Title Page

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

Chenxi Mi[1,2], Bo Gai[4], Xiangzhen Kong[1,3*], Yuzhe Jiang[3], Chun Ngai Chan[2], Karsten Rinke[3]

[1] State Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 211135, China

[2] Department of Biological Sciences, University of Lethbridge, Lethbridge, AB, T1K 3M4, Canada

[3] Department of Lake Research, UFZ - Helmholtz Centre for Environmental Research, Magdeburg 39114, Germany

[4] Swiss Federal Institute of Aquatic Science and Technology (Eawag), Surface Waters – Research and Management, Kastanienbaum, Switzerland

* Corresponding author

Abstract

Thermal structure shapes ecological dynamics in lakes and reservoirs. Yet full-profile temperature records over multi-decades remain scarce, constraining mechanistic understanding of depth-resolved thermal changes and subseasonal extremes (e.g., surface heat waves and late-season hypolimnetic warming). In this study, we focused on Rappbode Reservoir, Germany's largest drinking-water reservoir, and compiled four decades of high-resolution, full-depth temperature profiles with concurrent hydro-meteorological records that are rarely available for stratified systems. Building on these data, we developed a novel two-step analytical framework that integrates long-term monitoring and process-based modelling to yield a high-resolution, internally consistent dataset of spatiotemporal temperature dynamics. We then applied interpretable machine learning to quantify depth-specific statistical attributions of hydro-meteorological and operational drivers, using process-based interpretation to assess mechanisms associated with late-stratification hypolimnetic warming. Our results suggested that driver contributions varied markedly with depth and stratification phase: stratification-strength metrics tied to atmospheric heat exchange (i.e., surface temperature, vertical temperature difference, Schmidt stability) showed the largest SHAP contributions from the 30-day antecedent moving average of shortwave radiation and air temperature. For hypolimnetic temperature, outflow discharge accounted for the largest contribution during late stratification. Additional process-based analyses supported a link between episodic hypolimnetic warming by up to 10 °C in four specific years and intensified deep withdrawals, which weakened the density gradient and promoted downward transport of warm upper-layer water into the hypolimnion. The dual-perspective framework developed here, which integrates process-based

and machine-learning approaches, is broadly transferable for analyzing ecological processes and supporting evidence-based management in stratified waters.

Keywords: Thermal stratification, Long-term temperature profiles, Hypolimnetic warming, CE-QUAL-W2, Interpretable machine learning methods

1. Introduction

Driven by solar heating, wind stress, and basin-scale hydrological exchanges, inland waters deeper than 7 m commonly develop seasonal thermal stratification, namely, a distinct vertical temperature gradient within the water column (Kirillin and Shatwell 2016). Solar warming of the surface layer, combined with limited turbulent mixing, generates a warm epilimnion while deep waters remain cold and form the hypolimnion. These two strata are separated by a region of sharp temperature gradients defined as metalimnion (Boehrer and Schultze 2009). This vertical structure controls oxygen and nutrient fluxes (Noori et al. 2023), governs redox-sensitive release of phosphorus from sediments (Deng et al. 2011), and shapes the availability of thermal habitat for temperature-sensitive aquatic organisms (Kraemer et al. 2021). It also modulates whole-lake emissions of carbon dioxide and methane, thereby shaping the lake's net role in the global carbon cycle (Mi et al. 2023a). Given the growing influence of climate warming and hydrological variability on stratification phenology (Mi et al. 2023b; Shatwell et al. 2019), a robust understanding of its long-term dynamics is indispensable for safeguarding water quality and designing adaptive management strategies.

Observations of high-resolution vertical lake water temperature dynamics over long periods are rarely available. Satellite sensors and nearshore infrared radiometers have provided continuous records of surface water temperature in lakes and reservoirs worldwide for over 40 years, enabling robust attribution of surface warming to air temperature, radiation, and wind forcing (Woolway et al. 2020). However, satellites measure only the surface skin (near-surface) temperature, and resolving vertical thermal structure still requires in-situ profilers. Globally, truly multi-decadal (>20 yr), depth-resolved temperature archives that meet basic comparability criteria (harmonized vertical spacing, sub-monthly sampling frequency and consistent instrumentation over time) remain the exception rather than the rule, with available profile records frequently sparse and summer-biased, particularly below the metalimnion (Akbari et al. 2017). For example, Sharma et al. (2015) analyzed summer water temperatures from 291 lakes across different climate zones, and found that only ~20% of those lakes had temperature profile records for more than one year. Although a recent synthesis study compiled vertical temperature profiles from 153 lakes (Pilla et al. 2021) and quantitatively analyzed their thermal structure during stratified seasons, for most sites the record only consists of few profiles obtained during mid-summer, leaving large seasonal and inter-annual gaps. This global paucity of long-term, depth-resolved temperature records substantially limits the understanding of changes in internal thermal structure and their external forcings.

Mechanistic hydrodynamic lake models offer a principled means to generate such long-term time series of lake water temperatures at high spatial and temporal resolution (Dong et al. 2019). Following site-specific validation against routine monitoring data, such models enable physically consistent gap-filling to sub-daily resolution and longer-term extensions of vertical profiles, thereby supplying the vertically resolved context that short and sparse observations lack. However, robust mechanistic attribution of subsurface temperature variability and stratification intensity still hinges on multidecadal in situ observations, even when aided by model-based reconstructions (Anderson et al. 2021). Furthermore, deep-water thermal structure shows significantly lagged responses to decadal climate variability (e.g., the Pacific Decadal Oscillation and Atlantic Multidecadal Oscillation) and small-scale shear-induced mixing (Kirillin and Shatwell 2016; Oleksy and Richardson 2021). If the monitoring record is shorter than a full decadal cycle, the corresponding analyses cannot reliably separate the influences of climate forcing and internal lake processes on the deep-water thermal structure (Boehrer and Schultze 2008). In turn, beyond complicating mechanistic diagnosis, the absence of sustained temperature profiles also undermines causal inference for vertically structured ecological responses, governed by the thermal regime (Meinson et al. 2016).

In addition, in deep, seasonally stratified lakes and reservoirs that cool toward the temperature of maximum freshwater density during winter circulation, hypolimnetic waters commonly remain cold and relatively stable during the subsequent stratified period (Boehrer and Schultze 2008; Lewis Jr et al. 2019; Schwefel et al. 2025). Recent studies, however, have documented late-stratification episodes in which the hypolimnion warms unexpectedly, with peak values occasionally exceeding 10 °C (Lewis Jr et al. 2019; Schwefel et al. 2025). Such deep-water thermal anomalies are being reported with increasing frequency, highlighting an emerging climate-driven phenomenon at depth (Woolway et al. 2025). Unlike the widely studied surface warming, these rare hypolimnetic warming events pose a disproportionate threat to ecosystem stability and drinking-water security. This is because hypolimnetic waters are characterized by low dissolved oxygen, strongly reducing conditions, and minimal water exchange, so that even slight warming can swiftly upset its delicate balance (Mi et al. 2023b). In fact, temperature increases in hypolimnion significantly accelerate temperature-sensitive mineralization, benthic oxygen uptake, and sediment–water solute exchanges (Dadi et al. 2023; Nkwalale et al. 2023). As a result, a small warming of hypoxic bottom waters can rapidly trigger cascading environmental responses: oxygen is consumed faster and the sequestered nutrients, heavy metals, and greenhouse gases are abruptly mobilized (LaBrie et al. 2023).

Given these risks, systematic analysis of deep-water warming is highly relevant to management. Yet, elucidating hypolimnetic thermal processes and identifying short-term anomalous events critically rely on sustained temperature measurements with sufficient vertical resolution (or highly resolved numerical simulations, see Boehrer and Schultze (2009)). As noted above, current observations of subsurface temperatures remain notably limited due to data availability and sampling frequency, hampering our understanding of the mechanisms underlying episodic hypolimnion warming events.

To fill these described research gaps, we used Rappbode Reservoir, the largest drinking-water reservoir in Germany, as a model system and compiled a unique 40-year, high-resolution archive of full-depth (0–70 m) temperature profiles. Leveraging the complementary strengths of process-based and data-driven approaches, we first calibrated a CE-QUAL-W2 model to generate a physically consistent, gap-filled record of daily full-profile temperatures. We then applied an interpretable machine-learning layer, XGBoost coupled with SHAP, to obtain depth- and phase-resolved driver attributions of climate and operational forcings. Combined with the in-situ monitoring, these components constitute a mechanistic–statistical framework that sharpens inference on interacting drivers, and identifies the dominant controls on deep-reservoir thermal dynamics. In this study, we seek to address the following key questions: (i) Which external forcings set full-column stratification strength, beyond surface warming?, (ii) How do the dominant drivers of water temperature dynamics vary with depth?, and (iii) Under what conditions is late-stratification hypolimnetic warming initiated and sustained? In addressing these questions, this study provides the first quantitative explanation of the mechanisms triggering episodic hypolimnetic warming in stratified waters. Together, these advances yield a transferable mechanistic–statistical template in analyzing thermal evolution for inland waters, with conclusions that generalize well beyond this single system to stratified lakes and reservoirs.

2. Methods

2.1. Study site description

Located in the Harz Mountains, the Rappbode Reservoir (Fig. 1) is Germany's largest drinking-water reservoir and supplies raw water to more than one million people in the surrounding region (Rinke et al. 2013). The reservoir attains a maximum depth of 89 m (mean = 29 m), and the dam crest lies at 423.6 m a.s.l., with the total volume of 1.13×10^8 m^3). The reservoir is impounded within a narrow valley in the Harz Mountains. The regional catchment is underlain mainly by consolidated Paleozoic bedrock, including greywacke, clay schist, and diabase (Rinke et al. 2013; Werner et al. 2021). In the present hydrodynamic application, the model geometry and time-varying wetted domain are prescribed from reservoir bathymetry and observed water-level variations; subsurface geological sections are therefore not used to define the computational domain.

Inflows of the Rappbode Reservoir are delivered by the Hassel and Rappbode pre-reservoirs and by a transfer tunnel from the Königshütte Reservoir (Mi et al. 2018). Six selective-withdrawal outlets are embedded in the dam at 345, 360, 370, 380, 390, and 400 m a.s.l.; the lowest outlet discharges to the downstream Wendefurth Reservoir, whereas the 360 m and 370 m intakes are routinely used for potable supply. With an annual residence time of roughly one year and the regional temperate climate, Rappbode exhibits a classic dimictic pattern: stable thermal stratification in summer and winter, and complete holomixis during spring and autumn. The national ban on phosphate-based detergents instituted three decades ago reduced total phosphorus from 0.16 mg L^{-1} in 1990 to the current level of approximately 0.02 mg L^{-1}, and the reservoir is now classified as meso- to oligotrophic status (Mi et al. 2022).

2.2. Long-term water temperature measurements

Since 1981, bi-weekly vertical temperature profiles have been taken at the deepest point immediately upstream of the Rappbode Dam (see Fig. 1), spanning the full water column from the surface to 70 m depth. Up to 2008, temperatures at discrete depths were measured with a thermometer inside an open cylinder sampler; from 2009 onward, measurements have been acquired with a Hydrolab DS5 multiparameter probe. Parallel deployments during the transition period confirmed statistical equivalence between the two methods (Wentzky et al. 2018). In this study, we used data from 1981 to 2019 (39 years in total), and after quality control, the record yields a depth–time matrix with year-round, across-season coverage over the routine profiling range of 0–70 m, which serves as the basis for all subsequent analyses.

2.3. CE-QUAL-W2 model configuration

We developed an integrated framework linking four decades of water-temperature observations with process-based modeling and interpretable machine learning to elucidate the thermal evolution of the Rappbode Reservoir (Fig.1). Here, the process-based simulations were performed with CE-QUAL-W2 (hereafter W2, version 4.2.1), a laterally averaged, two-dimensional hydrodynamic model originally developed by the U.S. Army Corps of Engineers at 1975 (Cole and Wells 2006) and now maintained by the Department of Civil and

Environmental Engineering at Portland State University. W2 is widely used as a standard tool for resolving hydrodynamic processes in lakes, reservoirs, and estuaries (e.g., Carr et al. 2019; Kobler et al. 2018).

For Rappbode, the computational mesh comprises four branches subdivided into 106 longitudinal segments ($\Delta x = 100$–400 m) and lateral sub-segments of 5–700 m, yielding 3876 grid cells with a uniform 1 m vertical resolution (Fig. S1). This discretization fully resolves the reservoir's bathymetry, enabling precise simulation of its morphological features. Consistent with the W2 formulation, the main physical assumption in the present application is that lateral variations within each segment are secondary to longitudinal and vertical gradients; each segment is therefore represented by a laterally averaged cross-section. This assumption is appropriate for the narrow, dendritic morphology of Rappbode Reservoir and for the present objective of resolving open-water thermal stratification and withdrawal-induced vertical exchange. Groundwater exchange and subsurface hydrogeological flow were not explicitly represented. Thus, the depths shown in the model output and figures denote bathymetric water-column depths within the reservoir, rather than geological or hydrogeological sections. The model was driven by two sets of boundary conditions: (i) Hydrological forcing, including daily inflow discharge, inflow temperature, outflow discharge, and the withdrawal elevation at the dam, was provided by the Rappbode Reservoir authority (Talsperrenbetrieb Sachsen-Anhalt). The outflow boundary was prescribed directly from operator records, rather than inferred from the simulated thermal structure or adjusted during temperature calibration; (ii) Meteorological forcing, including hourly air temperature, relative humidity, wind speed and direction, incoming short-wave radiation, and cloud cover, was obtained from a monitoring buoy moored near the reservoir's centre (Rinke et al. 2013), with gaps infilled with records from the German Weather Service station at Harzgerode, ~15 km away from the reservoir. An intercomparison of Harzgerode and Rappbode Reservoir indicates a high degree of correlation in the meteorological records (see Mi et al. 2019 for more details). Simulations started on 1st January 1981 with an isothermal initial profile of 4 °C, representing a horizontally and vertically homogeneous state.

The present W2 application builds on earlier Rappbode Reservoir configurations (see Mi et al. 2023b; Mi et al. 2020), but all simulations reported here were re-run and evaluated for the 1981–2019 temperature archive used in this study. The model was calibrated manually using a trial-and-error procedure. Three site-specific parameters were adjusted because they depend on local optical conditions and meteorological exposure: the shading coefficient (SHADE), wind-sheltering coefficient (WSC), and pure-water light-extinction coefficient (EXH2O). The final values were SHADE = 1, WSC = 1, and EXH2O = 0.55 m^{-1}. All remaining parameters were retained at standard CE-QUAL-W2 settings because they are physically based and not typically subject to site calibration (Cole and Wells 2006). The complete parameter set and numerical options are provided in Text S1 and Table S1. The main calibration targets were the temperatures at different depths, and water level was used as an additional system-scale check of the water-balance representation. To close the reservoir water balance, residual inflows required to match the observed water-level evolution were represented as distributed tributary inflows, following the water-balance procedure described in the official CE-QUAL-W2 User Manual (Wells 2026). These residual inflows were used only for water-balance closure and were distributed along the main reservoir branch to minimize localized thermal effects. No separate split-sample validation period was defined. Instead, the full 1981–2019 observational record was used for continuous calibration and evaluation across a broad range of hydroclimatic and operational conditions. This choice follows the calibration guidance in the official manual of W2 (Wells 2026), which emphasizes evaluating model behavior across broad observed conditions rather than relying on an arbitrary calibration/verification split.

2.4. XGBoost description and modeling framework

Data-driven approaches used in Earth-system prediction span a broad spectrum, including regression models, kernel methods such as Support Vector Regression, tree ensembles such as Random Forest and XGBoost, neural networks, and metaheuristic frameworks designed for optimization. Genetic algorithms and related metaheuristics are particularly valuable for search and optimization tasks, including parameter calibration, feature selection, and model structure exploration (Rajwar et al. 2023; Schiavo and Pedretti 2026). The present study required a supervised regression model that could emulate CE-QUAL-W2 thermal targets from tabular hydrometeorological predictors and support transparent attribution across depths and stratification phases. We therefore selected XGBoost as the primary model because this tree ensemble framework is well suited to nonlinear tabular data, computationally efficient, and directly compatible with SHAP interpretation (Mi et al. 2024; Zhu et al. 2022). To place this choice in context, we further benchmarked XGBoost against three other data-driven approaches under the same experimental protocol.

XGBoost model description

Extreme Gradient Boosting (XGBoost) is an ensemble method that iteratively fits Classification-and-Regression Trees (CART) within a gradient-boosting framework. At each iteration t, the algorithm minimizes a regularized objective:

$$L^{\mathrm{t}} = \sum_{\mathrm{i=1}}^{\mathrm{n}} l\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \sum_{\mathrm{k=1}}^{\mathrm{t}} \Omega(f_k) \qquad (1)$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{\mathrm{j=1}}^{\mathrm{T}} w_j^2 \qquad (2)$$

where $l(\cdot)$ is a twice-differentiable loss (squared error for regression), f_t is the t-th tree, T is the number of leaves and w_{j} denotes leaf weights; γ and λ provide structural and weight regularization, respectively. Here, $\Omega(f)$ represents the regularization term that penalizes model complexity to prevent overfitting. By employing a second-order Taylor expansion of L^t, XGBoost obtains closed-form gain scores that guide an exact or histogram-based greedy search for the optimal split, while column subsampling and block-wise caching accelerate computation. This design yields three decisive benefits: (i) high accuracy, since second-order information markedly improves converge4nce; (ii) strong generalization, owing to L1/L2 shrinkage, learning-rate decay and stochastic sub-sampling; and (iii) computational efficiency, as multi-threaded and out-of-core execution scale to millions of observations and features on commodity hardware. Built-in cross-validation, early stopping and SHAP-based interpretability further underline XGBoost's suitability for handling small- to medium-sized and highly non-linear environmental data sets (Li et al. 2024; Lyu et al. 2019).

XGBoost model structure

We developed XGBoost models for predicting reservoir thermal structure according to daily outputs (1981–2019) from CE-QUAL-W2. Target variables comprise epilimnetic (5 m depth), metalimnetic (15 m depth), hypolimnetic (30 m depth) and bottom (50 m depth) temperatures, Schmidt Stability, the bottom-to-surface temperature difference and mixed-layer depth (see the Supplementary Materials for details on the last three indices and the post-processing methods). Since the above indices related to stratification dynamics are applicable only during thermal stratified period, therefore, the XGBoost model was constructed and analyzed exclusively for May–October. Guided by recent advances in machine-learning modeling of lentic thermal structures (e.g., Bertone et al. 2015; Kreakie et al. 2021), the predictor set includes both the instantaneous and 30-day moving-average values of observed air temperature, relative humidity, wind speed and direction, incoming short-wave radiation, cloud cover and reservoir outflow discharge. Descriptive statistics of the predictor and target variables used for machine-learning model development are summarized in Tab. S2. Here, we included 30-day moving-average predictors to represent the cumulative component of atmospheric and hydraulic influences across depths, with the physical basis for this setting (epilimnetic thermal inertia τ) further detailed in Section 4.2. To be noted, our previous studies indicated that inflow discharge exerted only a marginal influence on the reservoir's thermal structure (Mi et al. 2020); thus, it was excluded as an input feature from the XGBoost model. All predictors included in the analysis were linearly de-trended, *Z*-score standardized and screened for collinearity ($|\rho| < 0.75$) to minimize feature redundancy (Deka and Weiner 2024).

The full machine-learning dataset was randomly split into training and testing subsets in an 80:20 ratio. Comparisons of the empirical cumulative distributions of the training and testing subsets, quantified by the Kolmogorov–Smirnov statistic together with Jensen–Shannon and energy distances, indicated no appreciable distributional shift for any target-specific model (Tab. S3). Within the training subset, a 5-fold rolling-origin cross-validation combined with Bayesian optimization was used to optimize the XGBoost hyperparameters (i.e., eta, max_depth, subsample, and colsample_bytree), and early stopping was applied to minimize the root mean square error (RMSE) on the validation folds. Model performance was then evaluated using RMSE and the coefficient of determination (R^2), with training and testing results reported separately. The optimal values of eta, max_depth, subsample, and colsample_bytree were 0.2, 6, 0.7, and 0.7, respectively.

Other data-driven models for comparison

Comparative experiments were conducted to benchmark the predictive performance of XGBoost against three widely used data-driven models for reservoir thermal-structure prediction, including Support Vector Regression (SVR), Random Forest (RF), and Multivariate Adaptive Regression Splines (MARS). To ensure a rigorous and directly comparable benchmark, the three baseline models were trained and evaluated under an identical experimental protocol to XGBoost, covering predictors, train–test partitioning, and performance metrics. Statistical differences in predictive performance were evaluated by applying Kruskal–Wallis tests (Kruskal and Wallis 1952) to the test-set squared-error distributions of XGBoost, SVR, RF, and MARS for each target variable, followed by pairwise Diebold–Mariano tests (Diebold and Mariano 2002) comparing XGBoost with each benchmark model. All models were performed on a desktop workstation equipped with an Intel(R) Core(TM) Ultra 9 285K CPU @ 3.70 GHz, 256 GB RAM, and an NVIDIA T400 4 GB GPU.

SVR is a classical supervised regression algorithm developed within the support vector machine framework proposed by Cortes and Vapnik (Vapnik 2013). The formulation follows the structural risk minimization (SRM) principle. The flexibility of SVR is provided by kernel functions, which transform nonlinear relationships in the original predictor space into linear representations in a higher-dimensional feature space. Here, the radial basis function (RBF) was used as the kernel function, with the regularization parameter C and ε-insensitive loss parameter set to 1 and 0.1, respectively.

RF is a tree-based ensemble learning framework (Yousefi and Toffolon 2022). It constructs an ensemble of decorrelated decision trees by repeatedly drawing bootstrap samples from the training data and selecting random subsets of predictors at each split, and the final prediction is obtained by averaging the outputs of all trees. In this study, the model was configured with 500 decision trees, a minimum terminal node size of 50 samples, and bootstrap sampling with replacement for tree construction.

MARS is a nonparametric regression method that approximates covariate effects with adaptive piecewise linear splines. The fitted equation is built by forward addition of hinge pairs followed by backward pruning, thereby retaining nonlinear effects and predictor interactions in a parsimonious form (Heddam et al. 2020). In the present analysis, the maximum interaction degree was set to 2, the generalized cross validation penalty was 3, and the forward fitting threshold was 0.001.

SHAP-based feature contribution analysis

In the interpretability component of the modelling framework, we quantified feature contributions using SHapley Additive exPlanations (SHAP) values. SHAP is a model-agnostic interpretability framework rooted in cooperative game theory (Lundberg et al. 2020). It assigns each predictor a Shapley value calculated by evaluating the predictor's marginal contribution to model output across all possible subsets of input features (>0 pushes the prediction upward, <0 pulls it downward). We report the mean absolute SHAP for each feature as a measure of global importance. Here, SHAP distributions and effect sizes were visualized with beeswarm plots: the horizontal axis shows the SHAP value, and point color redundantly encodes the same measure on a single, shared scale (e.g., warm colors on the right indicate positive contributions, cool colors on the left indicate negative contributions, see Fig.9). Feature attributions computed via SHAP satisfy exact additivity, local accuracy, and consistency (null-feature property), providing a rigorous quantitative basis for evaluating the relative contributions of hydrometeorological predictors on reservoir thermal dynamics. These SHAP values are interpreted here as statistical attributions to model predictions rather than as standalone evidence of physical causality. For more background on SHAP values, please refer to the Supplementary Material (Text S2).

As complementary model-agnostic diagnostics, we further evaluated predictor relevance and response structure using permutation feature importance (PFI), partial dependence plots (PDPs), and individual conditional expectation (ICE) curves. PFI was calculated on the held-out test data by randomly permuting each predictor and quantifying the resulting increase in prediction error, following the permutation-based variable-importance framework introduced by Breiman (2001) and its model-agnostic formalization by Fisher et al. (2019). PDPs were used to summarize the average marginal response of hypolimnetic water temperature to dominant predictors, following Friedman (2001), whereas ICE curves were used to visualize observation-level heterogeneity in these fitted responses, following Goldstein et al. (2015). Together, PFI, PDPs, and ICE curves served as complementary model-agnostic diagnostics to the SHAP analysis, allowing predictor importance and observation-level response heterogeneity to be evaluated within a consistent interpretation framework.

3. Results

3.1. Model performance in simulating long-term reservoir hydrodynamics

The calibrated W2 accurately reproduced the interannual and seasonal water-level dynamics of Rappbode Reservoir over the 39-year period (R^2 = 0.99, see Fig. S2), indicating that the hydrological boundary conditions and water-balance closure were implemented consistently. The water level typically attains its annual maximum in spring (April–May, ~420 m a.s.l.) in response to increased rainfall, then declines sharply through summer driven by intensified evaporation and elevated drinking-water withdrawals. During winter, marked reductions in precipitation and inflows caused the water-level to continually decline, reaching annual minima at the end of the season (e.g., 395 m a.s.l. in 2002). All of these characteristic fluctuations were satisfactorily captured by the model.

The model successfully captured spatiotemporal dynamics of the reservoir's thermal structure (such as thermocline depth and stratification onset (offset), see Fig.2). Simulated temperatures closely matched observations, with most points clustering along the 1:1 line (R^2 = 0.95; RMSE = 0.97 °C, see Fig. 2b). The simulated mean temperature (6.61°C) closely matched the observations (6.76°C) over the study period. More specifically, the discrepancy between observed and simulated temperatures generally decreased with depth, as indicated by RMSE of 1.27°C, 1.31°C, 0.75°C, and 0.59°C at depths of 5 m, 15 m, 30 m, and 50 m, respectively (Fig. 3, Tab. S4). Both simulated and observed profiles show that the intra-annual temperature amplitude diminished sharply with depth: it exceeded 15 °C at 5 m depth, but declined to <10 °C at 30 m and <8 °C at 50 m (Fig. 3).

Additionally, we separated long-term monotonic trends from interannual variability using annual temperature series at representative depths (Fig. S3). At 5m depth, both series warmed significantly, at rates of 0.05 K yr^{-1} for the observations and 0.06 K yr^{-1} for the simulations (Mann–Kendall test, $p < 0.05$), with the mean epilimnetic temperature rising from ~15 °C in 1981–1985 to ~18 °C at the end of the record. By contrast, temperatures at 15–50 m exhibited pronounced year-to-year variability but no statistically significant monotonic trend (Mann–Kendall test, $p > 0.05$). These depth-specific trends provide the long-term thermal context for the event-focused analysis of late-stratification bottom-water warming presented below.

3.2. Performance of XGBoost in simulating reservoir stratification dynamics

XGBoost consistently delivered the best point performance across all depth-specific temperature models, with the highest test R^2 and lowest test RMSE between W2 and XGBoost at all four depths (Tab. 1). Statistical tests further showed that its advantage over SVR and MARS was significant throughout, whereas the improvement over RF was significant at 15 and 50 m depth, but not at 5 or 30 m depth (Tab. S5). Using daily W2 outputs as targets, XGBoost exhibited progressively smaller prediction deviations with increasing depth, as RMSE between W2 and XGBoost declined from 0.70 °C at 5 m to 0.28 °C at 50 m. The seasonal temperature pattern was also accurately represented by XGBoost: surface temperatures (5 m depth) followed a pronounced unimodal, bell-shaped curve that peaked in mid-July to early August, when the median daily temperature in both W2 and XGBoost simulations was ~19 °C, and cooled markedly at the onset (May) and offset (October) of the stratified period (Fig. 4). With increasing depth, the timing of peak temperature lagged progressively. At depths of 30 m and 50 m, temperatures rose continuously throughout the stratified period, reaching their maxima in October.

XGBoost also outperformed the three benchmark models for the stratification metrics and reproduced the CE-QUAL-W2 outputs with close agreement (Fig. 5), yielding test R^2 values of 0.96, 0.82, and 0.97 and corresponding test RMSE values of 0.76 °C, 3.49 m, and 128.36 J m^{-2} for bottom-to-surface temperature difference, mixed-layer depth, and Schmidt Stability, respectively (Tables 1 and 2). Specifically, Schmidt Stability was generally $< 1 \times 10^3$ J m^{-2} at the onset of stratification in May, increased steadily to a peak of $\sim 3 \times 10^3$ J m^{-2} in August, and declined thereafter, returning to the initial levels by October; this seasonal pattern closely paralleled that of the bottom-to-surface temperature difference (Fig. 5a). In contrast, the mixed-layer depth deepened throughout the stratified period: it remained < 10 m from May to August but reached approximately 20 m by October (Fig. 5b).

Because XGBoost consistently outperformed the three benchmark models across all temperature and stratification targets, it was used as the reference model for subsequent SHAP-based interpretation of driver importance. SHAP analysis of the XGBoost model indicates that epilimnetic temperature (5 m) is governed chiefly by the antecedent 30-day means of air temperature, dew-point temperature, and incident short-wave radiation (Fig. 6). With depth, the relative SHAP contribution of air temperature decreases, ranking third, fourth, and fifth for the temperature at 15, 30, and 50 m depth, respectively, while short-wave radiation becomes the primary external driver below the surface layer. The importance of outflow discharge strengthens progressively down-profile, ranking eighth for epilimnetic temperature, fourth for temperature at 15 m depth, and second for temperature at 50 m depth. The relative importance of cloud cover as a driver of water temperature increased incrementally with depth, ranking sixth at depths of 5–30 m and fourth at 50 m. Wind speed and direction remain comparatively minor controls across all depths, ranking between sixth and eighth (Fig. 6). In sum, SHAP resolves a depth-structured forcing pattern: atmospheric control shifts from non-radiative terms in the epilimnion to short-wave radiation below the surface layer, while hydraulic withdrawal gains influence toward the hypolimnion. Complementary PFI analyses reproduced this vertical ordering of dominant controls (Fig. S4): antecedent air temperature ranked first at 5 m, whereas antecedent shortwave radiation dominated at 15 and 30 m and outflow discharge increased in importance with depth.

Our results also indicated that both the bottom-to-surface temperature difference and Schmidt Stability are governed principally by the antecedent 30-day means of air temperature, incident short-wave radiation, and dew-point temperature (Fig. 7). By comparison, the mixed-layer depth was primarily driven by short-wave radiation over the preceding 30-day period, with outflow discharge and dew-point temperature ranking second and third, respectively. Cloud cover also significantly affected stratification dynamics, ranking fourth in importance for mixed-layer depth and fifth for Schmidt stability. In addition to antecedent forcings, instantaneous (same-day) conditions also exerted substantial control: specifically, same-day air temperature ranks fourth for its effect on both the bottom-to-surface temperature difference and Schmidt Stability (Fig. 7), whereas the same-day outflow discharge ranks fifth in modulating mixed-layer depth.

3.3. Mechanisms underlying the occurrence of elevated bottom-water temperatures in specific years

During the stratified season, hypolimnetic (50 m depth) temperatures in Rappbode Reservoir generally remains within the range of 4–6 °C. However, both in-situ profiles and W2 simulations reveal episodic warming to ~10 °C at the end of stratification (September–October) in 1981, 2001, 2002, and 2007 (Fig. 8b). These years were therefore treated as episodic departures within a deep-water temperature record that showed strong interannual variability but no significant monotonic trend. To elucidate the mechanisms underlying these anomalies, daily hypolimnetic temperatures simulated by W2 were adopted as targets to develop specific XGBoost models for early (May–June), mid (July–August), and late (September–October) stratified periods. SHAP analysis was then applied to quantify the dominant predictors in each interval.

The period-specific XGBoost models satisfactorily reproduced CE-QUAL-W2 hypolimnetic temperatures with RMSE of 0.08, 0.09 and 0.13 °C for early, mid and late stratification, respectively ($R^2 = 0.96$–0.98). The results further demonstrate pronounced stage-dependent shifts in the controls on hypolimnetic temperature (Fig. 9). Specifically, during the early-stratified period (May–June) the 30-day antecedent mean air temperature, outflow discharge and wind direction emerge as the three leading drivers. By mid-stratification (July–August) the principal controls change over to the 30-day mean short-wave irradiance followed by outflow discharge, whereas in the late-stratified phase (September–October) the positive influence of outflow discharge becomes overwhelmingly dominant (Fig. 9): its SHAP contribution attains 0.513, over four times greater than that of air temperature (0.131, ranked second) and shortwave radiation (0.115, ranked third). To corroborate this finding, we examined years in which late-season hypolimnetic temperatures were anomalously elevated. In 1981, 2001, 2002, and 2007, the median discharge during late stratification was 5.24 $m^3 s^{-1}$, approximately 125% higher than the 40-year median (2.32 $m^3 s^{-1}$) for this interval, a difference that was statistically significant (Dunn's test, $p < 0.05$, see Fig.8a). Furthermore, during this period, the SHAP contribution of outflow discharge was also markedly greater than those of the other predictors (Fig. 9d). Complementary explainability analyses supported this interpretation (Figs. S4–S5). PFI again identified antecedent outflow as the dominant predictor of late-stratification bottom-water temperature. PDP and ICE curves likewise showed a strong positive and nonlinear warming response to increasing antecedent outflow, while antecedent air temperature and antecedent shortwave radiation exerted weaker and more heterogeneous effects. Local SHAP decompositions for two representative late-stratification cases further highlight this event-scale contrast (Fig. S6). In the strongest anomalous warming case (2007-10-02), antecedent outflow discharge increased the model-predicted 50 m water temperature by >3 °C, whereas antecedent air temperature, cloud cover, and shortwave radiation each contributed <0.5 °C. By contrast, the reference case was defined as the non-anomalous late-stratification day whose bottom-water temperature was closest to the median of all non-anomalous late-stratification cases (1988-09-11); on that day, no comparable positive outflow-discharge contribution was detected, and all feature-level contributions remained small (|SHAP| <0.12 °C). Collectively, the hydrological signature provides independent evidence that enhanced release discharge is the key driver of late-stratification hypolimnetic warming.

4. Discussion

4.1. Reconstruction of thermal structure from multi-decadal full-profile observations

In this study we assembled a uniquely long and vertically resolved data set comprising almost four decades (1981–2019) of temperature profiles (0–70 m) and concomitant meteorological forcing for Rappbode Reservoir, the largest drinking water reservoir in Germany. On this basis, we established an analytical framework, combining process-based (CE-QUAL-W2) and data-driven (XGBoost) models, to systematically assess their capabilities in reproducing the reservoir's thermal structure. Whereas most publicly available lake- and reservoir-temperature records span ≤ 10 years and provide limited long-term coverage of deep-water temperatures (Ladwig et al. 2021; Woolway et al. 2020), the 39-year full-profile archive assembled here offers a robust empirical basis for multi-decadal hydrodynamic modelling.

Here, the long-term observations and simulations consistently indicate a pronounced epilimnetic warming of the reservoir, with observed and simulated warming rates of 0.06 and 0.05 °C a^{-1}, respectively. These rates are comparable to, and slightly higher than, the concurrent regional air-temperature trend (0.04 °C yr^{-1}; Mann–Kendall test, $p = 0.001$), and are broadly consistent with warming rates reported for temperate lakes in previous large-scale syntheses (O'Reilly et al. 2015; Shatwell et al. 2019). In contrast, deep-water temperatures at 30 m and 50 m depth exhibited no significant trends, thereby reflecting a characteristic pattern of intensified surface warming concurrent with relatively stable deep-water temperatures. This non-monotonic deep-water background is important for interpreting the anomalously warm late-stratification years analyzed below: these events represent episodic departures from long-term variability rather than evidence of sustained hypolimnetic warming over the full record.

This vertical asymmetry in warming preserves the cold-water refuge but implies a strengthening of thermal stratification. Numerous studies have shown that such shifts in lake and reservoir thermodynamics would exacerbate hypolimnetic anoxia and enhance internal nitrogen and phosphorus loading (Sun et al. 2022; Weinke and Biddanda 2019), subsequently impairing source-water quality and downstream aquatic ecosystems (Jane et al. 2023). Accordingly, coupling the validated thermal models with biogeochemical modules to quantify present-day and future trajectories of dissolved oxygen, and to develop targeted water-quality management strategies, constitutes a logical and necessary extension of the current work.

4.2. Atmospheric and hydraulic drivers governing thermal dynamics in stratified reservoirs

The SHAP values indicated that the antecedent 30-day mean air temperature is the most influential predictor of epilimnetic temperature, as well as of Schmidt stability and the bottom-to-surface temperature difference, followed by dew-point temperature and shortwave radiation. This hierarchy corroborates earlier findings derived from short-term datasets for dimictic lakes by Livingstone and Lotter (1998) and Darko et al. (2019), and is now quantitatively confirmed by our 39-year observational record. All such results highlight the lagged response of the three surface-heat-forced stratification metrics to cumulative antecedent heat fluxes (i.e., the thermal inertia effect). Here, the epilimnetic thermal inertia was estimated from a mixed-layer heat budget. The areal heat capacity of the mixed layer is $\rho c_p h$, where ρ is water density, c_p is the specific heat capacity of water (4.18×10^3J kg^{-1} K^{-1}), and h is mixed-layer depth. The characteristic timescale of a 1 K mixed-layer temperature response under a mean net heat flux H_{net} can therefore be estimated as:

$$\tau_{1K} = \frac{\rho\, c_p\, h\Delta T}{|H_{net}|}\ , \Delta\mathrm{T} = 1\ \mathrm{K} \qquad (3)$$

This estimate follows the same physical basis as the lake temperature formulation of Piccolroaz et al. (2013). For Rappbode Reservoir, using ρ = 998 kg m^{-3} , h = 9-11 m, and H_{net} = 18-24 W m^{-2} for the stratified period (Mi et al. 2019) gives a characteristic response time on the order of several weeks, consistent with the 30-day averaging window that dominates the SHAP ranking. Because this response time exceeds the characteristic 1-day scale of surface heat-flux variability by an order of magnitude, daily atmospheric perturbations are strongly filtered by thermal inertia and thus contribute little to the statistical ranking.

By comparison, in the hypolimnion (30–50 m), short-wave radiation exceeds air temperature in the SHAP ranking, and the importance of outflow discharge concurrently increases, because sustained deep withdrawals can remove the coldest layers, erode thermocline stability and induce compensation flows that advect warmer water downward (Olsson 2021; Weber et al. 2017), thereby regulating hypolimnetic temperature. Our analysis further reveals that, for mixed-layer depth, outflow discharge emerges as the second-most influential predictor. As shown above, this can be attributed to sustained hypolimnetic withdrawals, which induce downward flows and consequently deepen the mixed layer (Deng et al. 2011; Mi et al. 2023a). Additionally, the role of cloud cover should not be overlooked (ranking fourth in SHAP values), since cloudiness can effectively modulate incoming shortwave and longwave radiation, thus altering surface buoyancy fluxes (Boehrer and Schultze 2008).

We further tested this SHAP-derived meteorological ranking using four targeted CE-QUAL-W2 scenarios in which shortwave radiation, air temperature, dew-point temperature, or cloud cover was separately replaced by its long-term median for the same calendar day during 1 April to 31 October, with all other forcings unchanged. Details of the scenario design and diagnostic calculation are provided in Text S4.1. The scenario results reproduced the key SHAP pattern: air temperature remained the strongest meteorological forcing for the water temperature at 5 m depth, whereas shortwave radiation ranked highest at the other subsurface layers. This agreement strengthens the physical credibility of the SHAP result that antecedent shortwave radiation has the highest relative contribution to thermal variability at greater water-column depths.

4.3. Late-season hypolimnetic warming — process interpretation of period-specific SHAP signals

During late stratification, increased deep-layer withdrawals are more effective at warming the hypolimnion because the reservoir is already in a mechanically vulnerable stratification state. Over this period, nocturnal cooling and declining net radiation substantially reduce Schmidt stability to $< 1 \times 10^3$ J m^{-2} (Fig. 5), compared with $> 2 \times 10^3$ J m^{-2} in midsummer (July–August). This weaker density structure lowers the buoyancy resistance across the metalimnion, allowing bottom-withdrawal forcing to enhance entrainment and compensatory downward transport of warmer upper-layer water. In parallel, the mixed layer deepens from about 6 m in early summer to 18 m by September-October (Fig. S8), thereby reducing the vertical distance between the mixed-layer base and the water-withdrawal intake (at the depths of 60–70 m) by ≈20%. Over the same period, hypsographic calculations combining the W2 volume–elevation curve, daily water level, and mixed-layer depth indicated that the mean water volume below the mixed-layer base decreased from 8.78×10^7 to 4.67×10^7 m^3, corresponding to a 46.8% reduction.

Thus, a given bottom-withdrawal discharge acted on a substantially smaller deep-water volume during late stratification, increasing the withdrawal-to-volume ratio and allowing the density interface to be displaced more efficiently. The concurrence of a weakened stratification intensity, reduced deep-water volume and a contracted transport pathway provides a physical basis for why high-discharge deep withdrawals can rapidly elevate bottom temperatures during late stratification. We further tested this interpretation using targeted CE-QUAL-W2 withdrawal scenarios for the four anomalous years. In the P50 and P90 scenarios, bottom withdrawal during 1 August to 31 October in 1981, 2001, 2002 and 2007 was replaced by the median and 90th percentile, respectively, of the corresponding calendar day values from the non-anomalous years. The thermal response was evaluated for September to October, when anomalous bottom warming occurred. The results suggested that reducing high bottom withdrawals strongly suppressed hypolimnetic warming: the mean 50 m temperature across the four anomalous years decreased from 8.05 °C in the baseline simulation to 6.30 °C in P90 and 6.00 °C in P50, while the corresponding maximum temperature decreased from 11.37 °C to 8.43 °C and 6.99 °C. These scenarios 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. Details of the scenario design and results are provided in Text S4.2.

Despite pronounced hypolimnetic warming induced by deep-water withdrawals in the late stratification periods of 1981, 2001, 2002, and 2007, the surface-to-bottom temperature difference remained above 1 °C for most of October (Fig. S9), indicating persistent and stable stratification conditions (Fang and Stefan 2009). Hypolimnetic warming under such weakened stratification poses a severe ecological threat since oxygen demand in the hypolimnion increases sharply with temperature while atmospheric re-oxygenation remains inhibited. For Rappbode Reservoir, our previous work suggested that raising bottom-water temperature from a typical 6 °C to 10 °C (October 2001) augments the sediment oxygen demand (SOD) from 0.46 to 1.04 g O_2 m^{-2} d^{-1} (Mi et al. 2020). Under a persistent density barrier, this additional demand cannot be offset by oxygen replenishment from surface waters, greatly amplifying the risk of hypolimnetic anoxia. Such adverse effects associated with a "warm yet stratified" state have also been systemically documented at the CW Bill Young Regional Reservoir (Bryant et al. 2024), and Cedar Lake (James et al. 2015), USA. Taken together, these observations underscore the management relevance of hypolimnetic warming during late-season stratification decay; accordingly, future work should quantify threshold conditions that link this warming to hypolimnetic oxygen debt, for example through stability and withdrawal-rate percentiles, and test operational adaptation strategies that minimize exposure to this state.

4.4. An integrated analytical framework for long-term thermal reconstruction and driver attribution

Our study not only clarifies the thermal dynamics of reservoirs and their external drivers, but also introduces a new methodology that is of broader interest. By integrating routine observations with calibrated and evaluated CE-QUAL-W2 hindcast simulations for Rappbode Reservoir, we reconstruct a four-decade, depth-resolved (0–70 m) temperature archive with high vertical and temporal resolution, transforming sparse monitoring into continuous vertical profiles for subsequent analyses. Second, an interpretable machine-learning workflow (XGBoost with rolling-origin cross-validation and early stopping) provides fast, accurate prediction across depths, which is directly relevant for forecasting use-cases increasingly requested by reservoir managers (Carey et al. 2022). Third, the SHAP analysis provides additive, model-consistent statistical attributions that decompose thermal anomaly into depth- and sample-specific contributions of individual predictors. In practice, this yields a quantitative explanation of reservoir thermal anomalies without running time-consuming large scenario ensembles with the process model. Collectively, these advantages constitute an integrated mechanistic–statistical framework that is readily transferable to other stratified lakes and reservoirs for long-term thermal reconstruction, driver attribution and operational evaluation.

When this framework is extended beyond a single monitoring station, geostatistical approaches become particularly relevant because spatial covariance and interpolation are then central to the analysis. Methods such as kriging, co-kriging, and hybrid geostatistical–machine-learning frameworks can be highly effective for spatial covariance modelling, interpolation among monitoring sites, and prediction at unsampled locations (Brcković et al. 2025; Grey et al. 2025). These methods were not implemented here because the present analysis focused on depth-resolved temporal attribution at a fixed reservoir station, using a physically consistent CE-QUAL-W2 reconstruction rather than spatial interpolation across monitoring locations. They would be particularly valuable in future applications that extend this framework to multi-station observations or reservoir networks, where spatial covariance and temporal driver attribution need to be evaluated jointly.

5. Conclusion

- We developed an integrated CE-QUAL-W2 and machine-learning framework that reliably reconstructs, over nearly four decades, the full-depth thermal structure of Rappbode

Reservoir, capturing the spatiotemporal variability of layer-specific temperatures and stratification-strength metrics.

- SHAP attribution based on the integrated framework indicated strong vertical heterogeneity in the predictors of reservoir thermodynamics: surface temperature was primarily associated with the antecedent 30-day means of air temperature, dew-point temperature, and incident shortwave radiation, whereas the influence of shortwave radiation and outflow discharge increased progressively with depth and the relative effect of air temperature strongly declined.
- The dominant predictors of bottom-water temperature shifted systematically across stratification phases: during early stratification (May–June), the leading drivers were the antecedent 30-day mean air temperature, outflow discharge, and wind direction; during mid stratification (July–August), shortwave radiation and outflow discharge jointly dominated; during late stratification (September–October), outflow discharge became the largest contributor, with a SHAP contribution more than four times that of air temperature and shortwave radiation.
- Observations and simulations consistently showed that bottom-water temperature rose anomalously to ~10 °C during late stratification in 1981, 2001, 2002, and 2007, whereas it typically remained at 4–6 °C in other years. Further analyses indicated that these events were largely linked to intensified deep withdrawals in late stratification, with outflow discharge ~2.25 times higher than in typical years. The enhanced discharge weakened the density gradient and strengthened downward compensatory transport, driving the bottom-water thermal anomaly.

Code and data availability

The source code of the model CE-QUAL-W2 can be freely downloaded at http://cee.pdx.edu/w2/. All the datasets that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' contributions

CM conducted the W2 modeling approach, with the help from BG. CM, XK and YJ analyzed data and wrote the manuscript. CNC and KR reviewed, edited and proofread the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by National Natural Science Foundation of China (42577078, 42177062), National Key Research and Development Program of China (2023YFF0807203), Science and Technology Projects of Xizang Autonomous Region, China (XZ202501ZY0091) and International Partnership Program of Chinese Academy of Sciences (IPP)(045GJHZ2023055FN). We would also like to thank the PIFI Project of the Chinese Academy of Sciences for financially supporting this work (Grant Nos. 2024PG0017, and 2024DC0005). XK was further supported by the “Hundred People Program” of the Chinese

Academy of Sciences.

Reference

Akbari, E., S. K. Alavipanah, M. Jeihouni, M. Hajeb, D. Haase & S. Alavipanah, 2017. A review of ocean/sea subsurface water temperature studies from remote sensing and non-remote sensing methods. Water 9(12):936.

Anderson, E., C. Stow, A. Gronewold, L. Mason, M. McCormick, S. Qian, S. Ruberg, K. Beadle, S. Constant & N. Hawley, 2021. Seasonal overturn and stratification changes drive deep-water warming in one of Earth's largest lakes. Nature Communications 12 doi:10.1038/s41467-021-21971-1.

Bertone, E., R. A. Stewart, H. Zhang, M. Bartkow & C. Hacker, 2015. An autonomous decision support system for manganese forecasting in subtropical water reservoirs. Environmental Modelling & Software 73:133-147.

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

Boehrer, B. & M. Schultze, 2009. Density Stratification and Stability.

Brcković, A., T. Malvić, J. Orešković & J. Kapuralić, 2025. Comparison of Neural Network, Ordinary Kriging, and Inverse Distance Weighting Algorithms for Seismic and Well-Derived Depth Data: A Case Study in the Bjelovar Subdepression, Croatia. Geosciences 15(6):206.

Breiman, L., 2001. Random forests. Mach Learn 45(1):5-32.

Bryant, L. D., N. Brockbank & D. Austin, 2024. To mix or not to mix? A holistic approach to stratification-preserving and destratification aeration of drinking-water supply reservoirs. Water Research 261:121974.

Carey, C. C., W. M. Woelmer, M. E. Lofton, R. J. Figueiredo, B. J. Bookout, R. S. Corrigan, V. Daneshmand, A. G. Hounshell, D. W. Howard & A. S. Lewis, 2022. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. Inland Waters 12(1):107-120.

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.

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

Dadi, T., K. Friese, K. Wendt-Potthoff, R. Marcé & M. Koschorreck, 2023. Oxygen Dependent Temperature Regulation of Benthic Fluxes in Reservoirs. Global Biogeochem Cy 37(4):e2022GB007647.

Darko, D., D. Trolle, R. Asmah, K. Bolding, K. A. Adjei & S. N. Odai, 2019. Modeling the impacts of climate change on the thermal and oxygen dynamics of Lake Volta. Journal of Great Lakes Research 45(1):73-86.

Deka, P. P. & J. Weiner, 2024. XGBoost for Regression Predictive Modeling and Time Series Analysis: Learn how to build, evaluate, and deploy predictive models with expert guidance. Packt Publishing Ltd.

Deng, Y., Y. Tuo, J. Li, K. Li & R. Li, 2011. Spatial-temporal effects of temperature control device of stoplog intake for Jinping I hydropower station. Science China Technological Sciences 54(1):83-88.

Diebold, F. X. & R. S. Mariano, 2002. Comparing predictive accuracy. J Bus Econ Stat 20(1):134-144.

Dong, F., C. Mi, M. Hupfer, K. E. Lindenschmidt, W. Peng, X. Liu & K. Rinke, 2019. Assessing vertical diffusion in a stratified lake using a 3D hydrodynamic model. Hydrological Processes.

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.

Fisher, A., C. Rudin & F. Dominici, 2019. All models are wrong, but many are useful: Learning a variable's

importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 20(177):1-81.
Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Ann Stat:1189-1232.
Goldstein, A., A. Kapelner, J. Bleich & E. Pitkin, 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat 24(1):44-65.
Grey, V., T. Fletcher, K. Smith-Miles, B. Hatt & R. Coleman, 2025. Harnessing the strengths of machine learning and geostatistics to improve streamflow prediction in ungauged basins; the best of both worlds. Journal of Hydrology:133936.
Heddam, S., M. Ptak & S. Zhu, 2020. Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. Journal of Hydrology 588:125130 doi:https://doi.org/10.1016/j.jhydrol.2020.125130.
James, W. F., P. W. Sorge & P. J. Garrison, 2015. Managing internal phosphorus loading and vertical entrainment in a weakly stratified eutrophic lake. Lake Reserv Manage 31(4):292-305.
Jane, S. F., J. L. Mincer, M. P. Lau, A. S. Lewis, J. T. Stetler & K. C. Rose, 2023. Longer duration of seasonal stratification contributes to widespread increases in lake hypoxia and anoxia. Global Change Biology 29(4):1009-1023.
Kirillin, G. & T. Shatwell, 2016. Generalized scaling of seasonal thermal stratification in lakes. Earth-sci Rev 161:179-190.
Kobler, U. G., A. Wüest & M. Schmid, 2018. Combined effects of pumped-storage operation and climate change on thermal structure and water quality. Climatic Change doi:10.1007/s10584-018-2340-x.
Kraemer, B. M., R. M. Pilla, R. I. Woolway, O. Anneville, S. Ban, W. Colom-Montero, S. P. Devlin, M. T. Dokulil, E. E. Gaiser & K. D. Hambright, 2021. Climate change drives widespread shifts in lake thermal habitat. Nature Climate Change 11(6):521-529.
Kreakie, B., S. Shivers, J. Hollister & W. Milstead, 2021. Predictive model of lake photic zone temperature across the conterminous United States. Frontiers in environmental science 9:707874.
Kruskal, W. H. & W. A. Wallis, 1952. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47(260):583-621.
LaBrie, R., M. Hupfer & M. P. Lau, 2023. Anaerobic duration predicts biogeochemical consequences of oxygen depletion in lakes. Limnology and Oceanography Letters 8(4):666-674.
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.
Lewis Jr, W. M., J. H. McCutchan Jr & J. Roberson, 2019. Effects of Climatic Change on Temperature and Thermal Structure of a Mountain Reservoir. Water Resources Research 55(3):1988-1999.
Li, Z., Z. Zhang, S. Xiong, W. Zhang & R. Li, 2024. Lake Surface Temperature Predictions under Different Climate Scenarios with Machine Learning Methods: A Case Study of Qinghai Lake and Hulun Lake, China. Remote Sensing 16(17):3220.
Livingstone, D. M. & A. F. Lotter, 1998. The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with pal\sgmaelig; olimnological implications. J Paleolimnol 19:181-198.
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.
Lyu, H., N. Sha, S. Qin, M. Yan, Y. Xie & R. Wang, 2019. Advances in neural information processing systems. Adv Neural Inf Process Syst 32.

Meinson, P., A. Idrizaj, P. Nõges, T. Nõges & A. Laas, 2016. Continuous and high-frequency measurements in limnology: history, applications, and future challenges. Environ Rev 24(1):52-62.

Mi, C., M. A. Frassl, B. Boehrer & K. Rinke, 2018. Episodic wind events induce persistent shifts in the thermal stratification of a reservoir (Rappbode Reservoir, Germany). International Review of Hydrobiology 103:71-82.

Mi, C., D. P. Hamilton, M. A. Frassl, T. Shatwell, X. Kong, B. Boehrer, Y. Li, J. Donner & K. Rinke, 2022. Controlling blooms of Planktothrix rubescens by optimized metalimnetic water withdrawal: a modelling study on adaptive reservoir operation. Environmental Sciences Europe 34(1):1-16.

Mi, C., K. Rinke & T. Shatwell, 2023a. Optimizing selective withdrawal strategies to mitigate hypoxia under water-level reduction in Germany's largest drinking water reservoir. JEnvS.

Mi, C., A. Sadeghian, K.-E. Lindenschmidt & K. Rinke, 2019. Variable withdrawal elevations as a management tool to counter the effects of climate warming in Germany's largest drinking water reservoir. Environmental Sciences Europe 31(1):19.

Mi, C., T. Shatwell, X. Kong & K. Rinke, 2023b. Cascading climate effects in deep reservoirs: Full assessment of physical and biogeochemical dynamics under ensemble climate projections and ways towards adaptation. Ambio doi:10.1007/s13280-023-01950-0.

Mi, C., T. Shatwell, J. Ma, V. C. Wentzky, B. Boehrer, Y. Xu & K. Rinke, 2020. The formation of a metalimnetic oxygen minimum exemplifies how ecosystem dynamics shape biogeochemical processes: A modelling study. Water Research.

Mi, C., A. B. Tilahun, M. Flörke, H. H. Dürr & K. Rinke, 2024. Climate warming effects in stratified reservoirs: Thorough assessment for opportunities and limits of machine learning techniques versus process-based models in thermal structure projections. Journal of Cleaner Production:142347.

Nkwalale, L., R. Schwefel, M. Yaghouti & K. Rinke, 2023. A simple model for predicting oxygen depletion in lakes under climate change. Inland Waters 13(4):576-595.

Noori, R., R. I. Woolway, C. Jun, S. M. Bateni, D. Naderian, S. Partani, M. Maghrebi & M. Pulkkanen, 2023. Multi-decadal change in summer mean water temperature in Lake Konnevesi, Finland (1984–2021). Ecol Inform 78:102331 doi:https://doi.org/10.1016/j.ecoinf.2023.102331.

O'Reilly, C. M., S. Sharma, D. K. Gray, S. E. Hampton, J. S. Read, R. J. Rowley, P. Schneider, J. D. Lenters, P. B. McIntyre & B. M. Kraemer, 2015. Rapid and highly variable warming of lake surface waters around the globe. Geophysical Research Letters 42(24):10,773-10,781 doi:http://dx.doi.org/10.1002/2015GL066235.

Oleksy, I. A. & D. C. Richardson, 2021. Climate change and teleconnections amplify lake stratification with differential local controls of surface water warming and deep water cooling. Geophysical Research Letters 48(5):e2020GL090959.

Olsson, F., 2021. Impacts of water residence time on lake thermal structure: Implications for management and climate change. Lancaster University (United Kingdom).

Piccolroaz, S., M. Toffolon & B. Majone, 2013. A simple lumped model to convert air temperature into surface water temperature in lakes. Hydrology and earth system sciences 17(8):3323-3338.

Pilla, R. M., E. M. Mette, C. E. Williamson, B. V. Adamovich, R. Adrian, O. Anneville, E. Balseiro, S. Ban, S. Chandra & W. Colom-Montero, 2021. Global data set of long-term summertime vertical temperature profiles in 153 lakes. scientific data 8(1):200.

Rajwar, K., K. Deep & S. Das, 2023. An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. Artificial Intelligence Review:1.

Rinke, K., B. Kuehn, S. Bocaniov, K. Wendt-Potthoff, O. Buttner, J. Tittel, M. Schultze, P. Herzsprung, H. Ronicke,

K. Rink, K. Rinke, M. Dietze, M. Matthes, L. Paul & K. Friese, 2013. Reservoirs as sentinels of catchments: the Rappbode Reservoir Observatory (Harz Mountains, Germany). Environmental Earth Sciences 69(2):523-536 doi:http://dx.doi.org/10.1007/s12665-013-2464-2.

Schiavo, M. & D. Pedretti, 2026. Genetic and iterative metaheuristics-informed algorithms for precision shallow groundwater modeling and drought inference. Journal of Geophysical Research: Machine Learning and Computation 3(1):e2025JH000854.

Schwefel, R., L. G. Nkwalale, S. Jordan, K. Rinke & M. Hupfer, 2025. Temperatures and hypolimnetic oxygen in German lakes: Observations, future trends and adaptation potential. Ambio 54(3):428-447.

Sharma, S., D. K. Gray, J. S. Read, C. M. O'reilly, P. Schneider, A. Qudrat, C. Gries, S. Stefanoff, S. E. Hampton & S. Hook, 2015. A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009. Scientific data 2(1):1-19.

Shatwell, T., W. Thiery & G. Kirillin, 2019. Future projections of temperature and mixing regime of European temperate lakes. Hydrology and Earth System Sciences 23(3):1533-1551.

Sun, C., S. Wang, H. Wang, X. Hu, F. Yang, M. Tang, M. Zhang & J. Zhong, 2022. Internal nitrogen and phosphorus loading in a seasonally stratified reservoir: Implications for eutrophication management of deep-water ecosystems. Journal of Environmental Management 319:115681.

Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.

Weber, M., K. Rinke, M. Hipsey & B. Boehrer, 2017. Optimizing withdrawal from drinking water reservoirs to reduce downstream temperature pollution and reservoir hypoxia. Journal of Environmental Management 197:96-105.

Weinke, A. D. & B. A. Biddanda, 2019. Influence of episodic wind events on thermal stratification and bottom water hypoxia in a Great Lakes estuary. Journal of Great Lakes Research 45(6):1103-1112.

Wells, S. A., 2026. CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Qual-ity Model, Version 4.5. 5 User Manual: Part 5–Model Utilities.

Wentzky, V. C., J. Tittel, C. G. Jäger & K. Rinke, 2018. Mechanisms preventing a decrease in phytoplankton biomass after phosphorus reductions in a German drinking water reservoir—results from more than 50 years of observation. Freshwater biology 63(9):1063-1076 doi:http://dx.doi.org/10.1111/fwb.13116.

Werner, B. J., O. J. Lechtenfeld, A. Musolff, G. H. de Rooij, J. Yang, R. Gründling, U. Werban & J. H. Fleckenstein, 2021. Small-scale topography explains patterns and dynamics of dissolved organic carbon exports from the riparian zone of a temperate, forested catchment. Hydrology and Earth System Sciences 25(12):6067-6086.

Woolway, R. I., M. B. Kayastha, Y. Tong, L. Feng, H. Shi & P. Xue, 2025. Subsurface heatwaves in lakes. Nature Climate Change 15(5):554-559 doi:10.1038/s41558-025-02314-0.

Woolway, R. I., B. M. Kraemer, J. D. Lenters, C. J. Merchant, C. M. O'Reilly & S. Sharma, 2020. Global lake responses to climate change. Nature Reviews Earth & Environment 1(8):388-403.

Yousefi, A. & M. Toffolon, 2022. Critical factors for the use of machine learning to predict lake surface water temperature. Journal of Hydrology 606:127418.

Zhu, M., J. Wang, X. Yang, Y. Zhang, L. Zhang, H. Ren, B. Wu & L. Ye, 2022. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health 1(2):107-116 doi:https://doi.org/10.1016/j.eehl.2022.06.001.

Figures

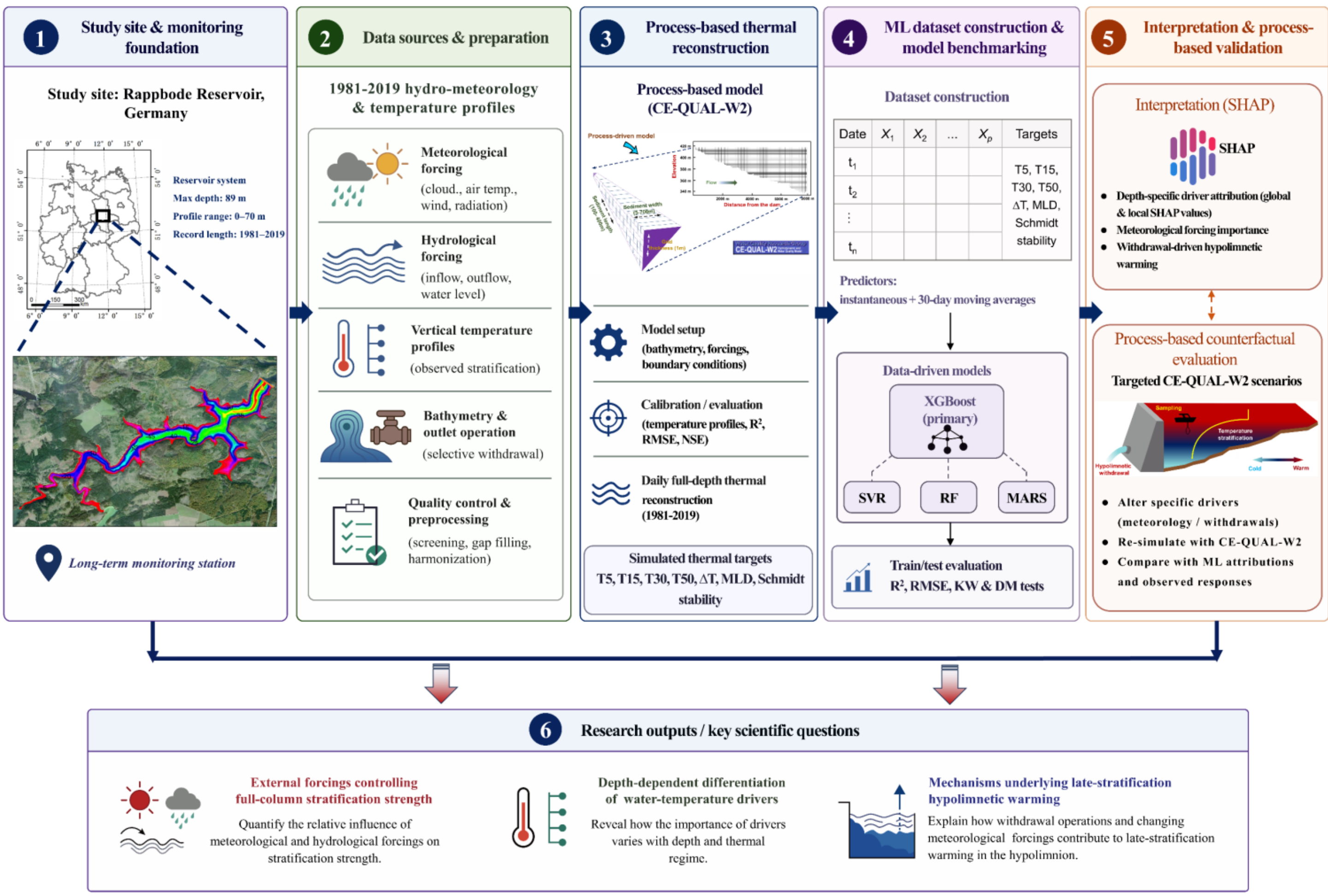

Fig. 1. Study site and methodological workflow of the integrated CE-QUAL-W2 and machine-learning framework. The workflow links long-term monitoring and hydro-meteorological forcing with CE-QUAL-W2 thermal reconstruction, data-driven model benchmarking, SHAP-based attribution, and targeted process-based validation. The final outputs address stratification-strength controls, depth-dependent thermal drivers, and mechanisms of episodic hypolimnetic warming.

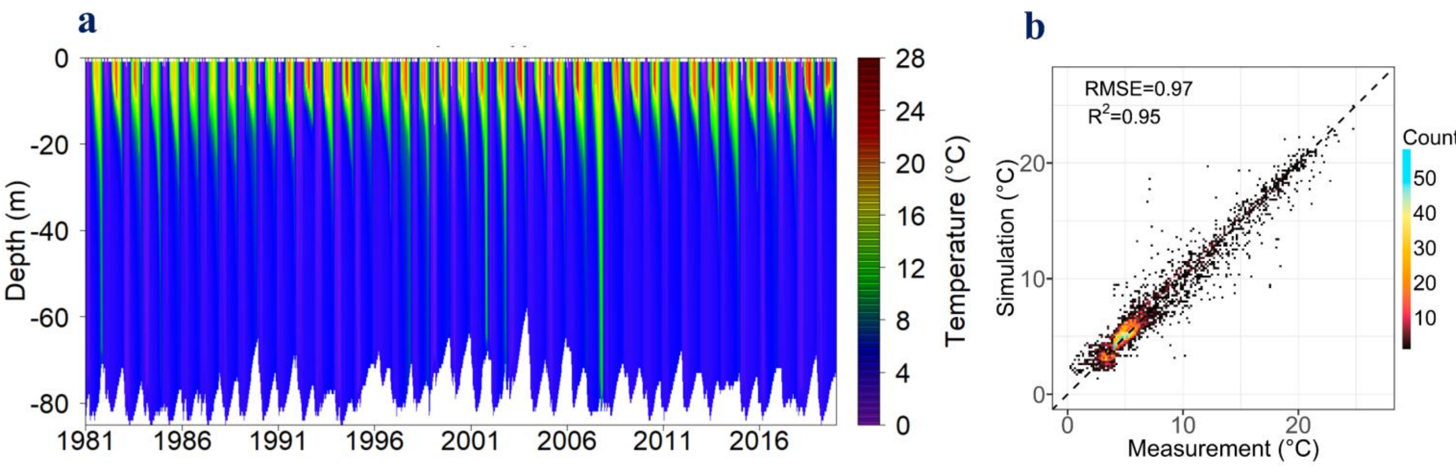

Fig. 2. (a) Thermal dynamics of Rappbode Reservoir, from W2 simulation results, during the study period; (b) Comparison between simulated and observed water temperature with the colour scale denoting the number of samples per hexagon. The dashed line represents the 1:1 line.

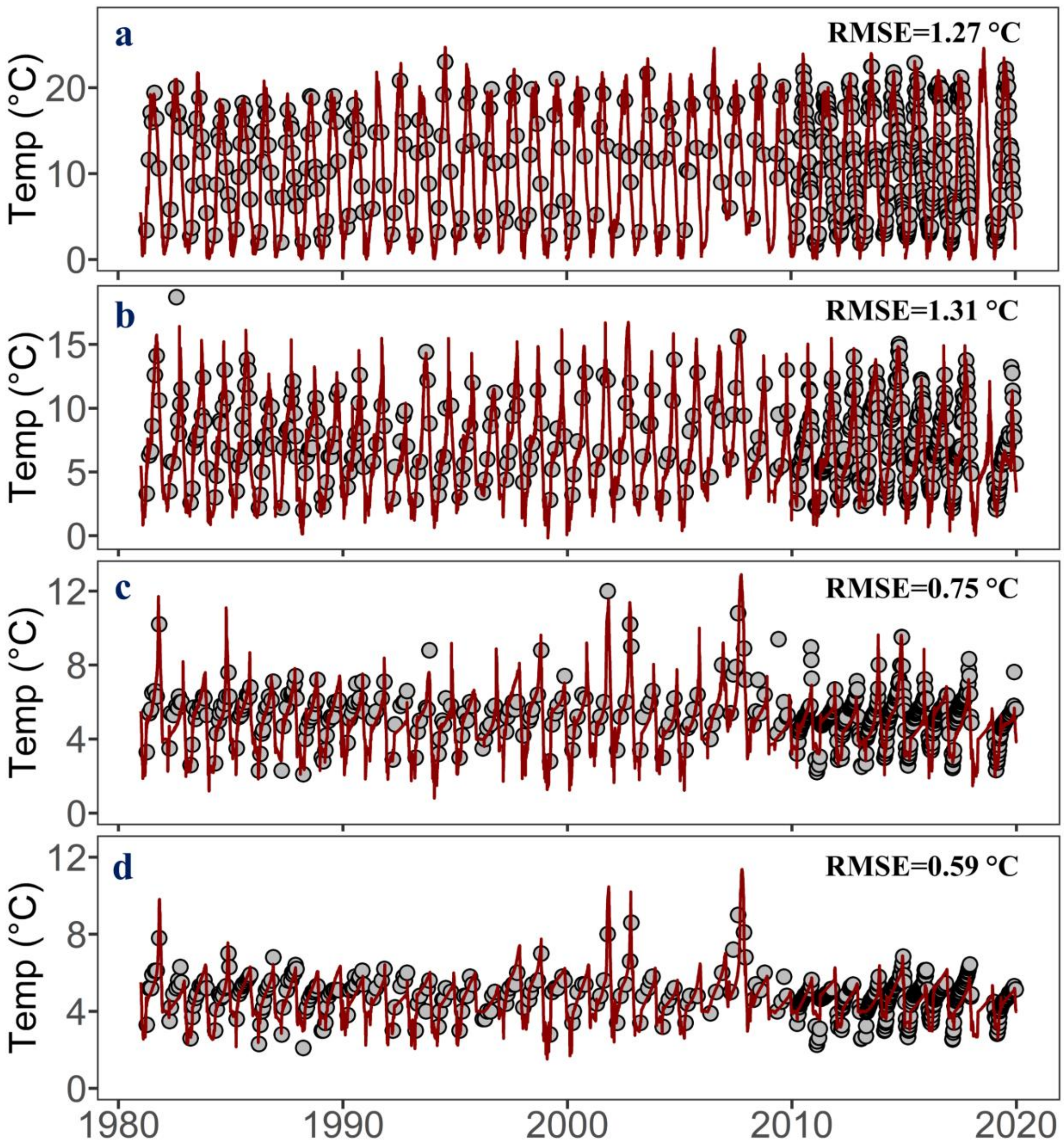

Fig. 3. Observed (grey circles) versus simulated (black line) water temperatures in the Rappbode Reservoir at 5 m (a), 15 m (b), 30 m (c), and 50 m (d). The RMSE for each panel is shown in the upper-right corner.

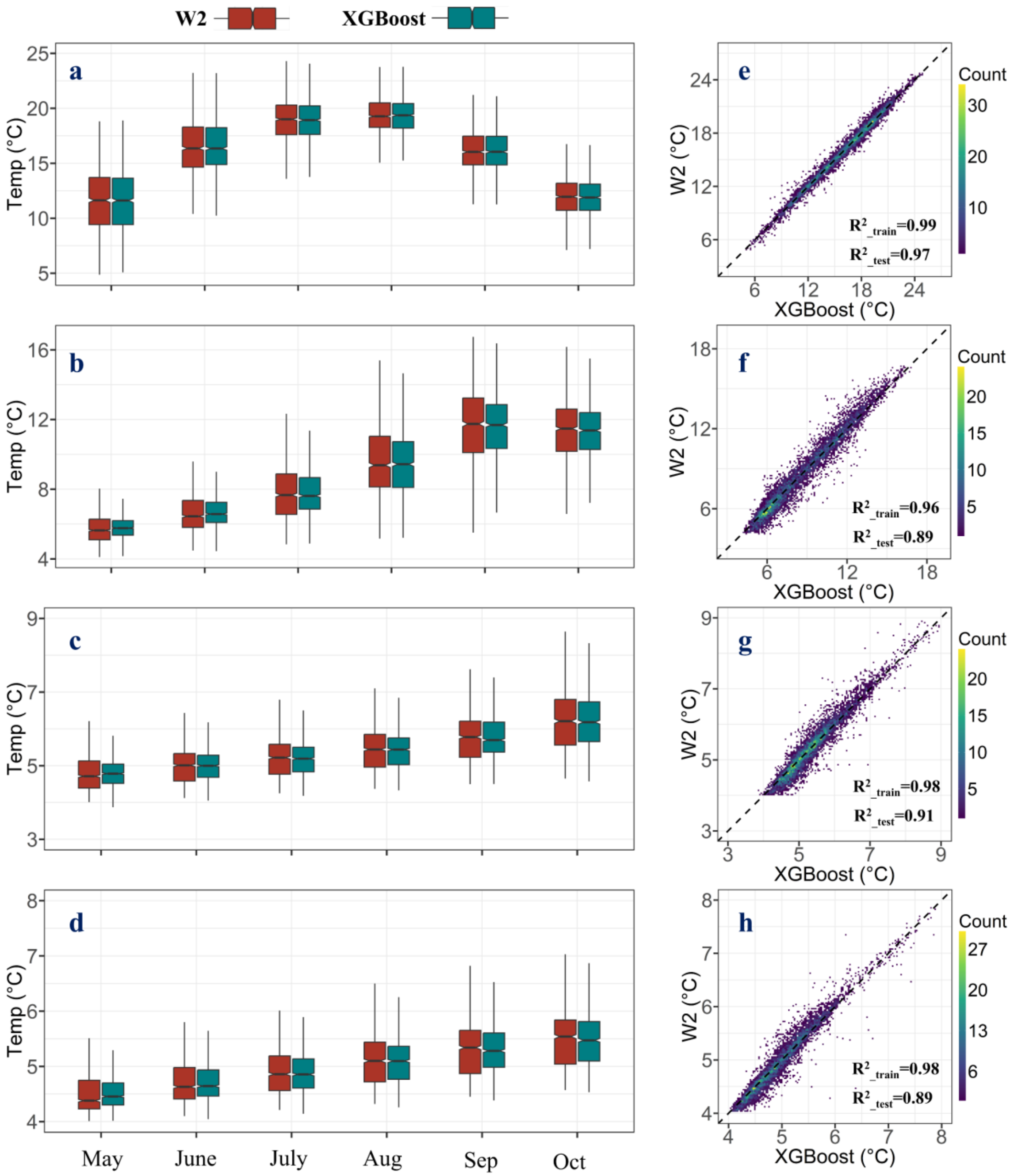

Fig. 4. Water temperature simulated by W2 and XGBoost in Rappbode Reservoir at the depth of 5m (a), 15m (b), 30m (c) and 50m (d), with the right column indicating the comparison for all specific depths. Training and testing R^2 values are reported separately in the corresponding panels.

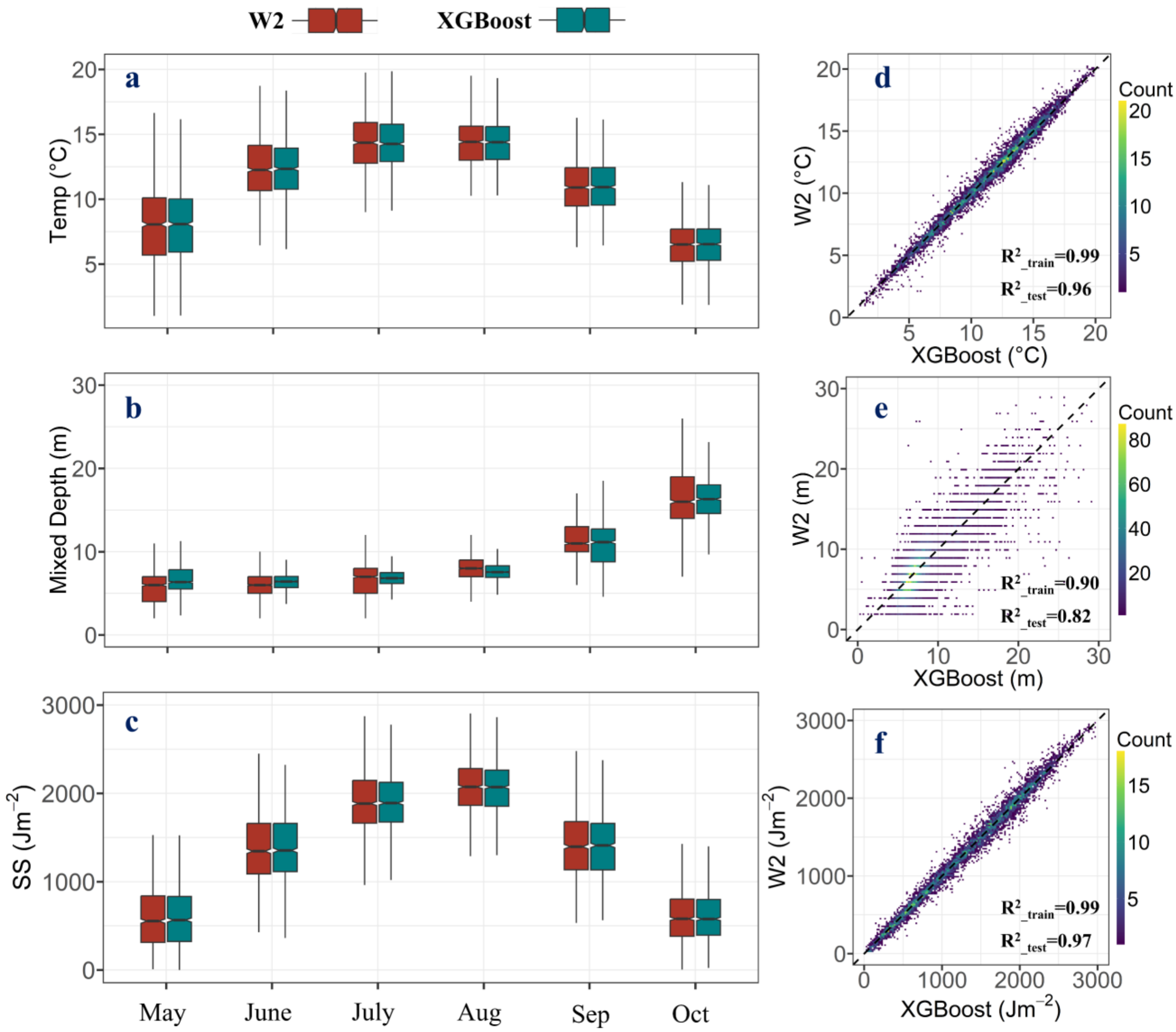

Fig. 5. Stratification dynamics simulated by W2 and XGBoost in Rappbode Reservoir for bottom-to-surface temperature difference (a), mixed-layer depth (b) and Schmidt Stability (c), with the right column indicating the comparison for the three indexes. Training and testing R^2 values are reported separately in the corresponding panels.

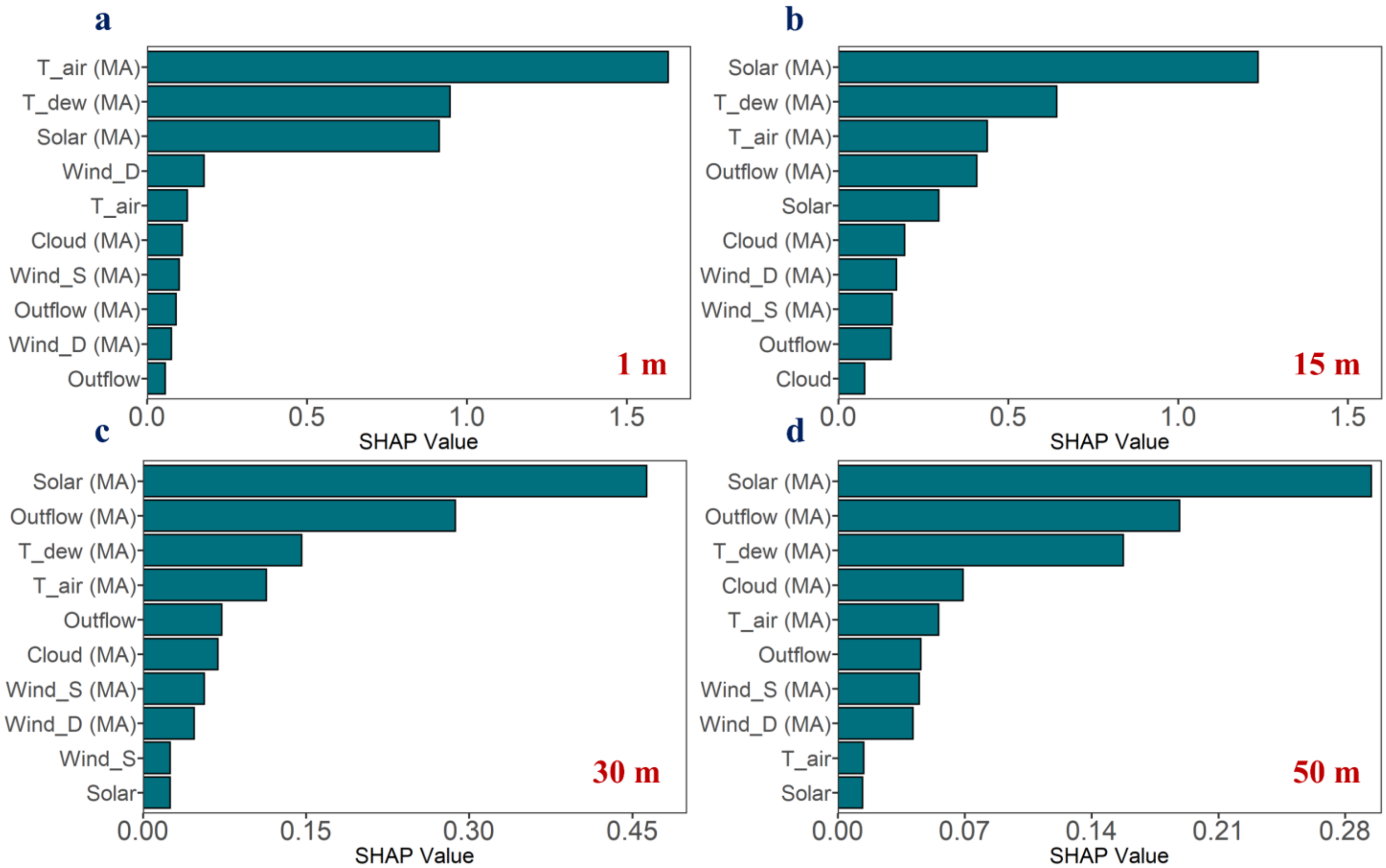

Fig. 6. Driving factors analysis for water temperature at the depth of 5m (a), 15m (b), 30m (c) and 50m (d) based on the SHAP values. MA indicates moving-average features.

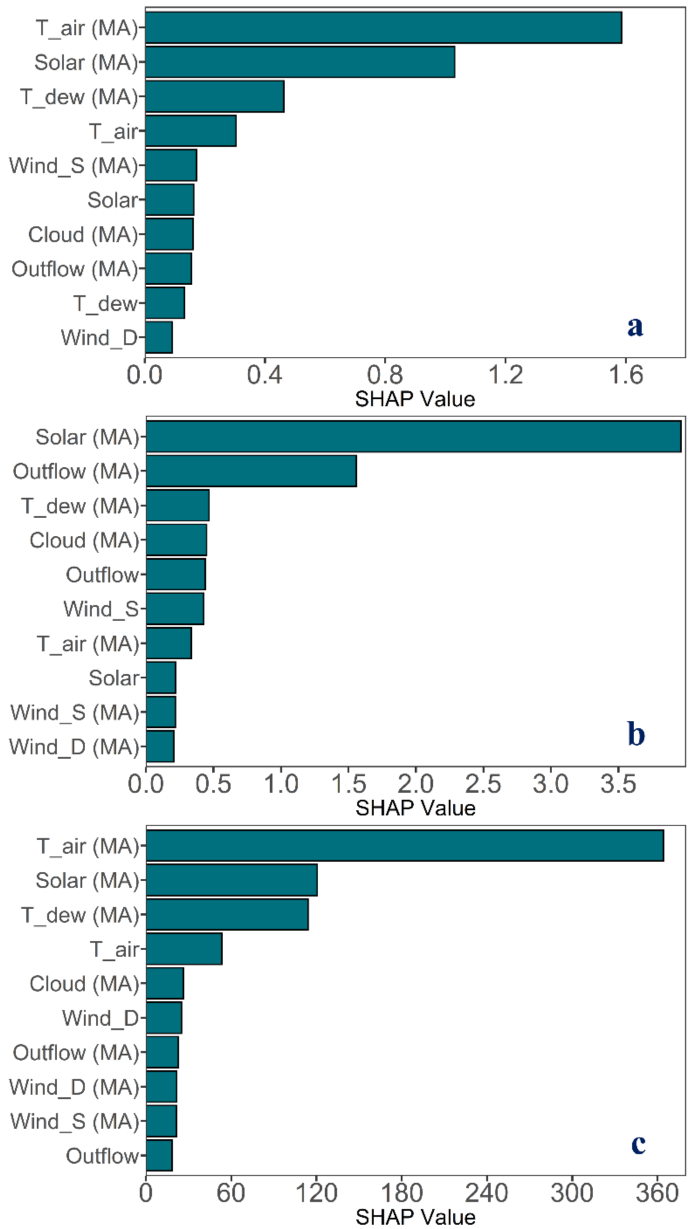

Fig. 7. Driving factors analysis for bottom-to-surface temperature difference (a), mixed-layer depth (b) and Schmidt Stability (c), based on the SHAP values. MA indicates moving-average features.

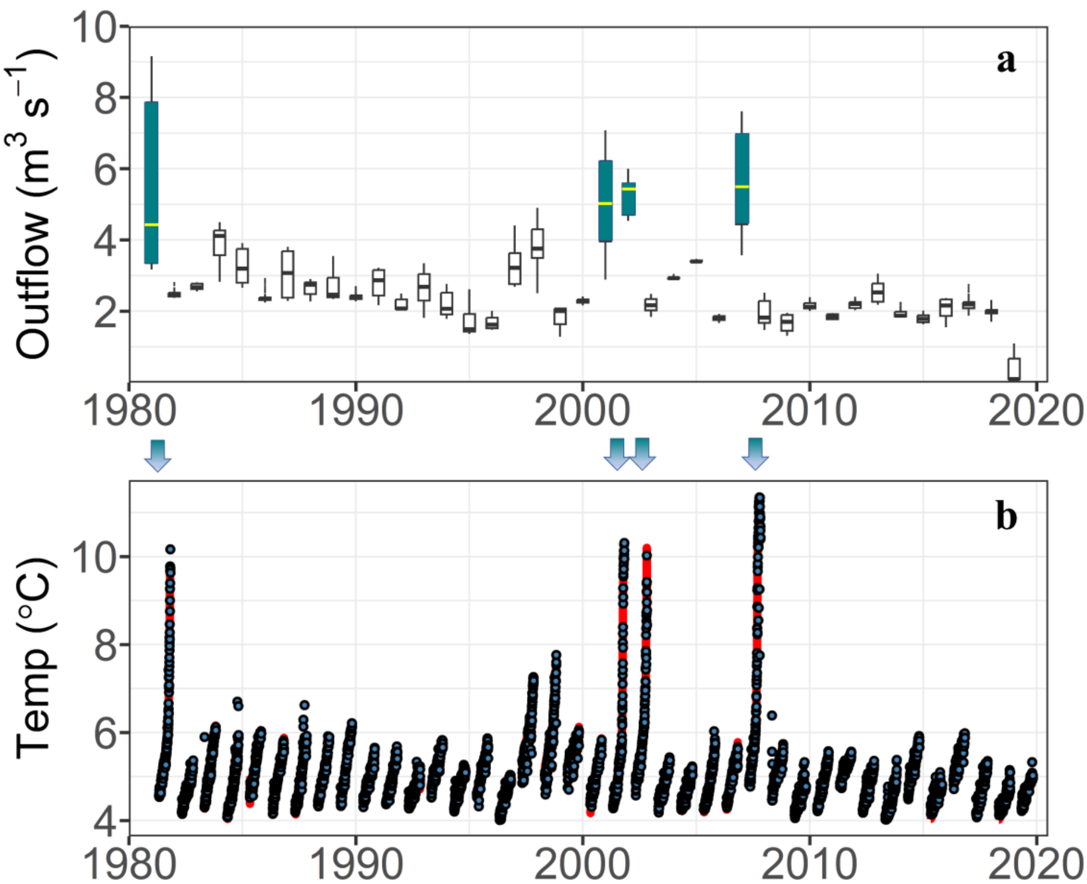

Fig. 8. (a) Moving-average of discharge during September–October for each year (with 1981, 2001, 2002, and 2007 highlighted in green); (b) water temperature, at the depth of 50m, from W2 (red lines) and XGBoost (black point) for each year.

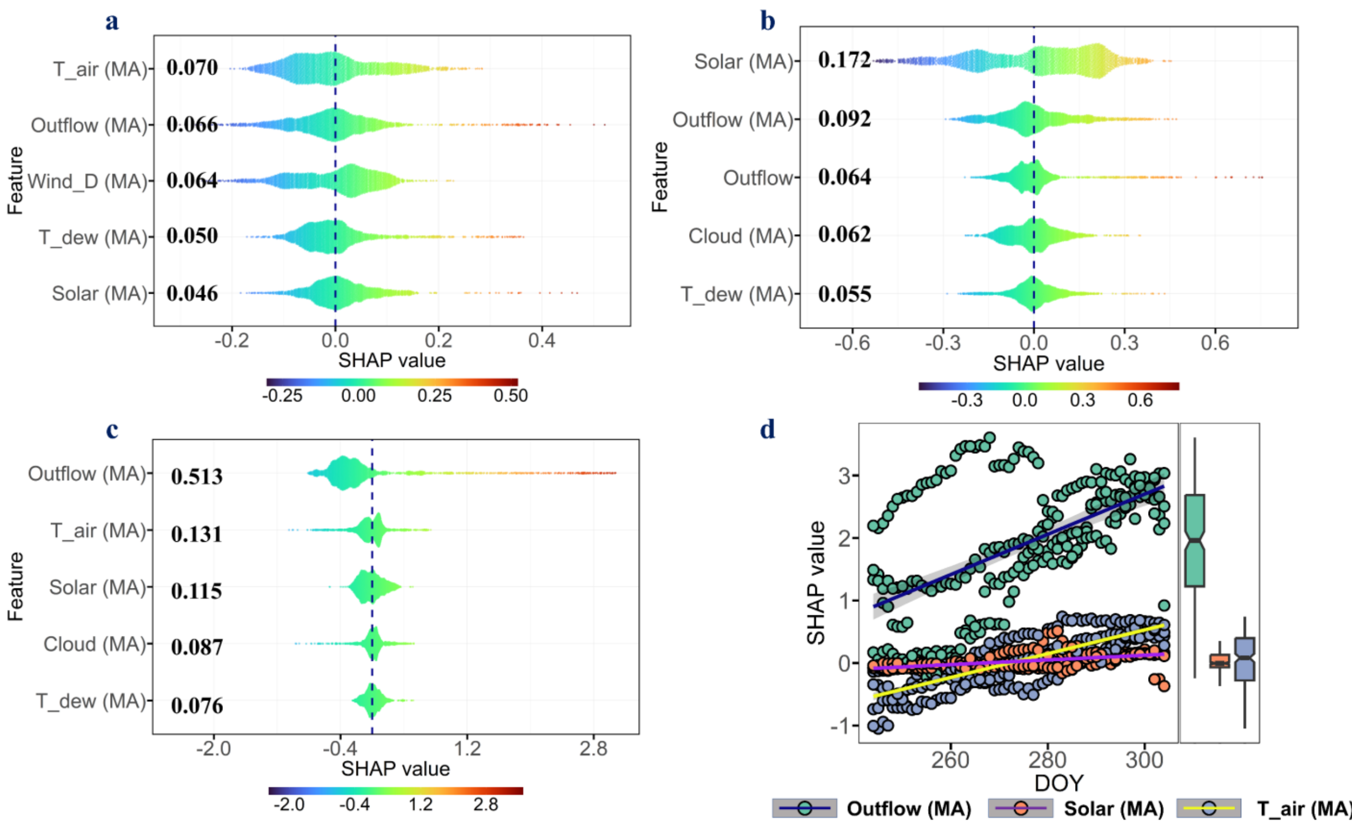

Fig. 9. Comparison of SHAP-derived driver importance for bottom-layer water temperature at 50 m depth. (a–c) SHAP value rankings showing how dominant drivers vary across the initial (May–June), middle (July–August), and final (September–October) stages of stratification. (d) Comparison of SHAP values during September–October for outflow (green), solar radiation (red), and air temperature (blue) in 1981, 2001, 2002, and 2007; points denote daily values plotted against day of year (DOY), with colored lines indicating linear regression fits, and adjacent boxplots summarizing the corresponding distributions. MA indicates moving-average features.

Tables

Table 1. Benchmark comparison of data-driven model performance in reproducing CE-QUAL-W2-simulated reservoir thermal-structure variables. Coefficients of determination (R^2) are reported separately for the training and testing datasets across all target variables, including depth-specific water temperatures and stratification metrics, for XGBoost, SVR, RF, and MARS.

Target variable	XGBoost		SVR		RF		MARS	
	Train	Test	Train	Test	Train	Test	Train	Test
Water temp (5 m depth)	0.99	0.97	0.96	0.95	0.98	0.96	0.94	0.93
Water temp (15 m depth)	0.96	0.89	0.78	0.75	0.89	0.82	0.73	0.71
Water temp (30 m depth)	0.98	0.91	0.82	0.79	0.88	0.83	0.75	0.73
Water temp (50 m depth)	0.98	0.89	0.88	0.77	0.86	0.76	0.74	0.67
Bottom-to-surface temperature difference	0.99	0.96	0.95	0.94	0.96	0.94	0.94	0.93
Mixed-layer depth	0.90	0.82	0.78	0.77	0.83	0.77	0.80	0.78
Schmidt stability	0.99	0.97	0.89	0.88	0.96	0.94	0.94	0.94

Table 2. Benchmark comparison of data-driven model performance in reproducing CE-QUAL-W2-simulated reservoir thermal-structure variables based on root mean square error (RMSE). RMSE values are reported separately for the training and testing datasets across all target variables, including depth-specific water temperatures and stratification metrics, for XGBoost, SVR, RF, and MARS. RMSE is expressed in °C for water temperature, m for mixed-layer depth, and J m^{-2} for Schmidt Stability.

Target variable	XGBoost		SVR		RF		MARS	
	Train	Test	Train	Test	Train	Test	Train	Test
Water temp (5 m depth)	0.34	0.70	0.78	0.83	0.35	0.71	0.93	0.94
Water temp (15 m depth)	0.33	0.67	1.34	1.43	0.93	1.21	1.51	1.52
Water temp (30 m depth)	0.19	0.41	0.51	0.62	0.40	0.53	0.62	0.66
Water temp (50 m depth)	0.11	0.28	0.30	0.41	0.33	0.42	0.44	0.48
Bottom-to-surface temperature difference	0.35	0.76	0.84	0.93	0.70	0.89	0.93	0.96
Mixed-layer depth	2.97	3.49	3.82	4.51	3.88	3.93	3.79	4.28
Schmidt stability	90.23	128.36	229.81	230.19	120.85	155.05	171.27	173.49