Supplementary Material

for

Regulatory pathways of phytoplankton regime shifts in shallow lakes: a new insight from non-equilibrium statistical mechanics

Yuting Yan[1], Feng Yu[1], Wenlong Zhang[1]*, Zikang Tang[1], Yanan Ding[1,2], Yi Li[1]

[1] State Key Laboratory of Water Cycle and Water Security, College of Environment, Hohai University, Nanjing, 210098, P. R. China

[2] Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment, Nanjing 210042, P. R. China

* Correspondence: Wenlong Zhang, 1223zhangwenlong@163.com

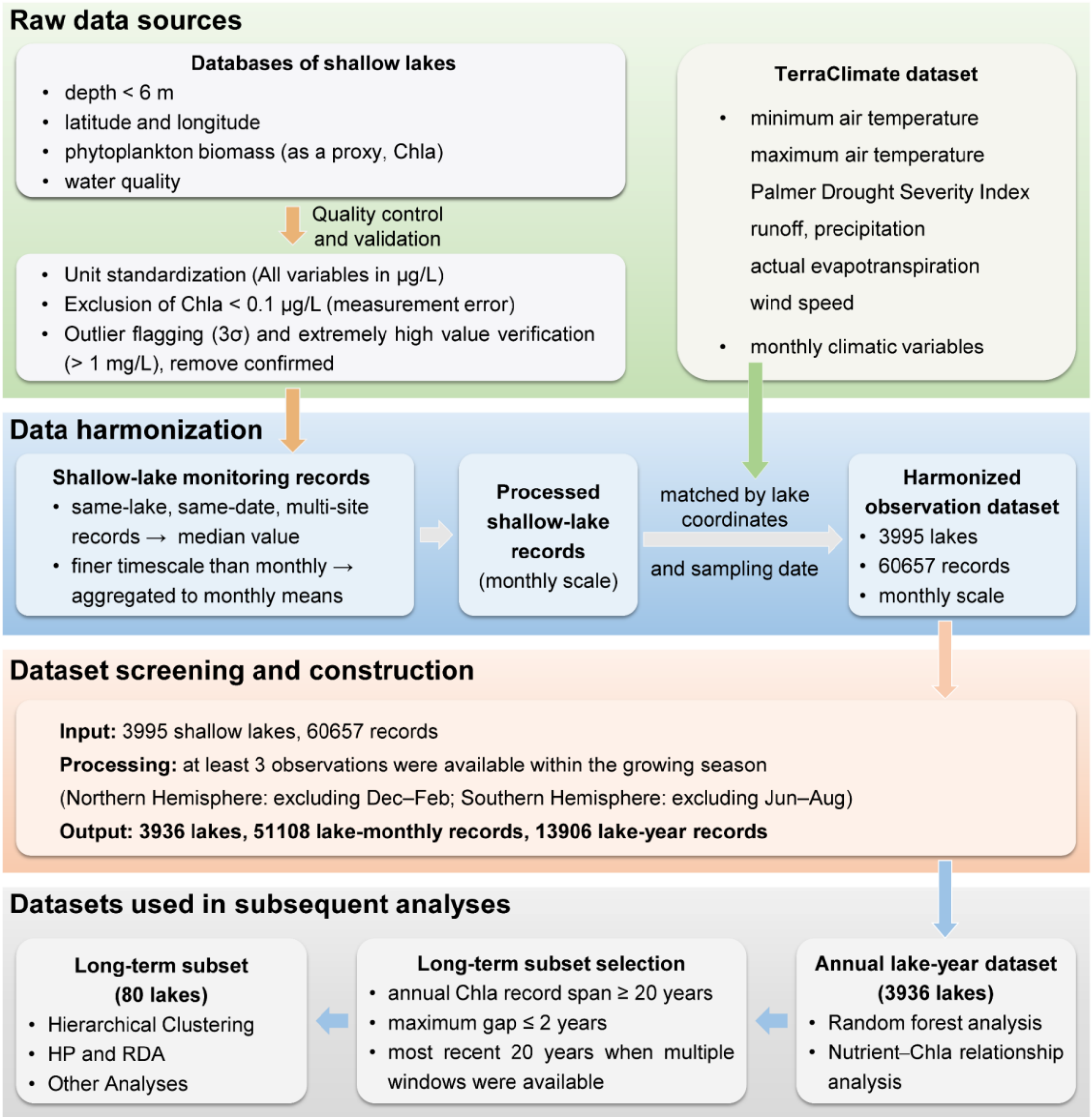

Figure S1 Workflow of data harmonization, screening, and construction of analysis datasets. During quality control and validation, all water quality variables (Chla, TN, TP) were recorded in consistent units (μg/L). Chla measurements below 0.1 μg/L were excluded due to the possibility of measurement error, as analytical equipment used within studies rarely had better precision. For all water chemistry variables, observations exceeding three standard deviations from the mean were flagged and cross-checked against the original data sources; extremely high values (> 1 mg/L) were additionally traced back to the original datasets to identify records potentially resulting from unit conversion errors or compilation errors. All flagged records confirmed as unreliable were subsequently discarded.

Water-quality observations from global shallow-lake databases, including phytoplankton biomass (as a proxy, chlorophyll-a), water quality, latitude and longitude, and climatic variables, were preprocessed and aligned to monthly resolution. Same-lake, same-date observations from multiple sampling sites were merged using the median value, and records collected at finer-than-monthly resolution were aggregated to monthly means; records without month information were excluded. The processed lake monitoring records were then matched with monthly climatic variables from the TerraClimate dataset according to lake coordinates and sampling month/year, yielding a harmonized lake–climate dataset containing 60657 records from 3995 lakes.

The harmonized monthly dataset was then screened to retain only growing-season records from lake-years with at least three observations within the growing season. This screening yielded 51108 growing-season lake-month records from 3936 lakes, from which 13906 lake-year records were generated by annual averaging. The growing season was defined as all months except December–February in the Northern Hemisphere and June–August in the Southern Hemisphere. The lake-year dataset was used for the random forest analysis.

For long-term analyses, 80 lakes were further selected by requiring annual Chla records spanning at least 20 years with a maximum gap of 2 years; when multiple eligible windows were available, the most recent 20-year window was retained. Annual records from these 80 lakes were used for hierarchical clustering, whereas the corresponding growing-season monthly records were used for HP, RDA, and other analyses unless otherwise specified.

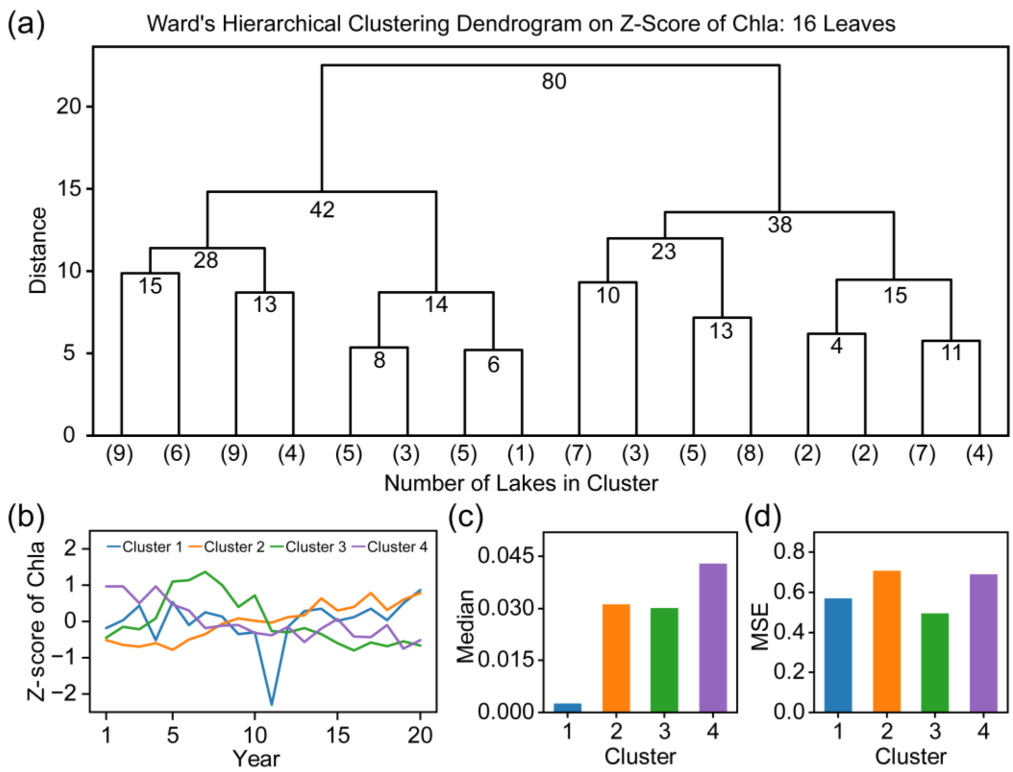

Figure S2 Hierarchical clustering tree (a), clustering trend (b), median of each cluster (c), mean square error of each cluster (d).

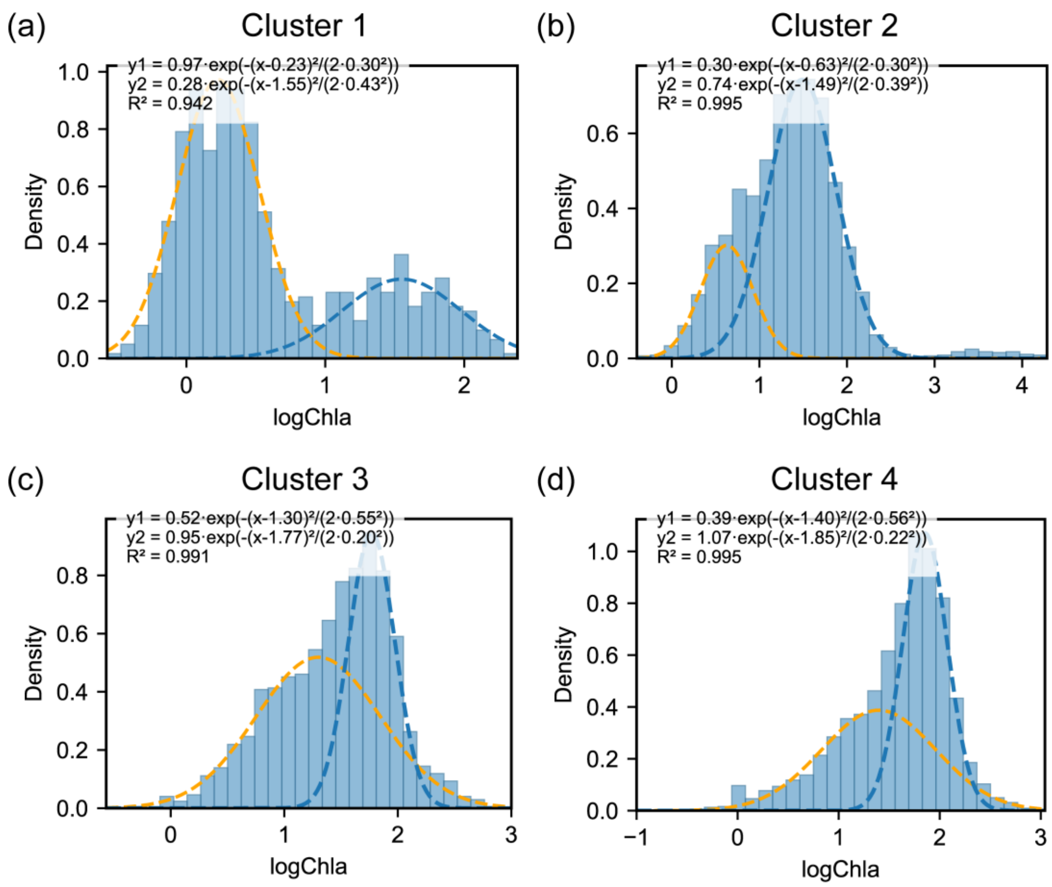

Figure S3 The frequency distributions of logChla for each cluster (a-d). Yellow dotted line, y1; blue dotted line, y2.

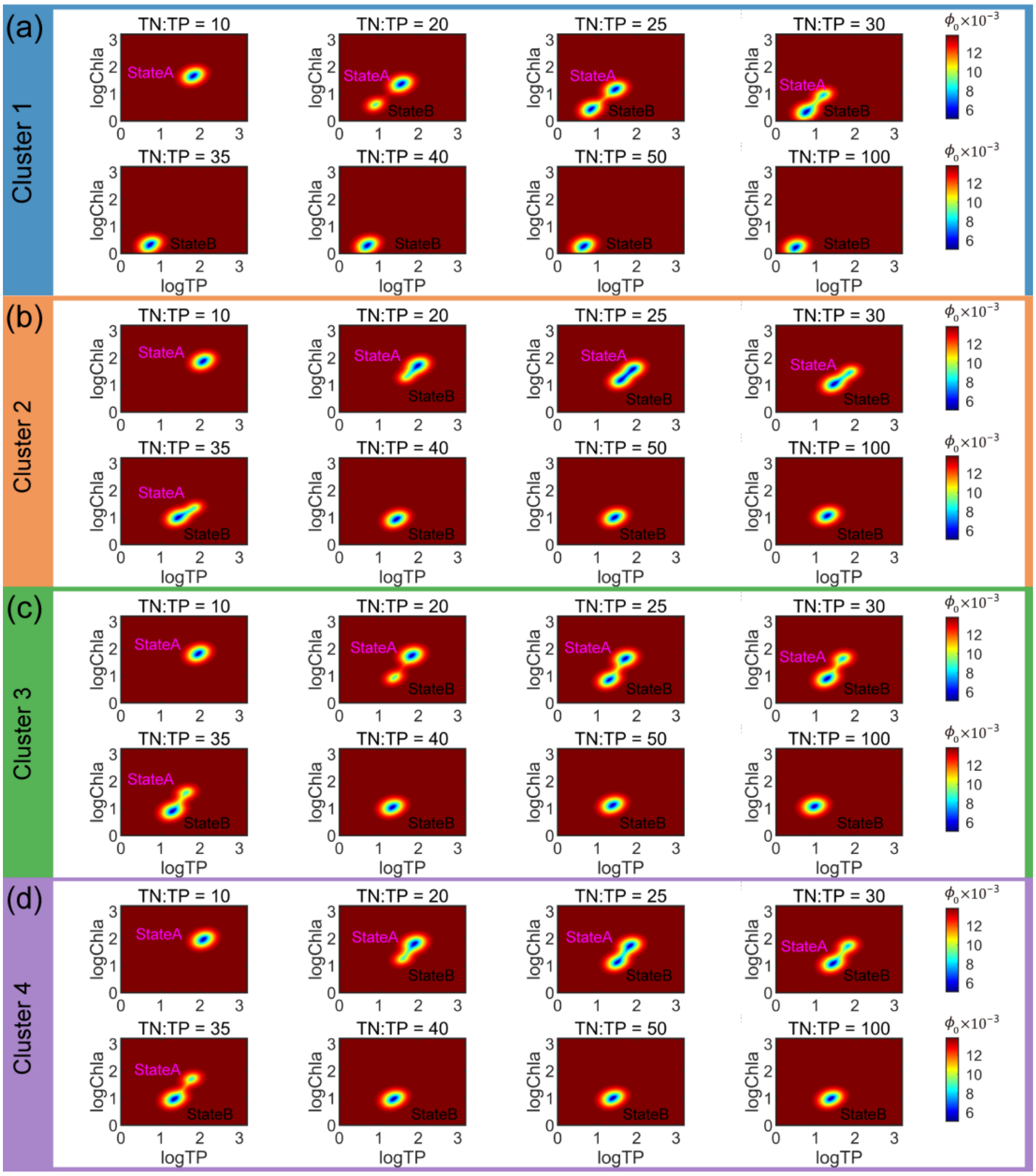

Figure S4 2D intrinsic potential landscapes under zero fluctuations (ϕ_0) for the SLPNF model across different clusters (a-d) along the TN:TP gradient (10–100). Colorbar on the right, ranging from blue to red, represents potential energy decreasing from high to low.

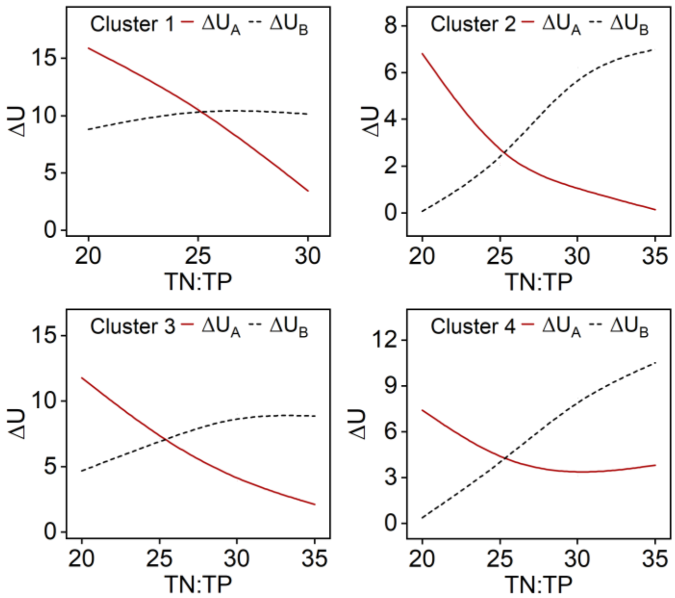

Figure S5 The barrier heights (ΔU) versus TN:TP for each cluster. ΔU_A, the barrier height from StateA to StateB; ΔU_B, the barrier height from StateB to StateA.

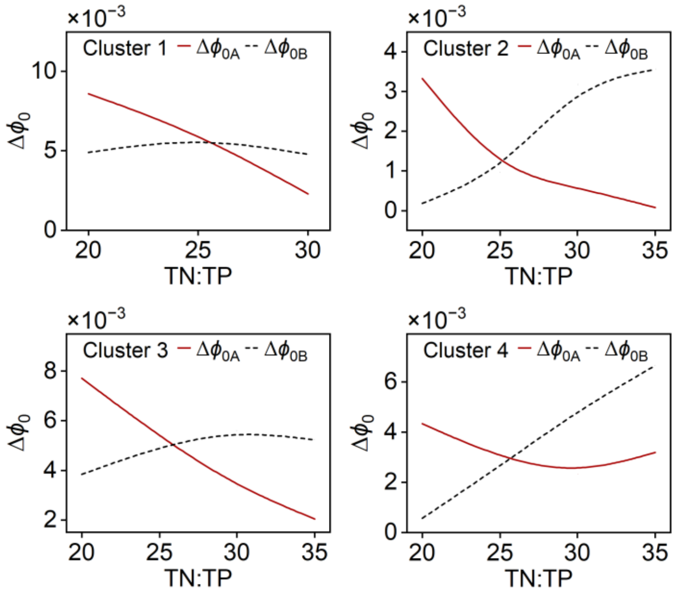

Figure S6 The intrinsic barrier heights (ϕ_0) versus TN:TP for each cluster. $\Delta\phi_{0A}$, the intrinsic barrier height from StateA to StateB; $\Delta\phi_{0B}$, the intrinsic barrier height from StateB to StateA.

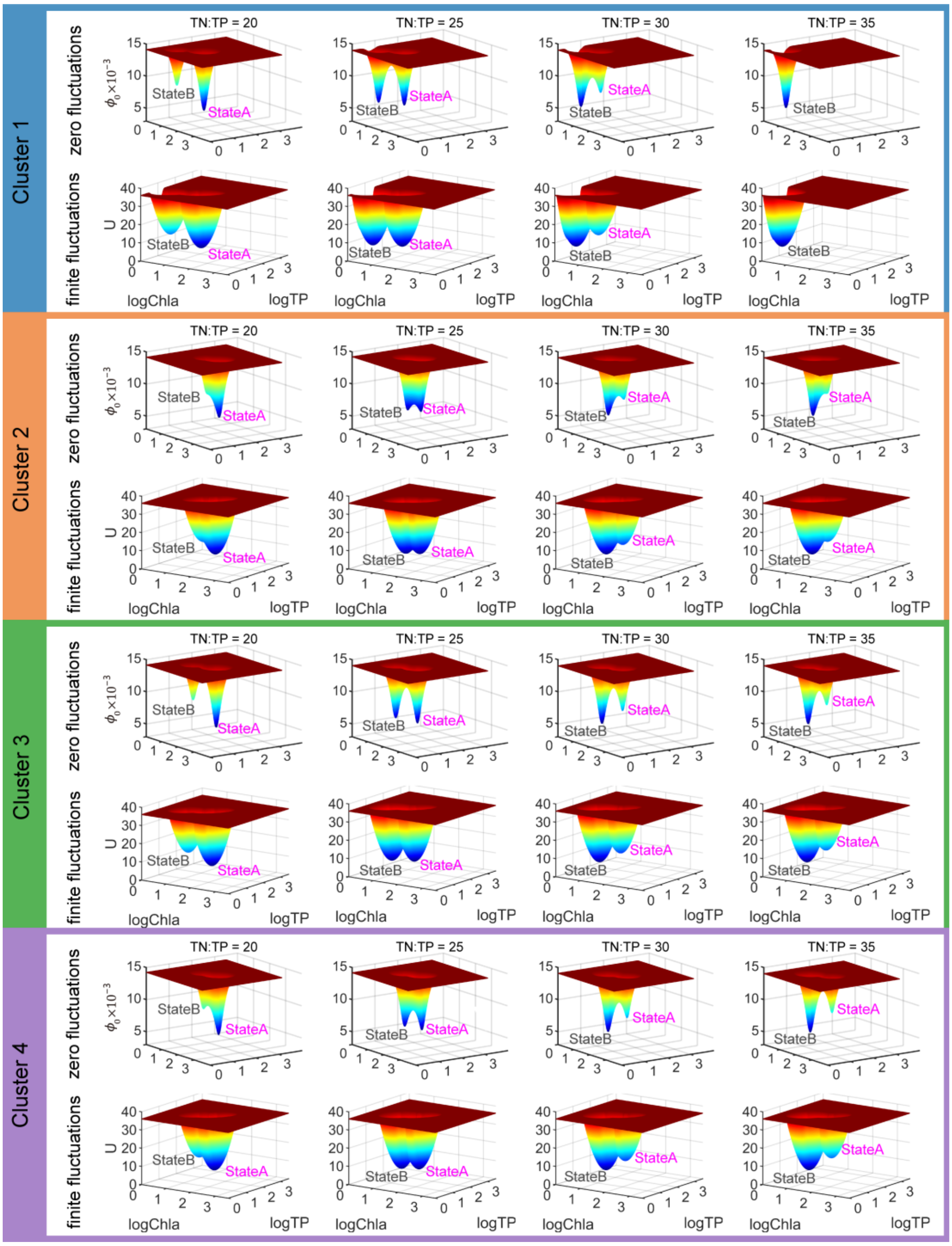

Figure S7 3D potential landscapes under zero fluctuations (ϕ_0) and finite fluctuations (U) for the SLPNF model under different clusters along the TN:TP gradient (20–35).

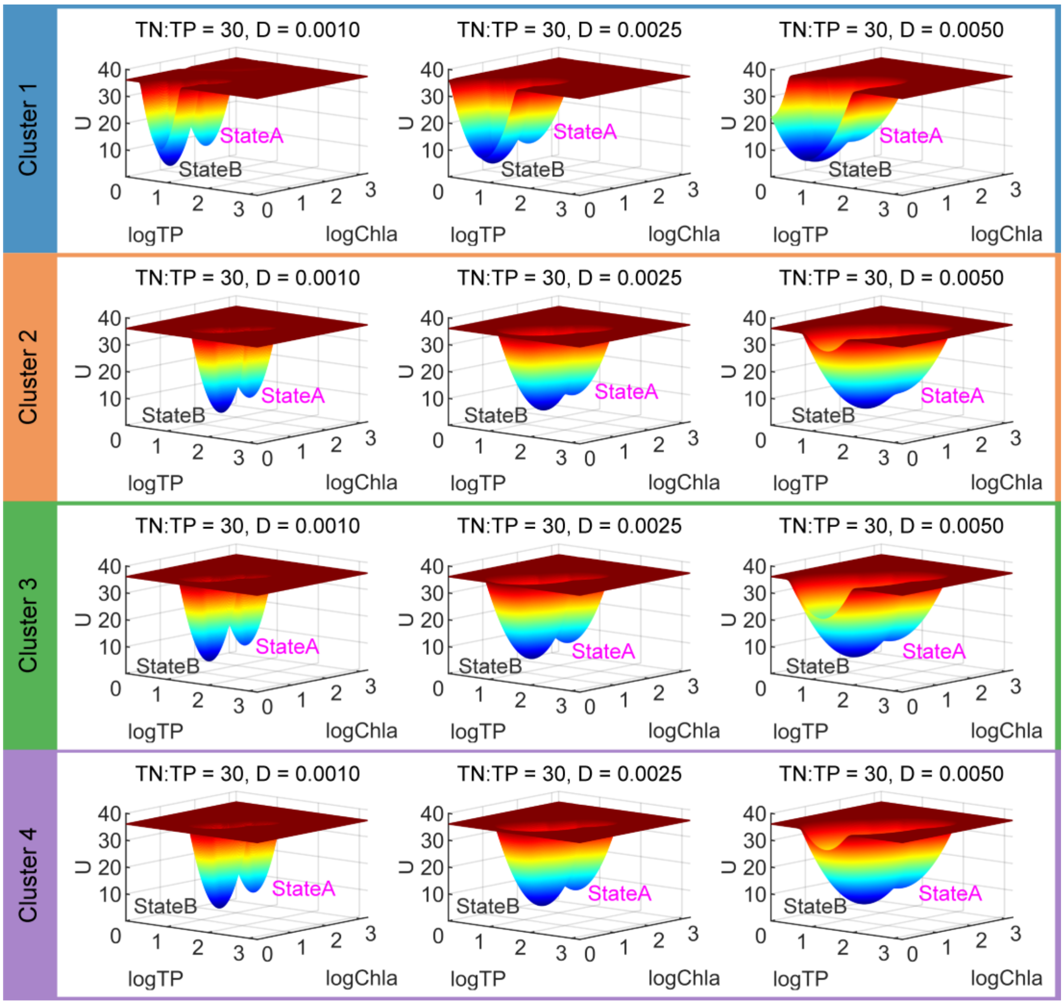

Figure S8 3D potential landscapes for the SLPNF model under different clusters with different parameters $D = 0.001$, 0.0025, 0.005 and TN:TP = 30. Parameter D was used to measure the strength of perturbations, which was defined as the diffusion coefficient of fluctuation intensity.

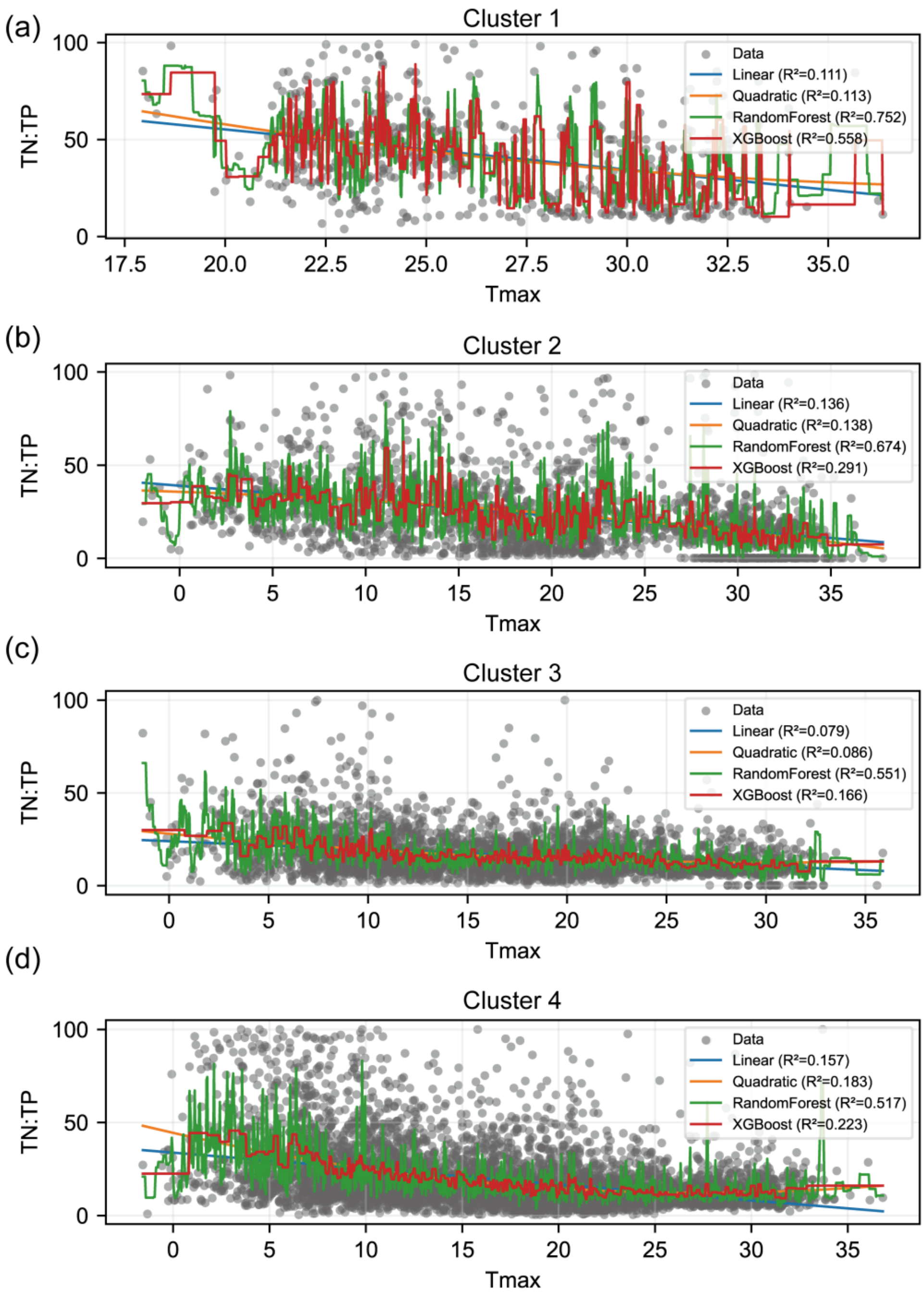

Figure S9 The model fitting performance for predicting the relationship between maximum temperature (Tmax) and TN:TP was evaluated to select the best surrogate model under different clusters (a-d), respectively. The random forest model was selected as the best surrogate model for subsequent sensitivity analysis.

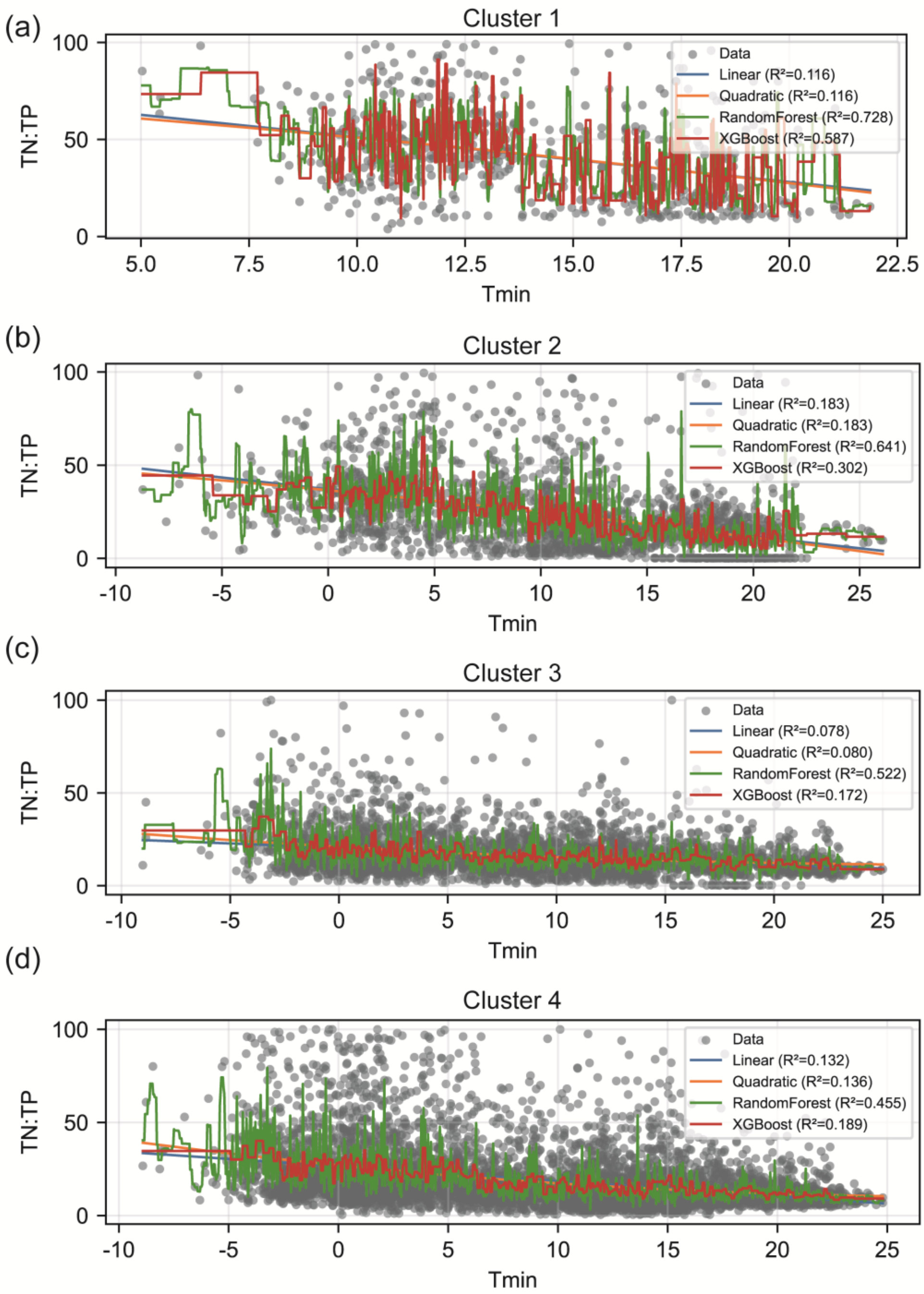

Figure S10 The model fitting performance for predicting the relationship between minimum temperature (Tmin) and TN:TP was evaluated to select the best surrogate model under different clusters (a-d), respectively. The random forest model was selected as the best surrogate model for subsequent sensitivity analysis.

Table S1 Information associated with each of the online databases.

Data type	Name	Data sources
Water quality data	A water quality database for global lakes	Naderian et al. (2024)
	The Danish Centre for Environment at the Aarhus University	Davidson et al. (2023) https://odaforalle.au.dk/
	Environmental Information Data Centre（UK）	https://catalogue.ceh.ac.uk/eidc/documents
	Long term and high frequency monitoring of lakes in the English Lake District	https://catalogue.ceh.ac.uk/documents/bf30d6aa-345a-4771-8417-ffbcf8c08c28
	Ministry of Land, Infrastructure, Transport and Tourism, Japan	https://www1.river.go.jp/
	The National Institute for Environmental Studies (NIES), Japan	https://www.nies.go.jp/chiiki/en/project10_en.html
	Government of New Brunswick(Canada) - Department of Environment and Local Government Lake Water Quality Data	https://www.elgegl.gnb.ca/LakesNB-NBLacs/en/SamplingLocation/Index
	A comprehensive time-series dataset linked to cyanobacterial blooms in Lake Taihu	Xue et al. (2024)
Climatic data	TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015	Abatzoglou et al. (2018)

Table S2 Distribution of collected shallow lake data by country.

Country	Lakes	Country	Lakes
Canada	114	Norway	3
Switzerland	4	Netherlands	27
China	9	New Zealand	2
Czech Republic	4	Poland	38
Germany	13	The Portuguese Republic	1
Denmark	2254	Sweden	20
Spain	11	Slovakia	1
Estonia	3	Slovenia	3
Finland	19	Turkey	13
Japan	9	The United States of America	1235
Mexico	153		

Table S3 Distribution of the 80 long-term monitoring shallow lakes by continent.

Continent	Lakes
North America	41
Asia	9
Oceania	2
Europe	28

Table S4 Summary of the parameters interpretation and default values for SLPNF model.

Symbol	Default values [a]	Definition and interpretation	Reference
A		Concentration of phytoplankton biomass ($\log_{10}(Chla(ug/L))$)	
N		Concentration of total nitrogen ($\log_{10}(TN(ug/L))$)	
P		Concentration of total phosphorus ($\log_{10}(TP(ug/L))$)	
R		Total nitrogen to total phosphorus ratios	
$\mu_{\max}$	0.85	Maximum growth rate of phytoplankton ($1/day$)	Reynolds (2006a), Laws (2013)
K_N	0.15	Half saturation constant of nitrogen concentration to phytoplankton (mg/L)	Mulder and Hendriks (2014)
K_P	0.08	Half saturation constant of phosphorus concentration to phytoplankton (mg/L)	Mulder and Hendriks (2014)
λ_1	0.25	Intrinsic decay coefficient of phytoplankton ($1/day$)	Tilman (1981), Reynolds (2006b)
λ_2	0.10	Competitive mortality coefficient of phytoplankton $1/(mg*L^{-1}*day)$	Reynolds (2006b)
α_N	0.50	Uptake rate of nitrogen by phytoplankton	Tilman (1981), Reynolds (2006c)
α_P	0.60	Uptake rate of phosphorus by phytoplankton	Tilman (1981), Reynolds (2006c)
θ_N	0.36	Phytoplankton conversion into nitrogen after death	Reynolds (2006b), Tiwari et al. (2019)
θ_P	0.55	Phytoplankton conversion into phosphorus after death	Reynolds (2006b), Tiwari et al. (2019)
I_N	0.002	Input rate of nitrogen ($ug*L^{-1}*day$)	Janse and Proefschrift (2005)
I_P	0.0015	Input rate of phosphorus ($ug*L^{-1}*day$)	Jørgensen (1988), Janse and Proefschrift (2005)

[a] For each cluster, due to differences in features, the default values were calibrated to ensure the accuracy and adaptability of the SLPNF model.

Table S5 Results of hierarchical partitioning for the RDA model.

Clusters	RDA maximum variance explained	Variable	Independent Contribution	p
Cluster 1	29.2%	minimum temperature	11.29%	0.001
		maximum temperature	11.03%	0.001
		wind speed	3.74%	0.001
		actual evapotranspiration	1.66%	0.001
		Palmer Drought Severity Index	0.65%	0.985
		precipitation	0.42%	0.891
		runoff	0.37%	0.072
Cluster 2	21.62%	minimum temperature	9.02%	0.001
		maximum temperature	5.58%	0.001
		actual evapotranspiration	2.43%	0.001
		runoff	2.02%	0.001
		wind speed	1.13%	0.001
		Palmer Drought Severity Index	0.72%	0.001
		precipitation	0.72%	0.001
Cluster 3	9.2%	maximum temperature	3.52%	0.001
		minimum temperature	2.79%	0.001
		actual evapotranspiration	1.16%	0.001
		wind speed	0.98%	0.001
		Palmer Drought Severity Index	0.32%	0.001
		precipitation	0.28%	0.001
		runoff	0.20%	0.002
Cluster 4	10.51%	minimum temperature	4.65%	0.001
		maximum temperature	3.12%	0.001
		actual evapotranspiration	1.26%	0.001
		wind speed	0.56%	0.001
		precipitation	0.44%	0.001
		Palmer Drought Severity Index	0.31%	0.001
		runoff	0.18%	0.001

Table S6 The regulation threshold for each cluster.

Cluster	Regulation threshold
Cluster 1	22.7820
Cluster 2	23.0681
Cluster 3	23.0681
Cluster 4	23.2583

References

Abatzoglou J T, Dobrowski S Z, Parks S A, and Hegewisch K C (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Scientific Data, 5(1): 170191

Davidson T A, Sayer C D, Jeppesen E, Søndergaard M, Lauridsen T L, Johansson L S, Baker A, and Graeber D (2023). Bimodality and alternative equilibria do not help explain long-term patterns in shallow lake chlorophyll-a. Nature Communications, 14(1): 398

Janse J H (2005). Model studies on the eutrophication of shallow lakes and ditches. Wageninge: Wageningen Universiteit

Jørgensen S E (1988). 10 - modelling eutrophication of shallow lakes. in: Mitsch W J, Straškraba M, and Jørgensen S E, eds. Developments in environmental modelling. Elsevier, 177–188

Laws E A (2013). Evaluation of In situ phytoplankton growth rates: a synthesis of data from varied approaches.

Mulder C, and Hendriks A J (2014). Half-saturation constants in functional responses. Global Ecology and Conservation, 2: 161–169

Naderian D, Noori R, Heggy E, Bateni S M, Bhattarai R, Nohegar A, and Sharma S (2024). A water quality database for global lakes. Resources, Conservation and Recycling, 202: 107401

Reynolds C S, ed. (2006a). Growth and replication of phytoplankton. The Ecology of Phytoplankton. Cambridge: Cambridge University Press, 178–238

Reynolds C S, ed. (2006b). Mortality and loss processes in phytoplankton. The Ecology of Phytoplankton. Cambridge: Cambridge University Press, 239–301

Reynolds C S, ed. (2006c). Nutrient uptake and assimilation in phytoplankton. The Ecology of Phytoplankton. Cambridge: Cambridge University Press, 145–177

Tilman D (1981). Tests of resource competition theory using four species of lake michigan algae. Ecology, 62(3): 802–815

Tiwari P K, Samanta S, Ferreira J D, and Misra A K (2019). A mathematical model for the effects of nitrogen and phosphorus on algal blooms. International Journal of Bifurcation and Chaos, 29(10): 1950129

Xue K, Ma R, Zhu G, Hu M, Cao Z, Xiong J, Zhang Y, Xu J, Huang Z, and Wu Y (2024). A comprehensive time-series dataset linked to cyanobacterial blooms in lake taihu. Scientific Data, 11(1): 1365