https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Hybrid Lake Model (HyLake) v1.0: unifying deep learning and physical principles for simulating lake-atmosphere interactions

Yuan He[1,2], Xiaofan Yang[1,2*]

[1] Guangdong Provincial Observation and Research Station for Coupled Human and Natural Systems in Land-ocean Interaction Zone, Beijing Normal University at Zhuhai, Zhuhai 519087, China
[2] State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Correspondence to: Xiaofan Yang (xfyang@bnu.edu.cn)

Abstract. Lake surface temperature (LST) serves as a crucial indicator of climate change in Earth systems. However, the challenge of improving LST and heat fluxes predictions remains due to the simplified physical principles inherent in traditional process-based models and the "black-box" structure of purely data-driven models. Accurate lake-atmosphere interaction modeling, which is essential for predicting LST and associated changes in latent heat (LE) and sensible heat (HE) fluxes, has yet to fully benefit from the integration of process-based and deep learning-based models. This study proposed Hybrid Lake Model v1.0 (HyLake v1.0), which integrates a Bayesian Optimized Bidirectional Long Short-Term Memory-based (BO-BLSTM-based) surrogate trained from Meiliangwan (MLW) site in Lake Taihu to approximate LST changes with surface energy balance equations. The performance of HyLake v1.0 was intercompared with FLake and hybrid lake models with different surrogates. Results demonstrated that HyLake v1.0 outperformed the others, with a R and RMSE of 0.99 and 1.08 °C in LST, a R and RMSE of 0.94 and 24.65 W/m^2 in LE and a R and RMSE of 0.93 and 7.15 W/m^2 in HE. To assess model generalization and transferability in ungauged lake sites, HyLake v1.0 exhibited superior performance, with a MAE of 0.85 °C, 21.56 W/m^2 and 6.63 W/m^2 in LST, LE and HE respectively, across all lake sites compared to FLake. Under ERA5 reanalysis datasets, HyLake v1.0 performed better for 14 of 15 variables (including LST, LE, and HE across 5 lake sites), with a MAE of 0.90 °C, 35.02 W/m^2 and 7.97 W/m^2 in LST, LE and HE respectively, indicating strong generalization and transferability. The results supported HyLake v1.0 exhibited an excellent capacity in estimating lake-atmosphere interactions for untrained lake sites, indicating a reasonable performance for extending the application in other ungauged lakes. Furthermore, the proposed model shows promising potential for predicting lake-atmosphere interactions, laying the solid basis for future improvements.

1 Introduction

Lakes are a critical component of Earth's system and serve as sensitive indicators of the interactions between climate change and land surface processes (O'Reilly et al., 2015; W. J. Wang et al., 2024). The surface water temperature (LST) of lakes plays as a key factor in lake-atmosphere modeling systems, which also regulates the hydro-biogeochemical processes (e.g., evaporation rate, ice cover, mixing regime, and thermal storage) (Tong et al., 2023; Culpepper et al., 2024; Woolway et al.,

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

2020). It has been observed that LST has risen globally at a rate of 0.34 °C per decade, driving shifts in biodiversity and altering ecosystem services within aquatic environments, with far-reaching consequences for Earth's climate (W. J. Wang et al., 2024; Woolway et al., 2020). These trends highlight the severe threats posed by climate change to global lake ecosystems in recent decades (Carpenter et al., 2011; Woolway et al., 2020).

Accurately predicting LST under pronounced climate change benefits from positioning the cascade of changes in lakes, both in terms of physical and biogeochemical processes, such as algal blooms, lake heatwave and cold spells (O'Reilly et al., 2015; X. W. Wang et al., 2024a; X. W. Wang et al., 2024b; Woolway et al., 2024). A variety of lake temperature models has been developed to predict LST, including statistical models, process-based models, and machine learning (ML) models. Process-based lake temperature models, such as the Freshwater Lake model (FLake), the General Lake Model (GLM), and WRF-Lake, are built on relationships between climate variables and LST, often employing simplified assumptions based on empirical physical principles (Mironov et al., 2010; Piccolroaz et al., 2024; L. J. Xu et al., 2016). However, these models are limited in their capacity to incorporate data-driven information and require extensive and diverse datasets for calibration and parameterization, which specifically pose significant challenges in data-scarce regions (Shen et al., 2023). In contrast, statistical models, such as the Air2Water model, are often applied in well-mixed lakes and rely on mathematical relationships between forcing variables and LST without incorporating mechanistic linkages (Piccolroaz et al., 2020; W. J. Wang et al., 2024). While these models are simpler and require fewer forcing data, they are constrained by the size of high-quality observational data, limiting their applicability across different lakes (Huang et al., 2021). ML lake temperature models, often considered a subset of statistical models, offer greater complexity and automation by leveraging large datasets (Piccolroaz et al., 2024; Wikle & Zammit-Mangion, 2023). ML models, such as Artificial Neural Networks (ANNs) and Long Short-Term Memory (LSTM) networks, have demonstrated superior performance in reconstructing LST compared to purely process-based models across global lakes (Almeida et al., 2022; Willard et al., 2022). These models excel at capturing complex relationships from datasets, however, their reliance on large training datasets, high computational costs, and the opaque "black-box" structure of their predictions present significant challenges. Specifically, their lack of transferability and explainability often make them less preferable than statistical and process-based models (Korbmacher & Tordeux, 2022; Piccolroaz et al., 2024). The limitations of these individual model types in predicting changes in lake-atmosphere interactions highlight the need for hybrid approaches that integrate the strengths of processed-based and data-driven model.

Hybrid models combine the physical principles of process-based models with the data-driven models, offering a multi-output structure that enhances both explainability and transferability while maintaining the flexibility and accuracy of data-driven models (Kurz et al., 2022; Piccolroaz et al., 2024). For example, Read et al. (2019) developed a hybrid deep-learning-based framework that integrates an energy balance-guided loss function from GLM with a Recurrent Neural Network (RNN) to reconstruct LST. This hybrid approach, which embedded the physical principles into deep-learning-based models, outperforms process-based models when transferred to unmonitored lakes (Willard et al., 2021). Nevertheless, such models face challenges related to computational expense, explainability and transferability in ungauged lakes and periods (Raissi et al., 2019; Willard et al., 2023). Alternatively, hybrid models can incorporate ML-based surrogates into process-based

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

backbones, where surrogates replace processes with insufficient physical understanding, enabling multiple outputs to be predicted simultaneously. These models, widely used in hydrological models, combine physical principles with ML-based surrogates, offering improved interpretability, generalizability, and transferability capabilities in ungauged regions (Shen et al., 2023). Specifically, Feng et al. (2022) embedded neural networks into the Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrological model to predict multiple physical outputs, achieving performance comparable to purely data-driven models. Similarly, L. J. Zhong et al. (2024) developed a distributed framework integrating ML and traditional river routing models to predict streamflow. These hybrid approaches outperform traditional process-based models and require relatively less training data by involving the integration of physical constraints, thereby providing a powerful tool for uncovering previously unrecognized physical relationships (Shen et al., 2023). Lake-atmosphere interactions represent a tightly coupled system (B. B. Wang et al., 2019), where process-based models traditionally approximate the interdependence between LST and heat fluxes (latent and sensible heat, LE and HE). The increasing LST influences these fluxes, while the emitted heat fluxes act as boundary conditions that in turn affect LST (R. I. Woolway et al., 2015). Thus, the development of novel hybrid lake models are promising in offering improved approximations and advance our understanding of lake-atmosphere interactions.

Lakes exhibit considerable regional variability, making it challenging to conduct long-term and large-scale observations (Zhang et al., 2019). Traditional lake models seem challenging to be generalized in ungauged lake or even regions in a large lake. Lake Taihu, the third largest freshwater lake in China, where differ significantly in its biological characteristics (Table 1), has experienced deterioration in water quality, thereby significantly threatening drinking water security (Zhang et al., 2020; Yan et al., 2024). These issues have prompted extensive field investigations and data collection in Lake Taihu, providing rich observational data for lake-atmosphere modeling. Specifically, the Lake Taihu eddy flux network has supported long-term observations of microclimate, radiation, and energy fluxes, offering a valuable dataset for the development of hybrid lake models (Zhang et al., 2020). The objectives of this study are to (1) develop a novel hybrid lake model (HyLake v1.0) by embedding LSTM-based surrogate into process-based model; (2) validate the performance of HyLake v1.0 in LST, LE, and HE; and (3) evaluate the generalizability and transferability of HyLake v1.0 in lake sites with different biological characteristics under ERA5 reanalysis datasets. The proposed HyLake v1.0 is beneficial to improve the performance in describing the lake-atmosphere interactions for ungauged lakes.

2 Materials and Methodology

2.1 Study area and datasets

Lake Taihu (30.12–32.22°N, 119.03–121.91°E), located in the Yangtze Delta, is the third-largest freshwater lake in China, covering an area of 2400 km² with an average depth of 1.9 m, with a rapid increasing rate of ~0.37 °C/decade (Yan et al., 2024; Zhang et al., 2020; Zhang et al., 2018). As a typical urban lake, situated in one of the most densely populated regions of China, has experienced significant eutrophication, marked by recurrent algae blooms, which threaten local drinking water

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

security (Yan et al., 2024). Given the pressing need to understand the challenges surrounding water quality improvement and hydro-biogeochemical processes in Lake Taihu, this study employs the lake models to assess these issues across 5 distinct sites from the Taihu Lake eddy flux network: Meiliangwan (MLW), Dapukou (DPK), Bifenggang (BFG), Xiaoleishan (XLS), and Pingtaishan (PTS) (Figure 1, Table 1). These sites span varying biological characteristics and eutrophication gradients, offering a comprehensive view of the lake's ecological diversity (Zhang et al., 2020), providing a solid data base for evaluating the generalizability and transferability of lake models. Specifically, MLW (31.4197°N, 120.2139°E) as the first lake site in Lake Taihu eddy flux network, located at the northern shore of the lake in MLW Bay, with a biological characteristic of eutrophication. BFG (31.1685°N, 120.3972°E) as a lake site, located in the eastern shore of Lake Taihu, features a submerged macrophyte community and relatively clean water. PTS (31.2323°N, 120.1086°E), situated centrally in the lake, has experienced significant algae blooms and lacks aquatic vegetation. DPK (31.2661°N, 119.9312°E), located on the western shore of Lake Taihu, is marked by severe eutrophication and deeper water, while XLS (30.9972°N, 120.1344°E), located on the southern shore of Lake Taihu, is a vegetation-free, clean water site (Zhang et al., 2020).

Each site is equipped with an eddy covariance system that continuously monitors latent (LE) and sensible heat fluxes (HE) using sonic anemometers and thermometers (Model CSAT3A; Campbell Scientific, Logan, UT, USA) positioned 3.5 to 9.4 m above the lake surface. Hydrometeorological variables, including air humidity and temperature (Model HMP45D/HMP155A; Vaisala, Helsinki, Finland), wind speed (Model 03002; R.M. Young Co., Traverse City, MI, USA), and net radiation components (Model CNR4; Kipp & Zonen, Delft, the Netherlands), are also measured. These variables, alongside inferred radiative LST, were collected at 30-minute intervals and are publicly accessible via Harvard DataVerse (https://doi.org/10.7910/DVN/HEWCWM) (Lee, 2004; Zhang et al., 2020). The dataset spans from 2012 to 2015, with any data gaps filled using ERA5 reanalysis datasets. For this study, variables from the MLW site are used to train the Long Short-Term Memory (LSTM)-based surrogates (Sect. 2.2), while data from the remaining sites serve to evaluate the generalization of HyLake v1.0 and train the LSTM-based surrogates.

To address the generalization of HyLake v1.0 across different forcing datasets, this study utilized ERA5 hourly reanalysis datasets from 2012 to 2015, with a spatial resolution of 0.25° at a single level to validate the performance of HyLake v1.0. These datasets, available from the Climate Data Store (cds.climate.copernicus.eu) (Hersbach et al., 2020), include variables such as air temperature, dew point temperature, surface pressure, wind speed, and surface net longwave and shortwave radiation. The ERA5 datasets are used to validate HyLake v1.0 in Lake Taihu, providing insights into the model's generalization, transferability and performance under different climatic forcing datasets.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

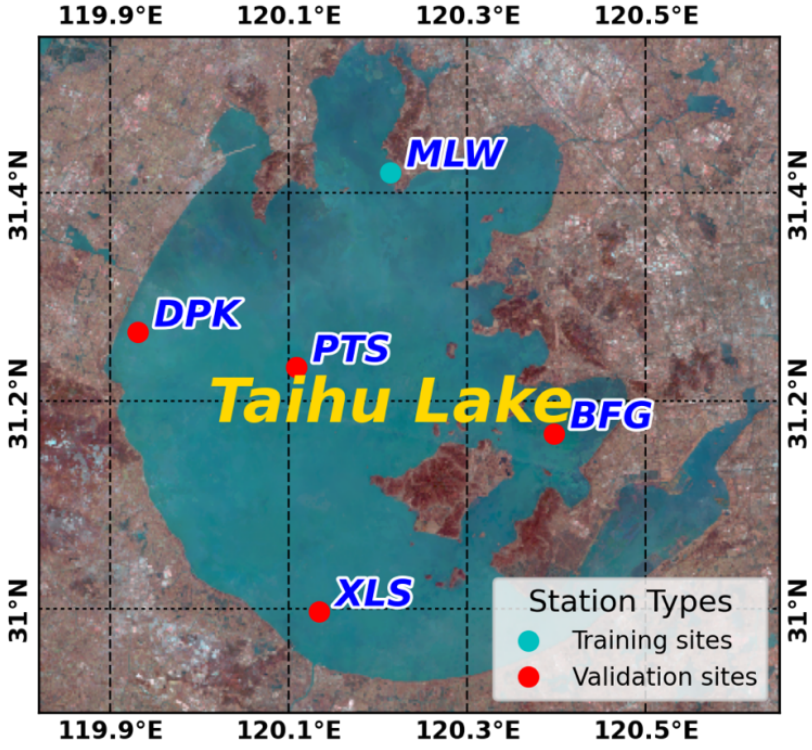


Figure 1: The locations of Lake Taihu and the five eddy covariance lake sites (MLW, DPK, BFG, XLS, and PTS) are shown in cyan and red bubbles, overlaid on a true-color image from Landsat 8. MLW as a training site was used to train BO-BLSTM-based surrogate, while the other validation sites were adapted as ungauged sites to valid the HyLake v1.0 performance.

Table 1. Information for selected lake sites in Lake Taihu.

Site	MLW	DPK	BFG	XLS	PTS
Lat. (◦N)	31.4197	31.2661	31.1685	30.9972	31.2323
Lon. (◦E)	120.2139	119.9312	120.3972	120.1344	120.1086
Start date	Jun 2010	Aug 2011	Dec 2011	Nov 2012	Jun 2013
Biology	Eutrophic	Super eutrophic	Submerged macrophyte	Transitional	Mesotrophic
Purpose	Train	Validation	Validation	Validation	Validation

2.2 The model architecture of HyLake v1.0

HyLake v1.0 was constructed in this study based on the backbone of physical principles from process-based lake models and then couple a LSTM-based surrogate for LST approximation to further solve the untrained variables (e.g., LE, HE), which is shown in Figure 2. Several variants of HyLake v1.0 that are composed of the same physical principles and distinct LSTM-based surrogates using different training strategies were used to intercompare with HyLake v1.0, named as Baseline and

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

TaihuScene hybrid models. The following sections will introduce the architecture of HyLake v1.0 from the physical principles, LSTM-based surrogates and their training strategies, respectively.

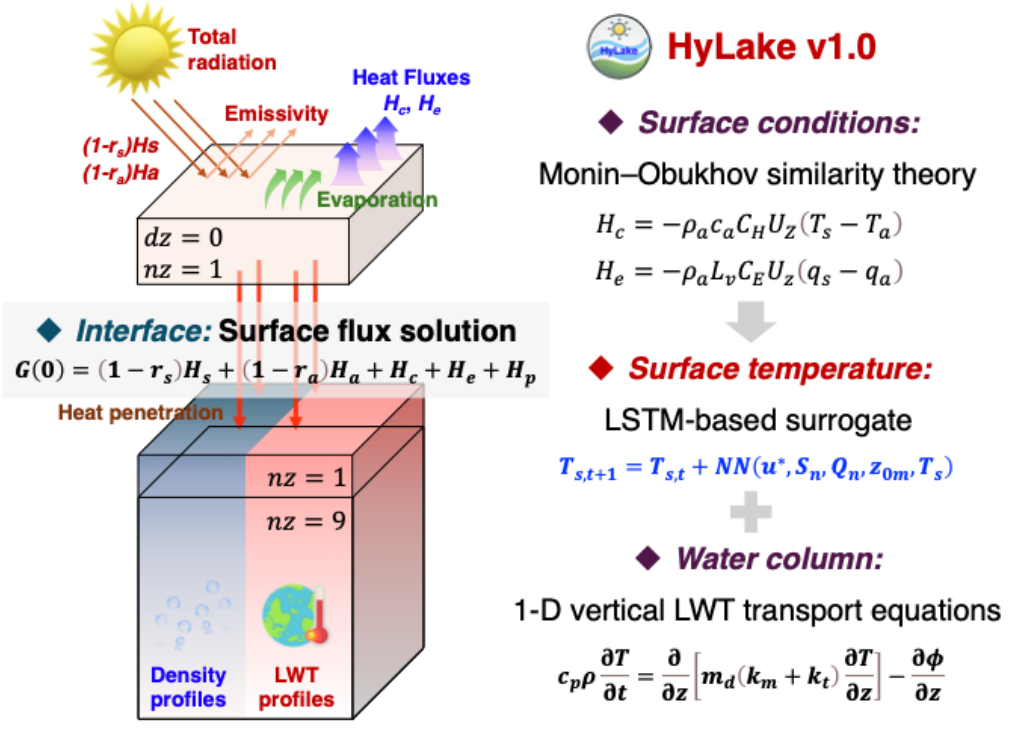


Figure 2: Conceptual model of HyLake v1.0. The LSTM-based surrogates were added to approximate the LST based on the surface conditions that calculated from Monin-Obukhov similarity theory, which further correct the LST in surface flux solution at the next timestep.

2.2.1 Physical principles for Lake-Atmosphere Modeling Systems

The backbone of HyLake v1.0 named process-based backbone model (PBBM) was constructed referred to the process-based lake models based on the governing equations and parameterization schemes of previously validated lake physical processes (Sarovic et al., 2022). This model serves as the foundational backbone to couple with various deep-learning-based surrogates. The conceptual model of PBBM is depicted in Figure 2 and Table 2. Specifically, the lake-atmosphere modeling system in PBBM primarily involves energy balance equations for solving surface fluxes at the lake-atmosphere interface and the 1-D vertical lake water temperature transport equations within the water column (Piccolroaz et al., 2024).

The changes in LST are primarily driven by the net heat fluxes entering the lake surface. Therefore, the net heat flux is imposed as a Neumann boundary condition at the upper boundary of the water column. Following Piccolroaz et al. (2024), the net heat flux $G(0)$ (W/m²) into the lake surface can be expressed by the energy balance equation:

$$G(0) = (1 - r_s)H_s + (1 - r_a)H_a + H_c + H_e + H_p \quad (1)$$

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

where H_s (W/m^2) and H_a (W/m^2) represent net downward shortwave and longwave radiation (also referred to the net solar and thermal radiation in ERA5), respectively; r_s and r_a account for the shortwave and longwave albedos of water; the HE and LE are denoted by H_c (W/m^2) and H_e (W/m^2); H_p represent the heat flux (W/m^2) brought from precipitation, often calculated via an empirical equation to quantify (Sarovic et al., 2022). All heat fluxes are considered positive in downward direction. The net shortwave and longwave radiation are derived from observation in Lake Taihu eddy flux network and ERA5 reanalysis datasets.

The sensible (HE, H_c) and latent (LE, H_e) heat fluxes follows the scheme by Verburg and Antenucci (2010)(Figure 2-3):

$$H_c = -\rho_a c_a C_H U_Z (T_s - T_a) \quad (2)$$

$$H_e = -\rho_a L_v C_E U_z (q_s - q_a) \quad (3)$$

where ρ_a (kg/m^3) donated air density; c_a = 1005 J/kg·K is the specific heat of air; $L_v \approx$ 2500 kJ/kg is the LE of vaporization; C_E and C_H are transfer coefficients for HE and LE derived iteratively with the Monin–Obukhov length based on bulk flux algorithms; U_Z (m/s) is the wind speed at observed height; T_s (°C) accounts for LST solved by 1-D vertical LWT transport equation; and T_a (°C) present the air temperature. Further details on the heat flux calculations can be found in Verburg & Antenucci (2010) and R. Iestyn Woolway et al. (2015).

At the bottom boundary of the 1-D lake model, the zero-temperature-gradient boundary and the zero-flux boundary are imposed as shown in Figure 2, which can be expressed as:

$$\partial T_z / \partial z = 0 \quad (4)$$

$$G(z_{max}) = 0 \quad (5)$$

where T_z (°C) present the lake water temperature; z means the mean lake depth; $G(z_{max})$ (W/m^2) account for the heat exchange between water column and sediment, which is set to 0 at the bottom boundary.

To simulate vertical temperature profiles in the water column (Figure 2), PBBM solves a 1-D vertical lake water temperature transport equations within a cylindrical water column assumption of constant surface area as follows:

$$c_p \rho \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left[m_d (k_m + k_t) \frac{\partial T}{\partial z}\right] - \frac{\partial \phi}{\partial z} \quad (6)$$

where c_p is the specific heat capacity of water, which is set to 1006 J/kg·K; ρ (kg/m^3) present water density, which can be calculated from lake water temperature (T, °C) at different depths; t (s) and z (m) is simulated time and depth of water column respectively; m_d is the enlarge coefficient of thermal conductivity, which is also set to 5 in PBBM; k_m and k_t (W/m·K) are the molecular and turbulent thermal conductivity, respectively; ϕ (W/m^2) is defined by the heat flux that penetrate into the lake from net solar radiation. The specific parameterization of PBBM follows the lake module in CLM v5.0 (Subin et al., 2012). The PBBM provides a backbone for HyLake v1.0, which adapted a LSTM-based surrogate to solve the LST instead of solving 1-D vertical lake water temperature transport equations by the implicit Euler scheme (Figure 3a).

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

2.2.2 LST Approximation Using LSTM-based surrogates

In PBBM, LST along with the LWT in the 1-D vertical lake water temperature transport equations (Equation (6)) are typically solved simultaneously using an implicit Euler scheme for numerical stability (Sarovic et al., 2022). This can be expressed in matrix form as:

$$MT^{n+1} = AT^{n} + B \tag{7}$$

where, M, A and B jointly form the tridiagonal system of equations based on implicit Euler scheme, which can be further decomposed into equations (8) to (10):

$$M = \begin{cases} \frac{-k_m - k_{t,j-1/2}}{z_j - z_{j-1}}, & z = j-1 \\ \frac{\Delta z_j c_p \rho_j}{\Delta t} + \frac{k_m + k_{t,j+1/2}}{z_{j+1} - z_{j-1}} + \frac{k_m + k_{t,j-1/2}}{z_j - z_{j-1}}, & z = j \\ \frac{-k_m - k_{t,j+1/2}}{z_{j+1} + z_j}, & z = j+1 \end{cases} \tag{8}$$

$$A = \frac{\Delta z_j c_p \rho_j}{\Delta t} \tag{9}$$

$$B = \phi^{n}_{j+1/2} - \phi^{n}_{j-1/2} \tag{10}$$

where j denotes the index of each water-column layer (from 1 to 10 in this study); n *re*presents the time step. Δz_j and ρ_j are the thickness and density at j^{th} water column, respectively. The terms $k_{t,j-1/2}$ and $\phi^{n}_{j-1/2}$ refer to the molecular diffusivity and residual radiation at middle location between j-1^{th} and j^{th} at n step. All other variables in these equations follow the same notations and definitions as given previously.

HyLake v1.0 and its variants, including Baseline and TaihuScene employed LSTM-based surrogates to solve the LST rather than the implicit Euler scheme for each time step (Figure 3a). Specifically, several sequence-to-one LSTM-based surrogates are adapted to be trained to approximate increments in LST based on dynamic inputs, including time series of historical 24-step variables of LST, friction velocity (u^*, m/s), surface roughness length (z_{0m}, m), and $G(0)$. These dynamic parameters were calculated from the outputs of surface flux solutions based on the observations. Thus, to address the numerical stability of autoregressive prediction in iterations, the LST increments can be expressed by:

$$T_{s,t+1} = T_{s,t} + \Delta T = T_{s,t} + NN(u^*, S_n, Q_n, z_{0m}, T_s) \tag{11}$$

where $NN(\cdot)$ donates LSTM-based surrogates within different HyLake models, which will activate to approximate the increment of lake surface temperature for each time step. Such LSTM-based surrogates have shown stable autoregressive prediction capabilities in hydrological modeling (J. Liu et al., 2024) and can readily be coupled with PBBM to provide numerically robust predictions of lake surface temperature. The other untrained variables, such as LE, and HE, were derived from the module of surface flux solutions (equation (1-3)).

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

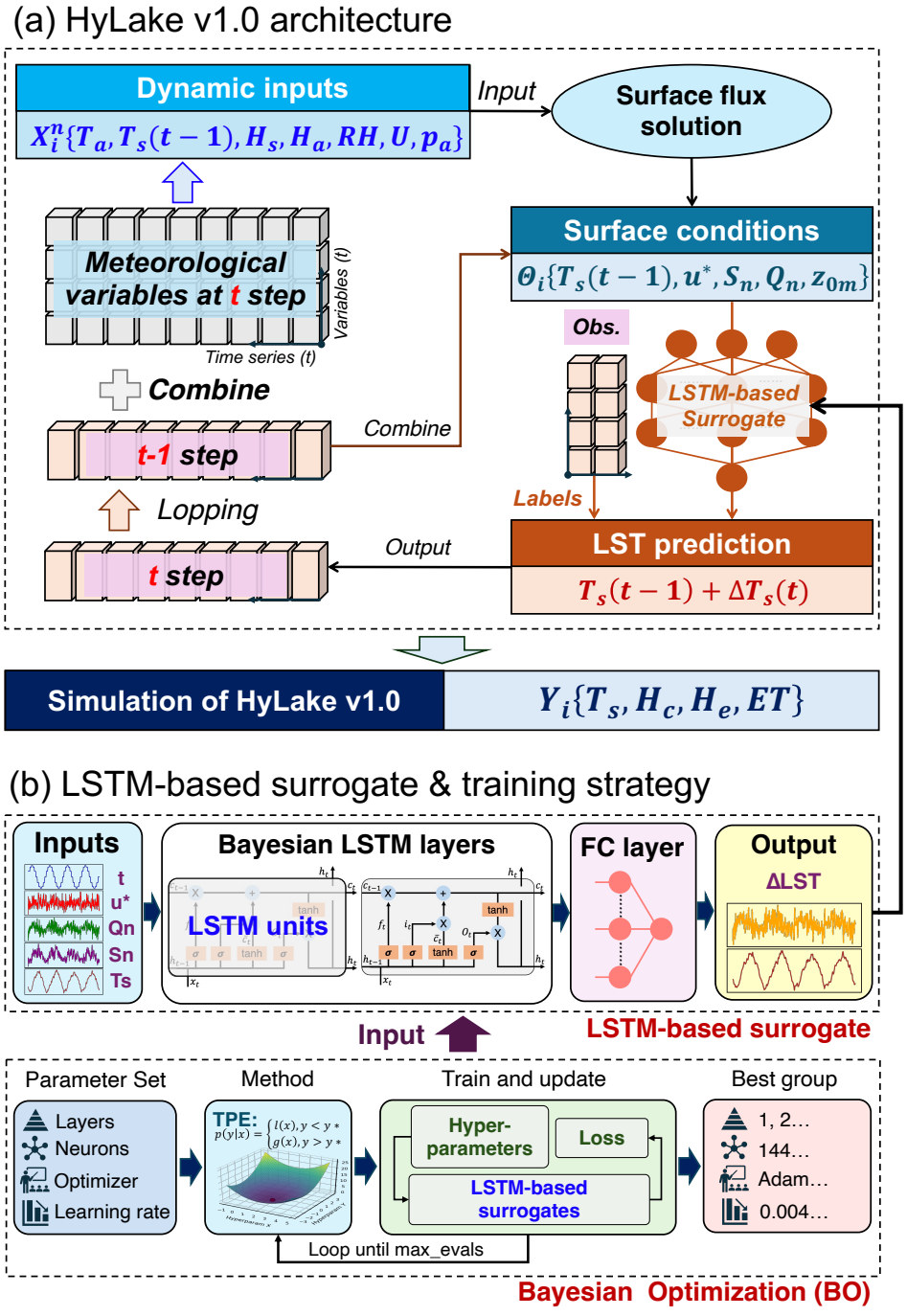


Figure 3: The general architecture of HyLake v1.0. (a) Coupling strategy of physical principles and LSTM-based surrogates and (b) training strategy of LSTM-based surrogates in HyLake v1.0. X_i^n, θ_i, and Y_i represent dynamic inputs for forcing surface flux solution in PBBM, surface conditions calculated from surface flux solution, and the outputs calculated from HyLake v1.0, respectively.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

2.2.3 The training strategy of LSTM-based surrogates

The LSTM-based surrogates in HyLake v1.0 are composed by the stacked LSTM or Bayesian LSTM (BLSTM) units and fully-connected layers (Figure 3b), including two different units to separately construct LSTM-based surrogates in HyLake v1.0 that trained on observations. LSTM units are a type of Recurrent Neural Network (RNN) designed to avoid vanishing gradients problem, making them particularly suited for time series forecasting (Sherstinsky, 2020). Here, this study constructed a LSTM surrogate for Baseline and a BLSTM surrogate for HyLake v1.0 to couple in PBBM (Table 2). Specifically, the LSTM unit comprises three gates: the forget gate, the input gate and the output gate, which controls whether information should be retained or updated (Hochreiter, 1997). The forget gate was firstly introduced by Gers et al. (2000), which can be expressed as follow:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (12)$$

$$\tilde{c}_t = \tanh\left(W_{\tilde{c}} x_t + U_{\tilde{c}} h_{t-1} + b_{\tilde{c}}\right) \quad (13)$$

where f_t is a resulting vector of the forget gate; $\sigma(\cdot)$ and $\tanh(\cdot)$ are the logistic sigmoid and hyperbolic tangent functions; W_f, U_f and b_f represent the trainable parameters in two weight matrices and a bias vector of the forget gate; $W_{\tilde{c}}$, $U_{\tilde{c}}$ and $b_{\tilde{c}}$ are another set of trainable parameters to calculate the next hidden state in LSTM unit. x_t and h_{t-1} are the current input and last hidden state, respectively, to calculate a potential update vector $\tilde{c}_t$. The input gate determines which information of $\tilde{c}_t$ will be used to update the cell state in the current time step:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (14)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (15)$$

where i_t is a resulting vector in the input gate, determining which new information will store in $\tilde{c}_t$ (Kratzert et al., 2018); W_i, U_i and b_i are trainable parameters in input gate. The output vector of the input gate c_t is updated by equation (15). Specifically, $\odot$ represents element-wise multiplication. The last gate is the output gate controlling the information of c_t that flows into the new hidden state h_t, which can be calculated from:

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_0) \quad (16)$$

$$h_t = \tanh(c_t) \odot O_t \quad (17)$$

where O_t is a resulting vector; W_o, U_o and b_0 are the trainable parameters for the output gate. The new hidden state h_t is calculated by combining the results from equations (14-15), allowing for an effective learning from long-term dependencies in historical time series (Kratzert et al., 2018). By stacking multiple LSTM layers on the top of the neural networks, LSTM-based surrogates used a fully-connected layer or Bayesian fully connected layer to connect the results from the last hidden state in LSTM to a single output neuron to acquire the final prediction. The basic formula of these layers is given by the following equation:

$$y = W_d h_n + b_d \quad (18)$$

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

where y is the prediction variable, which is LST in this study; h_n is the output from the last LSTM layer. In the fully-connected layer, W_d and b_d are deterministic constants learned during training, while in the Bayesian fully-connected layer, W_d and b_d are instead modeled as random variables from Gaussian distribution to capture uncertainty of parameters.

The hyperparameters of these LSTM-based surrogates are both adjusted to be optimal. Specifically, the LSTM-based surrogate in Baseline consists of two layers with 256 LSTM units and 1 fully-connected layer, with a batch size of 32, a learning rate of 0.01, and an Adam optimizer. This surrogate was adjusted manually to achieve the optimal performance. The BLSTM-based surrogate was composed by 4 layers with 467 LSTM units and 1 Bayesian fully-connected layer, with a batch size of 64, a learning rate of 0.00096, and an optimizer of RMSprop. The BLSTM surrogate in HyLake v1.0 was adjusted using Bayesian Optimization (BO-BLSTM), which is a powerful tool for the joint optimization of design choices using less computational power to compute better solution (Shahriari et al., 2016), with a hyperparameter space of the number of layers and units, learning rate and optimizer to search for the optimal group of hyperparameters (Figure 3b). The BO-BLSTM-based surrogate in TaihuScene comprised 7 layers with 836 BLSTM units and 1 Bayesian fully-connected layer, with a batch size of 145, a learning rate of 0.2538, and an optimizer of AdamW using Bayesian Optimization to search for the optimize group of hypermeters. A Tree-structured Parzen Estimator (TPE) is adopted in BO, performing 20 to 100 iterations of surrogate training and updates. Train, validation, and test datasets for each lake site were divided by 80%, 10% and 10%, into 2013-01-01 00:00:00 to 2015-05-26 04:00:00, 2015-05-26 04:00:00 to 2015-09-12 14:00:00, and 2015-09-12 14:00:00 to 2015-12-30 23:00:00.

2.3 Numerical experiments designment and evaluation metric

2.3.1 Numerical experiments for model intercomparison

To address the generalization and transferability of HyLake v1.0 in studied (MLW) and ungauged lake sites (DPK、BFG、XLS and PTS) (Table 1), this study further developed three numerical experiments using distinct architecture and forcing datasets (Table 2), including FLake, Baseline, and TaihuScene on the basis of PBBM to intercompare. The briefly introduction purposes for these additional numerical experiments are described as follows:

- PBBM as a backbone of HyLake v1.0 is a simplified process-based lake model was constructed based on the energy balance equations and the 1-D vertical lake water temperature transport equations;
- FLake is a bulk model based on a two-layer parametric representation of the evolving temperature profile and on the integral budgets of heat and of kinetic energy for the layers, which is widely used as a lake module for simulating lake-atmosphere interactions in Earth System Models (ESMs) (Huang et al., 2021; Mironov et al., 2010). FLake served as a well-known traditional process-based lake model is suitable for model intercomparison;
- Baseline is an experiment of hybrid lake model that coupled a LSTM-based surrogate trained on outputs of PBBM, which is proposed to intercompare the performance and assess the improvement of HyLake v1.0 by using observation informed surrogate;

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

- TaihuScene is another numerical experiment for hybrid lake model that coupled a BO-BLSTM-based surrogate trained on all observations from Lake Taihu, which is different from the HyLake v1.0. The purpose of TaihuScene is to intercompare the performance by using larger train datasets against small dataset of HyLake v1.0.

The PBBM performed similar to FLake in MLW site, indicating a high reliability and accuracy (Figure S1). Except for PBBM, all of the experiments were initially intercompared in each lake site from Lake Taihu, while FLake and TaihuScene was additionally intercompared using the forcing datasets from ERA5 reanalysis datasets. The performance of these numerical experiments was almost validated by the observations from Lake Taihu eddy covariance network. The specification can be found in Table 2.

Table 2. The specification of numerical experiments for model intercomparison.

Experiment	Forcing datasets	Surrogate	Train datasets	Description
PBBM	\	\	\	Backbone for HyLake v1.0
FLake	ERA5; observations	\	\	A freshwater lake model for intercomparison
Baseline	MLW	LSTM	PBBM outputs	A baseline experiment using PBBM outputs for model intercomparison
TaihuScene	ERA5; observations	BO-BLSTM	observations	A numerical experiment using large train dataset to train surrogate
HyLake v1.0	ERA5; observations	BO-BLSTM	MLW observations	Proposed hybrid lake model in this study

2.3.2 Metrics for model validation and intercomparison

To evaluate the performance of these numerical experiments, a Pearson correlation coefficient (R), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) was used in this study to compare the accuracy of LST and heat fluxes between simulations and observations. Specifically, the R is proposed to measure the linear correlation of the observed and modeled values, RMSE and MAE calculate if the results from models overestimate or underestimate the observations with the same data units (Piccolroaz et al., 2024). The calculation of R, RMSE, and MAE can be expressed by:

$$R = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y-\bar{y})^2} \quad (18)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\bar{y}-y| \quad (19)$$

where x and $\bar{x}$ are the observations and its average; while y and $\bar{y}$ are the results of model and its average; n represents the length of time series.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

3 Results

3.1 Evaluation of LSTM-based surrogates in HyLake v1.0

The PBBM, a backbone of HyLake v1.0, has already been validated in Figure S1 by comparing with FLake model and observations from MLW site, indicating a robust prediction with a R of 0.98 and RMSE of 1.78 °C in LST, a R of 0.85 and RMSE of 38.34 W/m² in LE and a R of 0.89 and RMSE of 9.37 W/m² in HE. FLake demonstrated a slightly better performance to PBBM in LST and HE while performed poorer in LE (LST: R = 0.98, RMSE = 1.76 °C; LE: R = 0.82, RMSE = 42.73 W/m²; HE: R = 0.84, RMSE = 7.24 W/m²). These results fully indicated that the backbone provided from PBBM for HyLake v1.0 is reasonable for all variables.

After that, to evaluate the integration of deep-learning-based algorithms within the HyLake v1.0 backbones, this study comprehensively evaluated the performance of several LSTM-based surrogates in both discrete and integrated modeling experiments for Baseline, HyLake v1.0 and TaihuScene (Figure 4 and Figure S2). Specifically, this study separately assessed the accuracy of the BO-BLSTM-based surrogate in the train, validation, and test sets, aiming to evaluate their ability to describe the physical principles between climate change and LST (Figure 4a-c). The ΔLST (the difference between current and previous time step LST) obtained from observations was used to train surrogate in HyLake v1.0. For the discrete BO-BLSTM-based surrogate in HyLake v1.0, a higher consistency between predictions and observations was observed (Figure 4a-c). Specifically, the results for the BO-BLSTM-based surrogate showed RMSE values of 0.1945 °C and MAE of 0.1306 °C in the train set, RMSE of 0.3359 °C and MAE of 0.1924 °C in the validation set, and RMSE of 0.2271 °C and MAE of 0.1461 °C in the test set, for ΔLST, respectively.

Considering the intercomparison for numerical experiments, the LSTM-derived ΔLST in Baseline experiment, which is trained from process-based simulations of PBBM, also performed superior performance, indicating a great capacity for learning physical principles (Figure S2a-c). The results indicated a RMSE of 0.0580 °C and a MAE of 0.0112 °C in the train dataset, RMSE of 0.0079 °C and MAE of 0.0058 °C in the validation dataset, and RMSE of 0.0161 °C and MAE of 0.0094 °C in the dataset. These results suggest that the LSTM-based surrogate can capture approximately about 90% of the physical principles even during validation and testing. When applied to Lake Taihu, another BO-BLSTM-based surrogate in TaihuScene was used to train with observations from all lake sites. It demonstrated a close match to observations, with RMSE values of 0.2363 °C and MAE of 0.1537 °C in the training dataset, RMSE of 0.3342 °C and MAE of 0.1880 °C in the validation dataset, and RMSE of 0.2281 °C and MAE of 0.1480 °C in the test dataset. These results were relatively lower than HyLake v1.0 due to the larger dataset size in training for ΔLST. The BO-BLSTM-based surrogates provide a feasible model to approximate LST with reasonable and robust way. Therefore, these surrogates, improved on the basis of purely LSTM-based surrogates, ensure robust capacity for autoregressive prediction of LST changes while maintaining numerical stability, laying a solid foundation for algorithms coupled with HyLake backbone.

Following the separately validation of LSTM-based surrogates, this study evaluated the predicted LST, LE and HE in FLake and HyLake by integrating these surrogates, including Baseline and HyLake v1.0 (Figure 4d-f). Compared to FLake,

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Baseline which utilized the LSTM-based surrogate trained on outputs from PBBM performed slightly poorer, with an R of 0.96 and RMSE of 2.71 °C for LST, an R of 0.74 and RMSE of 51.77 W/m² for LE, and an R of 0.75 and RMSE of 14.63 W/m² for HE. The physical principles learned from these simulations is limited but enabled the surrogate to provide predictions similar to those of PBBM. For LST, HyLake v1.0 outperformed both FLake and Baseline, with an R of 0.99 and RMSE of 1.08 °C (Figure 4d). For heat fluxes calculated from the energy balance equations of surface flux solution, HyLake v1.0 also outperformed FLake and Baseline, with an R of 0.94 and RMSE of 24.65 W/m² for LE and an R of 0.93 and RMSE of 7.15 W/m² for HE (Figure 4e-f). These results demonstrate the HyLake v1.0 that using BO-BLSTM-based surrogate as a module of the HyLake backbone to solve LST and thereby compute LE and HE in the subsequent time step offers numerical stability and predictability. Furthermore, HyLake v1.0 proves capable of estimating lake-atmosphere interactions, surpassing FLake in the integration of deep-learning-based and process-based models, which offered a robust way for applying in ungagged locations. TaihuScene was used to intercompare with HyLake v1.0 in model generalization and transferability across all lake sites and different forcing datasets, which will be discussed in the section 3.3.

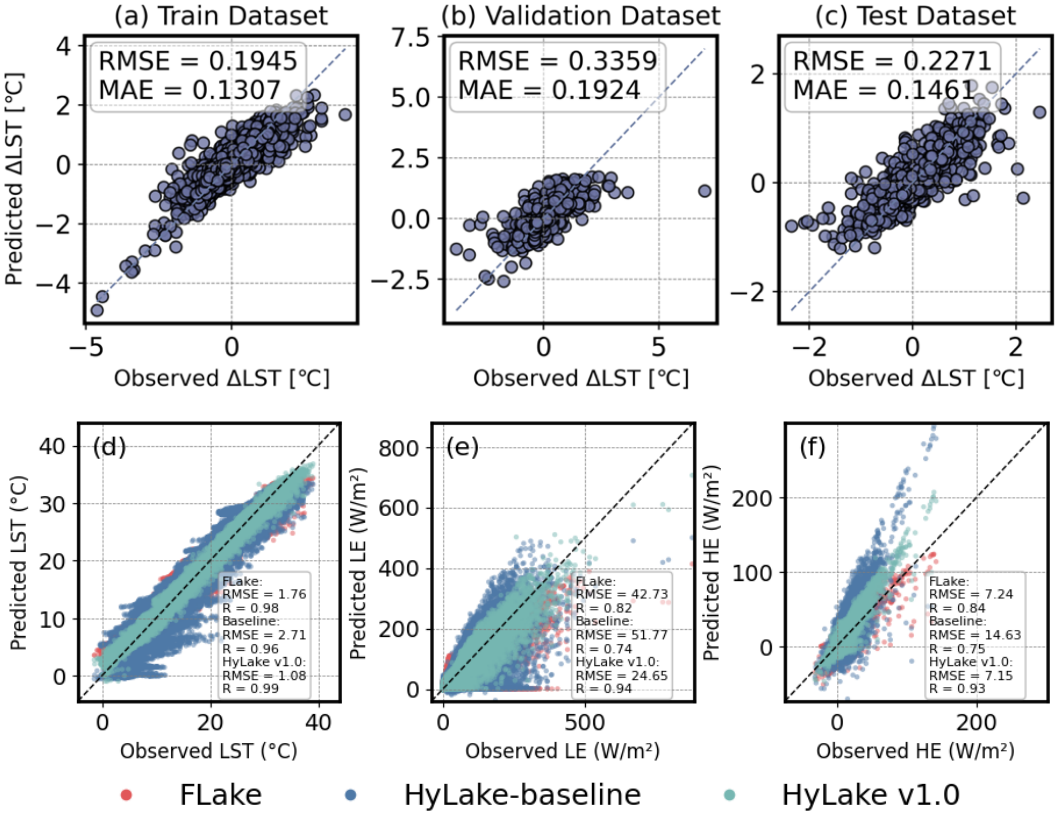


Figure 4: Model performance of HyLake v1.0. (a-c) present the comparison of separate prediction in ΔLST of BO-BLSTM-based surrogate in HyLake v1.0 for train, validation and test dataset, respectively. (d-f) Comparison of (g) LST, (h) LE and (i) HE with observations for FLake, Baseline and HyLake v1.0.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

3.2 Intercomparisons of LST, LE and HE from 2013 to 2015

This study conducted a comprehensive intercomparison of daily and hourly trends in LST, LE and HE from several experiments in the MLW site during the period from 2013 to 2015, including FLake, Baseline, and HyLake v1.0 (Figures 5–7).

As shown in Figure 5, the temporal changes in LST for the period of surrogates training, validation, and test datasets were compared. For daily changes in LST, HyLake v1.0 and FLake showed a closer match to observations, whereas Baseline, trained with process-based simulations, exhibited a larger error in comparison to the observed values (R = 0.96, RMSE = 2.71 °C, Figure 4d). HyLake v1.0 demonstrated a greater capability in capturing the daily changes in LST, particularly in mid-winter for each dataset, thus indicating long-term stability in LST modeling (R = 0.99, RMSE = 1.08 °C, Figure 4d). Specifically, FLake provided a good match to observations at a daily scale, but showed poorer performance in capturing diurnal variations of LST (R = 0.98, RMSE = 1.76 °C, Figure 4d). This study randomly selected two subperiods from the training, validation, and test periods (Figure 5b and c), where the diurnal variations of LST observations exhibited a significant bias compared to FLake, indicating that FLake is not able to accurately describe variations at the diurnal scale. Meanwhile, Baseline primarily captured the long-term trends of LST from PBBM but did not effectively represent diurnal variations due to the limitations of the datasets and physical principles provided by PBBM. In contrast, HyLake v1.0 was able to capture more information about the diurnal variations of LST from the observations, thereby outperforming both FLake and Baseline. Overall, HyLake v1.0, coupled with the BO-BLSTM-based surrogates trained on observations, offers a robust way for predicting LST trends at a finer temporal resolution.

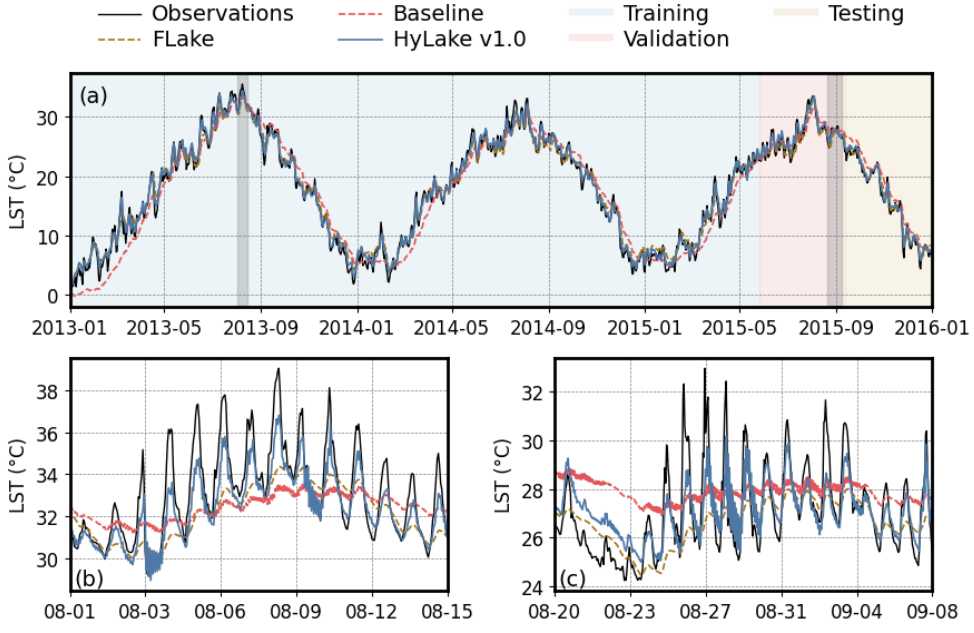
The LE and HE were calculated using the energy balance equations, where the LST, updated by LSTM-based surrogates in HyLake, served as an essential input. Consequently, it is necessary to validate the variations in these heat fluxes outputted by HyLake v1.0 to assess its capacity for modeling lake-atmosphere interactions. This study validated the observed LE and HE at the MLW site on both daily and hourly scales (Figures 6–7).

Regarding the changes in LE (Figure 6), HyLake v1.0, which used LST calculated from the BO-BLSTM-based surrogate, demonstrated a minor bias in estimating LE (R = 0.94, RMSE = 24.65 W/m², Figure 4e), outperforming both FLake (R = 0.82, RMSE = 42.73 W/m², Figure 4e) and Baseline (R = 0.74, RMSE = 51.77 W/m², Figure 4e). Notably, using an improved LSTM-based surrogate resulted in a slightly and significant improvement of LE compared to the FLake and Baseline. Specifically, Baseline showed more similar performance to FLake, capturing the major trends of these heat fluxes. However, HyLake v1.0 derived LE was closer to the observation in peak and valley values (Figure 6bc), indicating its overall superior capacity for describing the diurnal variations. Still, some biases persisted in the validation and test periods. For example, HyLake v1.0 exhibited such as between 2015-08-20 and 2015-08-23 where the period showed an overestimation to the observations, despite the improvements in LE estimation provided by HyLake v1.0.

For HE, which exhibited relatively insignificant diurnal and seasonal variations during the studied period, HyLake v1.0 (R = 0.93, RMSE = 7.15 W/m², Figure 4f) outperformed both FLake (R = 0.84, RMSE = 7.24 W/m², Figure 4f) and Baseline (R =

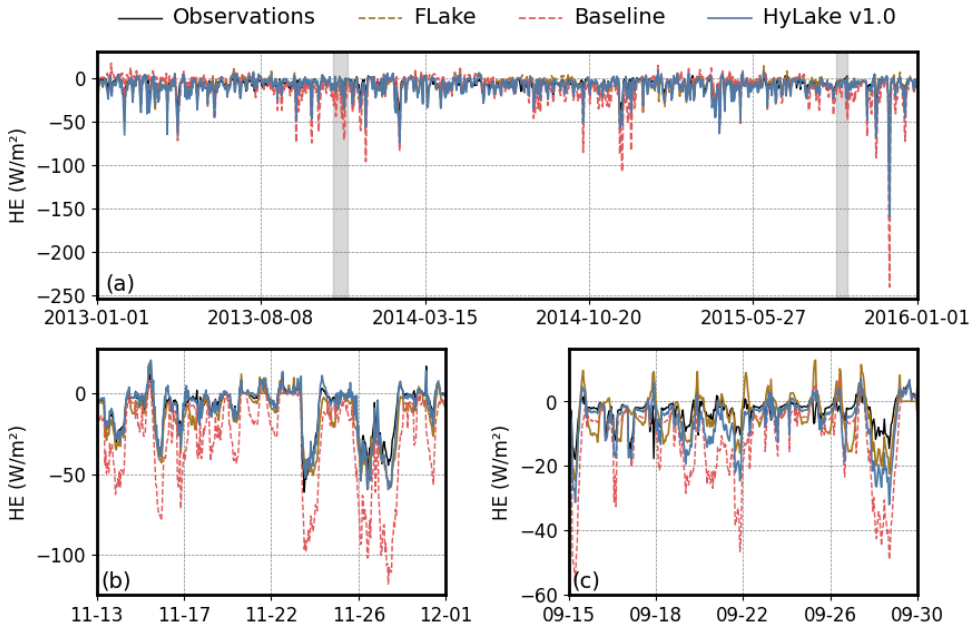
https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

380 0.75, RMSE = 14.63 W/m², Figure 4f) in simulating variations of both hourly and daily trends. The results were found that Hylake v1.0 is capable of correcting some of the partial biases in HE estimation by integrating the BO-BLSTM-based surrogate (Figure 7b and c), leading to more accurate simulations of these heat fluxes. Besides that, the HE calculated from both FLake and Baseline was difficult to accurately estimate the minor variations in hourly scale during simulations, which could accumulate bias in subsequent time steps. This issue was especially evident in the validation and test datasets (Figure
385 7c), whereas HyLake v1.0 showed minimal bias due to its improved representation of LST. In summary, HyLake v1.0 that coupled with a BO-BLSTM-based surrogate provided a more robust and reasonable prediction of LST, leading to better corrections for untrained variables (LE and HE) produced by the other modules. This improvement ensures HyLake v1.0's capability in accurately describing lake-atmosphere interactions with improved performance.



390 **Figure 5: Comparison of HyLake models in temporal trends of LST. Comparison of (a) the full time series and (b-c) partial time series of models derived LST and observations from 2013 to 2015. Blue, red, and yellow regions represent the period for the train, validation, and test datasets, respectively. Black solid, brown dashed, red dashed, and blue solid lines represent LST from observations, FLake, Baseline, and HyLake v1.0, respectively.**

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025



395 **Figure 6: Comparisons of FLake, Baseline, and HyLake v1.0 for LE. Comparison of (a) full and (b-c) partial time series of model derived LE and observations from 2013 to 2015.**

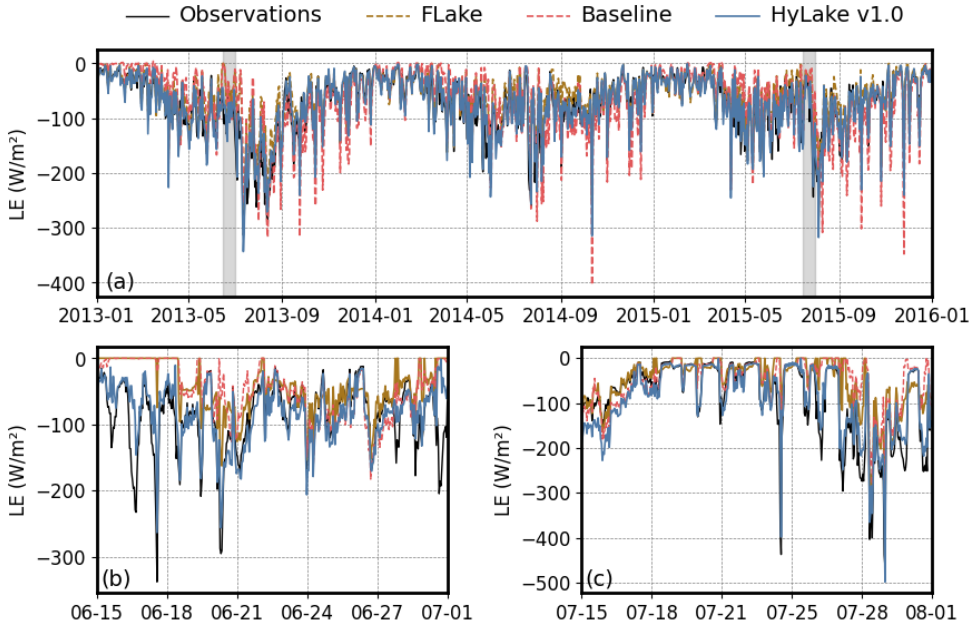


Figure 7: Comparisons of FLake, Baseline, and HyLake v1.0 for HE. Comparison of (a) full and (b-c) partial time series of HyLake derived HE and observations from 2013 to 2015.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

3.3 Validation across observational sites in Lake Taihu

The transferability and generalization of traditional deep-learning-based models remain challenging (T. F. Xu & Liang, 2021). However, integrating process-based models with deep learning-based models can mitigate these issues to some degree. To address these challenges with HyLake v1.0, this study specially developed a TaihuScene (Table 2), as a variant which enlarges the size of train datasets by incorporating data from all lake sites in Lake Taihu to train BO-BLSTM-based surrogates and evaluate the potential difference. The primary objectives of TaihuScene are to (1) offer a theoretically optimal coupled model (previous studies allocated that larger train datasets improve deep-learning-based models' performance (Halevy et al., 2009)) for simulating lake-atmosphere interactions in Lake Taihu, and (2) compare with HyLake v1.0 in generalization and transferability for ungauged regions.

Given to the experiment TaihuScene trained with large datasets, HyLake v1.0 still performed the best in predicting LST (MAE = 1.03 °C), HE (MAE = 24.79 W/m²), and LE (MAE = 7.88 W/m²) among FLake and TaihuScene (Figure 8). While FLake performed the worst in each variable, with a MAE of 1.69 °C in LST, a MAE of 41.95 W/m² in LE and a MAE of 10.44 W/m² in HE, respectively. TaihuScene performed moderately, with a MAE of 1.23 °C in LST, a MAE of 29.53 W/m² in HE and a MAE of 10.63 W/m² in LE, respectively.

Involving the relative bias (the difference between simulation and observation), the median biases (the dashed lines) and distribution of outputs by TaihuScene indicated an overestimation of LST (Figure 8a), which may contribute to the underestimation of heat fluxes derived from the common physical principles learned from large datasets during step-by-step iterative calculations (Figure 8b and c). While HyLake v1.0 exhibited an opposite estimation, with a slightly underestimation in LST and overestimation in HE and LE. This suggests that BO-BLSTM-based surrogates trained with observations from MLW site provide more reliable results than those trained with data from all sites due to the more clearly physical principles for training. But it is worthy that TaihuScene still far outperformed FLake, as evidenced by a denser distribution of biases. These results challenge the assumption that larger datasets always improve the performance of deep-learning-based models (T. F. Xu & Liang, 2021; W. Zhong et al., 2020), with the results suggested that HyLake v1.0, trained on relatively smaller datasets, performs better than TaihuScene in Lake Taihu.

The results of intercomparison in each lake site for FLake and TaihuScene experiments further explain the reasons for this phenomenon (Figure S3). HyLake v1.0 performed best at the MLW, PTS, and XLS sites but showed poorer results at the BFG and DPK sites. Specifically, HyLake v1.0 outperformed FLake and TaihuScene at the MLW site, with a MAE of 0.85 °C, 15.18 W/m², and 4.73 W/m² for LST, LE, and HE, respectively. In contrast, TaihuScene performed relatively worse, with a MAE of 1.38°C, 15.54 W/m², and 8.33 W/m². FLake showed moderate performance, with MAE values of 1.35 °C, 24.76 W/m², and 5.01 W/m² for LST, LE, and HE, respectively (Figure S3a-c). A similar pattern was appeared at the PTS and XLS sites. At PTS, HyLake v1.0 also showed the best performance (LST: MAE = 0.79 °C; LE: MAE = 20.12 W/m²; HE: MAE = 6.90 W/m²), while FLake performed moderately (LST: MAE = 1.22 °C; LE: MAE = 24.77 W/m²; HE: MAE = 6.93 W/m²), and TaihuScene performed the worst (LST: MAE = 1.47° C; LE: MAE = 39.66 W/m²; HE: MAE = 15.11 W/m²)

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

(Figure S3j-i). Similarly, at the XLS site, HyLake v1.0 performed the best (LST: MAE = 1.86°C; LE: MAE = 20.40 W/m²; HE: MAE = 6.61 W/m²), while FLake performed moderately (LST: MAE = 1.33 °C; LE: MAE = 30.00 W/m²; HE: MAE = 7.69 W/m²), and TaihuScene performed the worst (LST: MAE = 1.29°C; LE: MAE = 32.20 W/m²; HE: MAE = 12.19 W/m²) (Figure S3m-o). However, at the BFG and DPK sites, TaihuScene outperformed the other models in estimating LST, LE, and HE, with MAE values of 1.06°C, 27.92 W/m², and 9.73 W/m² at BFG and 1.00°C, 26.05 W/m², and 8.43 W/m² at DPK (Figure S3d-i). Specifically, TaihuScene performed slightly better than HyLake v1.0 (BFG: LST: MAE = 1.32°C, LE: MAE = 32.88 W/m², HE: MAE = 10.47 W/m²; DPK: LST: MAE = 1.29°C, LE: MAE = 34.71 W/m², HE: MAE = 10.54 W/m²) but was far superior to FLake (BFG: LST: MAE = 2.32 °C, LE: MAE = 65.05 W/m², HE: MAE = 16.53 W/m²; DPK: LST: MAE = 2.16 °C, LE: MAE = 62.71 W/m², HE: MAE = 15.56 W/m²).

It is clear that HyLake v1.0 demonstrated outstanding transferability and generalization across ungauged regions, surpassing traditional lake-atmosphere interaction models such as FLake in prediction accuracy for each variable. TaihuScene, though capable of predicting changes across all sites in Lake Taihu, also exhibited a superior overall performance at specific sites when compared to HyLake v1.0. This highlights HyLake v1.0 offers promising potential for extending its application to these ungauged lakes or sites with similar characteristics by effectively learning physical principles.

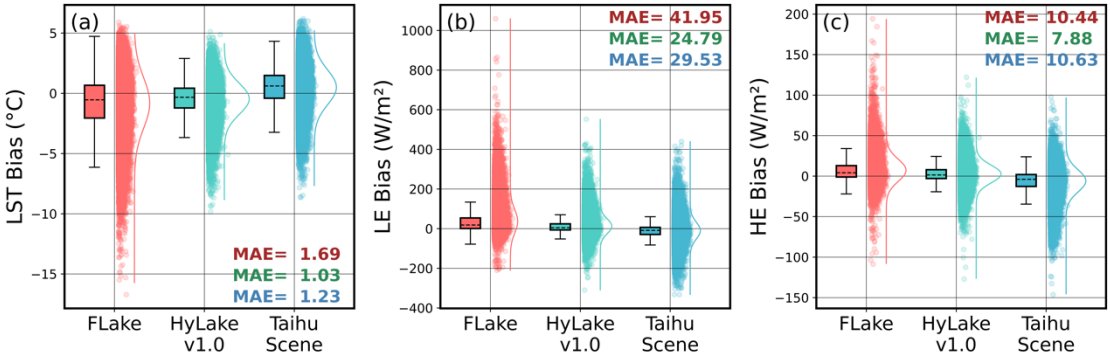


Figure 8: Comparisons of overall datasets for FLake, HyLake v1.0 and TaihuScene for each variable in all lake sites of Lake Taihu. Comparison of (a) LST, (b) LE, and (c) HE for FLake, HyLake v1.0, and TaihuScene in all lake sites. These figures compared the performance of datasets from BFG, DPK, PTS, MLW, and XLS sites. Dashed lines represent median biases of FLake, HyLake v1.0, and TaihuScene, respectively. The scatterplots and probability distribution curves illustrate the data distribution of LST, LE and HE. The Numbers at the top or bottom right of subfigures with same color to boxes indicate the MAE of outputs for FLake, HyLake v1.0, and TaihuScene experiments, respectively.

3.4 Comparison in Lake Taihu under ERA5 reanalysis datasets

The aforementioned numerical experiments indicated that HyLake v1.0 demonstrates a superior capability of transferability and generalization in predicting LST, LE, and HE across the ungauged regions in Lake Taihu. However, the limitations associated with the availability of adequate and accurate forcing datasets from observations impede the broader application of HyLake v1.0. To address these challenges, this study also validated the performance of FLake, HyLake v1.0, and TaihuScene using ERA5 reanalysis datasets, which are widely used as forcing datasets for process-based models (Albergel et al., 2018; Hersbach et al., 2020). This intercomparison aimed to evaluate the generalization and transferability of HyLake

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

v1.0 across different forcing datasets in ungauged locations. The spatial resolution of the ERA5 dataset covers merely 5 grid cells that encompass the studied lake sites, among the 11 grids for the entire Lake Taihu. These grids include portions of the land surface surrounding the lake, inevitably introducing uncertainty due to the scale mismatch between climatic forcing datasets and the lake model (Hersbach et al., 2020).

Despite these limitations, it was surprising to find that the evaluation showed that HyLake v1.0 exhibited performance similar to or even superior to that of FLake for each lake site, with consistent spatial patterns for LST, LE, and HE (Figure 9 and Figure S4). From the statistical properties, HyLake v1.0 still exhibited an incomparable performance in overall datasets, with MAE values of 0.90 °C, 35.02 W/m², and 7.94 W/m² for LST, LE and HE, respectively, which following with TaihuScene performed with MAE values of 1.29 °C, 47.78 W/m², and 10.11 W/m² for LST, LE and HE, far outperforming FLake with MAE values of 1.73 °C, 53.21 W/m², and 11.84 W/m² for LST, LE and HE, respectively (Figure 9a-c). From spatial patterns observed, LST in the middle of the lake was relatively higher for all experiments (Figure 9d, and g), while LE was higher in the southern and western shore (Figure 9e, and h), and HE showed higher values in the northwestern shores of Lake Taihu (Figure 9f, and i). Both HyLake v1.0 and TaihuScene revealed similar patterns of average across all variables, except for a slight overestimation of LE and underestimation of HE in the southeastern and northeastern shores of Lake Taihu in HyLake v1.0, respectively (Figure 9d-f). However, TaihuScene predicted relatively higher values in LE and HE, and lower in LST than the other models although it still followed similar spatial patterns for LST, LE, and HE (Figure 9g-i). This indicates that HyLake v1.0, coupled with a small dataset-trained BO-BLSTM-based surrogate, can still provide robust and reasonable predictions for estimating spatial patterns of Lake Taihu.

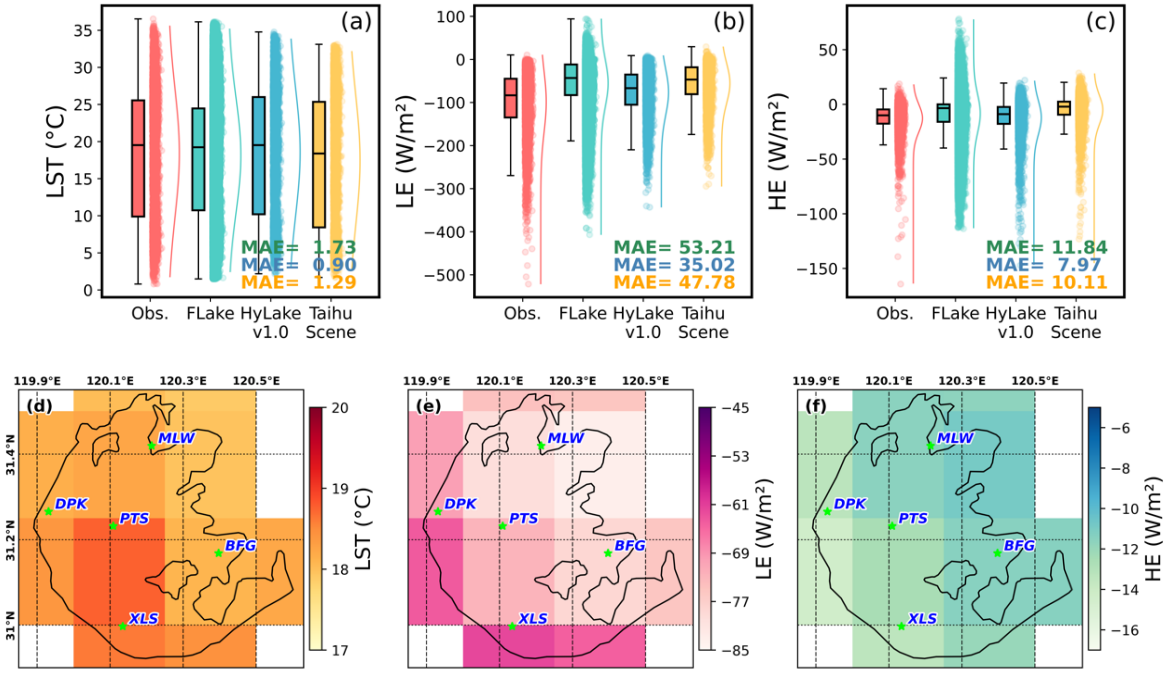
HyLake v1.0 demonstrated excellent generalization and transferability, as evidenced by its superior performance for each lake site (Figure S4). At the MLW site (Figure S4a-c), which located on the northern shore of Lake Taihu, HyLake v1.0 outperformed both FLake (LST: MAE = 1.68 °C; LE: MAE = 33.84 W/m²; HE: MAE = 9.68 W/m²) and TaihuScene (LST: MAE = 1.17 °C; LE: MAE = 22.53 W/m²; HE: MAE = 8.98 W/m²), with MAE values of 1.05 °C, 31.46 W/m², and 9.12 W/m² for LST, LE, and HE, respectively. The train datasets used for the BO-BLSTM-based surrogate in HyLake v1.0 contributed to its powerful performance at this site, while predictions from TaihuScene performed farther from the observations.

As for the ungauged sites, HyLake v1.0 still performed considerable generalization and transferability by learning physical principles from datasets in MLW site (Figure S4m-o). It performed excellently across most variables at these lake sites. TaihuScene showed robust predictions but outperformed HyLake v1.0 only at the XLS and MLW sites. For the BFG site (Figure 9d-f), HyLake v1.0 outperformed both FLake and TaihuScene, with MAE values of 0.94 °C, 42.30 W/m², and 9.94 W/m², respectively. TaihuScene performed the worst among these models, with MAE values of 1.85 °C, 60.32 W/m², and 15.73 W/m². FLake exhibited a moderately performance with MAE values of 1.15 °C, 49.52 W/m², and 10.77 W/m² for LST, LE and HE, respectively. At the DPK site (Figure 9g-i), HyLake v1.0 performed better than FLake for LST but performed slightly worse for LE and HE, with MAE values of 0.68°C, 52.82 W/m², and 8.43 W/m². TaihuScene performed the worst in this site (LST: MAE = 1.49 °C; LE: MAE = 69.67 W/m²; HE: MAE = 12.40 W/m²). FLake performed the moderate in this

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

site (LST: MAE = 1.14 °C; LE: MAE = 56.12 W/m²; HE: MAE = 9.11 W/m²). At the PTS site (Figure 9j-l), HyLake v1.0 (LST: MAE = 0.75°C; LE: MAE = 22.28 W/m²; HE: MAE = 6.43 W/m²) outperformed FLake for LST, LE, and HE (LST: MAE = 1.89 °C; LE: MAE = 57.12 W/m²; HE: MAE = 9.84 W/m²) and TaihuScene (LST: MAE = 0.94°C; LE: MAE = 31.75 W/m²; HE: MAE = 6.74 W/m²). HyLake v1.0 showed a slightly better performance at the XLS site (Figure 9m-o), with MAE values of 1.05 °C, 24.29W/m², and 5.69 W/m², while TaihuScene performed slightly worse for LE (MAE = 37.94W/m²) and HE (MAE = 6.79W/m²), FLake performed the worst with MAE values of 1.49 °C, 58.78 W/m², and 9.01 W/m².

Overall, both HyLake v1.0 and TaihuScene showed reliable transferability and generalization across lake sites in Lake Taihu. Specifically, HyLake v1.0 performed the best in 14 of 15 variables in Lake Taihu among these 3 experiments; TaihuScene outperformed HyLake v1.0 in 1 of 15 variables and outperformed FLake in 8 of 15 variables in Lake Taihu. HyLake v1.0 providing superior results in most cases, proving its potential for extensive application in ungauged lakes under different forcing datasets. However, the prediction accuracy of these models based on ERA5 reanalysis datasets were almost reduced due to the potential uncertainty in lake-atmosphere modeling systems. The current coupling strategies of HyLake v1.0 ensured the numerical stability and superior performance in validation due to the robust capability of auto-regressive predictions by proposed LSTM-based surrogates and to avoid the numerical divergence or error accumulation in step-by-step iteration loops. The results evidenced that HyLake v1.0 coupled with BO-BLSTM-based surrogate is suitable for discovering the potential physical principles for lake-atmosphere interactions systems, indicating that the integration of deep learning-based surrogates as a module with process-based models is effective for improving predictions in ungauged lakes.



https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

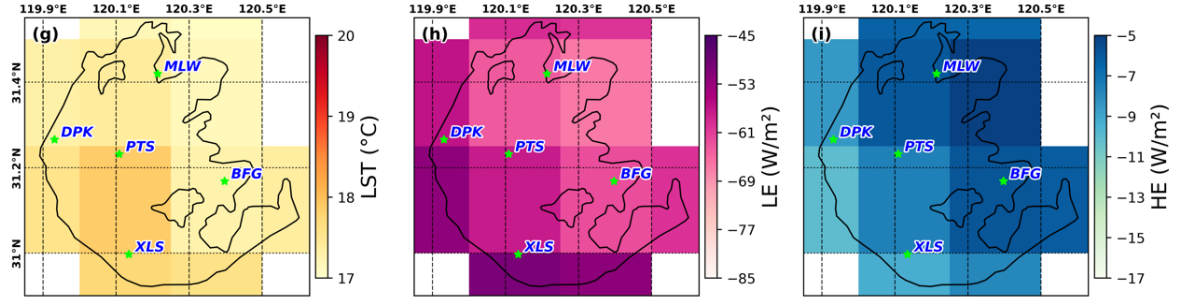


Figure 9: The statistical characteristics and spatial average of LST, LE and HE for HyLake v1.0 and TaihuScene drove by ERA5 reanalysis datasets. (a-c) represent the accuracy of LST, LE and HE for each model, respectively; (d-f) represent LST, LE and HE for HyLake v1.0, respectively; (g-i) represent LST, LE and HE for TaihuScene, respectively. The green stars noted in all figures are lake sites in Lake Taihu. The MAEs in legend of (a) from top to bottom mean the MAE for Obs., FLake, HyLake v1.0, and TaihuScene experiments, respectively.

4. Discussion

4.1 Limitation of HyLake v1.0

In this study, we firstly utilized the in-house programed python-based 1-D process-based lake model PBBM, and then replaced traditional LST approximation methods with a LSTM-based surrogate trained on observations from the Lake Taihu eddy flux network to construct a novel lake model named HyLake v1.0, which is a hard-coupled and auto-regressive hybrid model. To address issues related to model performance, generalization, and transferability in ungauged locations, three additional numerical experiments, including FLake, Baseline, and TaihuScene, were proposed to intercompare. These experiments were compared across different forcing datasets and validated for both spatial and temporal patterns at Lake Taihu (Table 2). The results demonstrated that HyLake v1.0 indicated outstanding performance in validation and robust predictions in ungauged lake sites under ERA5 reanalysis datasets, highlighting its superior capacity for lake-atmosphere interaction modeling systems. However, several limitations exist, primarily involving data scarcity, model architecture, explainability, and coupling strategies.

One challenge is the scarcity of observed radiation and energy fluxes, as it is expensive to collect long-term, multi-site microclimate and water temperature datasets from a large number of lakes for model forcing (Erkkilä et al., 2018; Nordbo et al., 2011). Although several reanalysis datasets (e.g., ERA5, NLDAS-2, GLSEA) have been used to drive lake models (Kayastha et al., 2023; Monteiro et al., 2022; Notaro et al., 2022; J. L. Wang et al., 2022), the lack of extensive observational data impedes the understanding of the relationships within lake-atmosphere interactions on a ground-truth scale. Due to the sparse observations, attempting to accurate predict the changes in lake-atmosphere modeling systems or parameterize the process-based models based on deep learning models and reanalysis datasets has to be popular (Almeida et al., 2022; Guo et al., 2021; Read et al., 2019). As a result, the relationships between reanalysis datasets and observation that learned by deep-learning-based surrogates may introduce additional biases, potentially influencing the learned relationships between climate change and lake surface conditions in models. Despite this, the results of this study still demonstrate that the relationships

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

revealed by observations using deep-learning-based surrogates can be robust and generalizable to ungauged regions with reasonable training strategies. Incorporating deep-learning-based surrogates with process-based models based on adequate and high-quality observations holds promise for developing a general HyLake model on a global scale, improving the understanding of lake-atmosphere interactions.

Moreover, simplified parameterizations in traditional process-based lake models are common (Golub et al., 2022; Mooij et al., 2010), which can influence the coupling strategies in HyLake. For example, the calculation of friction velocity (u^*) and surface roughness length (z_{0m}) in surface flux solutions has improved over time from constant empirical models to iterative routines (Hostetler et al., 1993; R. Iestyn Woolway et al., 2015), but substantial discrepancies still exist between simulation results and observations (Figure S5), which in turn influence the physical principles between land surface conditions and LST. These potential differences in physical processes lead to uncertainties in training deep-learning-based surrogates, contributing to the loss of knowledge during model training and thereby introducing large uncertainties in hybrid models. Furthermore, the long-term trends and diurnal variations in lake water temperature profiles remain challenging to accurate approximate using the finite difference method (e.g., Crank-Nicholson solution, implicit Euler scheme) (Piccolroaz et al., 2024; Sarovic et al., 2022; Subin et al., 2012). Several hybrid models that integrate deep-learning-based and process-based models have been constructed in previous studies, achieving improved performance in model comparisons (Ladwig et al., 2024; Read et al., 2019). However, these models and their training strategies generally perform better on training and test datasets, while their generalization and transferability need further validation. Lake Taihu, as a shallow, eutrophic, and large lake with almost complete mixing throughout the year and subject to complex chemical and biological influences in its aquatic ecosystem, requires a suitable model as part of the temperature-solving module in the water column to predict lake water temperature and estimate other implications under thermodynamic changes.

LSTM networks have been regarded as community benchmarks for daily hydrograph metrics (J. Liu et al., 2024). HyLake v1.0 integrates a BLSTM network with Bayesian Optimization, which enhances robustness and stability during long-term auto-regressive predictions compared to traditional LSTM networks. However, LSTM-based surrogates are more suited for short-term predictions compared to Transformers, which are suit for long-term predictions and commonly used in global weather forecasting systems (K. F. Bi et al., 2023; L. Chen et al., 2023). This limitation restricts the performance and robustness of LSTM-based surrogates for long-term predictions. Future improvements should incorporate more powerful deep-learning-based surrogates with enhanced long-term prediction, generalization, and transferability capabilities. Recent advances in state-of-the-art deep-learning models, such as Transformers, Graph Neural Networks, Temporal Convolutional Networks, TimesNet, and Kolmogorov-Arnold Networks, have shown exceptional performance in time series forecasting (Bai et al., 2018; Z. Liu et al., 2024; Wen et al., 2022; Wu et al., 2022), and are increasingly being applied in hydrological modeling (Sun et al., 2021; Z. C. Wang et al., 2024). To address the robustness of auto-regressive predictions, these novel deep-learning models need to be further validated and improved with additional modifications (Koya & Roy, 2024; J. Liu et al., 2024). However, developing powerful surrogates such as the Fuxi and Pangu models, which involve large parameters,

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

requires not only large size of observations but also substantial computational power and efficiency (K. Bi et al., 2023; Lei Chen et al., 2023).

While HyLake v1.0 performs better in terms of model explainability than purely data-driven models due to its hard-coupling structure, it still lags behind process-based models. LSTM-based surrogates, while offering powerful performance, introduce a "black-box" aspect that reduces their explainability (Chakraborty et al., 2021; de la Fuente et al., 2024). Developing deep-learning-based surrogates that incorporate knowledge from physical principles is an ongoing effort and has gained more attention recently (Piccolroaz et al., 2024; Read et al., 2019; Willard et al., 2023). For example, Read et al. (2019) proposed a process-guided deep-learning model with a loss function based on the simplified energy budget model from GLM and tested it on 68 lakes. Ladwig et al. (2024) developed a modular compositional learning framework for a 1-D Hydrodynamic Lake model, integrating four deep-learning models for each term and an eddy diffusion approach to improve performance in lake water temperature predictions at Lake Mendota. However, the simplified parameterizations, unclear relationships within the modules, and high computational demands hinder further development and limit the generalization and transferability of these models.

4.2 Future improvements

Future improvements to HyLake v1.0 should focus on investigating the scaling laws of datasets, development of surrogate architectures, and extension of coupled modules. Currently, HyLake v1.0 has been validated primarily in Lake Taihu, utilizing high-quality training data provided by the Lake Taihu eddy flux network (Zhang et al., 2020). Future development of HyLake v1.0 should focus on collect more observations, including heat fluxes and water temperature, and searching for more variables in datasets to train LSTM-based surrogates and acquire more general models at a larger scale. However, it is important to note that performance of HyLake may not always improve with an increase in the training data size. As HyLake is extended to larger scales or more lakes, the computational architecture will need to accommodate large training datasets, which may limit performance for specific lakes. Specifically, the scaling laws for deep-learning models indicate that model performance does not continue to increase indefinitely with the stacking of neurons and layers (Hestness et al., 2017). Adapting more powerful deep-learning-based surrogates will further improve HyLake v1.0 performance, leading to a better representation of lake-atmosphere interactions in ungauged lakes.

Moreover, the proposed hard-coupling strategies in HyLake v1.0 provide possibilities for coupling additional modules, such as lake water temperature, methane and carbon dioxide dynamics, oxygen dynamics, and lake-watershed interactions. Previous 1-D process-based lake models have often been developed as a backbone for coupling with these various modules (Saloranta & Andersen, 2007; Stepanenko et al., 2016). For example, Lake 2.0 was proposed by Stepanenko et al. (2016) first coupled methane and carbon dioxide modules to investigate oxygen, methane, and carbon dioxide dynamics in Kuivajärvi Lake, and MyLake proposed by Saloranta & Andersen (2007) gradually coupled modules such as DOC degradation, DO dynamics, microbial respiration, several gas exchange models, sediment-water interactions, dynamic light attenuation, nitrogen uptake and floating solar panels (Couture et al., 2015; Exley et al., 2022; Holmberg et al., 2014; Kiuru

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

et al., 2019; Markelov et al., 2019; Pilla & Couture, 2021; Salk et al., 2022). These hydro-biogeochemical processes have typically been described by PDEs or empirical models, which remain simplified and introduce uncertainty to some degrees (Li et al., 2021). The flexible structure of HyLake v1.0 allows for better capability to couple with different deep-learning-based surrogates for approximating lake water temperature changes, gas exchange, and DO dynamics, likely achieving better performance than purely process-based or data-driven models.

Overall, HyLake v1.0 holds significant potential for improvement, as current process-based lake models are overly simplistic and fail to capture the full range of hydro-biogeochemical processes accurately. Future efforts should mainly focus on improving the representation of heat and gas transport in the water column to better simulate aquatic ecosystems.

5. Conclusion

This study introduced a novel hybrid lake model, HyLake v1.0, by hard-coupling a Bayesian LSTM-based surrogate with Bayesian Optimization (BO-BLSTM-based surrogate) using observations from the MLW lake site in Lake Taihu. It replaced the implicit Euler scheme typically used in traditional process-based lake models with this BO-LSTM-based surrogate, enabling the prediction of LST, LE, and HE at the lake-atmosphere interface through energy balance equations. The HyLake v1.0 was proposed to offer more accurate prediction and flexibility in training. Specifically, several numerical experiments were conducted to validate and evaluate the performance of HyLake v1.0, including FLake, Baseline, and TaihuScene, which adopt different surrogates and training strategies. Additionally, this study used forcing datasets in 5 lake sites and ERA5 reanalysis datasets from Lake Taihu to examine the effects of different datasets on the generalization and transferability of HyLake v1.0 in ungauged regions. The experiments demonstrated that HyLake v1.0 effectively learns the physical principles governing lake-atmosphere interactions, highlighting its potential for application in ungauged lakes. Major conclusions are summarized as follows:

(1) The BO-BLSTM-based surrogate in HyLake v1.0 performed well in representing changes in ΔLST of test dataset (RMSE = 0.2587 °C, MAE = 0.1594 °C), outperforming TaihuScene while underperforming Baseline;

(2) HyLake v1.0 showed superior performance in LST (R = 0.99, RMSE = 1.08), LE (R = 0.94, RMSE = 24.65), and HE (R = 0.93, RMSE = 7.15) at both daily and hourly scales compared to observations, FLake and Baseline model, indicating that integrating physical principles with LSTM-based surrogates improves model accuracy and better captures changes in lake-atmosphere interactions;

(3) HyLake v1.0 performed well in capability of generalization and transferability in ungauged regions of Lake Taihu and low-resolution ERA5 forcing datasets. The results of intercomparison across lake site showed HyLake v1.0 presented the best capability in representation of LST (MAE = 1.03), LE (MAE = 24.79) and HE (MAE = 7.88) than FLake and TaihuScene. Specifically, it performed the best in MLW, PTS, and XLS, but slightly poorer in BFG and DPK sites than TaihuScene. Regarding the capability of spatial generalization using ERA5 forcing datasets, results indicated HyLake v1.0

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

performed the most closely matched the observations in Lake Taihu compared to FLake and TaihuScene in 14 of 15 variables.

These intercomparison experiments highlighted that HyLake v1.0, when coupled with a BO-BLSTM-based surrogate, offers excellent flexibility and is capable of capturing the underlying physical principles, providing more accurate predictions than traditional process-based models in ungauged regions. This also demonstrates the model's promising potential for application in ungauged lakes. Future work should focus on expanding HyLake v1.0 by exploring different architectures, utilizing larger training datasets, and incorporating additional coupled modules.

Code and data availability. The datasets and codes of HyLake v1.0 and other models used in this study are available at https://doi.org/10.5281/zenodo.15289113 (He et al., 2025). The ERA5 reanalysis datasets can be downloaded from the Climate Data Store (https://cds.climate.copernicus.eu/). Observations of lake surface water temperature, latent and sensible heat fluxes at Lake Taihu are available at Harvard Dataverse (https://doi.org/10.7910/DVN/HEWCWM; Zhang et al., 2020).

Author contributions. YH designed the HyLake model, wrote the code, organized the experiments, ran the simulations, performed the results, and wrote the original manuscript. XY contributed to the design and writing of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study was performed under the project "Coupled Human and Natural Systems in Land-Ocean Interaction Zone" founded by Guangdong Provincial Observation and Research Station, China and "Key technologies and demonstration for synergistic governance of aquatic environment and ecology in urban lakes of typical cities in the Yangtze River basin" founded by National Key R&D Program of China.

Financial support. This work was financially supported by the Guangdong Provincial Observation and Research Station (Grant No. 2024B1212040003) and National Key R&D Program of China (Grant No.2023YFC3208905).

References

Albergel, C., Dutra, E., Munier, S., Calvet, J. C., Muñoz-Sabater, J., de Rosnay, P., and Balsamo, G.: ERA-5 and ERA-Interim driven ISBA land-surface model simulations: which one performs better?, *Hydrol. Earth Syst. Sci.*, 22, 3515–3532, https://doi.org/10.5194/hess-22-3515-2018, 2018.

Almeida, M. C., Shevchuk, Y., Kirillin, G., Soares, P. M. M., Cardoso, R. M., Matos, J. P., et al.: Modelling reservoir surface temperatures for regional and global climate models: a multi-model study on the inflow and level-variation effects, *Geosci. Model Dev.*, 15, 173–197, https://doi.org/10.5194/gmd-15-173-2022, 2022.

Bai, S., Kolter, J. Z., and Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modelling, *arXiv* [preprint], arXiv:1803.01271, https://doi.org/10.48550/arXiv.1803.01271, 2018.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3-D neural networks, *Nature*, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Carpenter, S. R., Stanley, E. H., and Vander Zanden, M. J.: State of the world's freshwater ecosystems: physical, chemical and biological changes, *Annu. Rev. Environ. Resour.*, 36, 75–99, https://doi.org/10.1146/annurev-environ-021810-094524, 2011.

Chakraborty, D., Basagaoglu, H., and Winterle, J.: Interpretable vs. non-interpretable machine-learning models for data-driven hydro-climatological process modelling, *Expert Syst. Appl.*, 170, 114498, https://doi.org/10.1016/j.eswa.2020.114498, 2021.

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H.: FuXi: a cascade machine-learning forecasting system for 15-day global weather forecast, *npj Clim. Atmos. Sci.*, 6, 190, https://doi.org/10.1038/s41612-023-00512-1, 2023.

Culpepper, J., Jakobsson, E., Weyhenmeyer, G. A., Hampton, S. E., Obertegger, U., Shchapov, K., et al.: Lake-ice quality in a warming world, *Nat. Rev. Earth Environ.*, 5, 671–685, https://doi.org/10.1038/s43017-024-00590-6, 2024.

Couture, R. M., de Wit, H. A., Tominaga, K., Kiuru, P., and Markelov, I.: Oxygen dynamics in a boreal lake respond to long-term changes in climate, ice phenology and DOC inputs, *J. Geophys. Res.-Biogeosci.*, 120, 2441–2456, https://doi.org/10.1002/2015JG003065, 2015.

de la Fuente, L. A., Ehsani, M. R., Gupta, H. V., and Condon, L. E.: Toward interpretable LSTM-based modelling of hydrological systems, *Hydrol. Earth Syst. Sci.*, 28, 945–971, https://doi.org/10.5194/hess-28-945-2024, 2024.

Deng, B., Liu, S., Xiao, W., Wang, W., Jin, J., and Lee, X.: Evaluation of the CLM4 lake model at a large and shallow freshwater lake, *J. Hydrometeorol.*, 14, 636–649, https://doi.org/10.1175/JHM-D-12-067.1, 2013.

Erkkilä, K. M., Ojala, A., Bastviken, D., Biermann, T., Heiskanen, J. J., Lindroth, A., et al.: Methane and carbon-dioxide fluxes over a lake: comparison between eddy-covariance, floating-chambers and boundary-layer method, *Biogeosciences*, 15, 429–445, https://doi.org/10.5194/bg-15-429-2018, 2018.

Exley, G., Page, T., Thackeray, S. J., Folkard, A. M., Couture, R. M., Hernandez, R. R., et al.: Floating solar panels on reservoirs impact phytoplankton populations: a modelling experiment, *J. Environ. Manage.*, 324, 116410, https://doi.org/10.1016/j.jenvman.2022.116410, 2022.

Feng, D. P., Liu, J. T., Lawson, K., and Shen, C. P.: Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic-prediction accuracy, *Water Resour. Res.*, 58, e2022WR032404, https://doi.org/10.1029/2022WR032404, 2022.

Gers, F. A., Schmidhuber, J., and Cummins, F.: Learning to forget: continual prediction with LSTM, *Neural Comput.*, 12, 2451–2471, https://doi.org/10.1162/089976600300015015, 2000.

Golub, M., Thiery, W., Marcé, R., Pierson, D., Vanderkelen, I., Mercado-Bettin, D., et al.: A framework for ensemble modelling of climate-change impacts on lakes worldwide: the ISIMIP lake sector, *Geosci. Model Dev.*, 15, 4597–4623, https://doi.org/10.5194/gmd-15-4597-2022, 2022.

Guo, M. Y., Zhuang, Q. L., Yao, H. X., Golub, M., Leung, L. R., Pierson, D., and Tan, Z. L.: Validation and sensitivity analysis of a 1-D lake model across global lakes, *J. Geophys. Res.-Atmos.*, 126, e2020JD033417, https://doi.org/10.1029/2020JD033417, 2021.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Halevy, A., Norvig, P., and Pereira, F.: The unreasonable effectiveness of data, *IEEE Intell. Syst.*, 24, 8–12, https://doi.org/10.1109/MIS.2009.36, 2009.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al.: The ERA-5 global reanalysis, *Q. J. R. Meteorol. Soc.*, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., et al.: Deep-learning scaling is predictable, empirically, *arXiv* [preprint], arXiv:1712.00409, https://doi.org/10.48550/arXiv.1712.00409, 2017.

Hochreiter, S.: Long short-term memory, *Neural Comput.*, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Holmberg, M., Futter, M. N., Kotamäki, N., Fronzek, S., Forsius, M., Kiuru, P., et al.: Effects of changing climate on the hydrology of a boreal catchment and lake DOC – probabilistic assessment of a dynamic-model chain, *Boreal Environ. Res.*, 19, 66–82, 2014.

Hostetler, S. W., Bates, G. T., and Giorgi, F.: Interactive coupling of a lake-thermal model with a regional climate model, *J. Geophys. Res.-Atmos.*, 98, 5045–5057, https://doi.org/10.1029/92JD02843, 1993.

Huang, L., Wang, X., Sang, Y., Tang, S., Jin, L., Yang, H., et al.: Optimising lake-surface-water-temperature simulations over large lakes in China with FLake model, *Earth Space Sci.*, 8, e2021EA001737, https://doi.org/10.1029/2021EA001737, 2021.

Kayastha, M. B., Huang, C. F., Wang, J. L., Pringle, W. J., Chakraborty, T. C., Yang, Z., et al.: Insights on simulating summer warming of the Great Lakes: understanding the behaviour of a newly developed coupled lake-atmosphere modelling system, *J. Adv. Model. Earth Syst.*, 15, e2023MS003620, https://doi.org/10.1029/2023MS003620, 2023.

Kiuru, P., Ojala, A., Mammarella, I., Heiskanen, J., Erkkilä, K. M., Miettinen, H., et al.: Applicability and consequences of integrating alternative models for CO_2-transfer velocity into a process-based lake model, *Biogeosciences*, 16, 3297–3317, https://doi.org/10.5194/bg-16-3297-2019, 2019.

Korbmacher, R., and Tordeux, A.: Review of pedestrian-trajectory-prediction methods: comparing deep-learning and knowledge-based approaches, *IEEE Trans. Intell. Transp. Syst.*, 23, 24126–24144, https://doi.org/10.1109/TITS.2022.3205676, 2022.

Koya, S. R., and Roy, T.: Temporal-fusion transformers for stream-flow prediction: value of combining attention with recurrence, *J. Hydrol.*, 637, 131301, https://doi.org/10.1016/j.jhydrol.2024.131301, 2024.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using long short-term-memory networks, *Hydrol. Earth Syst. Sci.*, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kurz, S., De Gersem, H., Galetzka, A., Klaedtke, A., Liebsch, M., Loukrezis, D., et al.: Hybrid modelling: towards the next level of scientific computing in engineering, *J. Math. Ind.*, 12, 8, https://doi.org/10.1186/s13362-022-00123-0, 2022.

Ladwig, R., Daw, A., Albright, E. A., Buelo, C., Karpatne, A., Meyer, M. F., et al.: Modular compositional learning improves 1-D hydrodynamic lake-model performance by merging process-based modelling with deep learning, *J. Adv. Model. Earth Syst.*, 16, e2023MS003953, https://doi.org/10.1029/2023MS003953, 2024.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Lee, X.: *Handbook of Micrometeorology: A Guide for Surface-Flux Measurement and Analysis*, Kluwer Acad., Dordrecht, 2004.

Li, L., Sullivan, P. L., Benettin, P., Cirpka, O. A., Bishop, K., Brantley, S. L., et al.: Toward catchment hydro-biogeochemical theories, *WIREs Water*, 8, e1495, https://doi.org/10.1002/wat2.1495, 2021.

Liu, J., Bian, Y., Lawson, K., and Shen, C.: Probing the limit of hydrologic predictability with the transformer network, *J. Hydrol.*, 637, 131389, https://doi.org/10.1016/j.jhydrol.2024.131389, 2024.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., et al.: KAN: Kolmogorov–Arnold networks, *arXiv* [preprint], arXiv:2404.19756, https://doi.org/10.48550/arXiv.2404.19756, 2024.

Markelov, I., Couture, R. M., Fischer, R., Haande, S., and Van Cappellen, P.: Coupling water-column and sediment biogeochemical dynamics: modelling internal phosphorus loading, climate-change responses and mitigation measures in Lake Vansjø, Norway, *J. Geophys. Res.-Biogeosci.*, 124, 3847–3866, https://doi.org/10.1029/2019JG005254, 2019.

Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., and Terzhevik, A.: Implementation of the lake-parameterization scheme FLake into the numerical-weather-prediction model COSMO, *Boreal Environ. Res.*, 15, 218–230, 2010.

Monteiro, M. J., Couto, F. T., Bernardino, M., Cardoso, R. M., Carvalho, D., Martins, J. P. A., et al.: A review on the current status of numerical-weather prediction in Portugal 2021: surface–atmosphere interactions, *Atmosphere*, 13, 1356, https://doi.org/10.3390/atmos13091356, 2022.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., et al.: Challenges and opportunities for integrating lake-ecosystem-modelling approaches, *Aquat. Ecol.*, 44, 633–667, https://doi.org/10.1007/s10452-010-9339-3, 2010.

Nordbo, A., Launiainen, S., Mammarella, I., Leppäranta, M., Huotari, J., Ojala, A., and Vesala, T.: Long-term energy-flux measurements and energy balance over a small boreal lake using eddy-covariance technique, *J. Geophys. Res.-Atmos.*, 116, D02119, https://doi.org/10.1029/2010JD014542, 2011.

Notaro, M., Jorns, J., and Briley, L.: Representation of lake–atmosphere interactions and lake-effect snowfall in the Laurentian Great Lakes Basin among HighResMIP global-climate models, *J. Atmos. Sci.*, 79, 1325–1347, https://doi.org/10.1175/JAS-D-21-0249.1, 2022.

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., et al.: Rapid and highly variable warming of lake-surface waters around the globe, *Geophys. Res. Lett.*, 42, 10773–10781, https://doi.org/10.1002/2015GL066235, 2015.

Piccolroaz, S., Woolway, R. I., and Merchant, C. J.: Global reconstruction of twentieth-century lake-surface-water temperature reveals different warming trends depending on the climatic zone, *Clim. Change*, 160, 427–442, https://doi.org/10.1007/s10584-020-02663-z, 2020.

Piccolroaz, S., Zhu, S., Ladwig, R., Carrea, L., Oliver, S., Piotrowski, A. P., et al.: Lake-water-temperature modelling in an era of climate change: data sources, models and future prospects, *Rev. Geophys.*, 62, e2023RG000816, https://doi.org/10.1029/2023RG000816, 2024.

Pilla, R. M., and Couture, R. M.: Attenuation of photosynthetically active radiation and ultraviolet radiation in response to changing dissolved-organic carbon in browning lakes: modelling and parametrisation, *Limnol. Oceanogr.*, 66, 2278–2289, https://doi.org/10.1002/lno.11753, 2021.

Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics-informed neural networks: a deep-learning framework for solving forward and inverse problems involving nonlinear partial-differential equations, *J. Comput. Phys.*, 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045, 2019.

Read, J. S., Jia, X. W., Willard, J. D., Appling, A. P., Zwart, J. A., Oliver, S. K., et al.: Process-guided deep-learning predictions of lake-water temperature, *Water Resour. Res.*, 55, 9173–9190, https://doi.org/10.1029/2019WR024922, 2019.

Salk, K. R., Venkiteswaran, J. J., Couture, R. M., Higgins, S. N., Paterson, M. J., and Schiff, S. L.: Warming combined with experimental eutrophication intensifies lake-phytoplankton blooms, *Limnol. Oceanogr.*, 67, 147–158, https://doi.org/10.1002/lno.11982, 2022.

Saloranta, T. M., and Andersen, T.: MyLake – a multi-year lake-simulation-model code suitable for uncertainty- and sensitivity-analysis simulations, *Ecol. Model.*, 207, 45–60, https://doi.org/10.1016/j.ecolmodel.2007.03.018, 2007.

Sarovic, K., Buric, M., and Klaic, Z. B.: SIMO v1.0: simplified model of the vertical-temperature profile in a small, warm, monomictic lake, *Geosci. Model Dev.*, 15, 8349–8375, https://doi.org/10.5194/gmd-15-8349-2022, 2022.

Shahriari, B., Swersky, K., Wang, Z. Y., Adams, R. P., and de Freitas, N.: Taking the human out of the loop: a review of Bayesian optimisation, *Proc. IEEE*, 104, 148–175, https://doi.org/10.1109/JPROC.2015.2494218, 2016.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al.: Differentiable modelling to unify machine learning and physical models for geosciences, *Nat. Rev. Earth Environ.*, 4, 552–567, https://doi.org/10.1038/s43017-023-00450-9, 2023.

Sherstinsky, A.: Fundamentals of recurrent-neural-network and long-short-term-memory network, *Physica D*, 404, 132306, https://doi.org/10.1016/j.physd.2019.132306, 2020.

Stepanenko, V., Mammarella, I., Ojala, A., Miettinen, H., Lykosov, V., and Vesala, T.: LAKE 2.0: a model for temperature, methane, carbon-dioxide and oxygen dynamics in lakes, *Geosci. Model Dev.*, 9, 1977–2006, https://doi.org/10.5194/gmd-9-1977-2016, 2016.

Subin, Z. M., Riley, W. J., and Mironov, D.: An improved lake model for climate simulations: model structure, evaluation and sensitivity analyses in CESM1, *J. Adv. Model. Earth Syst.*, 4, M02001, https://doi.org/10.1029/2011MS000072, 2012.

Sun, A. Y., Jiang, P. S., Mudunuru, M. K., and Chen, X. Y.: Explore spatio-temporal learning of large-sample hydrology using graph-neural networks, *Water Resour. Res.*, 57, e2021WR030394, https://doi.org/10.1029/2021WR030394, 2021.

Tong, Y., Feng, L., Wang, X., Pi, X., Xu, W., and Woolway, R. I.: Global lakes are warming slower than surface-air temperature due to accelerated evaporation, *Nat. Water*, 1, 929–940, https://doi.org/10.1038/s44221-023-00148-8, 2023.

Verburg, P., and Antenucci, J. P.: Persistent unstable atmospheric boundary layer enhances sensible- and latent-heat loss in a tropical great lake: Lake Tanganyika, *J. Geophys. Res.-Atmos.*, 115, D11109, https://doi.org/10.1029/2009JD012839, 2010.

Wang, B. B., Ma, Y. M., Wang, Y., Su, Z. B., and Ma, W. Q.: Significant differences exist in lake–atmosphere interactions and evaporation rates of high-elevation small and large lakes, *J. Hydrol.*, 573, 220–234, https://doi.org/10.1016/j.jhydrol.2019.03.066, 2019.

Wang, J. L., Xue, P. F., Pringle, W., Yang, Z., and Qian, Y.: Impacts of lake-surface temperature on the summer climate over the Great-Lakes region, *J. Geophys. Res.-Atmos.*, 127, e2021JD036231, https://doi.org/10.1029/2021JD036231, 2022.

Wang, W. J., Shi, K., Wang, X. W., Zhang, Y. L., Qin, B. Q., Zhang, Y. B., and Woolway, R. I.: The impact of extreme heat on lake warming in China, *Nat. Commun.*, 15, 70, https://doi.org/10.1038/s41467-023-44404-7, 2024.

Wang, X. W., Shi, K., Qin, B. Q., Zhang, Y. B., and Woolway, R. I.: Disproportionate impact of atmospheric-heat events on lake-surface-water-temperature increases, *Nat. Clim. Change*, 14, 1172–1177, https://doi.org/10.1038/s41558-024-02122-y, 2024a.

Wang, X. W., Woolway, R. I., Shi, K., Qin, B. Q., and Zhang, Y. L.: Lake cold spells are declining worldwide, *Geophys. Res. Lett.*, 51, e2024GL111300, https://doi.org/10.1029/2024GL111300, 2024b.

Wang, Z. C., Xu, N. N., Bao, X. G., Wu, J. H., and Cui, X. F.: Spatio-temporal deep-learning model for accurate stream-flow prediction with multi-source data fusion, *Environ. Model. Softw.*, 178, 106091, https://doi.org/10.1016/j.envsoft.2024.106091, 2024.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L.: Transformers in time series: a survey, *arXiv* [preprint], arXiv:2202.07125, https://doi.org/10.48550/arXiv.2202.07125, 2022.

Wikle, C. K., and Zammit-Mangion, A.: Statistical deep learning for spatial and spatiotemporal data, *Annu. Rev. Stat. Appl.*, 10, 247–270, https://doi.org/10.1146/annurev-statistics-033021-112628, 2023.

Willard, J. D., Jia, X. W., Xu, S. M., Steinbach, M., and Kumar, V.: Integrating scientific knowledge with machine learning for engineering and environmental systems, *ACM Comput. Surv.*, 55, 1–37, https://doi.org/10.1145/3514228, 2023.

Willard, J. D., Read, J. S., Appling, A. P., Oliver, S. K., Jia, X. W., and Kumar, V.: Predicting water-temperature dynamics of unmonitored lakes with meta-transfer learning, *Water Resour. Res.*, 57, e2021WR029579, https://doi.org/10.1029/2021WR029579, 2021.

Willard, J. D., Read, J. S., Topp, S., Hansen, G. J. A., and Kumar, V.: Daily surface temperatures for 185 549 lakes in the conterminous United States estimated using deep learning (1980–2020), *Limnol. Oceanogr. Lett.*, 7, 287–301, https://doi.org/10.1002/lol2.10249, 2022.

Woolway, R. I., Jones, I. D., Hamilton, D. P., Maberly, S. C., Muraoka, K., Read, J. S., et al.: Automated calculation of surface-energy fluxes with high-frequency lake-buoy data, *Environ. Model. Softw.*, 70, 191–198, https://doi.org/10.1016/j.envsoft.2015.04.013, 2015.

Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., and Sharma, S.: Global lake responses to climate change, *Nat. Rev. Earth Environ.*, 1, 388–403, https://doi.org/10.1038/s43017-020-0067-5, 2020.

https://doi.org/10.5194/egusphere-2025-1983
Preprint. Discussion started: 21 May 2025

Woolway, R. I., Tong, Y., Feng, L., Zhao, G., Dinh, D. A., Shi, H. R., et al.: Multivariate extremes in lakes, *Nat. Commun.*, 15, 4559, https://doi.org/10.1038/s41467-024-49012-7, 2024.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., and Long, M.: TimesNet: temporal 2-D-variation modelling for general time-series analysis, *arXiv* [preprint], arXiv:2210.02186, https://doi.org/10.48550/arXiv.2210.02186, 2022.

Xu, L. J., Liu, H. Z., Du, Q., and Wang, L.: Evaluation of the WRF-lake model over a highland freshwater lake in southwest China, *J. Geophys. Res.-Atmos.*, 121, 13989–14005, https://doi.org/10.1002/2016JD025396., 2016.

Xu, T. F., and Liang, F.: Machine learning for hydrologic sciences: an introductory overview, *WIREs Water*, 8, e1533, https://doi.org/10.1002/wat2.1533, 2021.

Yan, X., Xia, Y. Q., Ti, C. P., Shan, J., Wu, Y. H., and Yan, X. Y.: Thirty years of experience in water-pollution control in Taihu Lake: a review, *Sci. Total Environ.*, 914, 169821, https://doi.org/10.1016/j.scitotenv.2023.169821, 2024.

Yuan He.: Code and datasets of paper "Hybrid Lake Model (HyLake) v1.0: unifying deep learning and physical principles for simulating lake-atmosphere interactions ", Zenodo [code and data set], https://doi.org/10.5281/zenodo.15289113, 2025.

Zhang, G., Yao, T., Chen, W., Zheng, G., Shum, C. K., Yang, K., et al.: Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes, *Remote Sens. Environ.*, 221, 386–404, https://doi.org/10.1016/j.rse.2018.11.038, 2019.

Zhang, G. Q., Yao, T. D., Xie, H. J., Yang, K., Zhu, L. P., Shum, C. K., et al.: Response of Tibetan-Plateau lakes to climate change: trends, patterns and mechanisms, *Earth-Sci. Rev.*, 208, 103269, https://doi.org/10.1016/j.earscirev.2020.103269, 2020.

Zhang, Y. L., Qin, B. Q., Zhu, G. W., Shi, K., and Zhou, Y. Q.: Profound changes in the physical environment of Lake Taihu from 25 years of long-term observations: implications for algal-bloom outbreaks and aquatic-macrophyte loss, *Water Resour. Res.*, 54, 4319–4331, https://doi.org/10.1029/2017WR022401, 2018.

Zhang, Z., Zhang, M., Cao, C., Wang, W., Xiao, W., Xie, C. Y., et al.: A dataset of microclimate, radiation and energy fluxes from the Lake Taihu eddy-flux network, *Earth Syst. Sci. Data*, 12, 2635–2645, https://doi.org/10.5194/essd-12-2635-2020, 2020.

Zhong, L. J., Lei, H. M., and Yang, J. J.: Development of a distributed physics-informed deep-learning hydrological model for data-scarce regions, *Water Resour. Res.*, 60, e2023WR036333, https://doi.org/10.1029/2023WR036333, 2024.

Zhong, W., Yu, N., and Ai, C. Y.: Applying big-data-based deep-learning system to intrusion detection, *Big Data Min. Anal.*, 3, 181–195, https://doi.org/10.26599/BDMA.2020.9020003, 2020.