PRIMARY RESEARCH PAPER

Multiple habitat templates for phytoplankton indicators within the functional group system

**Man Zhang · Xucong Lv · Jing Dong · Yunni Gao ·
Jingxiao Zhang · Mei Li · Xiaofei Gao · Xuejun Li**

Abstract The functional group (FG) system sensu Colin S. Reynolds of freshwater phytoplankton has been widely applied in many aspects of freshwater ecological assessments. However, there is no current consensus as to how and to why some functional coda shows multiple habitat templates in different water bodies. To test whether there are multiple habitat templates for some functional coda in the FG system, we used two mesotrophic deep reservoirs in China as examples, analysing phytoplankton taxonomic and environmental data collected in 4 seasons over two years. Results showed that a large proportion of the indicators had similar habitat templates within the FG system, whilst some species had multiple or different habitat templates according to their niche differentiation. The random forest (RF) model filtered a minimum set of environmental conditions and provided two potential habitat templates for *Oocystis lacustris*. One explanatory habitat template of clear water provided by the FG system was verified in the Nanwan Reservoir, and the other habitat template of high dissolved inorganic nitrogen (DIN) was found in the Nianyushan Reservoir. Further studies of the potential multiple habitat templates for some functional coda are recommended for its proper use.

Keywords Freshwater phytoplankton · Random forest · Habitat template · Ecological assessment · Colin S. Reynolds

Introduction

Assemblages of phytoplankton, the most important primary producers in freshwater, change rapidly in response to environmental variations and therefore have been used in water assessments for a long time (Dufrene and Legendre, 1997; Reynolds et al., 2002; Hu et al., 2015). According to the ecological niche principle, ecologists ascribe phytoplankton species with identical or similar lifestyles and living strategies to a basic unit, a 'functional group' (FG) proposed by Colin S. Reynolds, when analysing their responses to environmental variations (Reynolds et al., 2002; Padisák et al., 2009). As FGs can reflect changes in aquatic environments, and the system (a functional classification system) has been widely applied in many aspects of freshwater phytoplankton assessments (Padisák et al., 2006; Borics et al., 2007). Two species co-occurrence hypotheses are the basis of the FG system: (1) Environmental filtering: in a habitat typically constrained by temperature regime, light availability, phosphorus, carbon or nitrogen, these factors become a series of filters for

M. Zhang · X. Lv · J. Dong · Y. Gao · J. Zhang · M. Li ·
X. Gao · X. Li (✉)
College of Fisheries, Henan Normal University,
Xinxiang 453007, China
e-mail: xjli@htu.cn

local species with appropriate adaptations to function in that habitat, and (2) Hierarchical competition: in comparison to a less well-adapted species, a functionally well-adapted species is likely to tolerate the local environmental conditions more successfully. As a consequence, a functional group assembles clusters of phytoplankton species and is sensitive to sets of environmental conditions.

The FG system scheme represents already a more mechanistic and more predictive approach than taxonomy. But it still has some problems, leading to inadequate groups being categorized by even experienced phytoplankton ecologists (Padisák et al., 2009). On the one hand, the species composition of a specific waterbody is a function of the number of processes operating at different spatial and temporal scales, ranging from historical and regional factors such as colonization and extinction to present-day abiotic and biotic interactions in the local environment. Thus, habitat templates for some phytoplankton species were somewhat different from the criteria proposed by Reynolds (Moreti et al., 2013; Santana et al., 2017). On the other hand, the increasing availability of statistical methods for sorting and discriminating numerical and graphical information invites additional contributions to consolidating the FG concept (Seip and Reynolds, 1995). With the popularity of new statistical methods, more methods can help accurately apply the FG system. Therefore, functional classification and habitat templates are affected by regional conditions and statistical methods.

In this paper, we examined two mesotrophic deep reservoirs in China, analysing phytoplankton taxonomic and environmental data in 4 seasons over two years. The objectives of this study were to assess the following: (1) the suitability of habitat templates on the selected lakes; and (2) the feasibility of using different statistical methods for habitat templates. These answers could help us to effectively use the FG system, providing useful information for ecological status assessments.

Materials and methods

Study sites

Two adjacent reservoirs on tributaries of the Huai River were chosen as sampling sites. Both Nanwan

(32° 4′ N, 113° 57′ E) and Nianyushan (31° 47′ N 115° 20′ E) are mesoeutrophic canyon reservoirs in China, 130 km from each other. The Nanwan Reservoir has a watershed area of 1100 km^2, a total volume of 16.33×10^8 m^3 and an average depth of 13.5 m (Zhang et al., 2019a). This reservoir is mainly fed by 3 rivers: the Wudao, Dongjia and Tanjia rivers. The Nianyushan Reservoir has a watershed area of 924 km^2, a total volume of 9.16×10^8 m^3 and an average depth of 15 m (Zhang et al., 2019b). The Guan River is the main river that feeds into the reservoir. The seasonal characteristics of the two reservoirs are as follows: average annual precipitation of 1200 mm and average annual temperature of 16 °C (Jia et al., 2015). The dry season last from October to May, whilst the wet season from June to September contains approximately 60% of the annual precipitation. Silver carp and bighead carp are the dominant fish species in both of the two reservoirs. To obtain a more comprehensive understanding of relationship between phytoplankton community and environmental conditions in the two reservoirs, 6 sampling sites were chosen covering lacustrine and transitional zone in both reservoirs (Fig. 1).

Sample collection and measurements

Sampling was carried out seasonally (four times in 1 year) at six sites in both reservoirs from August 21, 2014, to June 31, 2016. The monitored abiotic and biotic variables included water temperature (T), dissolved oxygen (DO), electrical conductivity (EC), pH, water depth, water level, precipitation, inflow, outflow, water level, total phosphorus (TP), soluble reactive phosphorus (SRP), total nitrogen (TN), dissolved inorganic nitrogen (DIN = N-NH$_4$ + N-NO$_3$ + N-NO$_2$), Secchi depth (SD), phytoplankton biomass and zooplankton biomass. T, DO, EC and pH were measured at the water surface by a multiparameter water quality monitor (YSI Incorporated, Yellow Springs Ohio, USA). Hydrological and meteorological data were provided by the two hydropower generation companies.

Water samples were collected with a 5-l iron sampler. The integrated water sample was collected from the surface to the bottom of the euphotic layer (SD × 2.5) at well-spaced intervals according to the different depths (5–18 m) of the sampling sites. The water sample was then subsampled for nutrients and
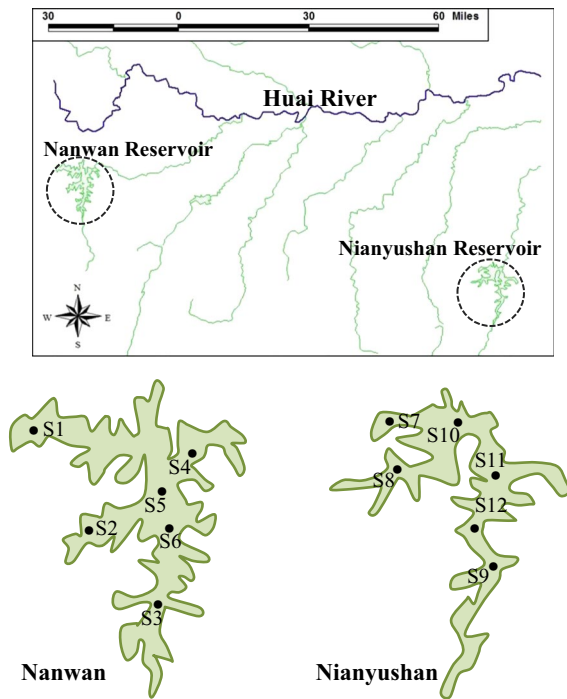
Fig. 1 Location of the Nanwan Reservoir and the Nianyushan Reservoir. Black dots show sampling sites in the two reservoirs

phytoplankton. Available forms of nutrients (SRP, N-NH$_4$, N-NO$_3$ and N-NO$_2$) were determined after filtration of the water (300 ml) through 0.45-μm Millipore filters. TP and TN were determined following persulfate digestion in an autoclave at 120 °C for 2 h. Chl-a was extracted from the water samples (300 ml) collected on Millipore filters using 90% acetone and measured by a spectrophotometer (HITACHI U2000). Nutrient and Chl-a concentrations were measured according to the Chinese National Standards for Water Quality and the Environmental Protection Agency of the USA (Association et al., 2012).

A total of 40 l of water was collected for each zooplankton sample from the water surface to the bottom of the reservoir at well-spaced intervals. The sampled water was filtered into a net with a 30-μm mesh, concentrated to 10–20 ml and preserved with 3% formaldehyde. Zooplankton biomass was estimated by measuring the length of at least 20 specimens whenever a sufficient number of animals was available. Individual biomass (μg) was estimated by body length following the equations of Dumont et al. (1975).

The integrated water samples for phytoplankton were collected with a 10 m length tube sampler. Phytoplankton samples were preserved with 3% formaldehyde and 1% Lugol's solution and stored under dark and cold conditions (4 °C). After sedimentation for at least 48 h, subsamples of 25 ml were concentrated in sedimentation chambers (Hydro-Bios Apparatebau GmbH, Germany), and the entire chamber bottom was examined at ×400 magnification under an inverted microscope (Nikon TS2, Japan) for phytoplankton counting. Taxa were identified to the species level when possible (Hu and Wei, 2006). To estimate phytoplankton biomass, at least 25 individuals for each species were measured and then approximations to geometric solids were applied to calculate individual biovolume (Hillebrand et al., 1999). Phytoplankton biomass was estimated from the biovolume, assuming that 10^6 μm^3 corresponding to 1 μg of biomass. Phytoplankton indicators were chosen on the basis of the relative biomass of a species and its relative frequency of occurrence according to the following indicator value method. Based on the species/genus classification, the indicators of these phytoplankton species were assigned into functional coda according to Reynolds et al. (2002) and Padisák et al. (2009).

Data analysis

Optimal number of clusters for the phytoplankton composition

Optimal clustering can help us better identify grouping of phytoplankton composition from all the samples. To obtain the optimal number of groups for identifying the phytoplankton composition, a silhouette-optimal rule was used to select an appropriate number of clusters (Borcard et al., 2018). The dissimilarity matrix was subjected to hierarchical cluster analysis using Ward's minimum variance agglomeration algorithm to detect compact, spherical clusters. The algorithm grouped sampling sites (including spatial and temporal samples) into clusters based on a specified number of clusters (*k*). We tested a range of cluster numbers from 2 to 10 and tested the stability of the clustering by calculating mean silhouette widths, a standard measure of cluster isolation (Cohen et al., 2016). Silhouette width measures the degree of membership of a site to its group, based on

the average distance between the site and all sites of the group to which it belongs, and this value is compared to the same measure calculated for the next closest group (Rousseeuw, 1987). The highest value of average silhouette width was used to identify the optimal number of clusters within which the data naturally fell. We calculated Ward's minimum variance and its silhouette information in the 'VEGAN' package (Okasen et al., 2007) and the 'CLUSTER' package (Junior, 2003) in R.

Identifying phytoplankton indicator species

After determining the cluster grouping, the indicator value method was used to identify the phytoplankton indicator species for each clustering group (Dufrene and Legendre, 1997). This approach combines the relative biomass of a species with its relative frequency of occurrence in various groups of sites. The indicator values of the species were tested via Monte Carlo simulation using 10,000 permutations. The accepted significance level was $P < 0.05$. The 'LABDSV' package (Roberts, 2006) in R was used for indicator species analysis.

Habitat templates for the phytoplankton indicators

To obtain a common habitat template for phytoplankton indicators, it is necessary to determine which statistical methods are most effective for analysing habitat templates. Three statistical methods were used: variation partitioning analysis (VPA), a generalized linear model (GLM) and a random forest (RF) classification model. Hellinger standardization (Legendre and Gallagher, 2001) of the phytoplankton biomass data was performed before carrying out the three statistical methods. Environmental explanatory conditions (including T, DO, EC, pH, water depth, water level, precipitation, inflow, outflow, water level, TP, SRP, TN, DIN, SD and zooplankton biomass) were normalized using feature scaling transformation to avoid differential weighting.

The VPA method based on the redundancy analysis (RDA) algorithm aims to seek the linear combination of environmental explanatory variables that best explained the variation in the phytoplankton indicators (Lai et al., 2022). VPA was used to reveal the explanatory rates of potential explanatory variables for the indicator species in each clustering

group. Variables with cumulative explanatory rate > 60% were considered habitat templates for the phytoplankton indicators. VPA was performed using the 'RDACCA.HP' package (Lai et al., 2022) in R.

A GLM can determine the importance of environmental explanatory variables for phytoplankton indicators. A full model was constructed using all the potential explanatory variables for the target phytoplankton biomass. A forward stepwise procedure was performed to build a final fitting model. Only factors remaining significant (significance level $P < 0.05$) throughout analysis were identified as significant environmental variables and retained in the final fitting model. The significant variables were then selected as habitat templates for the target phytoplankton. All GLM analyses were performed using the 'GLMULTI' package (Vandecandelaere, 2012) in R.

RF constructs decision trees whilst training and outputting classification trees (Breiman, 1999). Compared to the traditional machine learning technique (e.g. where artificial neural networks are always used for analysing the distribution of functional species), RF is much more efficient because it is suitable for demonstrating the nonlinear fittings and reflects the complex interactions between input variables and the response variable. To build a classification tree in the RF framework, all the potential explanatory variables for the target phytoplankton biomass were used. The optimal RF classification model was selected based on the Monte Carlo cross-validation error (CV error) (Wallhead et al., 2009; Guo et al., 2021). Then, we selected the tree size with the lowest CV error. The variation explained in the output classification tree provided a habitat template for the target phytoplankton. RF analysis was performed using the 'MVPART' package (De'Ath, 2006) in R.

Other statistical analyses

Independent T tests were used to assess differences in the annual limnological variables between the two reservoirs. To reflect the environmental filters for the target phytoplankton biomass, ridge plots were generated using the 'GGPLOT2' (Wickham, 2009) and 'GGRIDGES' (Wilke, 2017) packages in R.

Table 1 The phytoplankton biomass and potential environmental explanatory variables in the Nanwan and Nianyushan Reservoirs

Variables	Nanwan					Nianyushan				
	Annual	Autumn	Winter	Spring	Summer	Annual	Autumn	Winter	Spring	Summer
Precipitation (mm)	283.5±194.7	377.6±36.1	103.2±7.5	106.6±13.1	546.5±76.3	318.3±219.9	407.7±118.5	139.4±13.5	112.8±12.6	613.3±87.0
Inflow (million m³)	79.4±77.4 a	104.9±0.9	9.7±0.5	21.1±6.0	182.0±66.3	133.0±121.4 b	138.0±56.0	39.1±17.3	41.7±19.3	313.2±70.9
Outflow (million m³)	43.3±22.7 a	72.2±10.9	24.2±9.5	25.0±13.9	51.8±6.0	103.7±95.9 b	129.5±95.2	17.2±17.9	52.2±54.5	216.2±36.9
DIN (μg l^{-1})	222.2±159.4	167.6±44.7	174.7±34.2	404.3±215.6	142.1±100.0	155.0±136.8	71.1±49.8	124.4±26.8	279.9±217.9	144.8±58.6
TN (μg l^{-1})	695.2±495.4	754.8±287.3	590.0±215.5	1042.7±794.1	393.3±204.7	502.9±344.8	407.9±135.1	546.2±292.4	744.7±497.4	312.9±204.8
SRP (μg l^{-1})	5.9±4.7	6.9±5.2	9.5±5.0	2.1±1.4	4.8±2.7	7.5±11.8	7.0±4.8	13.6±22.1	2.2±1.8	7.1±2.9
TP (μg l^{-1})	20.9±7.3	17.6±6.6	26.4±10.0	18.2±4.6	21.5±3.4	20.9±7.2	14.9±4.6	24.4±9.3	19.7±5.5	24.6±3.6
Water temperature (°C)	18.3±7.2	24.4±1.2	9.5±1.0	13.2±1.0	26.1±1.0	18.5±6.6	24.7±0.3	10.5±1.3	13.8±0.4	25.2±0.4
Water depth (m)	9.7±6.5 a	8.8±6.0	8.9±7.0	11.5±7.3	9.5±6.1	12.1±9.4 b	11.5±10.9	10.4±9.1	13.9±6.9	12.4±11.0
SD (m)	2.1±1.1	1.3±0.3	2.1±0.7	3.4±1.3	1.6±0.5	2.2±1.2	1.4±0.4	2.2±0.5	3.8±1.6	1.8±0.5
DO (mg l^{-1})	7.5±1.4	6.6±1.0	8.7±1.0	8.2±1.2	6.5±0.9	7.4±1.6	5.8±0.4	8.8±1.0	8.7±1.0	6.1±0.4
EC (S/m)	171.7±26.4 a	189.0±6.2	137.3±3.9	159.9±8.8	200.7±13.0	115.0±14.8 b	126.8±3.1	93.7±5.6	113.3±9.8	126.1±4.4
pH	8.0±0.3	8.0±0.4	7.9±0.4	7.9±0.3	8.1±0.3	7.7±0.3	7.5±0.2	7.6±0.2	7.8±0.1	7.8±0.4
Water level (m)	292.6±8.6 a	292.0±8.9	289.8±7.6	294.2±6.1	294.4±11.1	387.5±5.9 b	389.1±3.3	383.5±8.4	386.5±2.9	390.8±4.9
Zooplankton biomass (μg l^{-1})	3850.0±600.5	5291.7±782.6	1299.3±272.2	5169.1±1925.9	3640.0±908.5	2636.5±545.9	3046.8±362.2	811.6±278.4	5321.8±1784.6	1468.2±204.6
Phytoplankton biomass (μg l^{-1})	5120.0±7429.6	11,295.5±10,983.3	2727.8±798.4	1118.0±1411.2	5338.5±6871.7	6417.5±13,777.7	21,452.1±21,888.8	1794.7±1024.0	446.5±333.1	1976.7±693.2

Values are shown as the mean ± standard deviation. Different letters indicate significant differences (independent t test, $P < 0.05$) in the annual variables between the two reservoirs

TP total phosphorus, *TN* total nitrogen, *SRP* soluble reactive phosphorus, *DIN* dissolved inorganic nitrogen, *SD* Secchi depth, *DO* dissolved oxygen and *EC* electrical conductivity
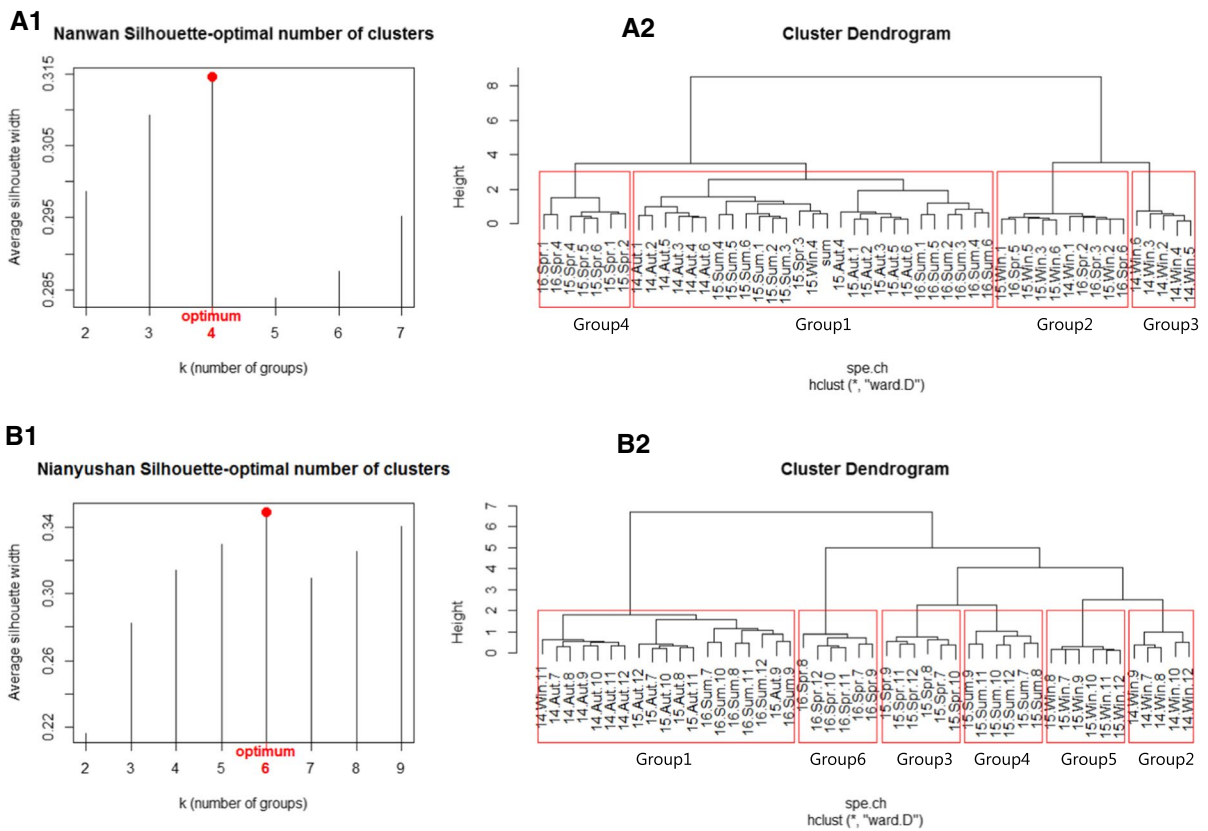
A1

A2

B1

B2

Fig. 2 Phytoplankton composition clusters in the two reservoirs. **A1** Silhouette width showing the optimal number of clusters in the Nanwan Reservoir; **A2** Nanwan phytoplankton composition cluster dendrogram. **B1** The optimal number of clusters in the Nianyushan Reservoir; **B2** Nianyushan phytoplankton composition cluster dendrogram. Phytoplankton samples were named by Year + Season + Sampling site. For example, 16.Spr.1 represents phytoplankton sampled in spring of 2016 at the 1st sites

Results

Phytoplankton indicators and habitat templates

The phytoplankton biomass and potential environmental explanatory variables of the two reservoirs over the four seasons are shown in Table 1. Both reservoirs are situated in monsoon climate zones, with most precipitation occurring in summer. Because of the urban water demand, the Nianyushan Reservoir has more outflow than the Nanwan Reservoir (independent T test, $P < 0.001$). No significant differences in the water trophic state (including TN, TP, DIN and SRP concentrations) were observed between the Nanwan and Nianyushan Reservoirs (independent T test, $P > 0.05$). The annual concentrations of TN, TP and SD in the Nanwan Reservoir were 695.2 μg l^{-1},

20.9 μg l^{-1} and 2.1 m, respectively, whilst those in the Nianyushan Reservoir were 502.9 μg l^{-1} and 20. 9 μg l^{-1} and 2.2 m, respectively, indicating that both reservoirs were mesoeutrophic (Carlson, 1977; Kratzer and Brezonik, 1981). Seasonal data showed that a eutrophic state occurred in summer, especially in the Nanwan Reservoir (Table 1). SD, total zooplankton biomass and total phytoplankton biomass also showed no significant differences between the Nanwan and Nianyushan Reservoirs (independent T test, $P > 0.05$).

In the Nanwan Reservoir, we recorded 118 species across the 48 sampling sites (spatial and temporal), distributed amongst 8 phyla: Chlorophyta (66), Bacillariophyta (22), Cyanophyta (17), Pyrrophyta (5), Xanthophyta (1), Chrysophyta (3), Cryptophyta (1) and Euglenophyta (4). The highest average silhouette

Table 2 Phytoplankton indicators of the clustering groups in the Nanwan Reservoir

Phytoplankton species	Cluster group	Indval	P value	Frequency	Habitat and season characteristics	Functional codon	Habitat template
Chroomonas sp.	Group 1	0.70	1.00E−04	40	Large amount of outflow, warm temperature, wet season. Autumn to summer	X2	Shallow, mesoeutrophic environments
Chlorella vulgaris Beijerinck	Group 1	0.74	1.00E−04	41		X1	Shallow, euhypertrophic environments
Pseudanabaena limnetica (Lemmermann) Komárek	Group 1	0.74	1.00E−04	41		S1	Turbid mixed environments
Ulnaria acus Kützing	Group 1	0.81	1.00E−04	48		D	Shallow turbid waters including rivers
Euglena sp.	Group 3	0.79	3.00E−04	12	Rich in total phosphorus. Winter	W1	Ponds, even temporary, rich in organic matter from husbandry activities or sewage
Anabaena minutissima Lemmermann	Group 3	0.94	1.00E−04	28		H1	Eutrophic, both stratified and shallow lakes with low nitrogen content
Oocystis lacustris Chodat	Group 4	0.82	1.00E−04	29	Clear water. Spring	F	Clear, deeply mixed meso-eutrophic lakes

Indval is the indicator value of species i at site cluster j. The functional coda are assigned according to the FG system (Reynolds et al., 2002; Padisák et al., 2009)

width was found at 4 clusters, showing that the phytoplankton composition from different sampling sites could be divided into 4 groups (Fig. 2A1). For the composition of phytoplankton assemblages, seasonal distinction was obvious in the hierarchical dendrogram (Fig. 2A2). Group 1, Group 2, Group 3 and Group 4 included phytoplankton samples in autumn-summer, winter-spring, winter and spring, respectively. Phytoplankton indicator species were found for Group 1, Group 3 and Group 4 according to the indicator value method (Table 2). Group 1 had four phytoplankton indicators: *Chroomonas* sp., *Chlorella vulgaris* Beijerinck, *Pseudanabaena limnetica* (Lemmermann) Komárek and *Ulnaria acus* (Kützing) Aboal with functional coda X2, X1, S1 and D, respectively. *Euglena* sp. and *Anabaena minutissima* Lemmermann were indicators for Group 3 and given functional coda W1 and H1, respectively. Only one indicator species (*Oocystis lacustris* Chodat) was found in Group 4 and was given functional codon F. After excluding explanatory variables with multi-collinearity, fourteen variables, namely, T, DO, pH, EC, water depth, precipitation, outflow, water level, TP, SRP, TN, DIN, SD and zooplankton biomass,

were selected to perform VPA. The results showed that warm water with substantial outflow was the explanatory habitat template for Group 1 (Fig. 3A1). High total phosphorus and clear water states were the explanatory habitat templates for Group 3 and Group 4, respectively (Fig. 3A2, A3).

In the Nianyushan Reservoir, 116 species across the 48 sampling sites were recorded and distributed amongst 8 phyla: Chlorophyta (66), Bacillariophyta (22), Cyanophyta (16), Pyrrophyta (5), Xanthophyta (1), Chrysophyta (3), Cryptophyta (1) and Euglenophyta (3). The highest average silhouette width was found at 6 clusters, showing that the phytoplankton composition from different sampling sites could be divided into 6 groups (Fig. 2B1). All phytoplankton indicators for each group are shown in Table 3. The VPA results for fourteen potential explanatory variables showed that the wet season with warm water was the explanatory habitat template for both Group 1 and Group 4 indicators (Fig. 3B1, B4). Few outflows with high water levels was the explanatory habitat template for Group 2 (Fig. 3B2). Few outflows with high total nitrogen was the explanatory habitat template for Group 3 (Fig. 3B3). High phosphorus and high DIN
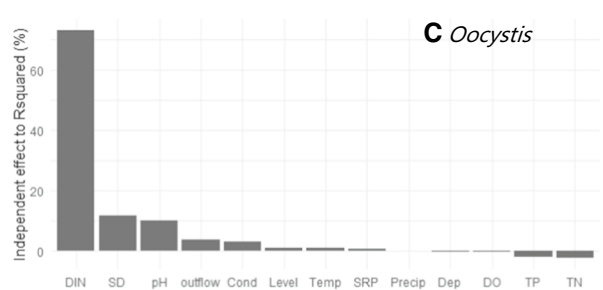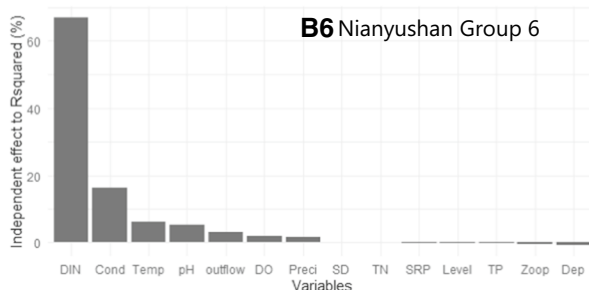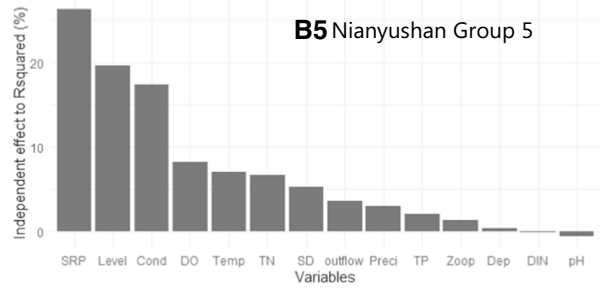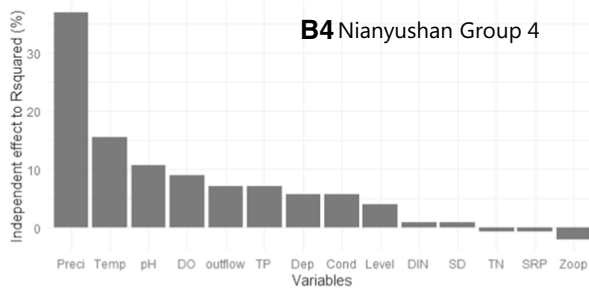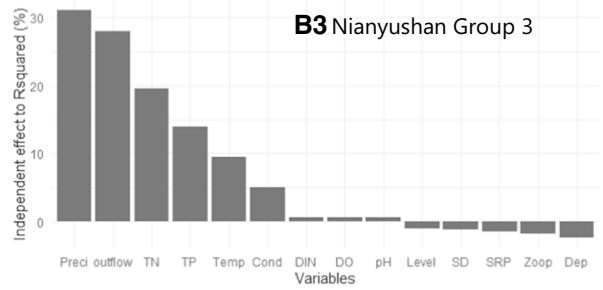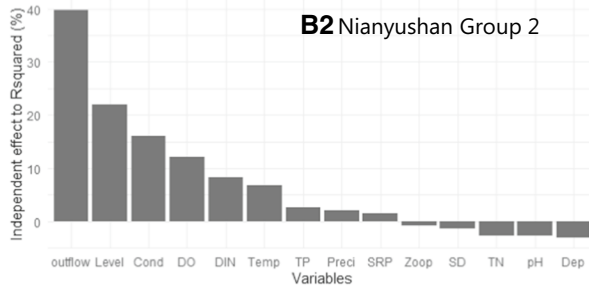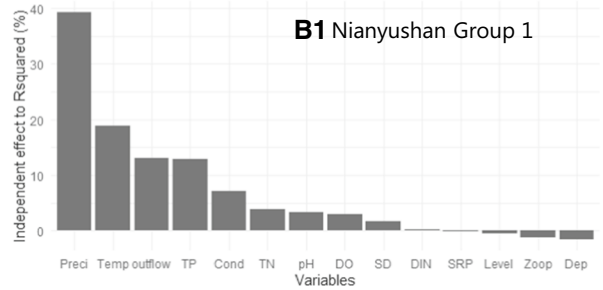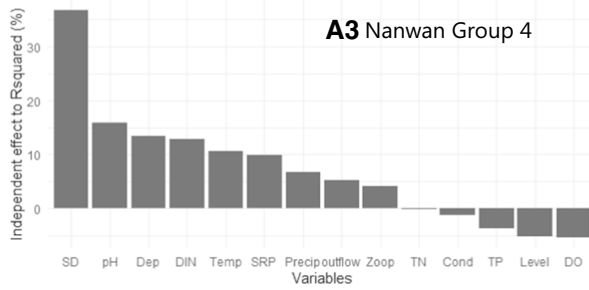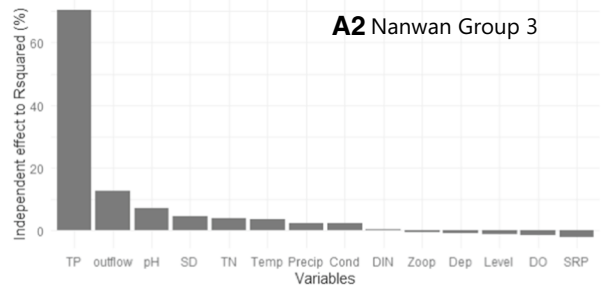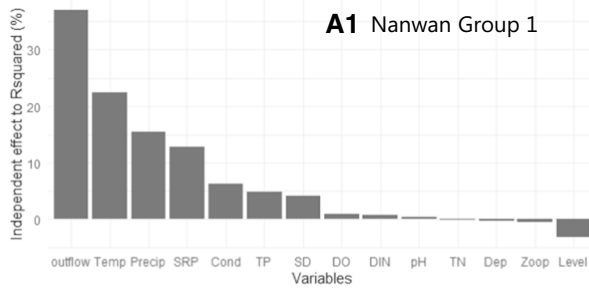
◄**Fig. 3** Contribution of the potential explanatory variables explaining the cluster groups of the two reservoirs and the biomass of *Oocystis lacustris*. **A1** Group 1 in the Nanwan Reservoir; **A2** Group 3 in the Nanwan Reservoir; **A3** Group 4 in the Nanwan Reservoir; **B1** Group 1 in the Nianyushan Reservoir; **B2** Group 2 in the Nianyushan Reservoir; **B3** Group 3 in the Nianyushan Reservoir; **B4** Group 4 in the Nianyushan Reservoir; **B5** Group 5 in the Nianyushan Reservoir; **B6** Group 6 in the Nianyushan Reservoir; **C** *Oocystis lacustris* in the two reservoirs. *TP* total phosphorus, *TN* total nitrogen, *SRP* soluble reactive phosphorus, *DIN* dissolved inorganic nitrogen, *SD* Secchi depth, *Precip* precipitation, *Temp* water temperature, *Level* water level, *Dep* water depth, *DO* dissolved oxygen, *Cond* electrical conductivity and *Zoop* zooplankton biomass

were the explanatory habitat templates for Group 5 and Group 6, respectively (Fig. 3B5, B6).

A large proportion of the indicators had similar habitat templates within the FG system. However, different habitat templates were found for some indicators (*Gomphonema parvulum* Kützing, *Elakatothrix* sp., *Chlorella vulgaris* and so on) from the corresponding coda within the FG system (Tables 2, 3). Some phytoplankton species were selected as indicators in both reservoirs, including *Chlorella vulgaris*, *Pseudanabaena limnetica*, *Ulnaria acus* and *Oocystis lacustris*. We found that both *Pseudanabaena limnetica* and *Ulnaria acus* were classified into Group 1 in the two reservoirs, with similar habitat templates according to VPA. *Chlorella vulgaris* was assigned to Group 1 in the Nanwan Reservoir and Group 4 in the Nianyushan Reservoir, which also had similar habitat templates according to VPA. Notably, *Oocystis lacustris* was classified into Group 4 in the Nanwan Reservoir (Table 2) and Group 6 in the Nianyushan Reservoir (Table 3), with different habitat templates for the two reservoirs (Nanwan: clear water state; Nianyushan: trophic state with high DIN). This indicator may have various habitat templates in different environments, and it is necessary for us to determine its common habitat template.

Common habitat template: *Oocystis lacustris* as an example

When using different statistical methods to analyse a common habitat template for an indicator, different results will be obtained. *Oocystis lacustris* was used as an example to determine which method best

reveals the habitat templates for a selected indicator in both reservoirs. The VPA results showed that the relative contribution of DIN was greater than 70%, suggesting that high DIN was the explanatory habitat template for biomass of *Oocystis lacustris* (Fig. 3C).

Based on GLM, we quantified the relative importance of all variables for biomass of *Oocystis lacustris*, with DIN and outflow expected to be the most important factors for explaining the biomass of *Oocystis lacustris*. The general linear equation produced from the regression model is shown in Eq. 1. The equation indicated that high DIN with few outflows was the habitat template for *Oocystis lacustris* (Fig. 4A).

$$y = 0.543 \times \mathrm{DIN} - 0.057 * \mathrm{outflow}$$
$$(R^2 = 0.31, P = 2.7e-8) \tag{1}$$

When analysing the possible habitat template in the RF classification model, the software asked us to choose the size of the classification tree. We selected a tree size of 3 with the lowest CV error to run the classification tree. The classification tree showed that highest biomass of *Oocystis lacustris* (only 2 samples, average biomass = 42.1 μg l^{-1}) occurred in eutrophic waters with DIN ≥ 676.3 μg l^{-1} (Fig. 4B). In addition, *Oocystis lacustris* was more commonly found in clear waters with SD ≥ 4.6 m when DIN < 676.3 μg l^{-1} (7 samples, average biomass = 8.7 μg l^{-1}). Thus, two habitat templates for *Oocystis lacustris* were provided in the RF analysis.

To demonstrate the effectiveness of habitat templates, ridge plots were generated to show the environmental filters for biomass of *Oocystis lacustris* directly (Fig. 5). The ridge plots suggested that DIN and SD can be environmental filters for *Oocystis lacustris* in the two reservoirs. Combing result from the RF model, one habitat template of clear water provided by the FG system was verified in the Nanwan Reservoir, and the other habitat template of high DIN was found in the Nianyushan Reservoir (Fig. 6). Therefore, the common habitat templates obtained from the VPA and RF models were reasonable.

Table 3 Phytoplankton indicators of the clustering groups in the Nianyushan Reservoir

Phytoplankton species	Cluster group	Indval	*P* value	Frequency	Habitat and season characteristics	Functional codon	Habitat template
Planktolyngbya limnetica (Lemmermann) Komárková-Legnerová & Cronberg	1	0.79	3.00E−04	23	Wet season, warm water. Summer to autumn	S1	Turbid mixed environments
Ulnaria acus Kützing	1	0.88	1.00E−04	48		D	Shallow turbid waters including rivers
Pseudanabaena limnetica (Lemmermann) Komárek	1	0.94	1.00E−04	36		S1	Turbid mixed environments
Trachelomonas sp.	2	0.59	1.00E−04	31	Few outflows with high water level. Winter	W2	Mesoeutrophic ponds, even temporary, shallow lakes
Characium sp.	2	0.80	1.00E−04	4		–	–
Stephanodiscus flabellatus Khursevich & Loginova	2	0.85	1.00E−04	6		C	Eutrophic small and medium lakes with species sensitive to the onset of stratification
Dactylococcopsis sp.	3	0.59	4.00E−04	30	Dry season with few outflows, high total nitrogen. Spring	–	–
Gomphonema parvulum Kützing	3	0.62	3.00E−04	10		T_B	Highly lotic environments
Asterionella formosa Hassall	3	0.74	2.00E−04	8		C	Eutrophic small and medium lakes with species sensitive to the onset of stratification
Rhizosolenia sp.	3	0.83	1.00E−04	15		A	Clear, deep, base-poor lakes, with species sensitive to pH increases
Dinobryon divergens O.E.Imhof	3	0.95	1.00E−04	9		E	Usually, small, shallow, base-poor lakes or heterotrophic ponds
Chlorella vulgaris Beijerinck	4	0.51	1.00E−04	41	Similar to Group 1, with more precipitation and warmer temperature. Summer	X1	Shallow, euhypertrophic environments
Elakatothrix sp.	4	0.63	3.00E−04	5		F	Clear, deeply mixed mesoeutrophic lakes
Cylindrospermopsis raciborskii Wołoszyńska	4	0.63	5.00E−04	5		S_N	Warm mixed environments
Anabaena minutissima Lemmermann	4	0.87	1.00E−04	22		S_N	Warm mixed environments
Melosira granulata (Ehrenberg) Ralfs	5	0.71	1.00E−04	42	Trophic state with high phosphorus. Winter	P	Continuous or semicontinuous mixed layer 2–3 m thick. Trophic states
Oocystis lacustris Chodat	6	0.76	1.00E−04	25	Trophic state with high dissolved inorganic nitrogen. Spring	F	Clear, deeply mixed mesoeutrophic lakes

Note: Indval is the indicator value of species i at site cluster j. The functional coda are assigned according to the FG system (Reynolds et al., 2002; Padisák et al., 2009)

A

B

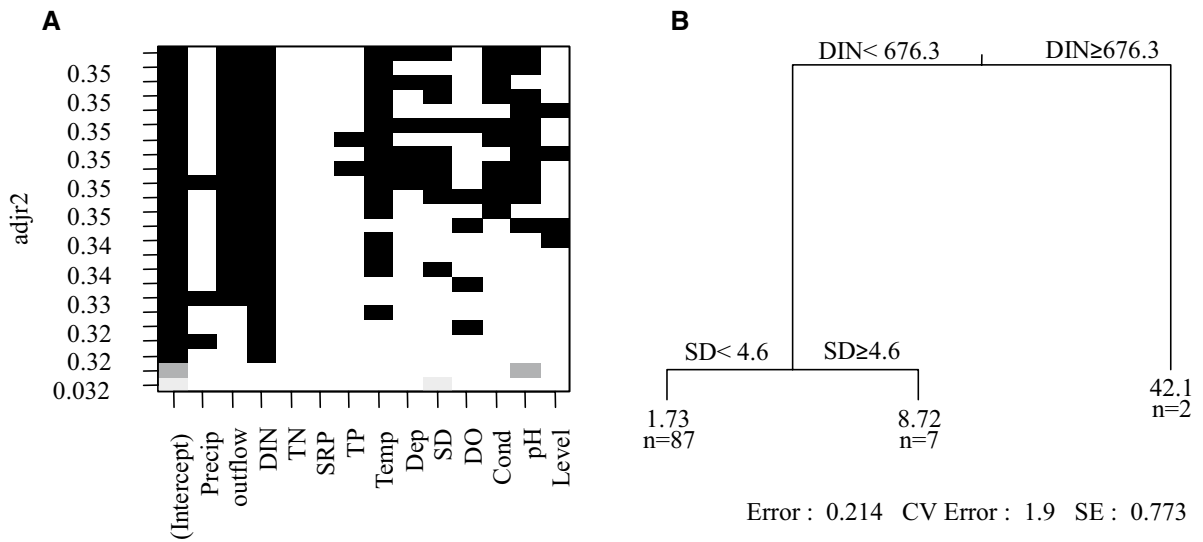Error : 0.214   CV Error : 1.9   SE : 0.773

Fig. 4 Environmental templates explained by the generalized linear model and random forest for the biomass of *Oocystis lacustris*. **A** The potential explanatory variables in the generalized linear model; **B** the explanatory variables in the random forest classification trees. *TP* total phosphorus, *TN* total nitro-gen, *SRP* soluble reactive phosphorus, *DIN* dissolved inorganic nitrogen, *SD* Secchi depth, *Precip* precipitation, *Temp* water temperature, *Level* water level, *Dep* water depth, *DO* dissolved oxygen and *Cond* electrical conductivity
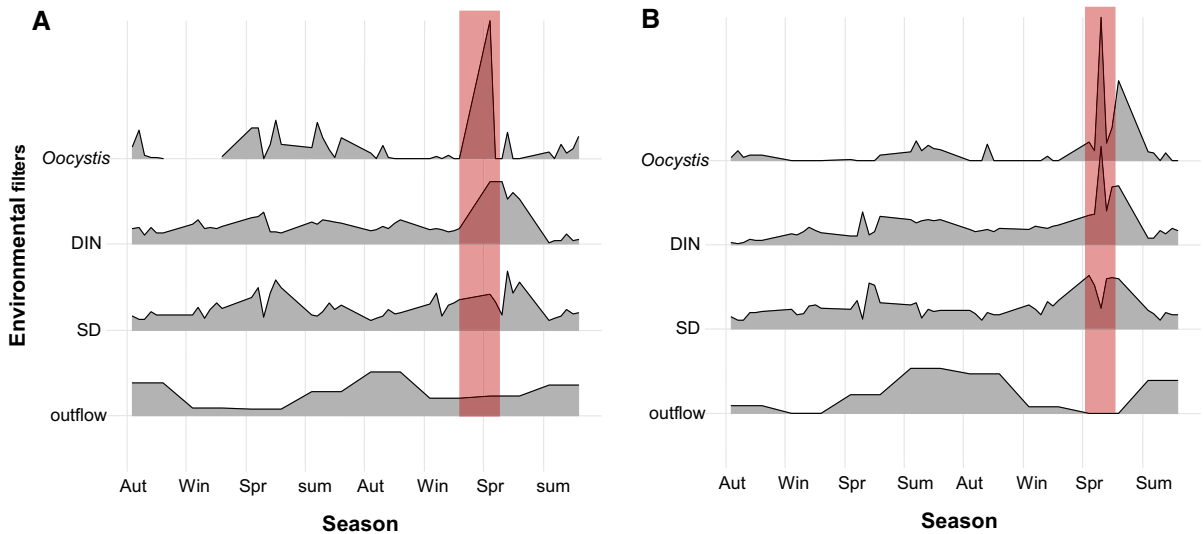
A

B

Fig. 5 Environmental filters for biomass of *Oocystis lacustris*. **A** Nanwan Reservoir; **B** Nianyushan Reservoir. *DIN* dissolved inorganic nitrogen, *SD* Secchi depth

Discussion

Different statistical methods were used to analyse the habitat templates of the FGs. Hierarchical agglomerative clustering of phytoplankton data and comparison with a principal component analysis (PCA) of corresponding environmental data were used by Fabbro and Duivenvoorden (2000), demonstrating that a syncretic two-part model can speculate the habitat templates for phytoplankton. Dissimilarity analysis
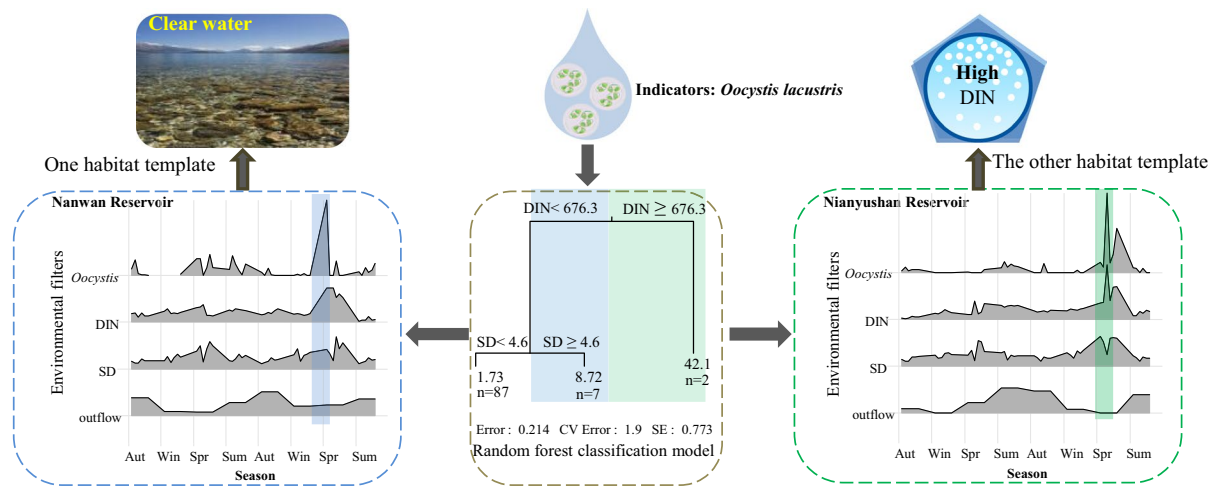
Fig. 6 Two habitat templates for *Oocystis lacustris* were provided in the random forest analysis

A *Pseudanabaena limnetica*

Error : 0.307   CV Error : 0.556   SE : 0.155

B *Chlorella vulgaris*
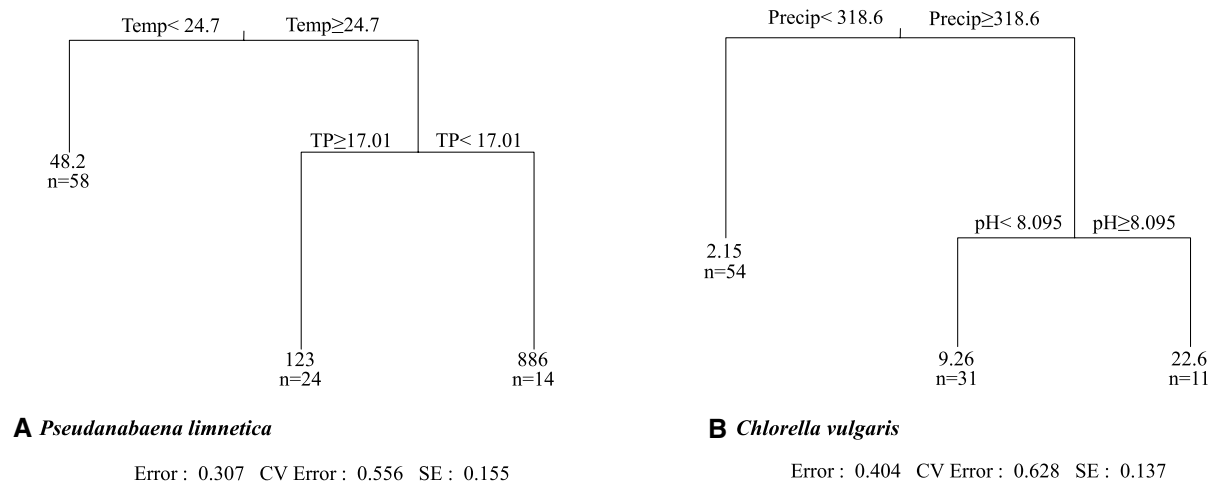
Error : 0.404   CV Error : 0.628   SE : 0.137

Fig. 7 Environmental templates explained by the random forest for *Pseudanabaena limnetica* and *Chlorella vulgaris* **A** *Pseudanabaena limnetica*; **B** *Chlorella vulgaris*. *Temp* water temperature; *TP* total phosphorus; *Precip* precipitation

was used by Bohnenberger et al. (2018) to determine the relationships between phytoplankton FGs and environmental dissimilarities. In the analysis of phytoplankton in a small urban lake, canonical variate analysis (CVA) was applied by Kruk et al. (2002) with several classification approaches, finding that the eigenvalue for the first canonical correspondence analysis (CCA) axis for the associations explained over 78% of the cumulative variance in the environmental variables. Traditional applications of canonical analysis (PCA, RDA, CCA, distance-based RDA) can be used to determine the habitat templates for

phytoplankton and have been used as essential frameworks. However, they pose challenges to model interpretation. VPA linking commonality analysis, variation and hierarchical partitioning, successfully estimate the explanatory variables' contributions in variation partitioning (Lai et al., 2022). In our research, several statistical methods were also used to identify phytoplankton clusters and habitat templates, including VPA, GLM and RF classification models. The results showed that both VPA and the RF classification model helped infer common habitat templates for phytoplankton indicators. Our example

using *Oocystis lacustris* showed that the RF classification model was quite successful in explaining the association between the phytoplankton indicator and habitat template. Kruk and Segura (2012) also successfully used an RF classification tree to identify the thresholds of environmental variables promoting FG dominance and found that the method can provide empirical values describing phytoplankton habitat templates. Our study further explored the possibility of using the RF model to obtain multiple potential habitat templates. The results showed that the RF model filtered a minimum set of habitat conditions and provided multiple possible habitat templates for the indicator.

Based on the phytoplankton FG concept, a phytoplankton assemblage index, Q(r), was developed within the Water Framework Directive (WFD) to assess the ecological status of different water types (Padisák et al., 2006; Borics et al., 2007; Abonyi et al., 2021). As demonstrated on the basis of case studies from many different waterbodies, the assemblage index Q or Qr has proven to be a promising tool for assessing the ecological status of waters (Padisák et al., 2006; Borics et al., 2007). To calculate the index, each species was assigned to the appropriate FGs. Then, all phytoplankton associations were evaluated and given a score between 0 and 5 (*F* value). However, different *F* values for the same codon were scored in different habitats (Padisák et al., 2006). One of the possible reasons may be because of multiple habitat templates for the phytoplankton assemblage. One example given by Padisák et al. (2003) is *Cylindrospermopsis raciborskii* (Wołoszyńska) Seenayya & Subba Raju, which, in most cases, is given codon S_N. However, if the species occurs in deep-layer chlorophyll maxima, labelling it as an R species is correct. In our case, *Pseudanabaena limnetica* (S1 codon) was given an *F* value of 0 in the evaluation system, suggesting that it was assigned to a hypertrophic state or turbulent waters (Padisák et al., 2006; Borics et al., 2007). However, *Pseudanabaena limnetica* was found to be an indicator of warm water (Temp ≥ 24.7 °C) with low total phosphorus (TP < 17.0 μg l^{-1}) in our research (Fig. 7A). *Chlorella vulgaris* (X1 codon) has been considered an indicator species of shallow and euhypertrophic waters (Reynolds et al., 2002; Padisák et al., 2009) but was found in alkaline water (pH ≥ 8.1) during the rainy season (Precip ≥ 318.6 mm) in our research (Fig. 7B). Therefore,

we should further study the potential multiple habitat templates for some functional coda if we want to use the FG system properly.

The success of the FG system suggests that most species within FGs may behave similarly enough to be described by a single habitat template (Peña, 2003; Le Quéré et al., 2010). However, species neutrality within an FG is a restricted case of the FG system: a species' biomass is a random walk not influenced by environmental conditions. For some species, wide niches weaken the effect of niche selection and mean that demographic stochasticity is a very important source of biomass variation at the species level (Hubbell, 2001). Mutshinda et al. (2016) indicated that the majority of diatom and dinoflagellate species in the Western English Channel were driven by neutral dynamics rather than by nonneutral factors (environmental filtering and hierarchical competition). This may help explain multiple habitat templates for these species. In our research, most of the indicators had the same habitat templates within the FG system, supporting the assumption of fixed niches for most species. Some species (e.g. *Oocystis lacustris*, *Pseudanabaena limnetica* and *Chlorella vulgaris*) with multiple habitat templates may have wide niches and ecological drift when they are found in different regions. We must account for some niche differentiation for these species within the FG system.

Data availability The data generated during the present study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Declarations

Conflict of 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.

Ethical approval Not applicable.

Consent to participate All authors have given their consent to participate.

Consent for publication All authors have given their consent to participate.

References

Abonyi, A., J.-P. Descy, G. Borics & E. Smeti, 2021. From historical backgrounds towards the functional classification of river phytoplankton sensu Colin S. Reynolds: what future merits the approach may hold? Hydrobiologia 848: 131–142.

Association, A.P.H., Association, A.W.W., Federation, W.E., 2012. Standard Methods for the Examination of Water and Wastewater. 22th edition. In: Rice, E.W., Baird, R.B., Eaton, A.D., Clesceri, L.S. (eds). American Public Health Nations Health, Washington DC

Bohnenberger, J. E., L. R. Rodrigues, M. Da & L. O. Crossetti, 2018. Environmental dissimilarity over time in a large subtropical shallow lake is differently represented by phytoplankton functional approaches. Marine & Freshwater Research 69: 95–104.

Borcard, D., Gillet, F., Legendre, P., 2018. Numerical Ecology with R. In: Gentleman, R., Parmigiani, G.G., Hornik, K. (eds). Springer, Montréal.

Borics, G., G. Várbíró, I. Grigorszky, E. Krasznai, S. Szabó & K. T. Kiss, 2007. A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers. Large Rivers 161: 466–486.

Breiman, L., 1999. Random forests–random features. Machine Learning 45: 5–32.

Carlson, R. E., 1977. A trophic state index for lakes. Limnology and Oceanography 22: 361–369.

Cohen, J. E., J. Lai, D. A. Coomes & R. B. Allen, 2016. Taylor's law and related allometric power laws in New Zealand mountain beech forests: the roles of space, time and environment. Oikos 125: 1342–1357.

De'Ath, G., 2006. The mvpart package. http://CRAN.R-project.org/package=mvpart.

Dufrene, M. & P. Legendre, 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67: 345–366.

Dumont, H., I. Van de Velde & S. Dumont, 1975. The dry weight estimate of biomass in a selection of cladocera, copepoda and rotifera from the plankton, periphyton and benthos of continental waters. Oecologia 19: 75–97.

Fabbro, L. & L. Duivenvoorden, 2000. A two-part model linking multidimensional environmental gradients and

seasonal succession of phytoplankton assemblages. Hydrobiologia 438: 13–24.

Guo, B., D. Zhang, L. Pei, Y. Su & L. Guo, 2021. Estimating PM25 concentrations via random forest method using satellite, auxiliary, and ground-level station dataset at multiple temporal scales across China in 2017. Science of the Total Environment 778: 146288.

Hillebrand, H., C. D. Durselen, D. Kirschel, U. Pollingher & T. Zohary, 1999. Biovolume calculation for pelagic and benthic microalgae. J. Phycol. 35: 403–424.

Hu, H. & Y. Wei, 2006. Chinese Freshwater Algae—System Classification and Ecology, Science Press, Beijing:

Hu, R., Y. Lan, L. Xiao & B. Han, 2015. The concepts, classification and application of freshwater phytoplankton functional groups. Journal of Lake Sciences 27: 11–23.

Hubbell, S. P., 2001. Unified Neutral Theory of Biodiversity & Biogeography, Princeton University Press, Princeton:

Jia, B., Simonovic, S., Zhong, P., Yu, Z., 2015. Improved Knowledge-based Cooperative Particle Swarm Optimization for Optimal Reservoir Flood Control Operation. CSCE 22nd Canadian Hydrotechnical Conference, Canada.

Junior, F., 2003. [R] Package cluster. https://cran.r-project.org/web/packages/cluster.

Kratzer, C. R. & P. L. Brezonik, 1981. A carlson-type trophic state index for nitrogen in Florida Lakes. Water Resources Bulletin 17: 713–715.

Kruk, C. & A. M. Segura, 2012. The habitat template of phytoplankton morphology-based functional groups. Hydrobiologia 698: 191–202.

Kruk, C., N. Mazzeo, G. Lacerot & C. S. Reynolds, 2002. Classification schemes for phytoplankton: a local validation of a functional approach to the analysis of species temporal replacement. Journal of Plankton Research 24: 901–912.

Lai, J., Y. Zou, J. Zhang & P. R. Peres-Neto, 2022. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. Methods in Ecology and Evolution 13: 782–788.

Le Quéré, C., S. P. Harrison, I. C. Prentice & E. T. Buitenhuis, 2010. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Global Change Biology 11: 2016–2040.

Legendre, P. & E. D. Gallagher, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271–280.

Moreti, L. O. R., L. Martos, V. M. Bovo-Scomparin & L. C. Rodrigues, 2013. Spatial and temporal fluctuation of phytoplankton functional groups in a tropical reservoir. Acta Scientiarum Biological Sciences 35: 359–366.

Mutshinda, C. M., Z. V. Finkel, C. E. Widdicombe & A. J. Irwin, 2016. Ecological equivalence of species within phytoplankton functional groups. Functional Ecology 30: 1–9.

Okasen, J., Kindt, R., Legendre, P., O'Hara, R.B., 2007. Vegan: community ecology package version 1.8–6. https://cran.r-project.org/web/packages/vegan.

Padisák, J., F. A. R. Barbosa, R. Koschel & L. Krienitz, 2003. Deep layer cyanoprokaryota maxima in temperate and tropical lakes. Advances in Limnology 58: 175–199.

Padisák, J., G. Borics, I. Grigorszky & É. Soróczki-Pintér, 2006. Use of phytoplankton assemblages for monitoring ecological status of lakes within the water framework directive: the assemblage index. Hydrobiologia 553: 1–14.

Padisák, J., L. O. Crossetti & L. Naselli-Flores, 2009. Use and misuse in the application of the phytoplankton functional classification: a critical review with updates. Hydrobiologia 621: 1–19.

Peña, M. A., 2003. Plankton size classes, functional groups and ecosystem dynamics: an introduction. Progress in Oceanography 57: 239–242.

Reynolds, C. S., V. Huszar, C. Kruk, L. Naselli-Flores & M. Sergio, 2002. Towards a functional classification of the freshwater phytoplankton. Journal of Plankton Research 24: 417–428.

Roberts, D.W., 2006. labdsv: Laboratory for Dynamic Synthetic Vegephenomenology. http://CRAN.R-project.org/package=labdsv.

Rousseeuw, P., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational & Applied Mathematics 20: 53–65.

Santana, L. M., L. O. Crossetti & C. Ferragut, 2017. Ecological status assessment of tropical reservoirs through the assemblage index of phytoplankton functional groups. Brazilian Journal of Botany 40: 695–704.

Seip, K. L. & C. S. Reynolds, 1995. Phytoplankton functional attributes along trophic gradient and season. Limnology and Oceanography 40: 589–597.

Vandecandelaere, M., 2012. [R] Package "glmulti": Include a variable in ALL models. http://CRAN.R-project.org/package=glmulti.

Wallhead, P. J., A. P. Martin, M. A. Srokosz & P. J. S. Franks, 2009. Skill assessment via cross-validation and Monte Carlo simulation: An application to Georges Bank plankton models. Journal of Marine Systems 76: 134–150.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, Houston. http://CRAN.R-project.org/package=ggplot2.

Wilke, C.O., 2017. ggridges: Ridgeline Plots in 'ggplot2'. http://CRAN.R-project.org/package=ggridges.

Zhang, M., R. A. Smyth, W. Zhu, L. Zhang, Y. Li, Y. Wang, X. Li, Q. Gu & Y. Gao, 2019a. Spatial distribution and filtering efficiency of *Daphnia* in a deep subtropical reservoir. Journal of Oceanology and Limnology 37: 1277–1288.

Zhang, M., Y. Wang, B. Gu, Y. Li, W. Zhu, L. Zhang, L. Yang & X. Li, 2019b. Resources utilization and trophic niche between silver carp and bighead carp in two mesotrophic deep reservoirs. Journal of Freshwater Ecology 34: 199–212.