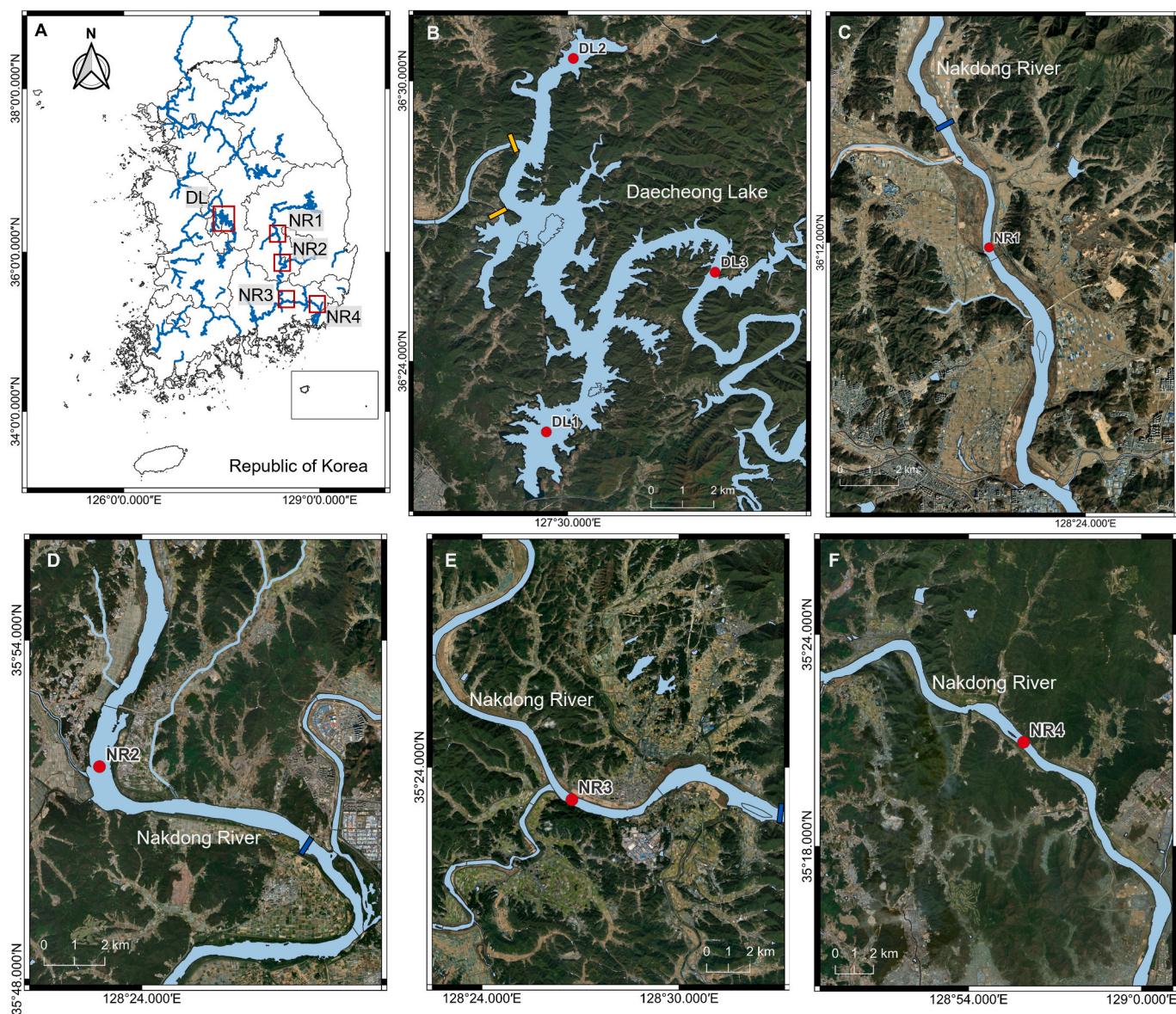
#### 2.1.1. Sampling sites and procedures

The Nakdong River (NR), with a length of 510 km, and Daecheong Lake (DL), covering an area of 73 km<sup>2</sup>, have the highest cyanobacteria densities among the continuously monitored surface waters in South Korea. The NR and DL are not hydrologically connected. The NR drains to the Korea Strait (South Sea). DL is part of the Geum River watershed which drains to the Yellow Sea on South Korea's west coast. Four and three CHAB alert sites used for drinking water intake sources on the NR and DL, respectively, were selected in this study (Fig. 1). These CHAB alert sites issue alerts based on weekly cyanobacteria monitoring. The alert thresholds are as follows: caution when cyanobacteria  $\geq 1000$  cells/mL, warning when  $> 10,000$  cells/mL (with monitoring increased to twice per week during this level), and severe bloom level when  $\geq 1,000,000$  cells/mL.

Composite water samples were taken using a Van Dorn water sampler at each study site and analyzed by local government agencies. At NR, samples were a mixture from two to three different depths at the point where the water was deepest: within 50-cm of the surface, 1/3 of the water depth, and 2/3 of the water depth (when depth  $\geq 2$  m). During algal alert "warning" periods, water samples were a mixture from three different points: the deepest point and the two points halfway to the left and right sides of the river. At each point, samples were also taken from two to three different depths. At DL, samples were a mixture from three different depths at the point where the water was deepest: within 50-cm of the surface, 1/2 of the water depth, and between the bottom and 5 m above the bottom. For cyanobacteria cell enumeration, a formalin solution (3–5 %) or Lugol's iodine (1–2 %) was added to 100 mL of a water sample. Water samples were stored at 4 °C for further analyses.

#### 2.1.2. Cyanobacterial, toxin, and water quality data

The data on cyanobacteria, MC-LR, geosmin, MIB, and water quality indicators were acquired from the Water Environment Information System (<https://water.nier.go.kr>), managed by the National Institute of Environmental Research, which is operated by the Ministry of Environment. Cyanobacterial cells were enumerated using a Sedgwick-Rafter counting chamber, and four genera (*Microcystis*, *Planktothrix*, *Dolichospermum*, and *Aphanizomenon*) were identified using a microscope at magnifications of 400  $\times$  to 1000  $\times$ . Total MC-LR concentrations were measured by liquid chromatography-tandem mass spectrometry (Guo et al., 2017; Shoemaker et al., 2015) after the cells were lysed through sonication (30 min, consisting of repetitions of 4 min and 30 s of sonication followed by a 30 s pause) in a vial placed in cold water, followed by filtration of the sample through a 1.2-μm GF/C filter. Dissolved geosmin and MIB concentrations were measured using a headspace-solid phase microextraction method coupled with gas chromatography-mass spectrometry (Hwan et al., 2023). The data related to cyanobacteria and water quality indicators were collected on



**Fig. 1.** Maps of (a) South Korea showing the study area, (b) Daecheong Lake (DL) sampling sites (DL1–3), and (c–f) Nakdong River (NR) sampling sites (NR1–4). Rectangles indicate installed dams in DL and weirs on the NR.

a weekly basis, except for the cyanobacteria data, which were collected more than two times a week when a “warning” alert was issued. In contrast, data for MC-LR, geosmin, and MIB were exclusively collected during these “warning” periods. Additionally, daily rainfall data were obtained from nearby weather monitoring stations at each study site. The period of data collection spanned from 2016 to 2022 for NR1–3 and DL, and from 2020 to 2023 for NR4. In 2024, measurements of MC congeners—MC-LR, -RR, -LA, -YR, -LF, and -LY—as well as geosmin and MIB, were obtained at the study sites and two additional recreational water sites approximately 18 and 25 km downstream of NR4. Notably, the amino acid residues differentiating these congeners include arginine-arginine in MC-RR, leucine and alanine in MC-LA, tyrosine and arginine in MC-YR, leucine and phenylalanine in MC-LF, and leucine and tyrosine in MC-LY.

## 2.2. Data analyses

### 2.2.1. DDM development and input variables

Data on water quality indicators (water temperature (WT, °C), pH,

dissolved oxygen (DO, mg/L), transparency (m), turbidity (NTU), Chl-a ( $\text{mg}/\text{m}^3$ ), biochemical oxygen demand (BOD, mg/L), chemical oxygen demand (COD, mg/L), dissolved total nitrogen (DTN, mg/L), dissolved total phosphorus (DTP, mg/L), electrical conductivity (EC,  $\mu\text{S}/\text{cm}$ ),  $\text{NH}_3\text{-N}$  (mg/L),  $\text{NO}_3\text{-N}$  (mg/L),  $\text{PO}_4^{3-}\text{P}$  (mg/L), suspended solids (SS, mg/L), total nitrogen (TN, mg/L), TP (mg/L), precipitation (7-day rainfall in mm), concentrations of MC-LR ( $\mu\text{g}/\text{L}$ ), geosmin (ng/L), and MIB (ng/L), as well as counts for four cyanobacteria genera (*Microcystis*, *Planktothrix*, *Dolichospermum*, and *Aphanizomenon*, measured in cells/mL), were collected from the study sites. A tabular data frame was created by aggregating these data points, ensuring alignment by their respective measurement dates for each study site. Data from April to November were used for model development, excluding periods when CHAB levels were minimal or absent.

Multiple linear regression (MLR), decision tree (DT), support vector machine (SVM), and random forest (RF) algorithms were employed to develop predictive DDMs. MLR, DT, and SVM models were developed for predicting concentrations of MC-LR, geosmin, and MIB, using cyanobacterial genera as independent variables. Subsequently, MLR, DT,

SVM, and RF models incorporated a variety of independent variables to predict a particular cyanobacteria genus associated with MC-LR production. Additionally, the ratio of TN to TP (TN/TP), NH<sub>3</sub>-N/NO<sub>3</sub>-N, and NH<sub>3</sub>-N/PO<sub>4</sub>-P were included as independent variables in the models. Furthermore, generalized linear model was used to examine the relationship between cyanobacteria genera and MC congeners, geosmin, and MIB using data in 2024.

### 2.2.2. Specifications of DDMs

In the development of DT, SVM, and RF models, various hyperparameters were tuned, resulting in the highest performance demonstrated in this study. For DT, the maximum depth of the tree was considered at three levels: 7, 10, and unlimited. The number of features to consider when looking for the best split included: the square root of the number of features, log2 of the number of features, and the actual number of features. The minimum number of samples required to split a node was set to 2, 3, or 4, and the minimum number of samples required to be at a leaf node was set to 1, 2, or 3. For SVM, the types of kernels considered were polynomial and radial basis function; and the regularization parameter C was experimented with at values of 0.01, 0.1, and 1. In the case of RF, the number of trees in the forest was fixed at one hundred. The maximum number of features considered for splitting a node was set at either five or the square root of the number of features.

### 2.2.3. Model evaluation and interpretation

To evaluate the performance of the developed models, the Nash-Sutcliffe Efficiency (NSE) was employed as the accuracy metric. The NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the observed data variance. An NSE value of 1 indicates perfect model predictions, 0 indicates that the model predictions are as accurate as the mean of the observed data, and a value less than 0 indicates unacceptable performance. To ensure a comprehensive and reliable evaluation, the leave-one-out cross-validation method was implemented to assess the test performance. This method involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This process is repeated such that each observation in the sample is used once as the validation data.

The important variables for predicting the toxic-producing

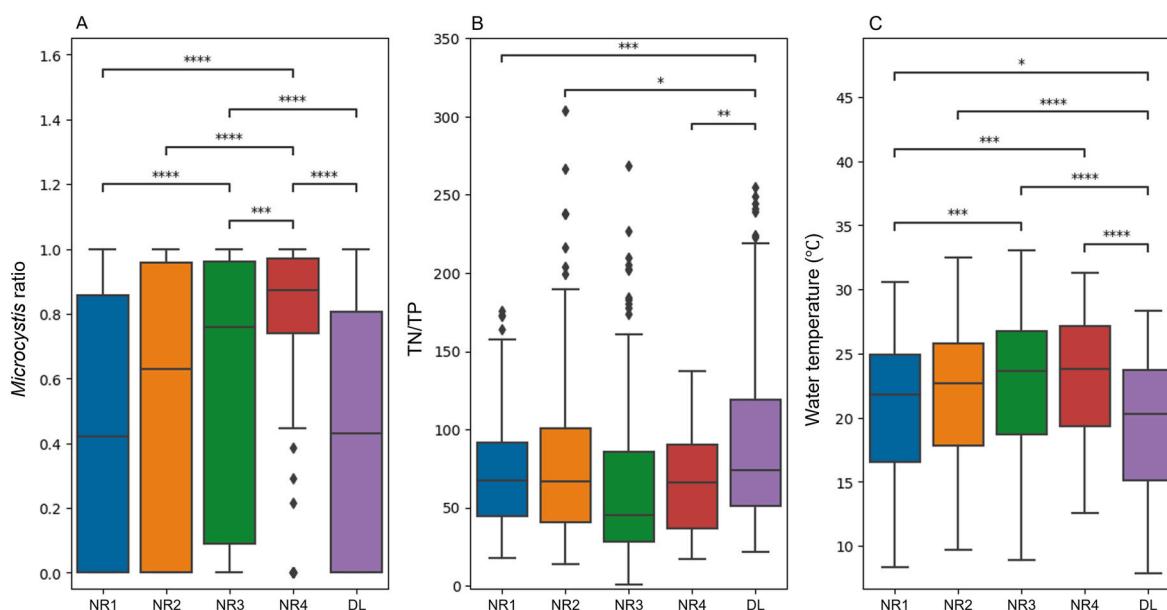
cyanobacteria for each site were analyzed with the highest-performing model, utilizing SHapley Additive exPlanations (SHAP), a technique from XAI (Lundberg and Lee, 2017; Li et al., 2022). Additionally, direct 1:1 relationships between subsets of data on TN/TP and occurrences of toxic-producing cyanobacteria were examined. To better understand the variation in the levels of TN and TP in relation to hydrological factors, a two-way Analysis of Variance (ANOVA) was conducted. Data pre-processing, the development of DDMs, and subsequent analyses were implemented using the Python programming language.

## 3. Results

### 3.1. Regional difference of CHABs in Nakdong River and Daecheong Lake

Community compositions of cyanobacteria genera in the NR differed across the four sites (Fig. 2a, S1). Overall, the ratio of *Microcystis* to total cyanobacteria was relatively lower in the upper region (NR1 and NR2) and higher in the downstream (NR3 and NR4). When caution levels occurred (cyanobacteria  $\geq 1000$  cells/mL) at NR1, the most upstream site, the ratio of *Microcystis* and frequency of *Microcystis*-dominance (7 out of 14 months) were relatively lower than other river sites. Among river sites, the largest proportions of *Dolichospermum* and *Aphanizomenon* were observed at NR1, followed by NR2 (Fig. S1) relative to the other river sites.

At NR1 and NR2, the largest increases in cyanobacteria cell density corresponded to an increase in *Dolichospermum* in 2022 and *Aphanizomenon* in 2017. Additionally, the ratio of *Microcystis* seemed to be smaller when the CHABs were more severe. In contrast, the density of cyanobacteria cells at NR3 and NR4 corresponded with *Microcystis* dominance. *Microcystis* was dominant when CHABs occurred at these downstream sites, except for three months. Particularly, most events when cyanobacteria densities were  $\geq 50,000$  cells/mL, these CHABs had a ratio of *Microcystis* of over 90 %. Interestingly, the most severe bloom event occurred in 2022, and while dominated ( $>90\%$ ) by *Dolichospermum* at the NR1 site, the relative dominance shifted downstream to nearly 100 % *Microcystis* at NR4.



**Fig. 2.** Boxplots of (a) the *Microcystis*-to-total cyanobacteria ratio, (b) the total nitrogen-to-total phosphorus ratio by mass (outliers at NR3 > 1200 omitted), and (c) water temperature at four sites on the Nakdong River (NR1–4) and at Daecheong Lake (DL). Annotated asterisks indicate statistically significant differences according to p-values obtained from two-sample t-tests (\*\*\*\*  $\leq 0.0001$ , 0.0001  $<$  \*\*\*  $\leq 0.001$ , 0.001  $<$  \*\*  $\leq 0.01$ , 0.01  $<$  \*  $\leq 0.05$ ; p-values larger than 0.05 are not shown).

### 3.2. Relationships between cyanobacteria, geosmin, MIB, and MCs

Concentrations of geosmin, MIB, and MC-LR at the study sites are presented in Fig. S2. Sites NR3 and NR4 show higher concentrations of T&O compounds and MC-LR and frequency of their detections, whereas sites NR1 and DL had lower concentrations and frequency. At NR1, where *Dolichospermum* dominated during bloom events, MC-LR was at very low concentrations and the concentrations of MIB were frequently higher than that of geosmin. At other sites, geosmin concentrations were significantly higher than MIB concentrations. At NR2, geosmin was abundant during *Aphanizomenon*-dominating bloom events in 2017, and geosmin and MC-LR concentrations increased in bloom events in 2022 when relative abundances of *Microcystis* and *Dolichospermum* were higher. At NR3 and NR4, concentrations of T&O compounds and MC-LR were significantly higher than previous sites, with larger density and dominance of *Microcystis*. Events with *Dolichospermum* presence exhibited higher geosmin concentrations (June–July 2022 at NR3 and NR4).

Correlation analysis revealed that MC-LR was highly correlated with *Microcystis* and total cyanobacteria. MC-LR was weakly correlated with *Planktothrix* and geosmin, and it had no relationships with *Dolichospermum*, *Aphanizomenon*, and MIB (Fig. S3). Moreover, geosmin was moderately correlated with log-transformed *Dolichospermum* and weakly correlated with log-transformed *Microcystis*. Further, MIB was moderately correlated with log-transformed *Microcystis*, and its correlations with other variables were very weak.

Taking the four cyanobacteria genera as independent variables, MC-LR was accurately predicted by the DDMs (Table 1). However, geosmin and MIB could not be accurately predicted (test NSE <0.6). Maximum predictive performances (test NSE) for predicting geosmin and MIB were 0.497 (DT) and 0.226 (MLR), respectively. The MLR results for predicting MC-LR (Table 2) demonstrated that *Microcystis* was the most significant variable (*p*-value <0.001), followed by *Planktothrix*, whereas *Dolichospermum* and *Aphanizomenon* were less significant.

For predicting geosmin, *Dolichospermum* and *Microcystis* were the dominant variables (by DT) among cyanobacteria genera. MLR demonstrated that *Dolichospermum* and *Aphanizomenon* were suitable for predicting MIB, but the model was not reliable (test NSE 0.226).

Additional data collected in 2024 including six congeners of MC, geosmin, and MIB were analyzed. Results of generalized linear models indicated *Microcystis* was significant for the productions of all MC congeners and MIB (Table S1). Geosmin concentrations were significantly influenced by *Dolichospermum* and *Planktothrix*. *Dolichospermum* exhibited significant negative relationships with MC-LR, MC-RR, MC-YR, and MIB, which may be due to the suppression of their growth by *Microcystis* or a very small density of *Dolichospermum*. Concentrations of MC-RR were highest among the six MC congeners, followed by MC-LR,

-YR, -LY, and -LF (Fig. S4). Measurements of MC-LA were all below the limit of quantification (0.05 µg/L). The proportions of MC congeners did not exhibit significant differences throughout the study sites.

### 3.3. Predicting *Microcystis* and examining major factors for *Microcystis*

Because MC-LR can be accurately predicted by *Microcystis*, as shown in the previous section, it is reasonable to predict *Microcystis* and examine the major factors affecting its increase. Transparency, turbidity, TN, TP, BOD, DTN, and DTP were omitted as independent variables to minimize multicollinearity in data-driven modeling. For example, TN and DTN were strongly correlated with NO<sub>3</sub>-N (*r* > 0.95) as well as TP and DTP with PO<sub>4</sub>-P (*r* > 0.9 for NR and *r* > 0.5 for DL). The performances of DDMs predicting cyanobacteria and *Microcystis* using various water quality indicators and precipitation data are presented in Table 3. Generally, RF had the best performance, whereas DT did not achieve the best performance in any of the sites. All sites had very good training performance, and most sites (NR2–4, DL1) had satisfactory test performance (NSE >0.5) for predicting cyanobacteria and *Microcystis*. Test performance values were always lower than the training performance values. SVM and RF models often exhibited better performance than MLR and DT models. In DL, RF achieved the highest performance of 0.522 (test NSE) for predicting *Microcystis* at DL1.

Incorporating NH<sub>3</sub>-N/NO<sub>3</sub>-N and NH<sub>3</sub>-N/PO<sub>4</sub>-P as independent variables for the models predicting *Microcystis* at NR1 and NR2 resulted in a slight increase in performance, indicating that these nutrient forms had an important effect on the proliferation of *Microcystis* upstream. The importance of these nutrient indicators was also highlighted in previous studies (Liu et al., 2011; Kim et al., 2017). Incorporating NH<sub>3</sub>-N/NO<sub>3</sub>-N and NH<sub>3</sub>-N/PO<sub>4</sub>-P as independent variables in the models for the other three sites, NR3, NR4 (the downstream regions of NR), and DL1, did not have a positive effect on the predictive performance. The performance of the models developed for DL3 and DL2 was lower than that for models for other sites. This could be due to the relatively small dataset size for multiple independent variables.

Important variables affecting the ratio of *Microcystis* are illustrated in Fig. 3. Generally, the number of cyanobacteria and WT were critical in all sites. WT was positively correlated to *Microcystis* for the models in NR1–3 and DL (Fig. 3a–c, e). Other critical independent variables were nutrients. Specifically, the nitrogen to phosphorus (N/P) ratios were major factors for all of the sites. Lower values of NH<sub>3</sub>-N/PO<sub>4</sub>-P and TN/TP increased the model output (*Microcystis* ratio) at NR1 and NR2 (Fig. 3a, b). NH<sub>3</sub>-N/PO<sub>4</sub>-P was as important as TN/TP at these two upstream sites. At NR3, TN/TP generally had a negative correlation with the *Microcystis* ratio. However, several events with high TN/TP had a positive effect on the *Microcystis* ratio, and some with low TN/TP had a negative effect (i.e., several events exhibited a positive correlation;

**Table 1**

Performance of the models for predicting MC-LR, geosmin, and 2-methylisoborneal (MIB) using the four genera of cyanobacteria as independent variables. DT, decision tree; MLR, multiple linear regression; NSE, Nash-Sutcliffe Efficiency; SVM, support vector machine.

Area	Dependent variable <sup>a</sup>	Independent variables	Train NSE			Test NSE		
			MLR	DT	SVM	MLR	DT	SVM
Nakdong River	MC-LR	Four genera of cyanobacteria (cells/mL)	0.840	0.895	0.838	0.826	0.636	0.837
	Geosmin		0.088	0.782	0.321	0.017	0.399	0.298
	MIB		0.029	0.338	0.389	0.001	0.045	0.177
	MC-LR	Four genera of cyanobacteria (Log <sub>2</sub> cells/mL)	0.257	0.895	0.404	0.168	0.578	0.422
	Geosmin		0.293	0.883	0.382	0.256	0.497	0.419
	MIB		0.126	0.451	0.380	0.081	0.162	0.214
Daecheong Lake	MC-LR	Four genera of cyanobacteria (cells/mL)	0.961	1.0	0.957	0.954	0.859	0.907
	Geosmin		0.325	0.823	0.279	0.283	0.105	0.267
	MIB		0.285	0.857	0.295	0.226	0.129	0.167
	MC-LR	Four genera of cyanobacteria (Log <sub>2</sub> cells/mL)	0.575	1.0	0.584	0.341	0.859	0.413
	Geosmin		0.315	0.823	0.206	0.273	0.105	0.173
	MIB		0.159	0.857	0.532	0.101	0.121	0.145

<sup>a</sup> MC-LR: µg/L, geosmin and MIB: ng/L.

**Table 2**

Results of multiple linear regression for predicting MC-LR ( $\mu\text{g/L}$ ) at the Nakdong River (train NSE = 0.840, test NSE = 0.826) and Daecheong Lake (train NSE = 0.961, test NSE = 0.954). DL, Daecheong Lake; NR, Nakdong River; RC, regression coefficient.

Area	Independent variables (cells/mL)				
	<i>Microcystis</i>	<i>Dolichospermum</i>	<i>Planktothrix</i>	<i>Aphanizomenon</i>	Intercept
NR	RC	0.00001696	0.00000384	0.0002000	-0.00000311
	P-value	0.000	0.259	0.056	0.381
DL	RC	0.00001911	0.00000756	-0.0002362	-0.00001649
	P-value	0.000	0.857	0.304	0.0039
					0.243

**Table 3**

Performance of data-driven models for predicting cyanobacteria (log cells/mL), *Microcystis* (log cells/mL), and the *Microcystis*-to-total cyanobacteria ratio (target A, B, and C, respectively). The highest test performance for each target in a site is in bold.

Target-site <sup>a</sup>	Dataset size (data points)	Train NSE				Test NSE			
		MLR	DT	SVM	RF	MLR	DT	SVM	RF
A-NR1	159	0.354	0.619	0.682	0.905	0.191	0.191	0.282	<b>0.349</b>
B-NR1	101	0.414	0.734	0.390	0.911	0.038	0.122	0.198	<b>0.284</b>
C-NR1	159	0.342	0.585	0.404	0.881	0.176	<0	0.266	0.223
C-NR1 <sup>b</sup>	159	0.422	0.454	0.818	0.883	0.275	<0	<b>0.306</b>	0.202
A-NR2	201	0.498	0.698	0.524	0.931	0.426	0.404	0.483	<b>0.514</b>
B-NR2	138	0.492	0.735	0.466	0.929	0.360	0.168	0.410	<b>0.447</b>
C-NR2	201	0.436	0.724	0.840	0.927	0.390	0.153	0.440	0.472
C-NR2 <sup>b</sup>	201	0.505	0.708	0.883	0.926	0.419	0.155	0.442	<b>0.488</b>
A-NR3	199	0.596	0.755	0.777	0.936	0.547	0.357	<b>0.591</b>	0.555
B-NR3	143	0.619	0.863	0.620	0.935	0.553	0.334	<b>0.577</b>	0.542
C-NR3	199	0.522	0.690	0.827	0.933	0.413	0.281	0.480	<b>0.517</b>
C-NR3 <sup>b</sup>	199	0.526	0.743	0.841	0.932	0.399	0.276	0.438	0.506
A-NR4	97	0.705	0.823	0.699	0.953	0.588	0.376	0.636	<b>0.656</b>
B-NR4	86	0.744	0.690	0.732	0.948	0.626	0.452	<b>0.652</b>	0.625
C-NR4	97	0.486	0.593	0.451	0.889	0.303	0.152	<b>0.329</b>	0.171
C-NR4 <sup>b</sup>	97	0.507	0.593	0.407	0.887	0.302	0.108	0.318	0.190
A-DL1	243	0.580	0.630	0.804	0.941	0.528	0.281	0.537	<b>0.572</b>
B-DL1	133	0.336	0.501	0.365	0.899	0.136	0.168	0.207	<b>0.271</b>
C-DL1	243	0.557	0.720	0.595	0.931	0.497	0.464	0.496	<b>0.522</b>
C-DL1 <sup>b</sup>	243	0.562	0.797	0.617	0.932	0.491	0.490	0.504	0.519
C-DL3	49	0.571	0.741	0.434	0.874	0.058	0.163	<b>0.279</b>	0.199
C-DL2	37	0.596	0.909	0.522	0.899	<0	<0	0.183	<b>0.221</b>

<sup>a</sup> Independent variables for the Nakdong River (NR): Cyanobacteria (only for target C), WT, pH, DO, COD, EC,  $\text{NH}_3\text{-N}$ ,  $\text{NO}_3\text{-N}$ ,  $\text{PO}_4^3-\text{P}$ , TN/TP, 7-day average rainfall, SS, and Chl-a; and Daecheong Lake (DL): TP along with the ones for the NR.

<sup>b</sup>  $\text{NH}_3\text{-N}/\text{NO}_3\text{-N}$  and  $\text{NH}_3\text{-N}/\text{PO}_4^3-\text{P}$  added as independent variables.

Fig. 3c). In contrast, the model for DL indicated that higher nitrogen and lower phosphorus levels were associated with a higher *Microcystis* ratio (Fig. 3e).

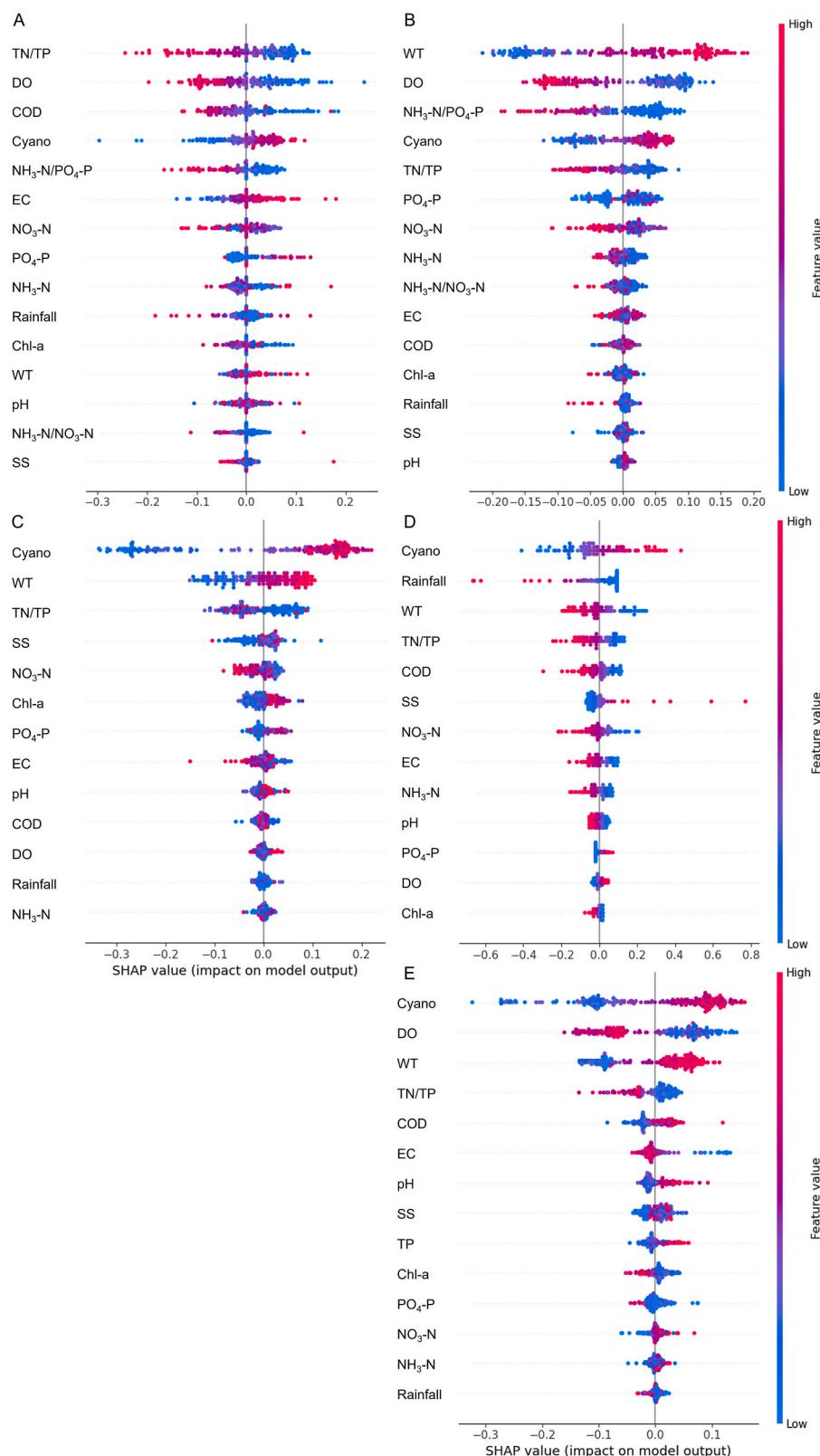
### 3.4. Investigating variations in nitrogen and phosphorus

Since nutrient variables were identified as important factors from the DDM interpretations and the directionality of their correlations varied across study sites, conflicting with the theory that nitrogen-rich environments promote microcystin production, more in-depth investigation of the variations in nitrogen and phosphorus at all locations is necessary. Data on TN/TP in Korean freshwaters exhibit clear seasonal patterns (Fig. S5). Specifically, TN/TP tended to decrease from April to July when the WT increased and increased from August to November when the WT decreased. Because TN/TP and WT were negatively correlated ( $r = -0.5$ ,  $p\text{-value} < 0.001$ ), the relationship between TN/TP and *Microcystis* was examined in the datasets with similar WT levels in the two sites (NR2 and NR3), where data were abundant, and the reliability of the models was sufficient. The negative correlation between *Microcystis* and TN/TP was significant when the WT  $\geq 30^\circ\text{C}$  (Fig. S6a). The correlations when the WT  $< 30^\circ\text{C}$  were not significant (Fig. S6b-h). The correlation coefficients between *Microcystis* and TN/TP were not significant despite limited levels of TN/TP (Fig. S7). The negative correlation when TN/TP was slightly high ( $47 > \text{TN/TP} \geq 40$ ) was relatively significant (Fig. S7d).

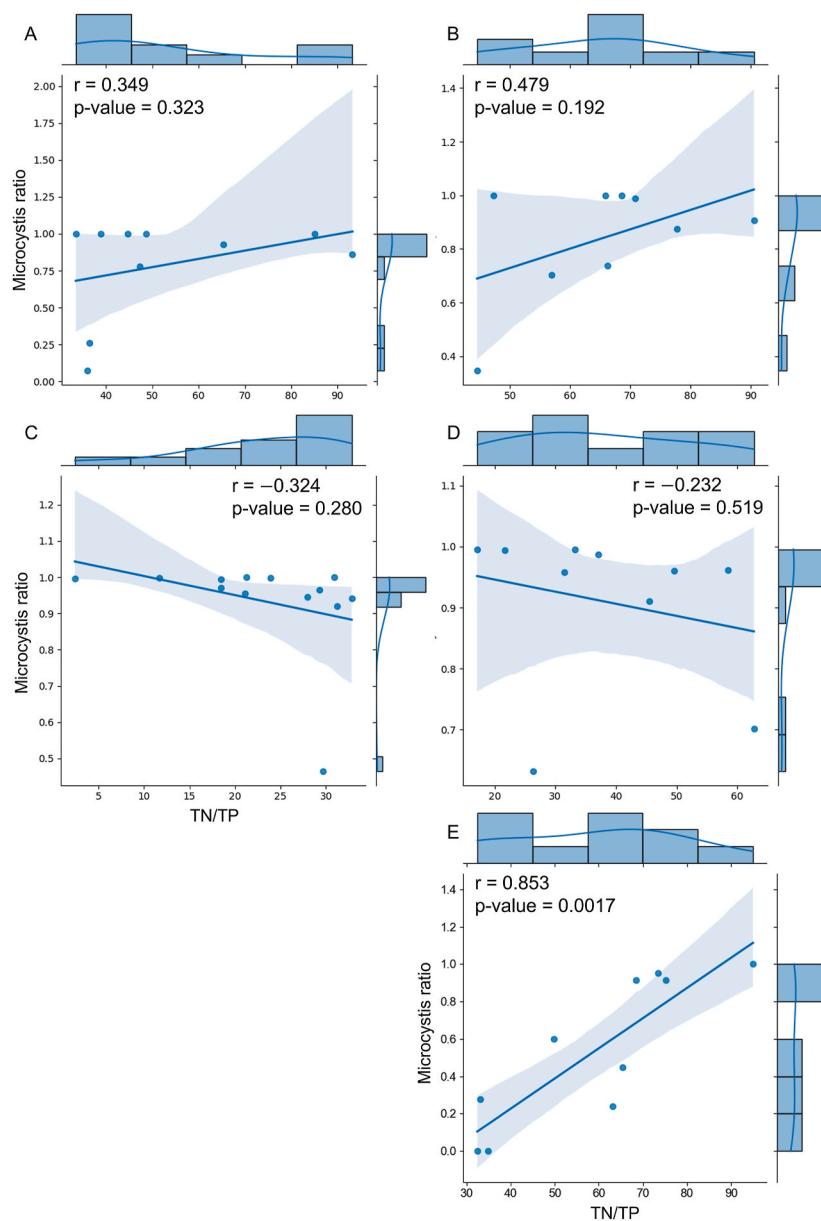
Correlations between TN/TP and the *Microcystis* ratio when WT was

high were analyzed at the study sites (Fig. 4). A significant correlation was only found at DL, where it was a positive and very strong. The correlations for NR sites tended to be positive at upstream sites and negative for downstream sites. However, the correlations were not significant for NR sites. When WT was low, the TN/TP was generally negatively correlated with *Microcystis* ratio at DL, which was similar to NR. These results suggest that a higher nitrogen-to-phosphorus ratio contributes to a larger proportion of *Microcystis* among the total cyanobacteria, particularly in stable waters such as lakes, rather than flowing waters. Furthermore, they indicate that the relative proliferation of *Microcystis* in flowing waters is attributed to a combination of multiple factors rather than solely nutrients.

To further confirm the effect of seasonality on TN and TP levels, a two-way ANOVA was performed using WT and rainfall as variables, testing their significance on TN and TP fluctuations (Fig. 5). The results indicate that TN levels were not impacted by rainfall; however, relatively low WT was significantly associated with elevated TN across all sites. Additionally, no WT-rainfall interaction effects were observed at all sites. In contrast, increased rainfall significantly affected the rise in phosphorus input within the NR. TP was also significantly positively correlated with WT. Moreover, the  $p$ -values for rainfall were generally lower than those for WT during TP analysis.



**Fig. 3.** Impact of independent variables on predicting dominance of *Microcystis* at (a) NR1 (SVM; test NSE = 0.306), (b) NR2 (RF; test NSE = 0.488), (c) NR3 (RF; test NSE = 0.517), (d) NR4 (SVM; test NSE = 0.329), and (e) DL1 (RF; test NSE = 0.522). Variables are ordered by importance, from most important (top) to least important (bottom).



**Fig. 4.** Scatter plots of the ratio of *Microcystis* versus the total nitrogen-to-phosphorus ratio (TN/TP) during high-water temperature events at (a) NR1 ( $\geq 26.8^{\circ}\text{C}$ ), (b) NR2 ( $\geq 29.7^{\circ}\text{C}$ ), (c) NR3 ( $\geq 30.0^{\circ}\text{C}$ ), (d) NR4 ( $\geq 29.0^{\circ}\text{C}$ ), and (e) DL1 ( $\geq 26.5^{\circ}\text{C}$ ).

#### 4. Discussion

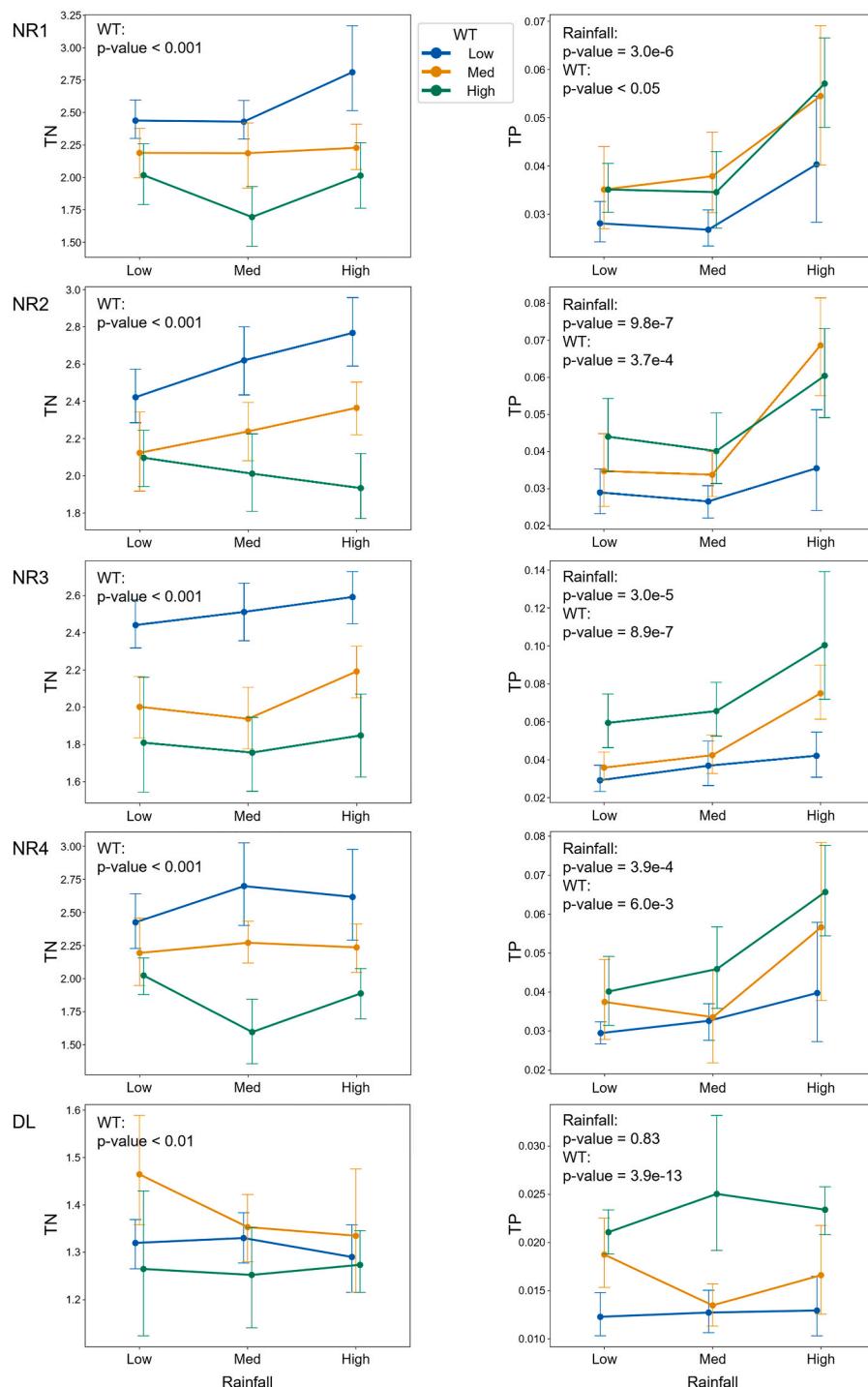
##### 4.1. Temperature-driven regional shifts in cyanobacterial communities

The difference in dominance of *Microcystis* tended to be associated with the frequency and magnitude of CHABs. The number of cyanobacteria increased, and CHAB occurred frequently along the downstream reaches of the NR because downstream (southern) regions have warmer climates than upstream (northern) regions. Specifically, the WT tended to be higher in the downstream locations (Fig. 2c), and the mean annual air temperature (from 2016 to 2022) at NR1 was  $13.3^{\circ}\text{C}$ , NR2:  $13.3^{\circ}\text{C}$ , NR3:  $13.4^{\circ}\text{C}$ , and NR4:  $15.2^{\circ}\text{C}$ . The higher dominance of *Microcystis* in the NR downstream is similar to the proliferation of *Aphanizomenon* in the Klamath River, USA, upstream, whereas *Microcystis* persists downstream after the river water flows through dams (Otten et al., 2015; Paerl et al., 2018). Similarly, *Aphanizomenon* abundance was greater in cooler NR1 and NR2 waters, periodically dominating in June in NR1–NR3 rather than warmer August months,

reaffirming other studies highlighting temperature driven succession from *Aphanizomenon* to *Microcystis* (Wu et al., 2016; Wen et al., 2022). The ratio of *Microcystis* in DL was similar to that upstream of NR (Fig. 2a). The N/P ratios in DL were higher, and WT was lower than those in NR (Fig. 2b, c). The N/P ratios did not differ significantly at the sites in NR. This suggests that the varying dominance of *Microcystis* in different regions was mainly due to the WT rather than the effect of the N/P ratio. Local authorities in the upstream regions should anticipate a potential increase in the *Microcystis* fraction due to rising temperature associated with climate change. Other models examining South Korean waterways have made similar temperature dependent predictions (Kim et al., 2024, 2026).

##### 4.2. Cyanobacteria-associated production of geosmin, MIB, and MCs

The stronger correlations of geosmin and MIB with log-transformed cyanobacteria suggest that T&O compounds may not increase linearly at high levels of cyanobacteria. The relationship between geosmin and



**Fig. 5.** Total nitrogen (TN) and total phosphorus (TP) (mg/L) at the study sites in response to different rainfall (mm, mean of 0–7 days ahead) and water temperature (WT, °C) levels (low <35th percentile ≤ med ≤ 65th percentile < high). P-values were obtained from ANOVA. No interaction effects were significant (i.e., p-values >0.05 for interaction effects of rainfall and WT on TN and TP).

*Dolichospermum* abundance has been observed in the U.S.; however, the geosmin correlations were stronger when measuring the *geoA* gene (Otten et al., 2016). MIB had no relationship with geosmin, which is corroborated by the findings of Hooper et al. (2023).

The MC-LR prediction model indicated that freshwater with 1300–1700 *Microcystis* cells/mL can produce 0.03 µg/L of MC-LR (a linear relationship), the recommended notification level (OEHHA, 2022) for levels that could adversely impact human health (Chen et al., 2011). This cell number is higher than the values reported in other studies, where 630 *Microcystis* cells/mL and 1000 *Microcystis* cells/mL

were reported to represent 0.3 MCs µg/L (Lu et al., 2019) and 0.44 MC-LR µg/L (Otten et al., 2015), respectively. SVM also predicted higher MC-LR when *Microcystis* was higher, but the values associated with other genera were negligible. These results are consistent with those in other studies (Sabart et al., 2010; Vasconcelos et al., 2011) that found MCs and *Microcystis* to be strongly correlated. This study quantitatively characterized the relationship between MC-LR and microscopically observed cyanobacteria genera with very high goodness of fit and model significance for surface waters for the first time. The findings imply that MCs can be minimized by regulating harmful cyanobacteria cells (Terin and

Sabogal-Paz, 2019).

The weak correlations between cyanobacteria genera and geosmin and MIB imply that environmental conditions, along with the quantity of cyanobacteria and other organisms, may influence these T&O compounds. A previous study identified *Dolichospermum*, *Planktothrix*, and COD as key predictors for geosmin, while *Pseudanabaena* and COD were significant for MIB (Qiu et al., 2023).

#### 4.3. Predicting *Microcystis*: Insights into influential environmental variables from DDMs

Cyanobacteria-predicting models had higher performance than *Microcystis*-predicting models. This suggests that the mechanism of *Microcystis*-bloom events may be more complicated than that of cyanobacterial blooms. The models for most sites exhibited positive correlation between WT and *Microcystis*. However, the model for NR4 indicated that lower temperature increased the ratio of *Microcystis* (Fig. 3d). This aberration in the behavior of the model for NR4 is speculated to be due to the dataset. Compared to other sites, NR4 had a small dataset size (i.e., 97 data points collected during 4 years), and the models exhibited unsatisfactory performance (test NSE  $\leq 0.5$ ). Moreover, it had numerous cyanobacteria during a year (i.e., a maximum of 450,000 cells/mL), which was unprecedented at all the sites. Therefore, the data may not have been enough to represent the volatility of *Microcystis* growth.

Higher TN/TP and/or NH<sub>3</sub>-N/PO<sub>4</sub>-P ratios contributed to a smaller proportion of *Microcystis* for the models at all sites. This is in line with previous study that MC levels were negatively correlated with TN/TP levels within 20–100 (Yu et al., 2014). In addition, models with satisfactory performance (test NSE  $\geq 0.5$ ) at NR2 and NR3 predicted the ratio of *Microcystis* to be higher when NH<sub>3</sub>-N was lower and PO<sub>4</sub>-P is higher (Fig. 3b, c). On the other hand, higher nitrogen and lower phosphorus were associated with a higher ratio of *Microcystis* at DL (Fig. 3e). These imply that different mitigation strategies should be applied at NR and DL to lower toxicity during CHAB occurrences.

The quantitative variable importance results indicated that COD was more closely related to *Microcystis* than to total cyanobacteria. This aligns with previous studies, which demonstrated relationships between water quality and *Microcystis* (Mori et al., 2022) or MCs (He et al., 2021). The positive correlation between COD and *Microcystis* may be attributed to the positive correlations among WT, TP, rainfall, and COD. Specifically, phosphorus inputs into the photic zone of the water column may be elevated in summer due to intense precipitation and elevated phosphorus-rich inflows from polluted tributaries. The increased inflows carry organic matter and have been implicated in destabilizing summer stratification in the NR (Park et al., 2021). The summer stratified NR (Jung et al., 2023), like eutrophic lakes, can also release readily bioavailable soluble reactive phosphorus from the hypolimnion and bottom sediments (Withers and Jarvie, 2008), of which destratification and releases have been linked to Korea's heavy summer rains (Tekile et al., 2015). In contrast, nitrogen indicators had a negative relationship with WT, and the models revealed a negative correlation between nitrogen indicators and cyanobacteria. No consistent trends in terms of differences in variable importance were observed except for COD in all sites. Although the order of importance varied between the models for predicting cyanobacteria and *Microcystis*, the impacts of the variables were not inverse. Namely, variables that positively affected cyanobacteria density also positively influenced *Microcystis* density, and vice versa.

#### 4.4. Nutrient dynamics and their implications for toxic bloom management

A positive correlation between TN/TP and *Microcystis* could not be observed even at approximately TN/TP = 7.2 (Fig. S7b), the Redfield nitrogen-to-phosphorus ratio on a mass basis, which is preferable for the growth of cyanobacteria. These results suggest that a phosphorus-

reducing nutrient management plan for surface water can help in decreasing the toxicity of cyanobacteria-rich water when it is difficult to implement nitrogen mitigation.

Previous studies reported that higher N/P induces more toxic-cyanobacteria-rich water (Gobler et al., 2016; Hellweger et al., 2022; Huisman et al., 2022) because *Microcystis* is a non-nitrogen-fixing cyanobacteria, whereas *Aphanizomenon* and *Dolichospermum* are nitrogen-fixers (Paerl et al., 2018). The results of the association between N/P and *Microcystis* may be different because the samples were acquired from the field in this study, and environmental conditions (e.g., water quality and temperature) were not controlled. Thus, the effect of sole N/P was not clearly derived in the models. On the other hand, the N/P ratios at the sites in this study (mean TN/TP at NR1: 70.9, NR2: 77.8, NR3: 63.1, NR4: 64.6, DL1: 109.2) were higher than those of surface waters in other countries (Lehman et al., 2008; Prater et al., 2017). Previous studies have reported that the content of MCs in *Microcystis* was higher, with nitrogen peaking at N/P = 16 and then decreasing (Lee et al., 2000). Similar to the results of this study, incidents where N/P was negatively correlated with *Microcystis* have been reported (Lehman et al., 2008; Liu et al., 2011; Schindler et al., 2016). Moreover, Hooper et al. (2023) reported that a N/P ratio of over 20 has a negative effect on the genes producing geosmin and MIB.

The results analyzing the effect of WT and rainfall on TN and TP (Fig. 5) imply that phosphorus primarily enters through non-point source pollution, especially during the rainy periods of the summer season. Nonetheless, rainfall did not significantly impact TP levels in DL. Therefore, the seasonal variance in TN/TP ratio is attributed to the heavy summer rainfall increasing phosphorus inputs into water bodies, as well as reduced TN levels in summer. Consequently, the negative correlation between *Microcystis* and TN/TP is primarily due to the seasonality of these nutrients. Additionally, the N-enriched ( $>0.8$  mg/L) compared to P threshold level for CHAB (0.05 mg/L) (Xu et al., 2014, 2017) in the study sites results in a dependency on TP. These findings suggest that the proliferation of toxic cyanobacteria is stimulated by phosphorus inputs from non-point source pollution in NR, in contrast to DL where the body of water is surrounded by mountains and nearby pollution sources are minimal (Fig. 1). This underscores the importance of managing pollution sources, such as livestock manure in open storage yards (non-point) and wastewater treatment facilities (point source), to decrease the likelihood of CHABs and minimize MC-LR.

## 5. Conclusions

Our results revealed increasing *Microcystis* dominance downstream into the warmer NR region, despite lower nitrogen levels. MC-LR, MC-RR, MC-YR, MC-LY, and MIB were significantly influenced by *Microcystis* abundance at all locations, with MC-LR and MC-YR being the most affected. The MC-LR prediction model indicated that even the caution cyanobacterial alert level (1000 cells/mL) may pose human risks. The less accurate DDMs for predicting T&O compounds suggest that future studies should explore more complex mechanisms or additional variables. RF models successfully predicted *Microcystis* presence using environmental predictors, though SHAP analysis revealed that both N and P were not positively associated with *Microcystis* except at DL. Further analysis showed that the NR is susceptible to P input from runoff, contrary to DL. Therefore, MC level is dependent upon non-point source pollution (P) and temperature at N-rich NR. Our findings demonstrate that XAI tools are effective in characterizing water bodies and predicting toxicity levels using environmental data. Although nitrogen is essential in MC-LR production at the cellular level, its influence on MC-LR concentrations was not significant at the study sites, whereas phosphorus levels were significant. If management strategies successfully reduce TP from point and non-point sources, with disproportionately less success in reducing TN, then the impact of TN on MC-LR concentrations may become more significant, especially if community composition shifts. Region-specific nutrient management—particularly

controlling non-point sources—is crucial for mitigating cyanotoxin contamination. Our findings highlight the importance of understanding microbial community composition, rather than overall CHAB magnitude alone, to better protect ecological health and public safety in temperate water resources.

## CRediT authorship contribution statement

**Jayun Kim:** Writing – review & editing, Writing – original draft, Visualization, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Jiyoung Lee:** Writing – original draft, Validation, Project administration, Investigation. **Jason W. Marion:** Writing – review & editing, Validation, Resources. **Joonhong Park:** Supervision, Project administration, Funding acquisition.

## 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.

## Acknowledgments

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (No. 2018R1A6A1A08025348). We also thank the National Institute of Environmental Research for providing publicly accessible data.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jenvman.2025.126616>.

## Data availability

Data will be made available on request.

## References

- Ai, Y., Lee, S., Lee, J., 2020. Drinking water treatment residuals from Cyanobacteria bloom-affected areas: investigation of potential impact on agricultural land application. *Sci. Total Environ.* 706, 135756. <https://doi.org/10.1016/j.scitotenv.2019.135756>.
- Brandenburg, K., Siebers, L., Keuskamp, J., Jephcott, T.G., Van de Waal, D.B., 2020. Effects of nutrient limitation on the synthesis of N-rich phytoplankton toxins: a meta-analysis. *Toxins* 12 (4), 221. <https://doi.org/10.3390/toxins12040221>.
- Busari, I., Sahoo, D., Das, N., Privette, C., Schlautman, M., Sawyer, C., 2025. Advancing harmful algal bloom predictions using chlorophyll-a as an indicator: combining deep learning and EnKF data assimilation method. *J. Environ. Manag.* 382, 125441. <https://doi.org/10.1016/j.jenvman.2025.125441>.
- Canale, R.P., Vogel, A.H., 1974. Effects of temperature on phytoplankton growth. *J. Environ. Eng. Div. ASCE* 100, 229–241. <https://doi.org/10.1061/JEEGAV.0000151>.
- Capblancq, J., Catalan, J., 1994. Phytoplankton: which, and how much? In: Margalef, R. (Ed.), *Limnology Now: a Paradigm of Planetary Problems*. Elsevier Science B. V., Amsterdam, pp. 9–36.
- Carmichael, W.W., Boyer, G.L., 2016. Health impacts from Cyanobacteria harmful algae blooms: implications for the north American Great Lakes. *Harmful Algae* 54, 194–212. <https://doi.org/10.1016/j.hal.2016.02.002>.
- Chen, Y., Xu, J., Li, Y., Han, X., 2011. Decline of sperm quality and testicular function in male mice during chronic low-dose exposure to microcystin-LR. *Reprod. Toxicol.* 31 (4), 551–557. <https://doi.org/10.1016/j.reprotox.2011.02.006>.
- Cheung, M.Y., Liang, S., Lee, J., 2013. Toxin-producing Cyanobacteria in freshwater: a review of the problems, impact on drinking water safety, and efforts for protecting public health. *J. Microbiol.* 51, 1–10. <https://doi.org/10.1007/s12275-013-2549-3>.
- Du, X., Liu, H., Yuan, L., Wang, Y., Ma, Y., Wang, R., Chen, X., Losiewicz, M.D., Guo, H., Zhang, H., 2019. The diversity of cyanobacterial toxins on structural characterization, distribution and identification: a systematic review. *Toxins* 11 (9), 530. <https://doi.org/10.3390/toxins11090530>.
- Facciponte, D.N., Bough, M.W., Seidler, D., Carroll, J.L., Ashare, A., Andrew, A.S., Tsongalis, G.J., Vaickus, L.J., Henegan, P.L., Butt, T.H., Stommel, E.W., 2018. Identifying aerosolized Cyanobacteria in the human respiratory tract: a proposed mechanism for cyanotoxin-associated diseases. *Sci. Total Environ.* 645, 1003–1013. <https://doi.org/10.1016/j.scitotenv.2018.07.226>.
- Faithfull, C.L., Bergström, A.K., Vrede, T., 2011. Effects of nutrients and physical lake characteristics on bacterial and phytoplankton production: a meta-analysis. *Limnol. Oceanogr.* 56 (5), 1703–1713. <https://doi.org/10.4319/lo.2011.56.5.1703>.
- Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., Van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. *Harmful Algae* 54, 87–97. <https://doi.org/10.1016/j.hal.2016.01.010>.
- Guo, Y.C., Lee, A.K., Yates, R.S., Liang, S., Rochelle, P.A., 2017. Analysis of microcystins in drinking water by ELISA and LC/MS/MS. *J. Am. Water Works Assoc.* 109, 13–25. <https://doi.org/10.5942/jawwa.2017.109.0027>.
- Gupta, N., Pant, S.C., Vijayaraghavan, R., Rao, P.V.L., 2003. Comparative toxicity evaluation of cyanobacterial cyclic peptide toxin microcystin variants (LR, RR, YR) in mice. *Toxicology* 188 (2–3), 285–296. [https://doi.org/10.1016/S0300-483X\(03\)00112-4](https://doi.org/10.1016/S0300-483X(03)00112-4).
- He, X., Wang, H., Zhuang, W., Liang, D., Ao, Y., 2021. Risk prediction of microcystins based on water quality surrogates: a case study in a eutrophicated urban river network. *Environmen. Pollut.* 275, 116651. <https://doi.org/10.1016/j.envpol.2021.116651>.
- Heisler, J., Gilbert, P.M., Burkholder, J.M., Anderson, D.M., Cochlan, W., Dennison, W. C., Dorch, Q., Gobler, C.J., Heil, C.A., Humphries, E., Lewitus, A., Magnien, R., Marshall, H.G., Sellner, K., Stockwell, D.A., Stoercker, D.K., Sudleson, M., 2008. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae* 8 (1), 3–13. <https://doi.org/10.1016/j.hal.2008.08.006>.
- Hellweger, F.L., Martin, R.M., Eigemann, F., Smith, D.J., Dick, G.J., Wilhelm, S.W., 2022. Models predict planned phosphorus load reduction will make Lake Erie more toxic. *Science* 376, 1001–1005. <https://doi.org/10.1126/science.abm6791>.
- Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nat. Geosci.* 15 (2), 130–134. <https://doi.org/10.1038/s41561-021-00887-x>.
- Hooper, A.S., Kille, P., Watson, S.E., Christofides, S.R., Perkins, R.G., 2023. The importance of nutrient ratios in determining elevations in geosmin synthase (geoA) and 2-MIB cyclase (mic) resulting in taste and odour events. *Water Res.* 232, 119693. <https://doi.org/10.1016/j.watres.2023.119693>.
- Huisman, J., Dittmann, E., Fastner, J., Schuurmans, J.M., Scott, J.T., Van de Waal, D.B., Visser, P.M., Welker, M., Chorus, I., 2022. Comment on ‘models predict planned phosphorus load reduction will make Lake Erie more toxic’. *Science* 378. <https://doi.org/10.1126/science.add9959>.
- Hwan, B.J., Kim, H.N., Kang, T.G., Kim, B.-H., Byeon, M.-S., 2023. Study of the cause of the generation of odor compounds (geosmin and 2-methylisoborneol) in the Han River system, the drinking water source, Republic of Korea. *Water Supply* 23 (3), 1081–1093. <https://doi.org/10.2166/ws.2023.037>.
- IARC (International Agency for Research on Cancer), 2010. Ingested nitrate and nitrite, and cyanobacterial peptide toxins. Geneva: IARC (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 94).
- Jung, E., Joo, G.J., Kim, H.G., Kim, D.K., Kim, H.W., 2023. Effects of seasonal and diel variations in thermal stratification on phytoplankton in a regulated river. *Sustainability* 15 (23), 16330. <https://doi.org/10.3390/su152316330>.
- Kehoe, M.J., Chun, K.P., Baulch, H.M., 2015. Who smells? Forecasting taste and odor in a drinking water reservoir. *Environ. Sci. Technol.* 49 (18), 10984–10992. <https://pubs.acs.org/doi/10.1021/acs.est.5b00979>.
- Kim, J., Jang, G., Jo, M., Park, J., 2026. Simulation of harmful cyanobacterial blooms under weir operations and climate change scenarios: a data-driven study of the Nakdong River. *Environ. Eng. Res.* 31 (1), 250081. <https://doi.org/10.4491/eer.2025.081>.
- Kim, H., Jo, B.Y., Kim, H.S., 2017. Effect of different concentrations and ratios of ammonium, nitrate, and phosphate on growth of the blue-green alga (cyanobacterium) *Microcystis aeruginosa* isolated from the Nakdong River, Korea. *ALGAE* 32 (4), 275–284. <https://doi.org/10.4490/algae.2017.32.10.23>.
- Kim, H.G., Cha, Y., Cho, K.H., 2024. Projected climate change impact on cyanobacterial bloom phenology in temperate rivers based on temperature dependency. *Water Res.* 249, 120928. <https://doi.org/10.1016/j.watres.2023.120928>.
- Kim, J., Jones, J.R., Seo, D., 2021. Factors affecting harmful algal bloom occurrence in a river with regulated hydrology. *J. Hydrol.: Reg. Stud.* 33, 100769. <https://doi.org/10.1016/j.ejrh.2020.100769>.
- Kim, J., Jung, W., An, J., Oh, H.J., Park, J., 2023. Self-optimization of training dataset improves forecasting of cyanobacterial bloom by machine learning. *Sci. Total Environ.* 866, 161398. <https://doi.org/10.1016/j.scitotenv.2023.161398>.
- Lau, S.S.S., Lane, S.N., 2002. Nutrient and grazing factors in relation to phytoplankton level in a eutrophic shallow lake: the effect of low macrophyte abundance. *Water Res.* 36 (14), 3593–3601. [https://doi.org/10.1016/S0043-1354\(02\)00059-3](https://doi.org/10.1016/S0043-1354(02)00059-3).
- Lee, J., Lee, S., Jiang, X., 2017. Cyanobacterial toxins in freshwater and food: important sources of exposure to humans. *Annu. Rev. Food Sci. Technol.* 8 (1), 281–304. <https://doi.org/10.1146/annurev-food-030216-030116>.
- Lee, S., Kim, J., Lee, J., 2021. Colonization of toxic Cyanobacteria on the surface and inside of leafy green: a hidden source of cyanotoxin production and exposure. *Food Microbiol.* 94, 103655. <https://doi.org/10.1016/j.fm.2020.103655>.
- Lee, S.J., Jang, M.-H., Kim, H.-S., Yoon, B.-D., Oh, H.-M., 2000. Variation of microcystin content of *Microcystis aeruginosa* relative to medium N:P ratio and growth stage. *J. Appl. Microbiol.* 89 (2), 323–329. <https://doi.org/10.1046/j.1365-2672.2000.01112.x>.
- Lehman, P.W., Boyer, G., Satchwell, M., Waller, S., 2008. The influence of environmental conditions on the seasonal variation of *Microcystis* cell density and microcystins concentration in San Francisco Estuary. *Hydrobiologia* 600, 187–204. <https://doi.org/10.1007/s10750-007-9231-x>.

- Li, L., Qiao, J., Yu, G., Wang, L., Li, H.Y., Liao, C., Zhu, Z., 2022. Interpretable tree-based ensemble model for predicting beach water quality. *Water Res.* 211, 118078. <https://doi.org/10.1016/j.watres.2022.118078>.
- Liu, X., Lu, X., Chen, Y., 2011. The effects of temperature and nutrient ratios on *Microcystis* blooms in Lake Taihu, China: an 11-year investigation. *Harmful Algae* 10 (3), 337–343. <https://doi.org/10.1016/j.hal.2010.12.002>.
- Lu, K.-Y., Chiu, Y.-T., Burch, M., Senoro, D., Lin, T.-F., 2019. A molecular-based method to estimate the risk associated with cyanotoxins and odor compounds in drinking water sources. *Water Res.* 164 (1), 144938. <https://doi.org/10.1016/j.watres.2019.114938>.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30, 4768–4777.
- Metcalf, J.S., Codd, G.A., 2012. Cyanotoxins. In: Whitton, B. (Ed.), *Ecology of Cyanobacteria II*. Springer, Dordrecht. [https://doi.org/10.1007/978-94-007-3855-3\\_24](https://doi.org/10.1007/978-94-007-3855-3_24).
- Mori, M., Flores, R.G., Suzuki, Y., Nukazawa, K., Hiraoka, T., Nonaka, H., 2022. Prediction of *Microcystis* occurrences and analysis using machine learning in high-dimension, low-sample-size and imbalanced water quality data. *Harmful Algae* 117, 102273. <https://doi.org/10.1016/j.hal.2022.102273>.
- Mrdjen, I., Lee, J., Weghorst, C.M., Knobloch, T.J., 2022. Impact of cyanotoxin ingestion on liver cancer development using an at-risk two-staged model of mouse hepatocarcinogenesis. *Toxins* 14 (7), 484. <https://doi.org/10.3390/toxins14070484>.
- Nguyen, H.Q., Arias, M.E., Zhang, Q., Tarabih, O.M., Armstrong, C., Sun, D., Philips, E.J., 2025. A comparative analysis of data-driven modeling approaches to forecast Cyanobacteria algal blooms in eutrophic lake discharge canals. *J. Environ. Manag.* 380, 124834. <https://doi.org/10.1016/j.jenvman.2025.124834>.
- OEHHA (California Office of Environmental Health Hazard Assessment), 2022. Recommendation for short-term notification level for microcystins. <https://oehha.ca.gov/water/chemicals/cyanotoxins-microcystins>.
- Otten, T.G., Crosswell, J.R., Mackey, S., Dreher, T.W., 2015. Application of molecular tools for microbial source tracking and public health risk assessment of a *Microcystis* bloom traversing 300km of the Klamath River. *Harmful Algae* 46, 71–81. <https://doi.org/10.1016/j.hal.2015.05.007>.
- Otten, T.G., Graham, J.L., Harris, T.D., Dreher, T.W., 2016. Elucidation of taste-and odor-producing bacteria and toxicigenic Cyanobacteria in a midwestern drinking water supply reservoir by shotgun metagenomic analysis. *Appl. Environ. Microbiol.* 82 (17), 5410–5420. <https://doi.org/10.1128/AEM.01334-16>.
- Pael, H.W., Otten, T.G., Kudela, R., 2018. Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum. *Environ. Sci. Technol.* 52 (10), 5519–5529. <https://doi.org/10.1021/acs.est.7b05950>.
- Park, H.-K., Lee, H.-J., Heo, J., Yun, J.-H., Kim, Y.-J., Kim, H.-M., Hong, D.-G., Lee, I.-J., 2021. Deciphering the key factors determining spatio-temporal heterogeneity of cyanobacterial bloom dynamics in the Nakdong River with consecutive large weirs. *Sci. Total Environ.* 755 (Part 2), 143079. <https://doi.org/10.1016/j.scitotenv.2020.143079>.
- Park, H.K., Lee, H.J., Heo, J., Yun, J.H., Kim, Y.J., Kim, H.M., Hong, D.G., Lee, I.J., 2021. Deciphering the key factors determining spatio-temporal heterogeneity of cyanobacterial bloom dynamics in the Nakdong River with consecutive large weirs. *Sci. Total Environ.* 755, 143079. <https://doi.org/10.1016/j.scitotenv.2020.143079>.
- Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Sci. Total Environ.* 502, 31–41. <https://doi.org/10.1016/j.scitotenv.2014.09.005>.
- Perkins, R.G., Underwood, G.J.C., 2000. Gradients of chlorophyll a and water chemistry along an eutrophic reservoir with determination of the limiting nutrient by in situ nutrient addition. *Water Res.* 34 (3), 713–724. [www.elsevier.com/locate/watres](http://www.elsevier.com/locate/watres).
- Prater, C., Frost, P.C., Howell, E.T., Watson, S.B., Zastepa, A., King, S.S.E., Vogt, R.J., Xenopoulos, M.A., 2017. Variation in particulate C : n : p stoichiometry across the Lake Erie watershed from tributaries to its outflow. *Limnol. Oceanogr.* 62, S194–S206. <https://doi.org/10.1002/lno.10628>.
- Qiu, P., Zhang, Y., Mi, W., Song, G., Bi, Y., 2023. Producers and drivers of odor compounds in a large drinking-water source. *Front. Ecol. Evol.* 11, 1216567. <https://doi.org/10.3389/fevo.2023.1216567>.
- Roussou, B.Z., Bertone, E., Stewart, R., Hamilton, D.P., 2020. A systematic literature review of forecasting and predictive models for Cyanobacteria blooms in freshwater lakes. *Water Res.* 182, 115959. <https://doi.org/10.1016/j.watres.2020.115959>.
- Sabart, M., Pobel, D., Briand, E., Combourieu, B., Salencon, M.J., Humbert, J.F., Latour, D., 2010. Spatiotemporal variations in microcystin concentrations and in the proportions of microcyst-in-producing cells in several *Microcystis aeruginosa* populations. *Appl. Environ. Microbiol.* 76 (14), 4750–4759. <https://doi.org/10.1128/AEM.02531-09>.
- Schindler, D.W., Carpenter, S.R., Chapra, S.C., Hecky, R.E., Orihel, D.M., 2016. Reducing Phosphorus to curb Lake eutrophication is a success. *Environ. Sci. Technol.* 50 (17), 8923–8929. <https://doi.org/10.1021/acs.est.6b02204>.
- Shin, H.-R., 2010. *Comprehensive Analysis of the 2011 Budget Plan, Budget Plan Analysis Series 8* (in Korean). Korean National Assembly Budget Office, Seoul, South Korea.
- Shoemaker, J., Tettenhorst, D., Delacruz, A., 2015. Method 544. Determination of Microcystins and Nodularin in Drinking Water by Solid Phase Extraction and Liquid Chromatography/tandem Mass Spectrometry (LC/MS/MS). U.S. Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, OH, USA.
- Srinivasan, R., Sorial, G.A., 2011. Treatment of taste and odor causing compounds 2-methyl isoborneol and geosmin in drinking water: a critical review. *Journal of Environmental Sciences* 23 (1), 1–13. [https://doi.org/10.1016/S1001-0742\(10\)60367-1](https://doi.org/10.1016/S1001-0742(10)60367-1).
- Svircev, Z., Lalić, D., Bojadžija Savić, G., Tokodi, N., Backović, D.D., Chen, L., Meriluoto, J., Codd, G.A., 2019. Global geographical and historical overview of cyanotoxin distribution and cyanobacterial poisonings. *Arch. Toxicol.* 93, 2429–2481. <https://doi.org/10.1007/s00204-019-02524-4>.
- Taranu, Z.E., Zurawell, R.W., Pick, F., Gregory-Eaves, I., 2012. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. *Glob. Change Biol.* 18 (12), 3477–3490. <https://doi.org/10.1111/geb.12015>.
- Tekile, A., Kim, I., Kim, J., 2015. Mini-review on river eutrophication and bottom improvement techniques, with special emphasis on the Nakdong River. *J. Environ. Sci.* 30, 113–121. <https://doi.org/10.1016/j.jes.2014.10.014>.
- Terin, U.C., Sabogal-Paz, L.P., 2019. *Microcystis aeruginosa* and microcystin-LR removal by household slow sand filters operating in continuous and intermittent flows. *Water Res.* 150, 29–39. <https://doi.org/10.1016/j.watres.2018.11.055>.
- Vasconcelos, V.M., Moraisa, J., Vale, M., 2011. Microcysts and Cyanobacteria trends in a 14 year monitoring of a temperate eutrophic reservoir (Aguieira Portugal). *J. Environ. Monit.* 13, 668–672. <https://doi.org/10.1039/COEM00671H>.
- Wang, C., Wang, Q., Ben, W., Qiao, M., Ma, B., Bai, Y., Qu, J., 2024. Machine learning predicts the growth of cyanobacterial genera in river systems and reveals their different environmental responses. *Sci. Total Environ.* 946, 174383. <https://doi.org/10.1016/j.scitotenv.2024.174383>.
- Wen, Q., Xiao, P., Li, H., Li, W., Yu, G., Li, R., 2022. Succession of *Aphanizomenon flos-aquae* and *Microcystis aeruginosa* in direct co-culture experiments at different temperatures and biomasses. *Journal of Oceanology and Limnology* 40 (5), 1819–1828. <https://doi.org/10.1007/s00343-022-2041-1>.
- Wilhelm, S.W., Farnsley, S.E., LeCleir, G.R., Layton, A.C., Satchwell, M.F., DeBruyn, J.M., Boyer, G.L., Zhu, G., Paerl, H.W., 2011. The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae* 10 (2), 207–215. <https://doi.org/10.1016/j.hal.2010.10.001>.
- Withers, P.J.A., Jarvie, H.P., 2008. Delivery and cycling of phosphorus in rivers: a review. *Sci. Total Environ.* 400 (1–3), 379–395. <https://doi.org/10.1016/j.scitotenv.2008.08.002>.
- Wu, Y., Li, L., Zheng, L., Dai, G., Ma, H., Shan, K., Wu, H., Zhou, Q., Song, L., 2016. Patterns of succession between bloom-forming Cyanobacteria *Aphanizomenon flos-aquae* and *Microcystis* and related environmental factors in large, shallow Dianchi Lake, China. *Hydrobiologia* 765, 1–13. <https://doi.org/10.1007/s10750-015-2392-0>.
- Xu, H., Paerl, H.W., Qin, B., Zhu, G., Hall, N.S., Wu, Y., 2014. Determining critical nutrient thresholds needed to control harmful cyanobacterial blooms in Eutrophic Lake Taihu, China. *Environ. Sci. Technol.* 49 (2), 1051–1059. <https://doi.org/10.1021/es503744q>.
- Xu, H., Paerl, H.W., Zhu, G., Qin, B., Hall, N.S., Zhu, M., 2017. Long-term nutrient trends and harmful cyanobacterial bloom potential in hypertrophic Lake Taihu, China. *Hydrobiologia* 787, 229–242. <https://doi.org/10.1007/s10750-016-2967-4>.
- Yang, F., Huang, F., Feng, H., Wei, J., Massey, I.Y., Liang, G., Zhang, F., Yin, L., Kacew, S., Zhang, X., Pu, Y., 2020. A complete route for biodegradation of potentially carcinogenic cyanotoxin microcystin-Lr in a novel indigenous bacterium. *Water Res.* 174, 115638. <https://doi.org/10.1016/j.watres.2020.115638>.
- Yu, G., Jiang, Y., Song, G., Tan, W., Zhu, M., Li, R., 2014. Variation of *Microcystis* and microcystins coupling nitrogen and phosphorus nutrients in Lake Erhai, a drinking-water source in Southwest Plateau, China. *Environ. Sci. Pollut. Res.* 21, 9887–9898. <https://doi.org/10.1007/s11356-014-2937-1>.
- Zhang, S., Du, X., Liu, H., Losiewicz, M.D., Chen, X., Ma, Y., Wang, R., Tian, Z., Shi, L., Guo, H., Zhang, H., 2021. The latest advances in the reproductive toxicity of microcystin-LR. *Environ. Res.* 192, 110254. <https://doi.org/10.1016/j.envres.2020.110254>.