Supplement of Hydrol. Earth Syst. Sci., 30, 4611–4628, 2026
https://doi.org/10.5194/hess-30-4611-2026-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Four decades of full-depth profiles reveal layer-resolved drivers of reservoir thermal regimes and episodic hypolimnetic warming

Chenxi Mi et al.

Correspondence to: Xiangzhen Kong (xzkong@niglas.ac.cn)

Tables

Table S1. CE-QUAL-W2 parameter set for the Rappbode Reservoir

Parameter	Description	Value
DLTMIN	Minimum timestep (sec)	1
BETA	Fraction of incident solar radiation absorbed at the water surface	0.45
WSC	Wind sheltering coefficient (-)	1
AX	Longitudinal eddy viscosity ($m^2\ sec^{-1}$)	1
DX	Longitudinal eddy diffusivity ($m^2\ sec^{-1}$)	1
TSED	Sediment temperature (°C)	10
TSEDF	Heat lost to sediments that is added back to water column	1
Z0	Wind roughness height (m)	0.001
FI	Interfacial Friction (-)	0.015
EXH2O	Light extinction for pure water (m^{-1})	0.55
HWI	Coefficient of water-ice heat exchange ($W\ m^{-2}\ °C^{-1}$)	10
SHADE	Shade fraction coefficient (-)	1

Table S2. Descriptive statistics of the hydro-meteorological predictor variables and thermal target variables used for machine-learning model development in Rappbode Reservoir. Summary metrics are reported as mean, median, standard deviation (SD), minimum, and maximum. The target variables comprise water temperature at 5, 15, 30, and 50 m depth, bottom-to-surface temperature difference, mixed-layer depth, and Schmidt stability.

Variable	Unit	Mean	Median	SD	Min	Max
Air temperature	°C	7.68	7.91	7.32	-21.90	26.97
Relative humidity	%	79.98	81.58	11.49	25.92	100.00
Wind speed	$m\ s^{-1}$	3.63	3.29	1.71	0.00	12.76
Wind direction	rad	3.79	4.04	1.28	0.00	6.28
Cloud cover	0–10 scale	6.73	7.34	2.63	0.00	10.00
Short-wave radiation	$W\ m^{-2}$	122.18	104.18	89.33	0.91	457.57
Outflow discharge	$m^3\ s^{-1}$	1.47	0.75	2.02	0.00	21.00
Water temp (5 m depth)	°C	15.76	16.16	3.77	4.86	24.71
Water temp (15 m depth)	°C	8.83	8.37	2.89	4.11	16.74
Water temp (30 m depth)	°C	5.55	5.32	1.23	4.01	8.64
Water temp (50 m depth)	°C	5.09	4.94	0.87	4.01	6.97
Bottom-to-surface temperature difference	°C	11.09	11.56	3.78	1.00	20.25
Mixed-layer depth	m	10.24	8.00	9.39	2.01	25.52
Schmidt stability	$J\ m^{-2}$	1339.52	1357.39	695.99	5.16	2896.95

Table S3. Distributional comparison of the target specific training and testing subsets. For each of the seven modeling targets, the Kolmogorov–Smirnov (KS) statistic and *p* value, Jensen–Shannon (JS) distance, and energy distance are reported based on the actual train–test partition used in model development. The KS statistic, *p* value, and JS distance are dimensionless. The energy distance has the same unit as the corresponding target variable.

Target variable	KS statistic	KS *p*-value	JS distance	Energy distance
Water temp (5 m depth)	0.018	0.891	0.005	0.002
Water temp (15 m depth)	0.021	0.706	0.004	0.001
Water temp (30 m depth)	0.017	0.883	0.013	0.001
Water temp (50 m depth)	0.016	0.944	0.012	0.001
Bottom-to-surface temperature difference	0.019	0.811	0.005	0.001
Mixed-layer depth	0.014	0.974	0.006	0.006
Schmidt stability	0.015	0.976	0.003	0.115

Table S4. Performance of CE-QUAL-W2 in reproducing observed water level and depth-specific water temperatures in Rappbode Reservoir. RMSE is expressed in m for water level and in °C for water temperature.

Target variable	R^2	RMSE
Water level	0.99	0.06
Water temp (5 m depth)	0.96	1.27
Water temp (15 m depth)	0.83	1.31
Water temp (30 m depth)	0.75	0.75
Water temp (50 m depth)	0.70	0.59

Table S5. Statistical comparison of the predictive performance of XGBoost, SVR, RF, and MARS across the seven target variables. Kruskal–Wallis tests were conducted on the test-set squared errors of the four models to evaluate overall differences in error distributions, followed by pairwise Diebold–Mariano tests comparing XGBoost with each benchmark model. All entries are *p*-values.

Target variable	KW *p*-value	DM *p*-value (XGB vs SVR)	DM *p*-value (XGB vs RF)	DM *p*-value (XGB vs MARS)
Water temp (5 m depth)	$< 2.2 \times 10^{-16}$	1.07×10^{-10}	0.72	$< 2.2 \times 10^{-16}$
Water temp (15 m depth)	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
Water temp (30 m depth)	$< 2.2 \times 10^{-16}$	5.27×10^{-10}	0.27	$< 2.2 \times 10^{-16}$
Water temp (50 m depth)	$< 2.2 \times 10^{-16}$	2.19×10^{-10}	4.23×10^{-4}	3.22×10^{-16}
Bottom-to-surface temperature difference	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	0.74	$< 2.2 \times 10^{-16}$
Mixed-layer depth	$< 2.2 \times 10^{-16}$	1.05×10^{-3}	0.02	2.92×10^{-5}
Schmidt stability	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Table S6. Coupling-weighted CE-QUAL-W2 counterfactual response index (CWCRI) for meteorological forcings across representative depths. MA indicates moving-average features.

Target variable	T_air (MA)	T_dew (MA)	Solar (MA)	Cloud (MA)
Water temp (5 m depth)	0.209	0.119	0.173	0.027
Water temp (15 m depth)	0.040	0.065	0.153	0.002
Water temp (30 m depth)	0.022	0.023	0.159	0.016
Water temp (50 m depth)	0.007	0.040	0.144	0.011

S1. CE-QUAL-W2 numerical options for the Rappbode Reservoir

In CE-QUAL-W2 (hereafter W2), surface heat exchange was computed component-wise, i.e., shortwave and longwave radiation, sensible, and latent heat fluxes were each evaluated with the model's bulk-aerodynamic formulations. For scalar transport, we adopted the ULTIMATE advection scheme, which uses flux limiting to reduce numerical diffusion while preventing spurious overshoots/undershoots (i.e., maintaining monotonicity and boundedness). Vertical mixing was represented with the turbulent kinetic energy (TKE) closure; consistent with the W2 manual, the maximum vertical eddy viscosity was capped at 1 m^2 s^{-1}. The calibration procedure, calibration targets, and evaluation strategy are described in Section 2.3 of the main text. The complete parameter set used for the Rappbode Reservoir is provided in Tab. S1.

S2. Background information on SHAP values

In our study, the final Shapley value for each predictor is obtained by taking a weighted average of these marginal contributions, thus yielding a unified and strictly additive measure of feature importance. For a model with M input variables, the Shapley contribution of feature i to a single prediction f(x) is defined as (Lundberg et al. 2020):

$$\varphi_i = \sum_{S \subseteq \mathcal{F}\{i\}} \frac{|S|!\,(M-|S|-1)!}{M!}\left[f_{S\cup\{i\}}\left(\mathbf{x}_{S\cup\{i\}}\right) - f_S(\mathbf{x}_S)\right]$$

where $\mathcal{F}$ $=\{1,\ldots,M\}$ is the full feature set, S is any subset that excludes i, and $f_S(\mathbf{x}_S)$=E$[f(x)|(\mathbf{x}_S)]$ denotes the conditional expectation of the model output when only the features in S are known. The combinatorial weight $\frac{|S|!(M-|S|-1)!}{M!}$ guarantees symmetry and fairness over all $M!$ permutations of the input vector, thereby satisfying the axioms of efficiency and additivity.

S3. Thermal indices and model interpretation

Three thermal indices (e.g., Schmidt Stability, the bottom-to-surface temperature difference and mixed-layer depth, see 2.4.2) were applied to further evaluate stratification intensity based on the water temperature output W2, which were also used to illustrate the capability of XGBoost in reproducing the thermal feature of the reservoir. All three have been commonly adopted in previous studies to characterize the thermal structure of inland waters (Fang and Stefan 2009; Ladwig et al. 2021). Here, Schmidt Stability was defined as the energy per unit area required to fully mix a vertically stratified water to a uniform density state (Boehrer and Schultze 2008):

$$S = \frac{g}{A_s} \int_0^{z_{\max}} (z - z_*)(\rho_z - \rho_*) A_z dz$$

where S is the Schmidt stability (in J m^{-2}), g is the gravitational acceleration (m s^{-2}), A_s is the reservoir surface area (m^2) and z_{max} is the maximum reservoir depth (m), ρ_z and A_z denote the water density (kg m^{-3}) and horizontal cross-sectional area (m^2) at depth z, respectively. The mixed-layer depth was considered as the depth of minimum curvature of the temperature profile, where d^2T/dz^2 is at a minimum (see Kirillin et al. 2013).

All the statistical analysis and post-processing were performed in R version 4.2.2 (R core Team 2022). The packages "glmtools" (v0.16.0, Read et al. 2014) and "rLakeAnalyzer" (v1.11.4.1, Winslow et al. 2019) were applied for calculating the stratification indices mentioned above. The package "xgboost" (v1.7.8.1, Chen et al. 2015) was used to construct the XGBoost model and

“SHAPforxgboost” (v 0.1.0, Liu and Just 2020) to calculate the SHAP value. The package "tidyr" (v1.3.0, Wickham H et al. 2023) was adopted to clean the original database, and “ggplot2” (v3.4.1, Wickham 2016) for visualization.

S4. Targeted CE-QUAL-W2 counterfactual validation of SHAP-derived attributions

We conducted targeted CE-QUAL-W2 counterfactual simulations to test whether the two key SHAP-derived patterns were consistent with process-based model responses: (i) the high relative importance of antecedent shortwave radiation below the surface layer and (ii) the importance of outflow discharge for late-stratification hypolimnetic warming.

S4.1. Meteorological anomaly-removal scenarios

To evaluate the SHAP-derived meteorological ranking, we constructed four anomaly-removal scenarios from the calibrated baseline simulation. In each scenario, only one meteorological forcing was modified: shortwave radiation, air temperature, dew-point temperature, or cloud cover. For every year from 1 April to 31 October, the daily value of the target forcing was replaced by its 1981–2019 median for the same calendar day, while all other meteorological and hydrological forcings were kept unchanged. The replacement period began on 1 April to account for the 30-day antecedent predictors used in the SHAP analysis, while the water temperature response was evaluated from May to October, consistent with the machine-learning analysis.

For each depth z and forcing k, we designed a coupling-weighted counterfactual response index (CWCRI) to compare the process-model response with the SHAP ranking:

$$CWCRI_k(z)=nRMSE_k(z)\times\mid\rho(T_{BASE,z},X_{k,30d})\mid \qquad \text{(S1)}$$

where $nRMSE_k(z)$ represents the standardized W2 temperature response to removing the anomaly of forcing k, and $|\rho(T_{BASE,z}, X_{k,30d})|$ represents the historical coupling between baseline W2 temperature and the corresponding 30-day moving-average forcing used in the SHAP model. CWCRI is a dimensionless index. A higher value indicates stronger process-consistent influence of the given meteorological forcing on simulated temperature at a given depth, reflecting both a larger standardized response to anomaly removal and tighter coupling with baseline-simulated temperature variations. The resulting W2-response ranking closely reproduced the key SHAP-derived pattern: air temperature ranked highest at 5 m, while shortwave radiation ranked highest at subsurface layers (Tab. S6). This agreement strengthens confidence that the elevated SHAP contribution of antecedent shortwave radiation at greater depths reflects a process-consistent thermal response.

S4.2. Withdrawal counterfactual scenarios

To evaluate the SHAP-derived dominance of outflow discharge during late stratification, we constructed two targeted withdrawal scenarios for the four anomalous years, 1981, 2001, 2002, and 2007. In the P50 and P90 scenarios, bottom withdrawal from 1 August to 31 October was replaced by the median and 90th percentile, respectively, calculated from the corresponding calendar-day values after excluding the four anomalous years. The difference in bottom withdrawal relative to the baseline was redistributed to the upper outlet, so that total daily outflow was conserved. Meteorological forcing and inflow boundary conditions were kept unchanged. The thermal response was evaluated from 1 September to 31 October, when anomalous hypolimnetic warming occurred.

The scenario results suggested that reducing anomalously high bottom withdrawals strongly suppressed late-stratification warming at 50 m (Fig. S7). Across the four anomalous years, mean

September–October temperature at 50 m decreased from 8.05 °C in the baseline simulation to 6.30 °C in P90 and 6.00 °C in P50. The corresponding maximum temperature decreased from 11.37 °C (baseline) to 8.43 °C (P90) and 6.99 °C (P50). These results provide support for the interpretation that intensified bottom withdrawals contributed to hypolimnetic warming, consistent with the SHAP-derived dominance of outflow discharge during late stratification.

Figures

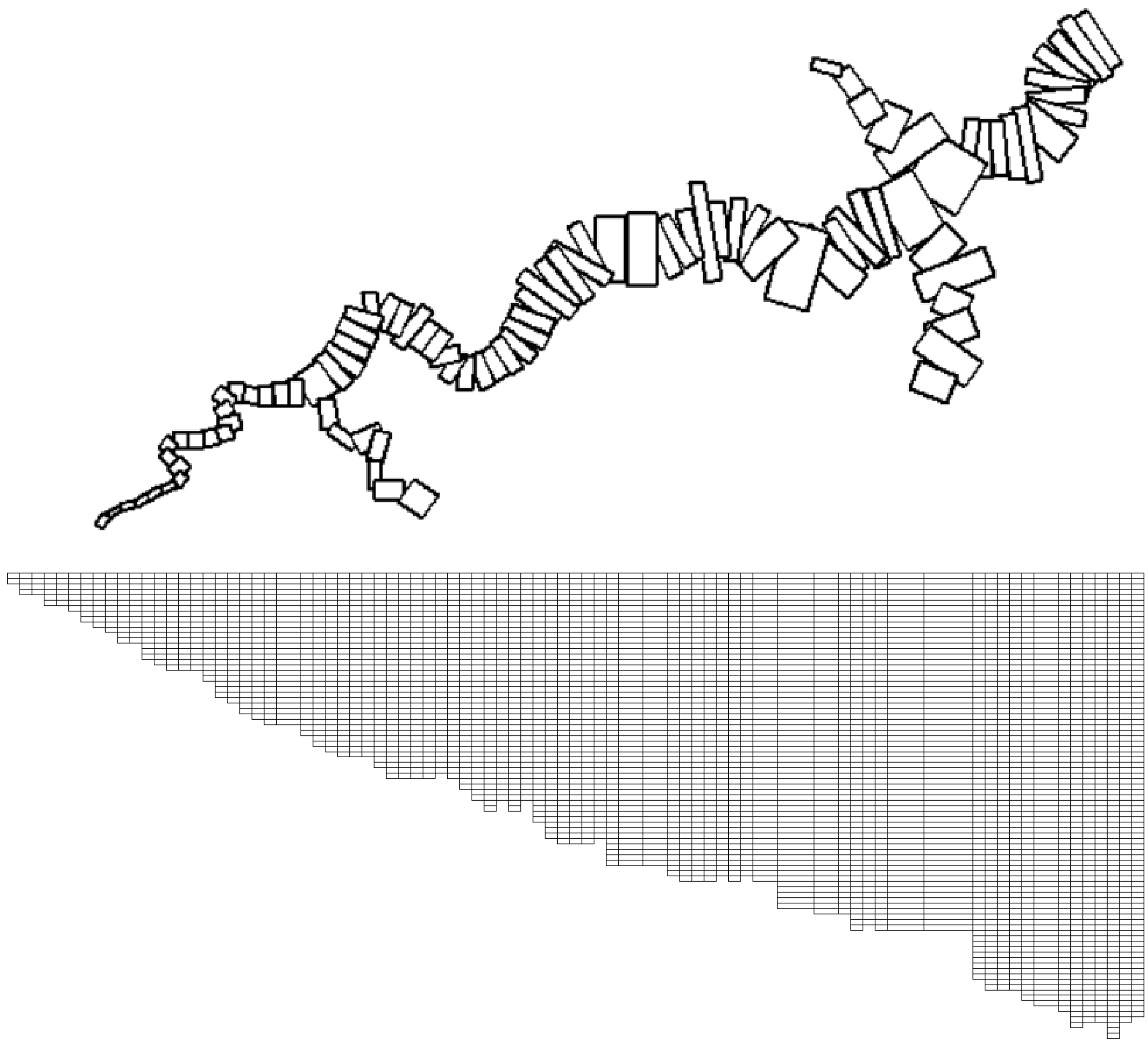

Fig. S1. CE-QUAL-W2 grid definition for Rappbode Reservoir in plan (top) and profile view (bottom).

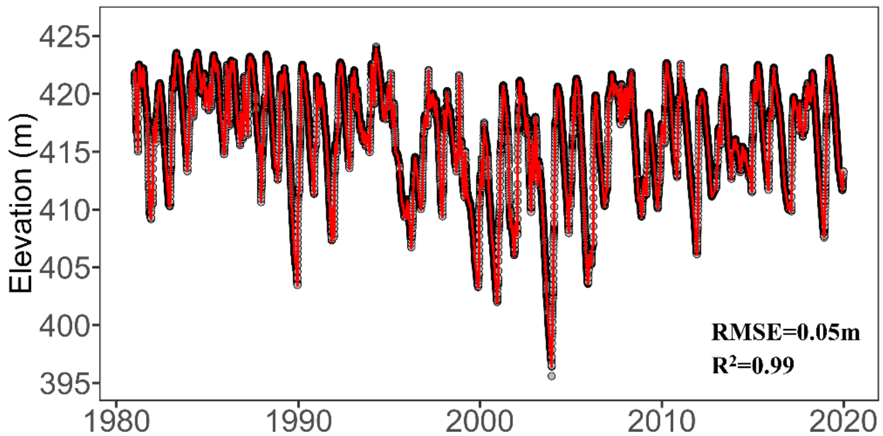

Fig. S2. Comparison of observed (the red line) and simulated (black points) water level, in Rappbode Reservoir, from 1981-2019

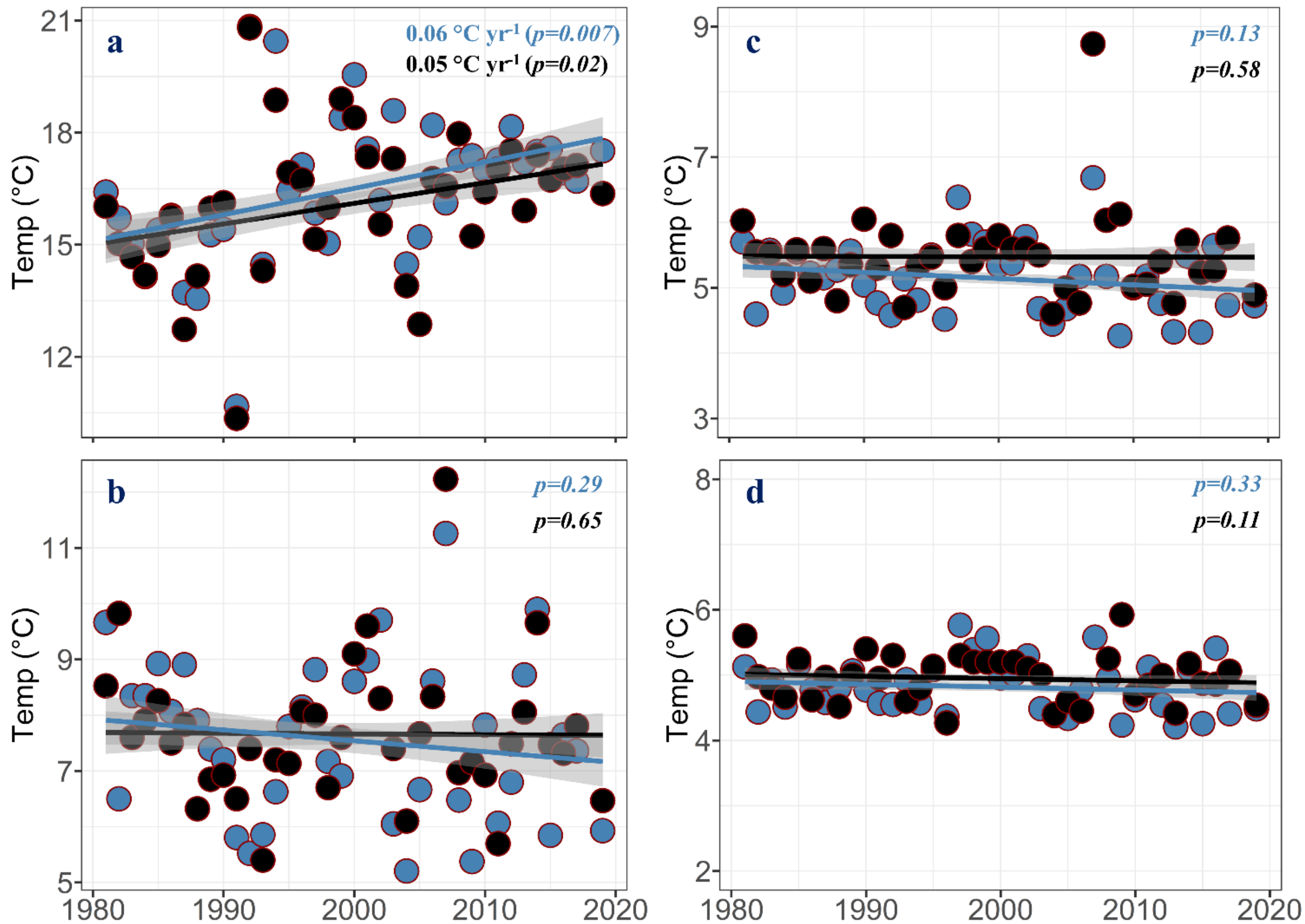

Fig. S3. Comparison of simulated (blue points) vs observed (black points) interannual variations in water temperature at different depths in the Rappbode Reservoir (a–d: 5 m, 15 m, 30 m, and 50 m, respectively)

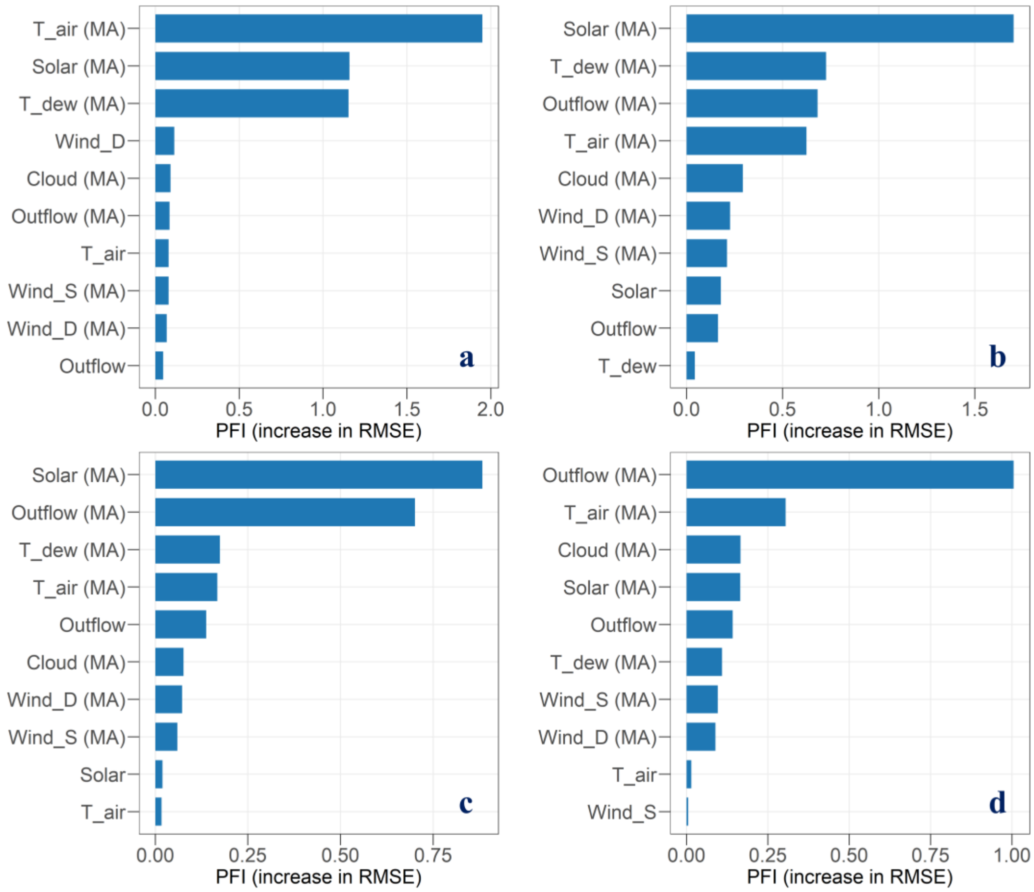

Fig. S4. Permutation feature importance (PFI) for the depth-specific water-temperature models at 5 m (a), 15 m (b), and 30 m (c), and for the late-stratification bottom-water temperature model at 50 m (d). PFI is expressed as the increase in test-set root mean square error (RMSE) after permutation of each predictor. Higher values indicate greater predictive importance. MA denotes 30-day moving-average features.

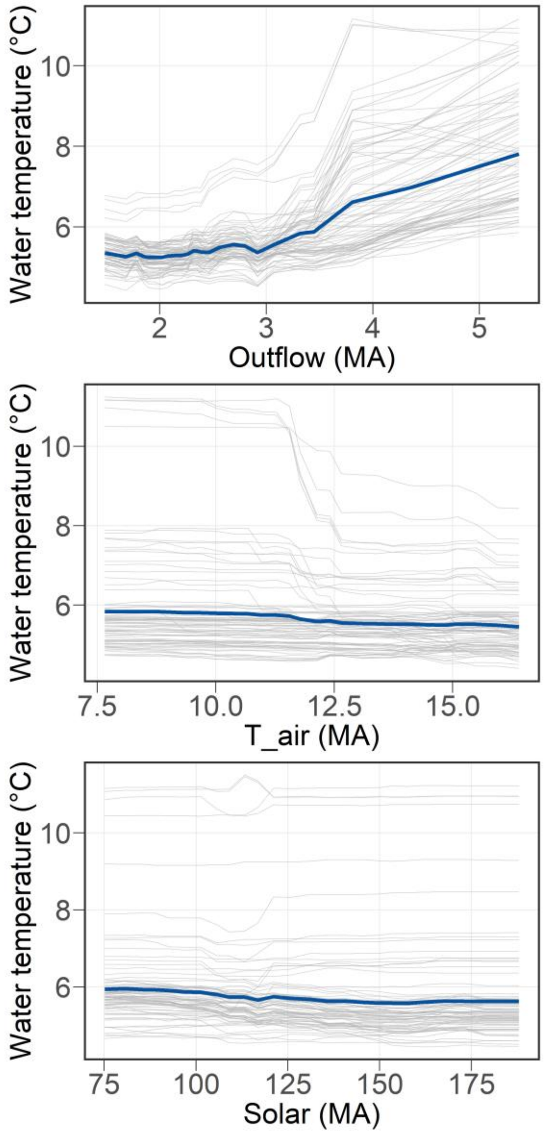

Fig. S5. Partial dependence plots (blue lines) and individual conditional expectation curves (gray lines) for the three dominant predictors of late-stratification bottom-water temperature at 50 m, namely antecedent outflow (top), antecedent air temperature (middle), and antecedent shortwave radiation (bottom). The panels show the modeled marginal response of bottom-water temperature to each predictor and the corresponding sample-specific heterogeneity.

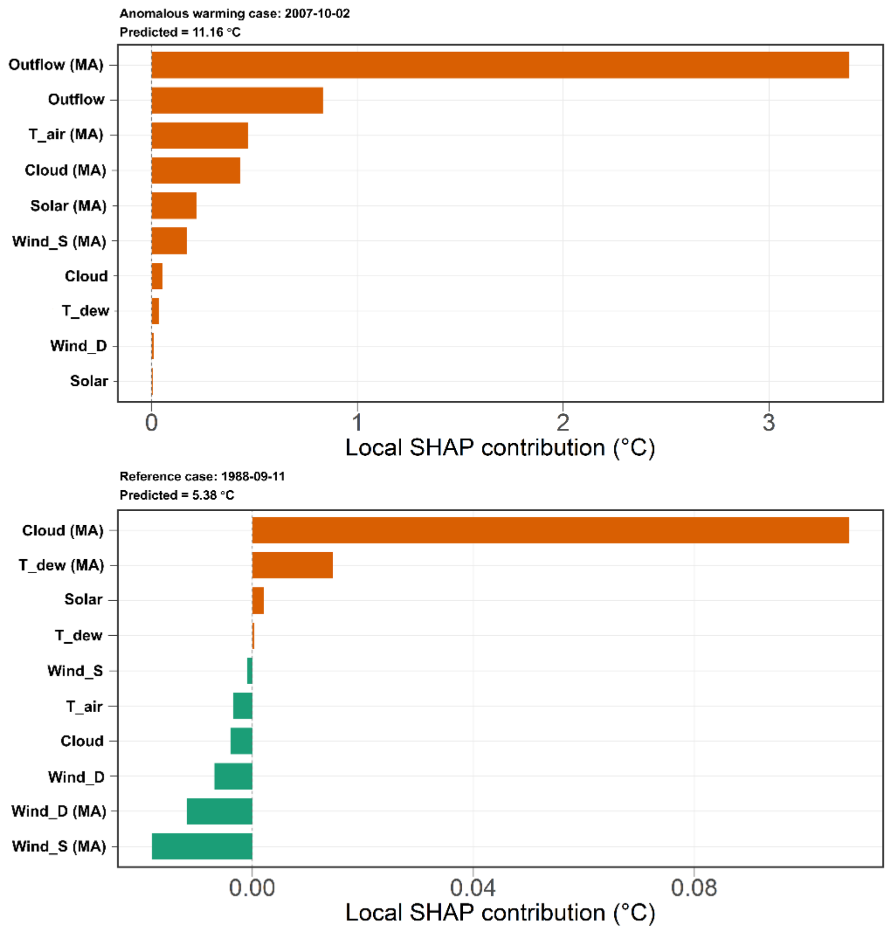

Fig. S6. Local SHAP decompositions for two representative late-stratification cases of bottom-water temperature at 50 m. The anomalous warming case (2007-10-02) was selected as the highest bottom-water-temperature day among the four anomalous years (1981, 2001, 2002, and 2007), whereas the reference case (1988-09-11) was selected as the non-anomalous late-stratification day whose bottom-water temperature was closest to the median of all non-anomalous cases. Bars show the contribution of each predictor to the model prediction relative to the base value; positive and negative SHAP values indicate upward and downward effects on predicted bottom-water temperature, respectively.

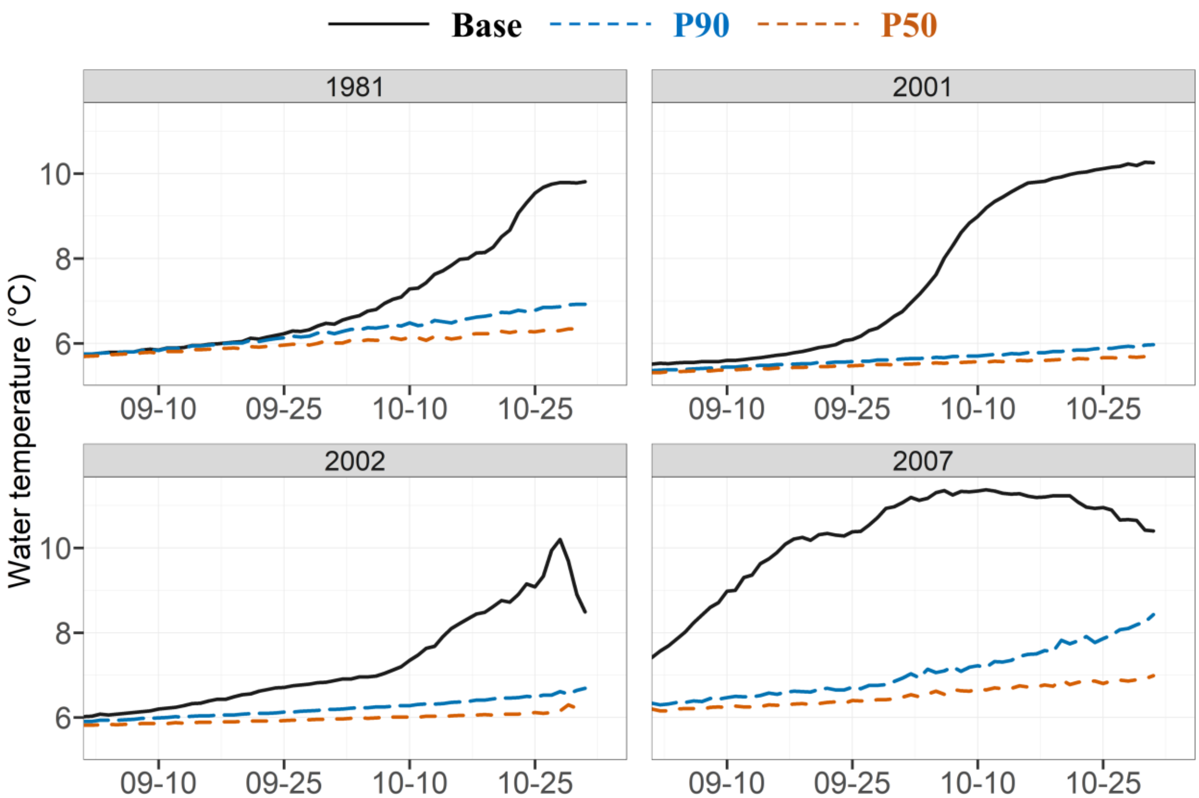

Fig. S7. CE-QUAL-W2 counterfactual simulations of September–October water temperature at 50 m under the baseline, P90, and P50 bottom-withdrawal scenarios in the four anomalous warming years. P90 and P50 indicate scenarios in which bottom withdrawal from 1 August to 31 October was replaced by the 90th percentile and median, respectively, of the corresponding calendar-day values from all non-anomalous years.

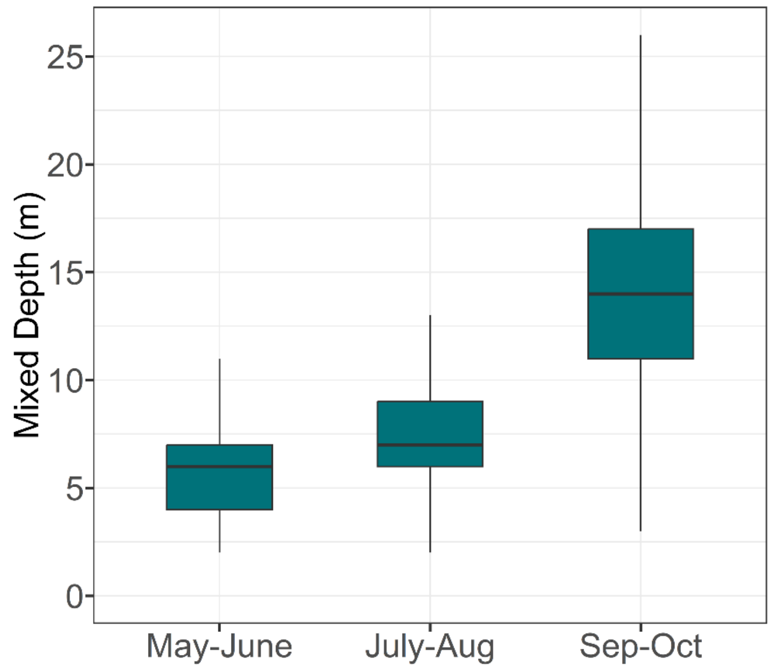

Fig. S8. Comparison of mixed-layer depth at different stages during the stratified period.

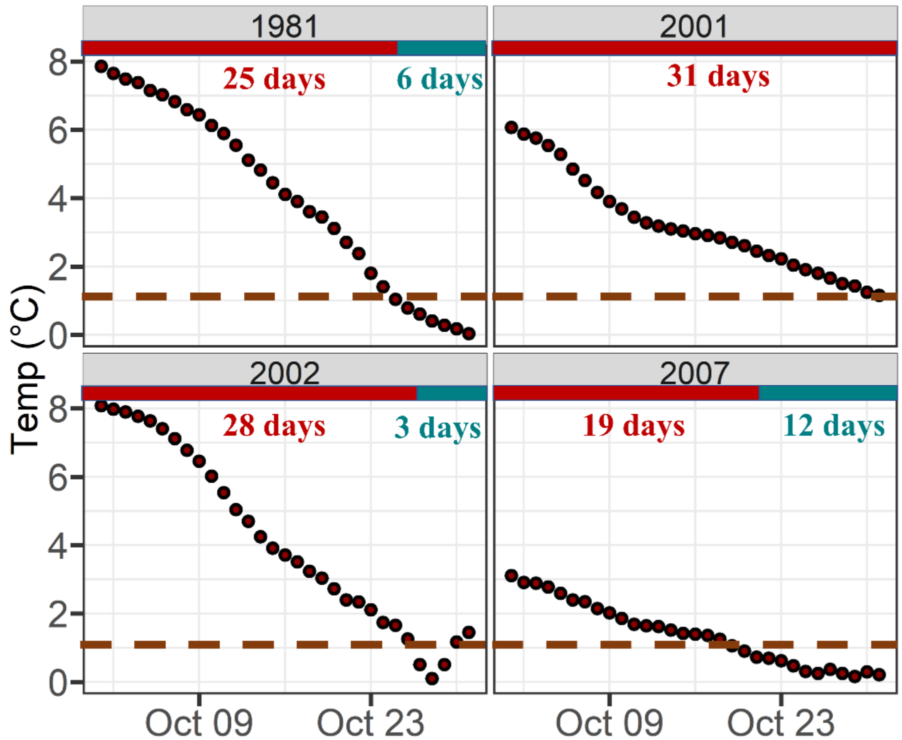

Fig. S9. Surface-to-bottom temperature difference at October during years of anomalously high bottom-layer temperatures (red text indicates stratification duration in that month; green text indicates mixing duration).

Reference

Boehrer, B. & M. Schultze, 2008. Stratification of lakes. Rev Geophy 46:RG 2005 doi:10.1029/2006RG000210.

Carr, M. K., A. Sadeghian, K.-E. Lindenschmidt, K. Rinke & L. Morales-Marin, 2019. Impacts of Varying Dam Outflow Elevations on Water Temperature, Dissolved Oxygen, and Nutrient Distributions in a Large Prairie Reservoir. Environ Eng Sci.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano & T. Zhou, 2015. Xgboost: extreme gradient boosting. R package version 04-2 1(4):1-4.

Cole, T. M. & S. A. Wells, 2006. CE-QUAL-W2: A two-dimensional, laterally averaged, hydrodynamic and water quality model, version 3.5.

Fang, X. & H. G. Stefan, 2009. Simulations of climate effects on water temperature, dissolved oxygen, and ice and snow covers in lakes of the contiguous United States under past and future climate scenarios. Limnology and Oceanography 54:2359-2370.

Kirillin, G., T. Shatwell & P. Kasprzak, 2013. Consequences of thermal pollution from a nuclear plant on lake temperature and mixing regime. Journal of Hydrology 496:47-56 doi:http://dx.doi.org/10.1016/j.jhydrol.2013.05.023.

Ladwig, R., P. C. Hanson, H. A. Dugan, C. C. Carey, Y. Zhang, L. Shu, C. J. Duffy & K. M. Cobourn, 2021. Lake thermal structure drives interannual variability in summer anoxia dynamics in a eutrophic lake over 37 years. Hydrology and Earth System Sciences 25(2):1009-1032.

Liu, Y. & A. Just, 2020. SHAPforxgboost: SHAP Plots for 'XGBoost'. R package version 0.1.0.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal & S.-I. Lee, 2020. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence 2(1):56-67.

Read, J. S., L. A. Winslow, G. J. A. Hansen, J. Van den Hoek, P. C. Hanson, L. C. Bruce & C. D. Markfort, 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. Ecological Modelling 291:142-150 doi:http://dx.doi.org/10.1016/j.ecolmodel.2014.07.029.

Sadeghian, A., S. C. Chapra, J. Hudson, H. Wheater & K.-E. Lindenschmidt, 2018. Improving in-lake water quality modeling using variable chlorophyll a/algal biomass ratios. Environmental Modelling & Software 101:73-85.

Team, R. C., 2022. R: A language and environment for statistical computing (Version r version 4.2. 2)(Vienna, Austria: R Foundation for Statistical Computing).

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H, Vaughan D & G. M, 2023. _tidyr: Tidy Messy Data_. R package version 1.3.0.

Winslow, L., J. Read, R. Woolway, J. Brentrup, T. Leach, J. Zwart, S. Albers & D. Collinge, 2019. Package 'rLakeAnalyzer'. Lake Physics Tools.