

Water Resources Research®

RESEARCH ARTICLE

10.1029/2025WR041062

Special Collection:

Advances in large-scale hydrological modeling and prediction under global change

Key Points:

- Multi-stage transfer learning (pre-training, re-training, and fine-tuning) is integrated into physics-informed deep learning (PIDL)
- Transfer-PIDL reduces local data demands and improves lake surface water temperature (LSWT) prediction
- Transfer-PIDL demonstrates generalizability to large-scale LSWT prediction

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Liu,
muyuan.liu23@imperial.ac.uk

Citation:

Liu, M., Woolway, R. I., Xu, C., Tong, Y., Wang, W., Shi, H., et al. (2026). Predicting lake surface water temperature with transfer-based physics-informed deep learning. *Water Resources Research*, 62, e2025WR041062. <https://doi.org/10.1029/2025WR041062>

Received 15 MAY 2025

Accepted 22 MAR 2026

Author Contributions:

Conceptualization: Muyuan Liu, R. Iestyn Woolway, Haoran Shi, Nada Alsulaiman, James Tlhomole, Matthew D. Piggott

Funding acquisition: Muyuan Liu, R. Iestyn Woolway, Matthew D. Piggott

Methodology: Muyuan Liu

Resources: R. Iestyn Woolway, Caicai Xu, Yan Tong, Weijia Wang, Matthew D. Piggott

Software: Muyuan Liu

Supervision: R. Iestyn Woolway, Matthew D. Piggott

© 2026. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Predicting Lake Surface Water Temperature With Transfer-Based Physics-Informed Deep Learning

Muyuan Liu¹, R. Iestyn Woolway², Caicai Xu^{3,4}, Yan Tong^{2,5}, Weijia Wang^{2,6}, Haoran Shi², Nada Alsulaiman^{7,8}, James Tlhomole⁹, Robert Ladwig¹⁰, and Matthew D. Piggott¹
¹Department of Earth Science and Engineering, Imperial College London, London, UK, ²School of Ocean Sciences, Bangor University, Bangor, UK, ³College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China, ⁴Institute of Zhejiang University-Quzhou, Quzhou, China, ⁵School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, ⁶Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, ⁷Department of Civil and Environmental Engineering, Imperial College London, London, UK, ⁸Environment and Life Sciences Research Center, Kuwait Institute for Scientific Research, Kuwait City, Kuwait, ⁹Department of Engineering, University of Cambridge, Cambridge, UK, ¹⁰Department of Ecoscience, Aarhus University, Aarhus, Denmark

Abstract Ongoing climate change has intensified lake surface warming. Enhanced accuracy in lake temperature modeling can better assess the risk of ecosystems being harmed by thermal tipping points, informing more sustainable water management. Recent progress in physics-informed deep learning (PIDL) has opened new avenues for improving such modeling. Yet, the site-specific focus of PIDL training continues to pose challenges to its broader applicability. Here, we demonstrate Transfer-PIDL, a transfer learning framework that enhances PIDL generalizability for lake surface temperature prediction. This approach employs a three-stage training strategy, consisting of pre-training based on large-scale satellite observations, re-training using process-based (PB) model simulations, and fine-tuning with local measurements. Our first experiment revealed that, given sufficient pre-training (e.g., on 40 source lakes), Transfer-PIDL outperformed local PIDL by 20%–39% in validation root-mean-square-error (RMSE). Additionally, to achieve comparable performance, Transfer-PIDL showed a reduced fine-tuning data requirement compared with both local PIDL and purely data-driven deep learning (DL) across three initial cases. Transfer-PIDL further demonstrated consistent accuracy across 43 additional lakes with in situ temperature observations (mean validation RMSE of 1.2°C), surpassing local PIDL (1.6°C), DL (1.8°C), and PB (1.9°C) models. In a global-scale experiment involving 869 lakes, Transfer-PIDL exhibited cross-thermal-system transferability, with the poorest-performing validation scenario still achieving a RMSE of $1.5 \pm 0.36^\circ\text{C}$, mean-absolute-error of $1.1 \pm 0.25^\circ\text{C}$, and R^2 of 0.85 ± 0.12 (mean $\pm$ SD). This study demonstrates the synergy between transfer learning and PIDL, offering a promising approach for large-scale lake temperature modeling.

1. Introduction

Lakes are undergoing intensified warming in a changing climate (Maberly et al., 2020; Merz et al., 2023; O'Reilly et al., 2015; Perkins-Kirkpatrick & Lewis, 2020), with growing evidence of extreme atmospheric events, such as heatwaves, impacting lake surface waters (Fleischmann et al., 2025; W. Wang et al., 2024; Woolway et al., 2021). The ongoing thermal stress in lake surface waters has been associated with an increased frequency and severity of ecological issues, for example, aquatic biodiversity loss (Rasconi et al., 2017), harmful algal blooms (Duan et al., 2024), and fish mortality events (Fleischmann et al., 2025; Till et al., 2019; Tye et al., 2022). Such temperature-induced disturbances pose a severe threat to lake functions and the services they provide, particularly the supply of safe drinking water, recreational benefits, and productive fisheries (Woolway et al., 2022). In turn, the rapid warming of lake surface water temperature (LSWT) has emerged as a critical global concern (O'Reilly et al., 2015; Woolway et al., 2022).

Modeling is essential for interpreting lake responses to climate change by capturing lake surface thermodynamics (Piccolroaz et al., 2024). Physics-based (PB) numerical models are typically advantageous for this purpose, as they offer a theory-based discretized approximation to real-world lake processes (Hipsey et al., 2019; Mironov, 2008). However, the main limitation with PB models lies in their site-specific parameterization and resource-intensive calibration, which often constrain their broader applicability (Bruce et al., 2018). This has led to a growing interest in data-driven approaches, particularly machine learning or deep learning (DL) techniques, given their adaptability and efficiency in tackling large-scale lake modeling problems (e.g., Liu et al., 2022, 2023;

Validation: Muyuan Liu
Visualization: Muyuan Liu, Caicai Xu
Writing – original draft: Muyuan Liu
Writing – review & editing:
R. Iestyn Woolway, Robert Ladwig,
Matthew D. Piggott

Piccolroaz et al., 2024; Shen, 2018; Xu et al., 2025). Yet, since classical machine learning prediction rules are abstracted from patterns seen in the data, it remains an open question whether such methods can still maintain strong performance in data-scarce lakes, or make reliable predictions beyond training contexts such as future climate scenarios. Indeed, there is an increasing need for integrating scientific knowledge and data-driven insights to advance research efforts (Karpatne et al., 2024).

Recent advances in physics-informed deep learning (PIDL) have led to new opportunities to overcome such modeling challenges. As a representative approach within the knowledge-guided machine learning framework (Karpatne et al., 2024), the cutting-edge PIDL integrates the strengths of PB models and DL techniques (Karniadakis et al., 2021). It utilizes PB simulations for pre-training and embeds physical laws to constrain the DL techniques (Karpatne et al., 2024). In practice, this hybrid approach has been shown to outperform PB models in simulating lake temperatures (Jia et al., 2019), while improving DL-based predictions when applied beyond the range of training data (Read et al., 2019). Despite these improvements, fully realizing the potential of PIDL remains challenging for most unrepresentative lakes. Given their lack of both site-specific characteristics and training data, PIDL models may exhibit reduced performance due to inadequate pre-training or fine-tuning (Jia et al., 2021; Lee et al., 2016; S. Wang et al., 2021). The high demand for the local information potentially limits the scalability of PIDL across a broad range of environmental settings (Karpatne et al., 2024).

Despite being affected by strong local factors such as wind, shading, and hydrological flows (Markfort et al., 2010; Read et al., 2014), lake systems often exhibit coherent physical behaviors across large spatial scales, particularly in their surface thermal patterns (Benson et al., 2000; Maberly et al., 2020; Woolway et al., 2020, 2021). Such large-scale patterns have shown great promise for incorporating transfer learning (TL) to enhance PIDL models. As a powerful technique often coupled with DL methods, TL seeks to improve modeling in new (target) domains by leveraging knowledge acquired from base (source) domains (Pan & Yang, 2010). Evidence suggested that pre-training DL models on data-rich source domains can enhance performance and reduce data requirements in similar target domains, as observed in various scientific applications such as soil moisture prediction (Q. Li et al., 2021), water quality management (Chen et al., 2024), and streamflow modeling (Ma et al., 2021). Recent advances in TL, coupled with meta-learning (Hospedales et al., 2021; Huisman et al., 2020), further indicated that lakes with similar characteristics (e.g., maximum depth) can share the same PIDL models (Willard et al., 2021). This specific-to-specific model transfer demonstrated the potential of incorporating TL with PIDL for lake temperature prediction. However, gaps remain in bridging large-scale lake temperature modeling with site-specific PIDL frameworks.

To fill this information gap, large-scale data sources, particularly satellite-derived LSWT observations, provide valuable input for this general-to-specific TL framework. In recent years, given their expanding spatial and temporal coverage (Carrea et al., 2023; Korver et al., 2024), satellite-derived LSWTs have underpinned a wide range of global-scale lake temperature studies, including long-term warming trends (Tong et al., 2023), thermal variability (Maberly et al., 2020), and responses to climatic drivers (Woolway et al., 2021). These achievements demonstrate the value of satellite-derived LSWTs as a data foundation for TL application, yet challenges remain. As factors such as the cool-skin effect and cloudiness still introduce inaccuracies into satellite-derived LSWTs (Piccolroaz et al., 2024), efforts are needed to improve TL effectiveness.

Therefore, this study proposed a three-stage TL framework to improve PIDL predictions for LSWT. This framework uses PB model simulations as a bridge, mitigating the impacts of satellite–field LSWT inconsistencies on final Transfer-PIDL performance. Specifically, the three stages comprise (a) large-scale pre-training using satellite observations, (b) site-specific re-training based on PB simulations, and (c) local fine-tuning with real measurements. We hypothesized that this Transfer-PIDL would (a) reduce local data requirements, (b) improve model generalizability across diverse lake environments, and (c) enhance transfer robustness against source–target discrepancies.

2. Materials and Methods

2.1. PIDL Model

2.1.1. PIDL Components

Prediction rules of pure DL models can be undesirable when generalizing outside of their training contexts, particularly if the training data lacks adequate representations (Rousso et al., 2020; Shen, 2018; Zhong

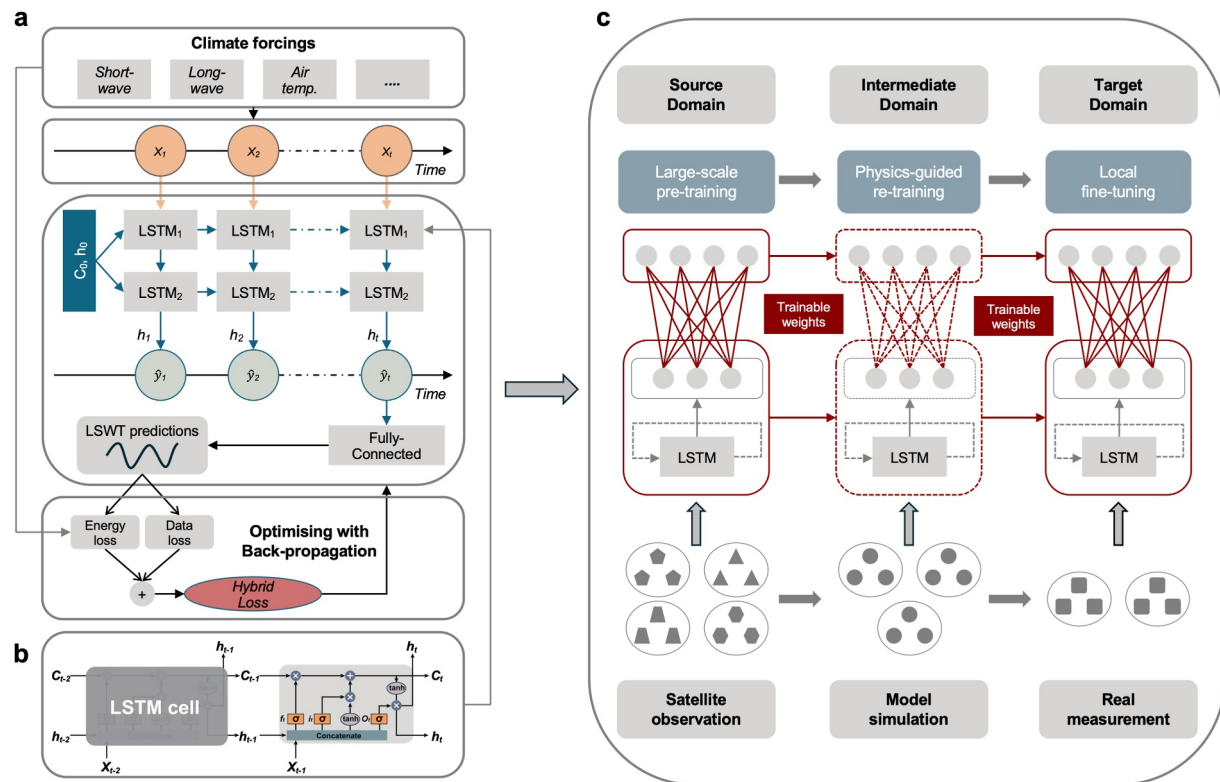


Figure 1. Physics-informed deep learning with transfer learning. (a, b) Structure of the physics-informed LSTM model for lake surface water temperature prediction. (c) Illustration of the multi-stage transfer learning process.

et al., 2021). To mitigate such limitations, an embedded physical constraint, for example, the law of energy conservation (Jia et al., 2019; Read et al., 2019), can often provide an effective solution for DL models in lake temperature prediction. In this study, the PIDL model consists of a long short-term memory (LSTM) network that integrates an energy-based loss term into the data-driven loss function (Figure 1a).

The LSTM is designed as an extension of the recurrent neural network (Hochreiter & Schmidhuber, 1997). Through its gating structure (Figure 1b), LSTM overcomes the challenges of gradient explosion and vanishing during loop training, facilitating the learning of both short- and long-term sequences (Liu et al., 2022; Xu et al., 2025). This provides a useful framework, particularly for modeling lake temperature (Jia et al., 2019; Read et al., 2019), as lakes have underlying thermodynamics that could evolve over time due to heat storage capacity (e.g., seasonal fluctuations). Additionally, the energy-based loss term was introduced via a simplified lake bulk model (see Text S1 in Supporting Information S1). This simple numerical equation divided modeling into two scenarios dependent on ice condition, that is, ice-free and ice-covered (Figure S1 in Supporting Information S1), both of which considered only the predominant lake energy budgets (all in W m^{-2}), such as shortwave radiation, longwave radiation, sensible heat, and latent heat (Read et al., 2019; Rogers et al., 1995; Wanders et al., 2019; Woolway et al., 2020). Other energy components such as sediment heating and advective heat transport were not included in this equation, as their contributions to the daily energy budget would typically be less than 1% (Lenters et al., 2005; Read et al., 2019). The input features for the model are the meteorological variables that can have major impacts on lake surface thermodynamics (Edinger et al., 1968; Schmid et al., 2014; Woolway et al., 2020), including shortwave and longwave radiation fluxes (in W m^{-2}), air temperature (in $^{\circ}\text{C}$), relative humidity (in %), and wind speed (in m s^{-1}). The day of the year (DOY) series was also included given the temporal fluctuations of lake temperatures. Additional details on the PIDL model components and training configurations are presented in Text S2 in Supporting Information S1 and Section 2.5.

2.1.2. PIDL Pre-Training

Pre-training serves as an essential phase in PIDL models. Commonly known as physics-guided initialization or physics-guided pre-training (Karpatne et al., 2024), the initial warm-up utilizes PB model simulation to obtain a weight basis that provides a priori physical consistency, helping ensure that the outputs satisfy the physical constraints (Tayal et al., 2022). This technique has been shown to significantly accelerate convergence and improve the predictive capabilities of PIDL models for lake temperatures, even when they are fine-tuned with extremely limited data (Jia et al., 2019). In this study, to perform the pre-training for PIDL models, we employed the General Lake Model (GLM; Hipsey et al., 2019). GLM is a widely used one-dimensional hydrodynamic model designed based on the balance of mass and energy, providing a robust framework for simulating lake thermodynamics at daily or sub-daily frequency (see <https://github.com/AquaticEcoDynamics/GLM>). Although GLM is a vertical one-dimensional model, we constrained the methodology by utilizing only its LSWT simulations, which most commonly achieves the best performance (Feldbauer et al., 2025). A more detailed description of the GLM configurations can be found in Section 2.5.

2.2. PIDL With Multi-Stage Transfer Learning

A potential limitation in PIDL pre-training, however, is the high demand for the proper parameterization or calibration of PB models. Incorrectly initialized PIDL models often lead to undesirable performance losses and require additional fine-tuning data to recover (e.g., hundreds to thousands of data points; Jia et al., 2021). For less monitored or even unmonitored lakes, which represent the majority, this strategy could constrain the generalizability of PIDL. In response to this challenge, rather than solely relying on site-specific PB model simulations, we jointly integrated global satellite observations for PIDL pre-training. Specifically, we implemented a multi-stage transfer learning to achieve this objective (Figure 1c).

For illustration, we first define the key notation in the transfer learning framework. Let S and T denote the source and target domains, respectively, and let P represent the model parameter space. Within this framework, the conventional PIDL training scheme can be interpreted as parameter-based transfer learning, where the source domain S corresponds to the PB model simulations, and target domain T corresponds to the real measurements (e.g., in situ data). Thus, we can define the following hypothesis to describe this training process:

$$\varphi_{ST}: P_S \rightarrow P_T, \quad (1)$$

where φ_{ST} represents the hypothesis to perform parameter transfer from S to T , that is, fine-tuning the PIDL pre-trained with PB simulations to adapt to real measurements. However, this one-step hypothesis may be directly affected by unreliable source data (e.g., uncertain site-specific PB simulations). We thus integrate global satellite observations to provide more consistent source information. To more effectively employ the parameter transfer, we extend this one-step hypothesis to a multi-stage setting (Farahani et al., 2020; Gu et al., 2020; Iman et al., 2023):

$$\varphi_{SMT}: P_S \rightarrow P_M \rightarrow P_T, \quad (2)$$

where φ_{SMT} represents the two-step hypothesis of parameter transfers, that is, from S to T via an intermediate domain M . In this setting, the source domain S now corresponds to large-scale satellite observations, whereas the intermediate domain M corresponds to site-specific PB model simulations. Specifically, three training stages are involved in φ_{SMT} : (a) pre-training using large-scale satellite observations, (b) re-training based on site-specific PB simulations, and (c) fine-tuning with local measurements. The details of each stage are provided in Figure 1c and Sections 2.2.1–2.2.3.

2.2.1. Large-Scale Pre-Training

In this stage, the PIDL model is pre-trained using large-scale satellite-derived lake data sets (see Section 2.3.1 for data set details). These data sets, such as the European Space Agency-Climate Change Initiative Lakes project (Carrea et al., 2023), provide long-term LSWT retrievals worldwide. This extensive spatial and temporal coverage helps the model learn general LSWT patterns across diverse meteorological conditions. This stage aims to develop a universal pre-trained PIDL model capable of transferring its parameter weights to downstream tasks

for individual lake targets. To better capture such transferable representations from the large-scale data set, we consider a data-driven mean-squared-error (MSE) loss function as:

$$\mathcal{L}_S = \mathcal{L}_{\text{MSE}-S}, \quad (3)$$

where $\mathcal{L}_{\text{MSE}-S}$ represents the MSE loss used in the large-scale pre-training stage (see “Loss Functions” for its formulation and “Training Configurations” for implementation details in Text S2 in Supporting Information S1).

2.2.2. Physics-Guided Re-Training

This stage is designed to facilitate the transition of the general pre-trained PIDL model to local lake environments, a process commonly referred to as domain adaptation. Domain adaptation is a special case of transfer learning, seeking to handle gaps under the “same task, different domain distributions” condition (Farahani et al., 2020). It is typically performed in an unsupervised manner by minimizing the distributional differences between the source and target domains. Often, this objective is achieved using the maximum-mean-discrepancy (MMD), a widely adopted metric for measuring domain distance (Farahani et al., 2020; Y. Li et al., 2015). MMD operates on the feature representations of the source and target data (i.e., their respective model inputs), serving as an unsupervised loss function to support domain shifts (Gretton et al., 2012; Y. Li et al., 2015). To better stabilize the shifts, the PB model simulations here are used as a guide (i.e., an intermediate domain), offering site-specific data patterns that facilitate such general-to-specific weight transfer. This design, trained with a supervised loss term, also helps reinforce the model's physical consistency during re-training. Additionally, another supervised loss based on available source data (i.e., satellite observations) is also incorporated to mitigate catastrophic forgetting of pre-learned general patterns. Therefore, a combined loss function at this stage is given by:

$$\mathcal{L}_M = \mathcal{L}_{\text{MSE}-M} + \mathcal{L}_{\text{MSE}-S} + \mathcal{L}_{\text{MMD}}, \quad (4)$$

where $\mathcal{L}_{\text{MSE}-M}$ represents the MSE loss based on the site-specific intermediate data (i.e., PB model simulations), $\mathcal{L}_{\text{MSE}-S}$ represents the MSE loss based on the large-scale source data (i.e., satellite observations), and $\mathcal{L}_{\text{MMD}}$ denotes the domain-discrepancy loss given by the MMD distance between source and intermediate features. The fixed weights assigned to these three loss terms are specified in the “Training Configurations” section of Text S2 in Supporting Information S1, with detailed calculation procedures provided in the “Loss Functions” section.

2.2.3. Local Fine-Tuning

In the final stage, the re-trained PIDL model was fine-tuned using local LSWT measurements. We made the parameter weights of PIDL modules (e.g., LSTM and fully connected layers) retrainable during the fine-tuning, which can take full advantage of physics information and better safeguard the calibration of the PIDL model against learning spurious patterns. This implementation employed both data-driven and energy-based losses as follows:

$$\mathcal{L}_T = \mathcal{L}_{\text{MSE}-T} + \mathcal{L}_{\text{Energy}}, \quad (5)$$

where $\mathcal{L}_{\text{MSE}-T}$ represents the MSE loss based on the target data (e.g., in situ measurements), and $\mathcal{L}_{\text{Energy}}$ is the energy-based loss described in Section 2.1.1. The weights of the two loss terms are specified in the Text S2 in Supporting Information S1, where the detailed calculations are also provided (see also “Training Configurations” and “Loss Functions”).

2.3. Data Sources

2.3.1. Global Lake Data Sets

Satellite-derived LSWT observations are obtained from the European Space Agency (ESA) Climate Change Initiative (CCI) Lakes project, which provides long-term monitoring for over 2,000 lakes worldwide (Carrea et al., 2023). The monitoring span 26 years (ranging from 1995 to 2020), containing mid-morning LSWT records derived across multiple satellite instruments each day. The daytime records are remapped onto a 0.05° spatial grid with an uncertainty level and include quality flags for each measurement. Only high-confidence satellite retrievals

(quality levels 4 or 5) were used to ensure accurate LSWT measurements, as recommended by the data product. The valid measurements in the lake mask were then averaged to yield the lake-wide mean LSWT retrievals.

The input meteorological variables for all models are provided by the land component of the fifth generation of European Reanalysis (or ERA5-Land reanalysis). The ERA5-Land is a gridded reanalysis data set produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides enhanced climate-forcing variables for global land applications (Muñoz-Sabater et al., 2021). In this study, daily meteorological variables were sourced and averaged from the ERA5-Land hourly data set, which provides 70 years of climate data at a 0.1° horizontal resolution (Muñoz-Sabater, 2019). Note that accumulated variables in ERA5-Land (i.e., total precipitation or fluxes) are interpreted as cumulative values over a certain period, their daily-averaged values were calculated following the conversion table released by ECMWF. We performed a nearest-neighbor search to obtain lake-specific meteorological variables, that is, extracting data from the grid point that is situated closest to the center of each lake. Although this approach may not fully capture the spatial variability of meteorological forcing (e.g., for lakes located in regions with strong climate heterogeneity), it has consistently been shown to provide reliable performance in large-scale lake modeling applications (Tong et al., 2023; W. Wang et al., 2024; Woolway et al., 2021).

Additional lake-specific information (e.g., surface area and hypsographic curves) is further needed for PB model parameterization. In this study, HydroLAKES is used to provide essential lake characteristic properties. HydroLAKES is a global lake database developed based on geostatistical models, which encompasses topographic estimates for 1.4 million lakes and reservoirs worldwide (Messager et al., 2016). We assigned property attributes (e.g., volume, depth, and surface area) to every studied lake via overlapping lake centroids with the geometric polygons provided by the HydroLAKES database.

Based on all lakes from the CCI project, we included those that were observed at least once per year and excluded lakes lacking corresponding ERA5-Land data (e.g., lakes located under sea masks). In addition, lakes for which the PB model could not be realistically parameterized using HydroLAKES estimates were also removed. Ultimately, 869 CCI lakes were retained for model experiments (see Table S1 and Figure S2a in Supporting Information S1).

2.3.2. In Situ Lake Data Sets

In addition, this study included 46 lakes with available in situ data (see Table S2 and Figure S2a in Supporting Information S1). Among them, three lakes, that is, Windermere (UK Centre for Ecology & Hydrology, <https://eidc.ac.uk>), Lough Feeagh (<https://data.gov.ie/organization/marine-institute>), and Lake Kasumigaura (National Institute for Environmental Studies, 2016), have collected long-term and high-completeness in situ daily LSWT observations, as well as field lake characteristics such as hypsographic curves. The remaining 43 lakes contained in situ LSWT observations only, with monitoring frequency varying substantially from minutes to months. As a supplement, their lake morphological characteristics were obtained from the HydroLAKES estimates. Site details and summaries of the observational data for each lake are provided in Table S2 and Figure S2a in Supporting Information S1.

2.4. Model Experiments

2.4.1. Effect of Data Availability on Model Performance

We hypothesized that the transfer learning strategy could improve the performance of lake-specific PIDL models, even in some cases where training data are extremely limited. To test its robustness, we designed this experiment by adjusting the data availability used for developing the Transfer-PIDL. Specifically, we focused on the effects of source data volume (i.e., the number of lakes combined for pre-training) and target data quantity (i.e., the amount of data used for fine-tuning).

The target domain for this experiment included three large lakes with in situ observations, that is, Windermere, Lough Feeagh, and Lake Kasumigaura (see Table S2 and Figure S2b in Supporting Information S1 for target lake details). The source domain consisted of available lakes from the CCI project, focusing on the top 5% (approximately 40 lakes in total) with relatively high completeness of satellite observations ($\geq 50\%$) to ensure data representativeness (excluding any of the target lakes; see Table S1 and Figure S2b in Supporting Information S1 for source lake details).

In this experiment, we used 10 years of daily LSWT observations (2009–2018) to model each target lake. The first 6 years (2009–2014) were allocated for model calibration (i.e., fine-tuning), while the remaining 4 years (2015–2018) were reserved for model validation. To quantify the impact of source data volume, we varied the pool of source lakes in increments of two, from two to the full set of 40, while keeping the fine-tuning data size consistent for each lake. A similar idea was applied to the experiment on target data quantity. We pre-trained the model with the data from the full 40 source lakes, while fine-tuning it on each target lake using consecutive daily local observations spanning different durations (ranging from a month to multiple months). For each lake, all sub-experiments were run for five independent iterations with random sampling.

2.4.2. Model Generalizability Under an Expanded Target Scope

To further evaluate the generalizability of the Transfer-PIDL model, this experiment expanded the scope of the target domain to an additional 43 lakes with publicly available in situ LSWT data (see Table S2 and Figure S2c in Supporting Information S1 for target lake details). This experiment was conducted following the first experiment, with source lakes ($n = 40$, excluding any target sites; see Table S1 and Figure S2c in Supporting Information S1 for source lake details) also selected from the CCI project with the highest satellite measurement completeness. Since the effects of data quantity were no longer a focus in this experiment, the Transfer-PIDL model was pre-trained using the complete source data sets and then fine-tuned exclusively with the full local calibration data for each target lake. We applied a comparable data-splitting scheme used in Experiment 1, using the first 60% of available observations for local fine-tuning and the remaining 40% for validation.

2.4.3. Impact of Thermal Dissimilarity on Model Transferability

Given the use of domain adaptation, multi-stage transfer learning is assumed to be capable of bridging the source-target differences. As such, this experiment was designed to test the transferability of the Transfer-PIDL model when the source and target domains were thermally distinct. To investigate this thermal dissimilarity further, we adopted the framework of Maberly et al. (2020), which defines distinct global lake thermal systems characterized by long-term LSWT dynamics and contemporary climate forcings.

To assess robustness, the modeling scope here was extended to all retained CCI lake sites ($n = 869$). These sites were grouped into five merged lake thermal systems following the classification of Maberly et al. (2020), namely, Northern Frigid ($n = 161$), Northern and Southern Cool ($n = 363$), Northern and Southern Temperate ($n = 149$), Northern and Southern Warm ($n = 82$), and Northern, Southern and Tropical Hot ($n = 114$). All of these lake thermal systems demonstrated significantly distinct LSWT patterns ($p < 0.001$, see Figure S3a in Supporting Information S1). For the experiment here, one of the lake thermal systems was chosen as the source domain, while each of the remaining four systems was independently used as a target domain. This process was repeated five times, ensuring that every system was selected as the source once. As in Experiment 1, the 26 years of daily CCI records (1995–2020) were also split into the first 60% (1995–2010, serving as the local fine-tuning or calibration set) and the last 40% (2011–2020, serving as the validation set). To maintain a consistent experimental setting, we selected the top 40 lakes in each thermal system for model pre-training (ranked by measurement completeness; see Figure S2d in Supporting Information S1 for source lake details), using only their calibration sets. This design also aimed to mitigate potential biases arising from imbalances in source data availability across thermal systems (see Figure S3b in Supporting Information S1).

2.5. Model Comparison and Implementation

In all experiments, PIDL, DL, and PB models were included for comparison. These models were trained exclusively on local data (collectively called local models), with each experiment calibrated using the same amount of fine-tuning data as the Transfer-PIDL model (see “Training Configurations” in Text S2 in Supporting Information S1). The DL model was developed under the same LSTM structure described in Section 2.1.1, while the PB model employed the GLM (described also in Section 2.1.2). The model performance was assessed primarily using the root-mean-square-error (RMSE), supplemented by the mean-absolute-error (MAE) and the coefficient of determination (R^2). We used RMSE as the calibration metric, and all model parameters were explicitly optimized to minimize the RMSE of the calibration data set. The calculation of these metrics, as well as the details of all model implementations specific to each experiment, are described in Text S2 in Supporting Information S1.

Additionally, to provide a consistent lake-specific performance baseline for these models, here we supplemented them with a global LSWT product given by the ERA5-Land, that is, the lake mixed-layer temperature data set (Muñoz-Sabater, 2019). This data set simulates hourly lake uppermost temperature based on the Fresh-water Lake (Flake) model (Mironov, 2008), offering global coverage at a 0.1° horizontal resolution. The ERA5-Land LSWT product has been independently validated against in situ observations across a broad range of lakes (Muñoz-Sabater et al., 2021), serving as a widely used benchmark for large-scale lake temperature modeling (Tong et al., 2023; Willard et al., 2022; Woolway & Maberly, 2020). It was therefore applied as a consistent baseline for evaluating model performance in this study. We extracted the lake-level daily ERA5-Land LSWT products using the same procedures described in Section 2.3.1.

For DL-based models (i.e., Transfer-PIDL, PIDL, and DL), we implemented neural network modules using the DL framework of *PyTorch* version 2 (Paszke et al., 2017). To support stable convergence of the neural networks, the Xavier uniform initialization was applied to the weights of all layer parameters (i.e., LSTM and fully connected layers) prior to the formal training (Glorot & Bengio, 2010). Hyperparameters were tuned using a grid-search scheme, with batch stochastic gradient optimization applied to improve computational efficiency (batch size = 128). The input meteorological variables for the DL-based models were derived from ERA5-Land. To preserve the original temporal variability of these climate-forcing variables, no further feature normalization was applied, as ERA5-Land has already undergone extensive data processing and technical validation procedures (Muñoz-Sabater et al., 2021). Further DL-based implementation details were provided in Text S2 in Supporting Information S1 (see “Model Architecture” and “Training Configurations”), with hyperparameter tuning details listed in Table S3 in Supporting Information S1.

GLM version 3 was utilized in this study (Hipsey et al., 2019), specifically by using packages *GLM3r* 3.1.18 (<https://github.com/GLEON/GLM3r>) and *glmtools* 0.16.0 (<https://github.com/GLEON/glmtools>) in the R environment (R Core Team, 2013). For parallel efficiency, we used the Latin hypercube sampling method to adjust GLM model parameters with the *LakeEnsemblR* 1.2.7 (<https://github.com/aemon-j/LakeEnsemblR>) package in R (Moore et al., 2021). Here we selected six physical parameters for model calibration, given their known importance either to the GLM simulations (Moore et al., 2021; Read et al., 2014, 2019; Willard et al., 2021), or to the lake surface thermodynamics (Woolway et al., 2020). These parameters included shortwave scaling factor, longwave scaling factor, windspeed scaling factor, wind stirring efficiency, hypolimnetic mixing efficiency, and light extinction coefficient (Table S4 in Supporting Information S1). For lakes in Experiment 1, field characteristic measurements (e.g., hypsography curve) were used for GLM parameterization. For lakes in Experiments 2 and 3 that lacked field measurements, estimates from the HydroLAKES database were used (Messenger et al., 2016). Their hypsography curves were estimated using cone shapes (Read et al., 2014), based on lake surface area and volume. GLM configuration details were provided within “Model Architecture” and “Training Configurations” in Text S2 in Supporting Information S1.

3. Results

3.1. Effects of Data Quantity on Transfer Learning

PIDL models consistently benefited from the knowledge gained across diverse source data sets in each lake studied in Experiment 1 (Figures 2–4 and Figure S4 in Supporting Information S1). Based on the full pre-training scale (i.e., 40 source lakes), the greatest improvement was observed in Lake Kasumigaura (Figure 2c), where validation RMSE decreased by 39%, from 1.25°C with local PIDL to 0.76°C with Transfer-PIDL. For the other two lakes, reductions were also achieved, with RMSE in Windermere (Figure 2a) dropping by 20% (from 0.82°C to 0.65°C) and that in Lough Feeagh (Figure 2b) showing a 26% decline (from 0.76°C to 0.56°C). Analogous patterns emerged from point-wise comparisons, where all three cases exhibited clear declines in prediction errors through transfer learning ($p < 0.001$, Figure 3), with the maximum RMSE decrease reaching 3.36°C (Figure 3c). As the number of source lakes increased, the performance gain tended to saturate, as observed from the curves of mean validation RMSEs (Figures 2a–2c). Beyond approximately 20–30 lakes, the curves flattened and exhibited relatively minor fluctuations (e.g., Lake Kasumigaura, Figure 2c). This trend suggested that diminishing returns may occur when source data sets became more abundant.

In each fine-tuning scenario (Figures 4a–4c), the results illustrated that the TL strategy can improve the prediction accuracy of PIDL models across three target lakes (see Figure S4 in Supporting Information S1). In cases where formal training data were relatively limited (e.g., less than 6 months), Transfer-PIDL models achieved

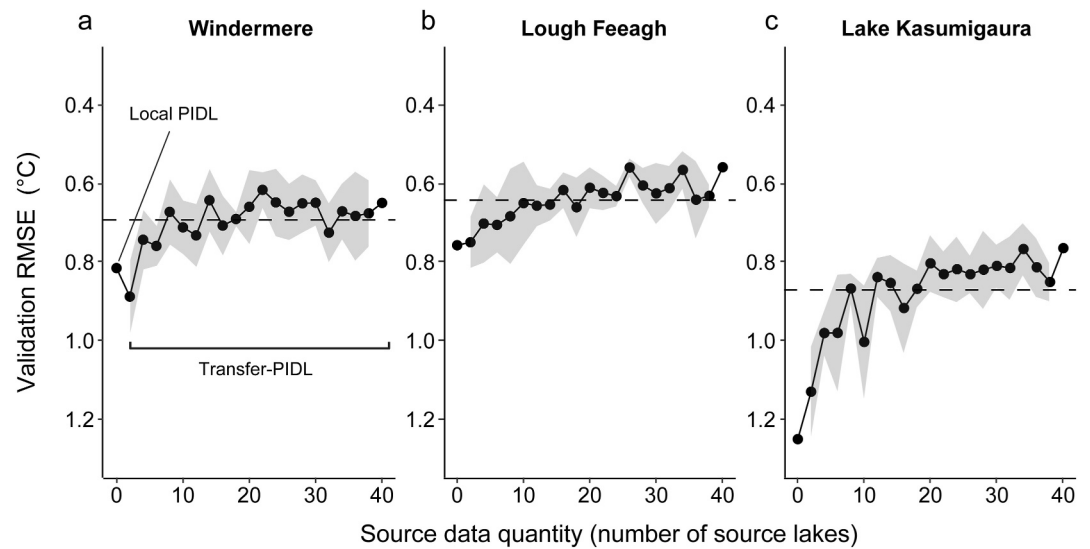


Figure 2. Effect of source data quantity on transfer learning performance for Windermere (a), Lough Feeagh (b), and Lake Kasumigaura (c). In (a–c), the horizontal dashed lines represent the mean value of validation RMSEs for Transfer-PIDL across all pre-training scenarios. Dots and shaded regions indicate the mean and standard deviation (SD) from five experimental iterations.

comparable performance to local PIDL models, while the latter improved the prediction of pure DL models to a large extent. This result can be seen at Windermere in the 6-month scenario, where the mean validation RMSEs for Transfer-PIDL, PIDL and DL were 1.67°C, 1.56°C, and 3.14°C, respectively (Figure 4a). To achieve lower validation RMSE levels (e.g., a threshold below 1.0°C), Transfer-PIDL appeared to require fewer fine-tuning data than local PIDL (e.g., 12-month Transfer-PIDL vs. 42-month PIDL in Windermere; see Figure 4a). Moreover, the TL benefits persisted as fine-tuning data became abundant (e.g., more than 24 months). In Lough Feeagh, using 30 months of data, the mean validation RMSEs for Transfer-PIDL, PIDL and DL were 0.67°C,

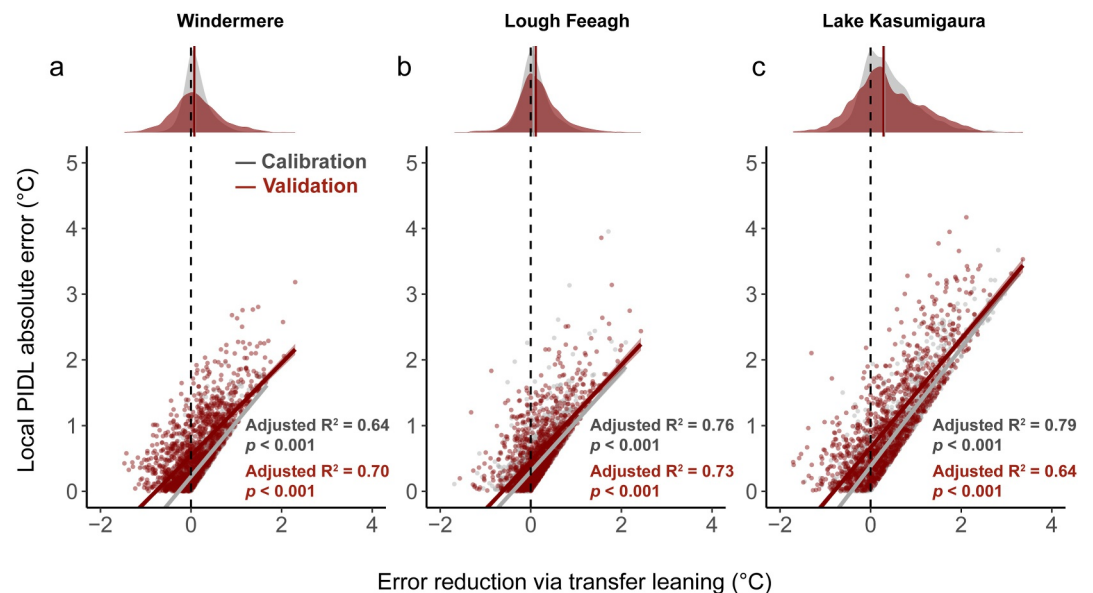


Figure 3. Transfer gains based on the full source data for Windermere (a), Lough Feeagh (b), and Lake Kasumigaura (c). In (a–c), the dots represent point-wise absolute prediction errors from local PIDL models. The solid-colored lines show the linear fit (adjusted R^2 and p -value) of error reduction via transfer learning, shaded by the 95% credible interval. The density histograms depict the distribution of error reductions, with vertical lines indicating the medians. The colors denote calibration (gray) and validation (red) sets.

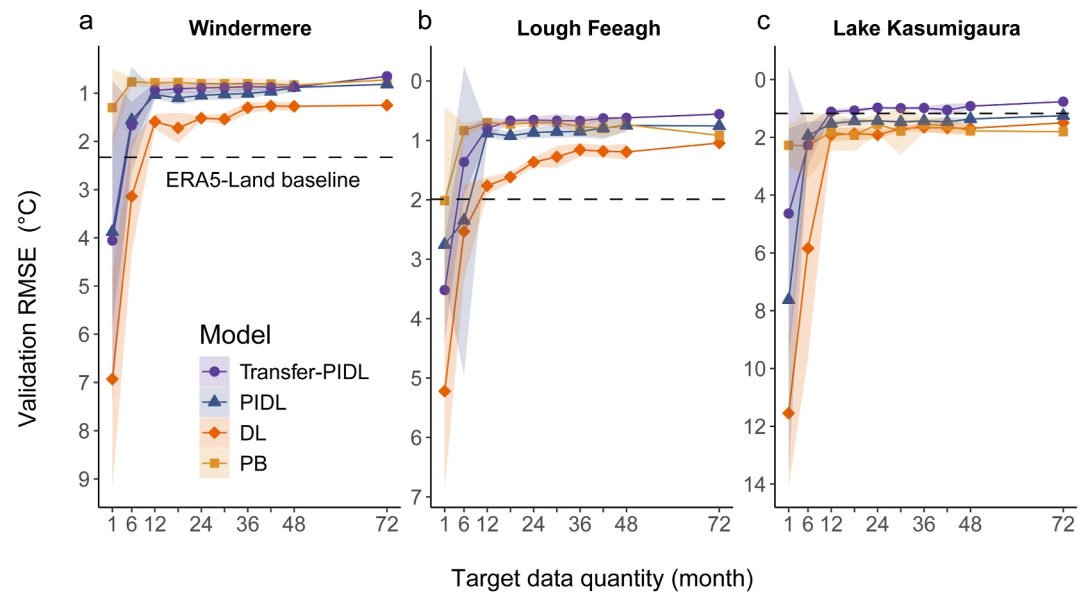


Figure 4. Effect of target data quantity on model performance for Windermere (a), Lough Feeagh (b), and Lake Kasumigaura (c). In (a–c), horizontal dashed lines show the baseline lake-specific performance based on ERA5-Land simulations, while dots and shaded regions indicate the mean and standard deviation (SD) from five experimental iterations.

0.86°C, and 1.28°C, respectively (Figure 4b). The PB models were robust to the quantity of calibrating data, for example, at Lake Kasumigaura (Figure 4c), the PB validation RMSE changed only from $2.28^{\circ}\text{C} \pm 0.70^{\circ}\text{C}$ (mean $\pm$ SD; 1-month data used) to 1.80°C (full 72-month data used). Across all three lakes studied, the PB models exhibited advantageous performance over pure DL models (Figures 4a–4c). This was especially significant when only 1 month of daily data points were available, for example, at Lough Feeagh (Figure 4b), the PB model achieved a validation RMSE of $2.02 \pm 1.57^{\circ}\text{C}$ (mean $\pm$ SD) compared to $5.22 \pm 1.60^{\circ}\text{C}$ for the pure DL model. However, as the amount of training data became increasingly sufficient, the superiority of PB models over pure DL models may disappear (e.g., with 24-month data at Kasumigaura; see Figure 4c).

3.2. Model Performance on an Expanded Target Scope

Predictions from the Transfer-PIDL model for an additional 43 target lakes were as accurate as those from the three lakes studied in Experiment 1 (see Figures 5 and 6 and Figures S5 and S9a in Supporting Information S1). Overall, Transfer-PIDL maintained high and stable performance, achieving a validation RMSE of $1.23 \pm 0.46^{\circ}\text{C}$ (mean $\pm$ SD), with 95% of the lake-specific validation RMSEs below 2.20°C (Figures 5a and 5b). Specifically, Transfer-PIDL achieved the best performance in Lake Luguhu (validation RMSE = 0.65°C , Figures 6a and 6b), median performance in Lake Qiandaohu (RMSE = 1.07°C , Figures 6c and 6d), and the poorest performance in Lake Hulunhu (RMSE = 2.70°C , Figures 6e and 6f). Across these lakes, predictions strongly aligned with observations in both calibration and validation sets (Figures 6a–6f). This consistency indicated that the large pre-

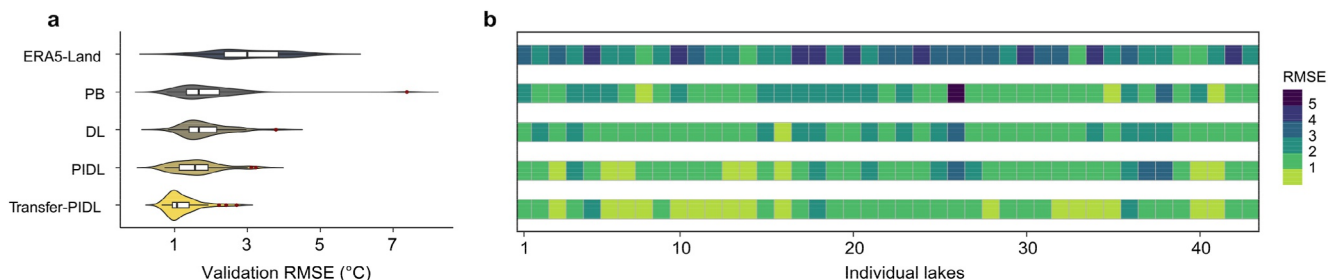


Figure 5. Model performance under an expanded target scope. (a) Model cross-comparisons, with red dots denoting the boxplot outliers. (b) Validation RMSE for each lake included in the experiment.

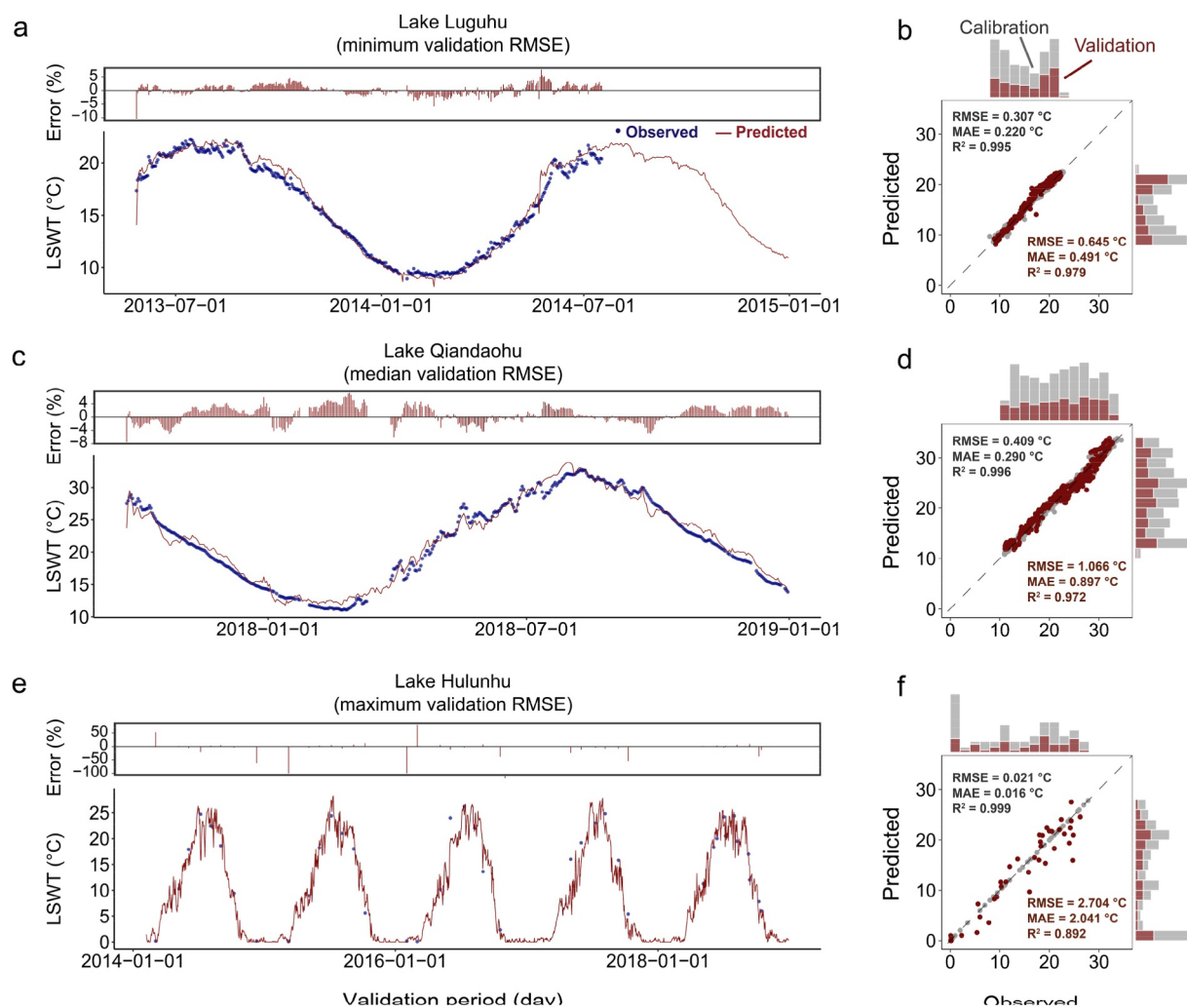


Figure 6. Time series of LSWTs predicted by the Transfer-PIDL model on the validation set in Experiment 2, corresponding to lakes that achieved the lowest (a, b), median (c, d), and highest (e, f) validation RMSEs.

trained model effectively adapted to individual lake targets, maintaining reliable accuracy when faced with diverse local environments.

Comparatively, models developed using only local data, namely, PIDL, DL, and PB models, yielded higher median validation RMSEs of 1.56°C, 1.67°C, and 1.66°C, respectively (see Figure 5 and Figures S6 and S9a in Supporting Information S1). Notably, while the PB model achieved a higher upper-bound performance compared to the other three DL-based models, its RMSE distribution spanned much wider variability (Figure 5a). This large variability suggested that PB models may exhibit limited scalability relative to the data-driven approaches when applied to different lake conditions. Nevertheless, PB models still demonstrated better overall performance than the baseline (i.e., the ERA5-Land LSWT product), as the latter demonstrated a higher median validation RMSE of 3.00°C (Figure 5a and Figure S9a in Supporting Information S1).

3.3. Transfer Learning Across Thermal Systems

The source-target thermal dissimilarity (see Figure 7 and Figure S3a in Supporting Information S1) posed a very limited impact on Transfer-PIDL accuracy (see Figures 7b–7g and Figures S7 and S9c in Supporting Information S1). In general, models pre-trained in similar thermal systems often performed better, as indicated by the lower RMSE values with reduced variability (e.g., intra-domain transfer within the hot lakes, see “H” in Figure 7g). When Transfer-PIDL was applied in cross-domain scenarios, the variability in RMSE tended to

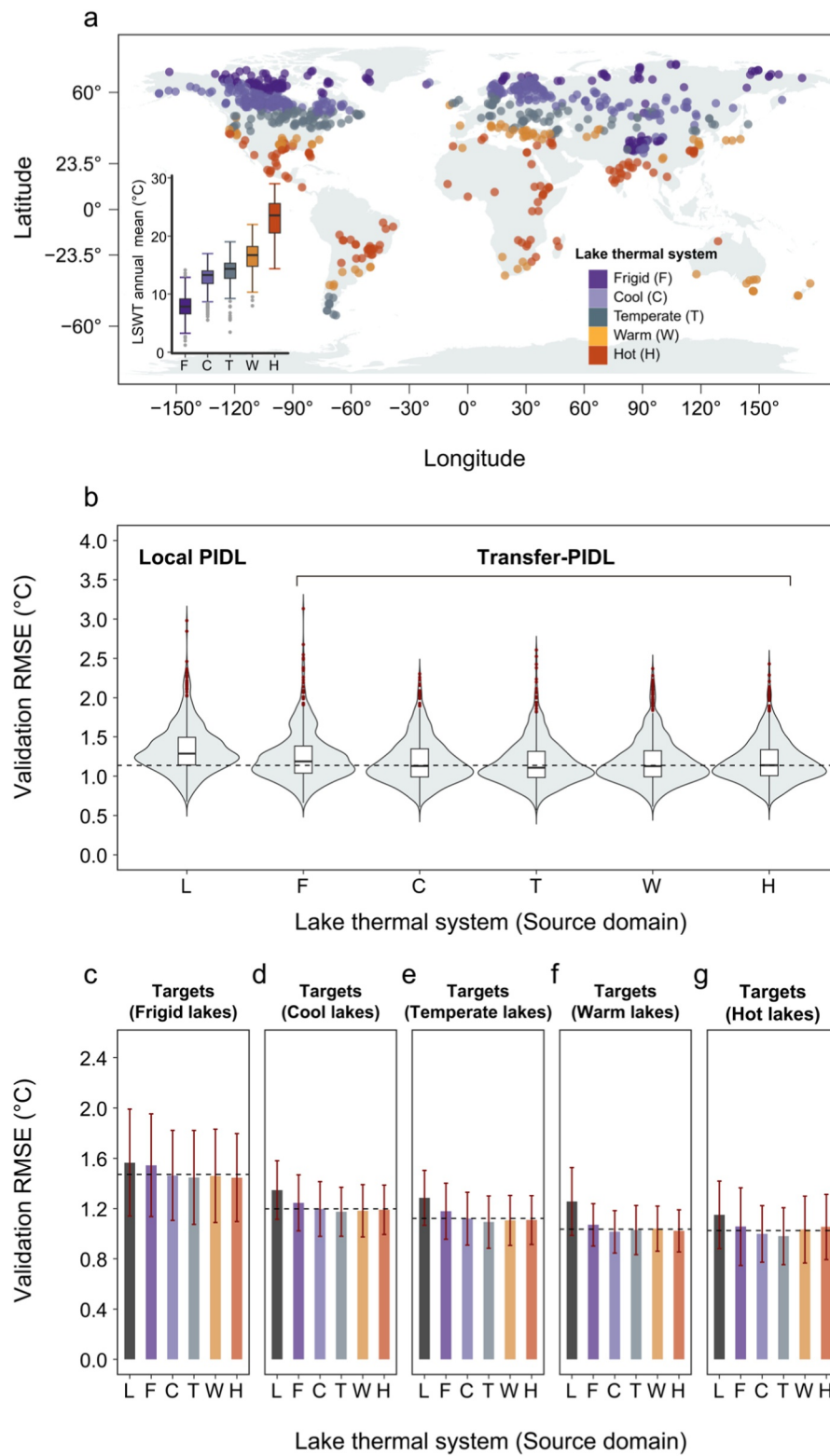


Figure 7.

increase (e.g., pre-training on a hot lakes while fine-tuning on frigid lakes, see “H” in Figure 7c). Nevertheless, Transfer-PIDL outperformed local PIDL models in each transfer scenario, as shown in Figures 7b–7g and Figures S8 and S9b in Supporting Information S1.

The most accurate global-scale application was achieved by the Transfer-PIDL pre-trained on the temperate lakes (see “T” in Figure 7b), which yielded the lowest lake-specific RMSEs, with a median of 1.11°C. This high performance remained consistent in cross-domain applications, where the mean validation RMSE fluctuated slightly among each thermal system, ranging from the minimum of 0.98°C on hot lakes (see “T” in Figure 7g) to the maximum of 1.45°C on frigid lakes (see “T” in Figure 7c). Furthermore, when focusing on temperate lakes as the targets, all source models exhibited comparable performance after localization (Figure 7e). The mean validation RMSE differed by less than 0.1°C, varying only between 1.11°C (warm source) and 1.18°C (frigid source). This accuracy was higher than that of the PIDL models trained solely on local information, which had a mean validation RMSE of 1.29°C (see “L” in Figure 7e).

When pre-trained on frigid or cool lakes, Transfer-PIDL exhibited slightly lower generalization performance (see “F” and “C” in Figure 7b). The frigid source produced the poorest-performing Transfer-PIDL in global-scale applications (see “F” in Figure 7b), as indicated by the largest median RMSE of 1.19°C with the widest range, spanning from 0.85°C (5% quantile) to 1.86°C (95% quantile). Yet, the poorest-performing Transfer-PIDL still surpassed the local PIDL models, showing a lower median validation RMSE (1.19°C vs. 1.29°C, see “L” in Figure 7b) and a significant improvement ($p < 0.05$, Figure S9b in Supporting Information S1). It also far outperformed the DL, PB, and ERA5-Land baselines (1.19°C vs. 1.47°C, 1.86°C, and 3.31°C, respectively, see Figures S8 and S9b in Supporting Information S1). Conversely, when using frigid lakes as the targets, Transfer-PIDL achieved poorer cross-domain applications (Figure 7c), particularly when compared its cross-domain performance on hot targets (see Figure 7g).

4. Discussion

4.1. Insights From Coupling General-to-Specific Transfer Learning to PIDL

As expected, leveraging the generalized knowledge gained from large-scale source data, Transfer-PIDL outperformed local PIDL in each experiment (Figures 2–7, Figures S4–S9 in Supporting Information S1). This advantage could be related to pre-training, which has been widely recognized in previous studies as an important factor contributing to the enhanced performance of PIDL models, whose RMSEs in lake temperature prediction often reached around 1.5°C (e.g., Jia et al., 2019; Read et al., 2019; Willard et al., 2021). A typical PIDL pre-training scenario is physics-guided initialization (Willard et al., 2024), which involves the use of PB model simulations (Jia et al., 2019). Given its theory-based numerical nature, a well-parameterized PB model can achieve reliable LSWT simulations without requiring large amounts of calibration data (e.g., Experiment 1; see Figure 4). These accurate simulations help mitigate problematic PIDL predictions, particularly when the final fine-tuning data set is sparse or unrepresentative, for example, only 30 days of training data points in Windermere (see PIDL vs. DL in Figure 4a). Despite this, such pre-training gains can be further amplified by utilizing a large-scale source data set (see Transfer-PIDL in Figure 4). One plausible factor contributing to this predictive enhancement could be the abundance of the source data sets, which may capture a wide variety of lake types and their potential responses to various climate scenarios. Such extensive data sets could feed data-hungry DL components (Pan & Yang, 2010; Shen, 2018), allowing them to act as powerful prediction engines under additional energy-based soft penalties (Figures 2–4). In addition, this comprehensive starting point can facilitate fine-tuning tailored to local lake environments, especially when they share relevant features with the large-scale source data, for example, the LSWT thermal patterns in this study (Figure 7).

Figure 7. Model transferability across lake thermal systems. (a) Distribution of studied lakes by thermal systems (Albers equal-area projection), with lake-specific annual mean LSWTs shown for each thermal system. (b) Overall performance of the Transfer-PIDL model across all lakes. (c–g) Transfer-PIDL model performance specific to each thermal system, indicated by validation RMSE with mean and SD. In (b–g), the local (L) denotes the local PIDL models, while the frigid (F), cool (C), temperate (T), warm (W), and hot (H) represent the thermal system of source data used for Transfer-PIDL pre-training. The horizontal dashed line in (b) represents the median performance of Transfer-PIDL over the five pre-training scenarios. The horizontal dashed lines in (c–g) represent the mean performance of Transfer-PIDL within each thermal system.

Another major benefit of general-to-specific transfer learning, in particular, is its ability to improve the generalizability of PIDL models (see Transfer-PIDL in Figures 5–7). A potential modeling challenge, however, is that PIDL performance may degrade when pre-training relies on inaccurate PB model simulations (Read et al., 2019). These inaccuracies often come from the inherent limitations of PB models, such as flawed parameterization (Read et al., 2014) and oversimplified governing equations (Huang et al., 2019). As seen in Experiments 2 and 3, these limitations, compounded by uncertainties in HydroLAKES-estimated lake characteristics (e.g., conically assumed hypsography for GLM), led to increased variability in PB model performance (Figure 5a; Figure S8 in Supporting Information S1), despite its superior results in Experiment 1 (Figure 4). Recent studies have shown that when a site lacks key parameters for PB models and sufficient data for bias correction, the corresponding PIDL performance often declines (e.g., Jia et al., 2021). Unfortunately, such dual information gaps in site-specific lake modeling remain widespread due to limited monitoring efforts (Piccolroaz et al., 2024; Read et al., 2014; Tong et al., 2023). Nevertheless, data availability increases considerably at broader geographical scales, such as continental or global. Large data sets can support pre-training, often yielding powerful general models (e.g., foundation models or FMs; the general-purpose models trained on diverse data sets, e.g., Mai et al., 2023; Xie et al., 2023). These universal models, in turn, can improve task-specific modeling robustness through effective parameter transfer (Karpatne et al., 2024), for example, multi-stage transfer learning in our study (Figure 1c). This benefit has been shown in many related works. For instance, Padarian et al. (2019) demonstrated that global spectral information was useful to improve the robustness of local soil spectroscopy predictions. Chen et al. (2024) also found that site-level river DL models can be enhanced via pre-training them on a basin-level hydrological data set. Moreover, once calibrated, a universal pre-trained PIDL model can transfer its parameter weights to each downstream local task. Compared with individually applied physics-guided initialization, this approach provides a more computationally efficient alternative (see Table S5 in Supporting Information S1). These findings highlight the efficiency of transfer learning in enhancing the scalability of PIDL modeling, extending its development beyond single-site constraints.

4.2. Trade-Offs Between Data Collection Effort and Transfer Gains

Across all three cases in our first experiment, the Transfer-PIDL exhibited diminishing returns in its final performance as the amount of pre-training data increased (see the mean validation RMSE curves in Figures 2a–2c). This trend is consistent with previous findings in related studies (Chen et al., 2024; Ma et al., 2021; Padarian et al., 2019). To further the transfer performance in hydrological applications, recent studies suggested that using more diverse source data sets may be preferable (Ma et al., 2021; Y. Wang et al., 2022). One potential reason for this could be the information redundancy from highly homogeneous source data set. Redundant information (e.g., homogeneous hydrological, geographic, or climatic characteristics) provides limited feature variability, which could constrain the pre-training model to abstract higher-level data representations (Padarian et al., 2019). Even with large pre-training data sets, insufficient additional variability may still hinder further transfer gains (Chen et al., 2024). By comparison, more diverse data can introduce a wider variety of environmental conditions, which has been suggested to reduce overfitting to the source data and improve the generalization of pre-trained model to unseen target scenarios (e.g., Ma et al., 2021). In this sense, data representativeness can be more influential than its absolute volume when considering a source domain for transfer learning applications.

We also observed reduced local data requirements for the final fine-tuning via transfer learning (Figures 4a–4c). Compared to the local PIDL, Transfer-PIDL typically required fewer fine-tuning data to achieve a comparable validation RMSE, which was several folds lower than the training demand of data-hungry DL (e.g., see Windermere in Figure 4a, to reach a threshold of 1.0°C). When local data were extremely sparse (e.g., Lake Hulunhu; Table S2 in Supporting Information S1), the Transfer-PIDL, driven by continuous climate-forcing inputs, was still able to robustly reconstruct LSWT dynamics (see Lake Hulunhu in Figures 6e and 6f). This reduced reliance on local data is a resultant benefit of large pre-training, as discussed above in Section 4.1. In practice, this efficiency is consistently valuable for hydrological applications, as high-fidelity field measurements for in situ models are often scarce, whereas pre-training data sets are typically diverse. These data sets can originate from remote sensing (Carrea et al., 2023; Korver et al., 2024), as demonstrated in this study, or reliable model simulations (Muñoz-Sabater et al., 2021; Tong et al., 2023) and even data borrowing techniques (Jia et al., 2021). Such a capability of Transfer-PIDL helps extract meaningful insights from limited inputs, showing great potential in bridging the gap between theoretical development and practical implementations.

4.3. Progressive Transfer in Lake Temperature Modeling

As reflected by Experiment 3 (Figure 7), the Transfer-PIDL remained accurate when source and target domains were thermally distinct (mean validation RMSEs below 1.46°C; see Figures 7c–7g and Figure S3a in Supporting Information S1). Limited accuracy losses were shown in cross-domain applications relative to intra-domain cases (Figure S9c in Supporting Information S1). We observed relatively poorer transfer performance during hot-to-frigid domain shifts (Figure 7c), likely due to the asymmetry LSWT dynamics between the two lake thermal systems (Figure S3a in Supporting Information S1). The frigid lakes significantly exhibit lower annual mean LSWTs than the hot regimes ($p < 0.001$; Figure S3a in Supporting Information S1), potentially reducing transfer efficiency due to the large thermal discrepancies. In addition, the frigid lake data set contained significantly fewer samples than the others (all $p < 0.001$; see Figure S3b in Supporting Information S1), suggesting that data sparsity may also contribute to its poorer performance both as sources (see “F” in Figure 7b) and as targets (Figure 7c), as already demonstrated in Experiment 1. Nevertheless, even in this challenging scenario, Transfer-PIDL maintained a low validation RMSE, at $1.45 \pm 0.35^\circ\text{C}$ (see “H” in Figure 7c), outperforming local PIDL models (see “L” in Figure 7c and Figure S9c in Supporting Information S1). The robust accuracy observed may be linked to the progressive transfer strategy. The large-scale pre-trained PIDL can capture coherent lake surface thermodynamics over broad scales, however, these patterns can be strongly influenced by local factors, such as wind and river inflows (Markfort et al., 2010; Read et al., 2014). Accordingly, site-specific re-training aligns the pre-trained PIDL with local lake environments, supporting its effective parameter transfer (Figures 7c–7g). Our findings are consistent with those of related studies. For instance, Samala et al. (2019) and Gu et al. (2020) found that two-step transfer can resolve the domain-shift challenges in medical imaging, while Mendes et al. (2020) and He et al. (2020) suggested that domain adaptation often achieved positive outcomes in transfer learning tasks. As in lake modeling, cross-domain generalization (particularly modeling between regions with differing data collection protocols) remains a long-standing challenge (Karpatne et al., 2024; Piccolroaz et al., 2024; Willard et al., 2024). Consistent evidence from cross-disciplinary studies provides methodological insights that indicate strong potential for future hydrological applications.

In addition to the cross-domain applications, the Transfer-PIDL model also demonstrated reliable overall performance in the global-scale application (Figure 7b), indicating its potential to address large-scale hydrological modeling challenges. Due to the limited availability of in situ measurements, global and regional hydrological models typically rely on satellite-derived observations (Piccolroaz et al., 2024). Such data have enabled numerous global-scale studies, particularly in lake temperature research, including assessments of climate change impacts on lake extreme thermal events (Woolway et al., 2021), the development of lake thermal regions (Maberly et al., 2020), and improvements in the understanding of global lake warming (Tong et al., 2023). However, the spatial and temporal coverage of satellite lake-temperature observations remains incomplete. This limitation is particularly pronounced in regions with frequent cloud cover (e.g., the tropics) or seasonal ice (e.g., high-latitude areas), as demonstrated in previous studies (e.g., Piccolroaz et al., 2024; Tong et al., 2023). In our study, the frigid lake (mostly located in northern high-latitude regions) and hot lake (mostly tropical) data sets contained significantly fewer CCI samples than the others (all $p < 0.001$; see Figure S3b in Supporting Information S1), again highlighting this limitation. By transferring the patterns learned from data-rich domains (e.g., temperate lakes, see Figure S3b in Supporting Information S1), Transfer-PIDL could help mitigate model bias in these data-sparse regions, as evidenced by its improved performance compared with local PIDL models (see Figures 7d and 7h). The global-scale performance of Transfer-PIDL indicates its potential to improve the comprehensive understanding of lake response to climate changes, such as supporting future studies on extreme thermal events in lake surface waters (e.g., Woolway et al., 2020).

4.4. Limitations and Perspectives

This study also highlights the current limitations and potential directions for future research on PIDL with transfer learning. From an architectural perspective, the present PIDL framework focuses on employing LSTM networks for lake temperature modeling, alternative architectures such as gated recurrent units (GRUs) and Transformers also merit investigation. In addition, the energy-based loss term is derived from a simplified lake bulk model, representing only partial physical knowledge. Future studies could incorporate more comprehensive thermodynamic processes, for example, snow and ice dynamics, through one-dimensional energy models to better enforce physical constraints. In transfer learning, although this study observed marginal benefits by increasing pre-training data (see Figure 2), further investigation is needed to better understand the roles of data

diversity and quantity. To more effectively measure the discrepancy between domains, other widely used metrics such as correlation alignment (CORAL; see B. Sun et al., 2016), also deserve exploration and comparison with the MMD. Advanced techniques in deep domain adaptation, for example, adversarial domain adaptation (Farahani et al., 2020), may improve transfer performance across domains that differ largely (e.g., hot–frigid regimes; Figure 7). From a data perspective, given the inherent constraints of satellite retrievals (e.g., due to weather conditions), emerging remote sensing technologies, such as drone-based observations, may help complement the representativeness of satellite data sets (e.g., in cloudy regions) and further enhance transfer performance. Moreover, incorporating uncertainty quantification (e.g., through ensemble or Bayesian approaches) would be a valuable extension. This could enhance model interpretability, for instance, by revealing sensitivities to additional meteorological variables such as humidity or precipitation. Finally, given improved predictions for lake surface temperatures, this approach holds potential for tackling more challenging tasks, such as sub-surface temperatures or stratification modeling. This could provide more practical insights into the lake responses to climate change.

5. Conclusions

In this study, the concept of multi-stage transfer learning was integrated into state-of-the-art physics-informed deep learning (PIDL) for predicting lake surface water temperature (LSWT). The proposed Transfer-PIDL included three training stages, that is, large-scale pre-training, physics-guided re-training, and local fine-tuning. Experimental results demonstrated that, when pre-training data sets were sufficient (e.g., 40 source lakes), Transfer-PIDL achieved improved performance relative to local PIDL, DL and physics-based (PB) approaches. To achieve comparable performance, it was associated with reduced site-specific data requirements for fine-tuning compared with local models. In addition, Transfer-PIDL demonstrated cross-thermal-system generalizability, showing applicability for large-scale LSWT prediction. These strengths may help advance understanding of lake responses to climate change and improve modeling for data-sparse lake systems.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The ESA-CCI lake data used in this study is available at <https://catalogue.ceda.ac.uk/uuid/7fc9df8070d34caca8092e45ef276f1/>. For in situ lake data, see Supporting Information S1 for extended data access. The ERA5-Land data sets can be accessed at <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land>. The HydroLAKES data set is provided at <https://www.hydrosheds.org/pages/hydrolakes>. The GLM source code is available at <https://github.com/GLEON/GLM3r>. The source code for the developed Transfer-PIDL is publicly available on GitHub at <https://github.com/louisl101/Transfer-PIDL>.

References

- Benson, B. J., Lenters, J. D., Magnuson, J. J., Stubbs, M., Kratz, T. K., Dillon, P. J., et al. (2000). Regional coherence of climatic and lake thermal variables of four lake districts in the Upper Great Lakes Region of North America. *Freshwater Biology*, 43(3), 517–527. <https://doi.org/10.1046/j.1365-2427.2000.00572.x>
- Bruce, L. C., Frassl, M. A., Arhonditsis, G. B., Gal, G., Hamilton, D. P., Hanson, P. C., et al. (2018). A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network. *Environmental Modelling & Software*, 102, 274–291. <https://doi.org/10.1016/j.envsoft.2017.11.016>
- Carrea, L., Crétau, J.-F., Liu, X., Wu, Y., Calmettes, B., Duguay, C. R., et al. (2023). Satellite-derived multivariate world-wide lake physical variable timeseries for climate studies. *Scientific Data*, 10(1), 30. <https://doi.org/10.1038/s41597-022-01889-z>
- Chen, S., Huang, J., Wang, P., Tang, X., & Zhang, Z. (2024). A coupled model to improve river water quality prediction towards addressing non-stationarity and data limitation. *Water Research*, 248, 120895. <https://doi.org/10.1016/j.watres.2023.120895>
- Duan, Z., Gao, W., Cheng, G., Zhang, Y., & Chang, X. (2024). Warming surface and lake heatwaves as key drivers to harmful algal blooms: A case study of Lake Dianchi, China. *Journal of Hydrology*, 632, 130971. <https://doi.org/10.1016/j.jhydrol.2024.130971>
- Edinger, J. E., Duttweiler, D. W., & Geyer, J. C. (1968). The response of water temperatures to meteorological conditions. *Water Resources Research*, 4(5), 1137–1143. <https://doi.org/10.1029/wr004i005p01137>
- Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. (2020). A brief review of domain adaptation. *arXiv*. <https://doi.org/10.48550/arXiv.2010.03978>
- Feldbauer, J., Mesman, J. P., Andersen, T. K., & Ladwig, R. (2025). Learning from a large-scale calibration effort of multiple lake temperature models. *Hydrology and Earth System Sciences*, 29(4), 1183–1199. <https://doi.org/10.5194/hess-29-1183-2025>
- Fleischmann, A., Papa, F., Hamilton, S., Melack, J., Forsberg, B., Val, A., et al. (2025). Extreme warming of Amazon waters in a changing climate. *Science*, 390(6773), 606–611. <https://doi.org/10.1126/science.adr4029>

Acknowledgments

The authors acknowledge support from the Imperial College London–China Scholarship Council funding program under Grant 202306320083. RIW was supported by a UKRI Natural Environment Research Council (NERC) Independent Research Fellowship [NE/T011246/1]. YT was supported by the National Natural Science Foundation of China (42501445). We appreciate the data sharing from the UK Centre for Ecology & Hydrology for Windermere (<https://eidc.ac.uk>), the Marine Institute in Ireland for Lough Feeagh (<https://data.gov.ie/organization/marine-institute/>), the National Institute for Environmental Studies in Japan for Lake Kasumigaura (<https://db.cger.nies.go.jp>). For additional lake data sharing, we equally appreciate the resources provided by the Environmental Data Initiative (<https://portal.edirepository.org/>), the Norwegian Water Resources and Energy Directorate (<https://sildre.nve.no/map>), the National Data Buoy Center (<https://www.ndbc.noaa.gov>), the North Temperate Lakes Long-Term Ecological Research Network (<https://lter.limnology.wisc.edu/>), the King County Water and Land Resource Division Hydrologic Monitoring Program (<https://green2.kingcounty.gov/hydrology>), the National Ecological Observatory Network (<https://data.neonscience.org>), the Swedish Infrastructure for Ecosystem Science (<https://www.fielddata.se/>), the Long-Term Ecological Research Austria (<https://deims.org/searh/sites/lter>), and the Environmental Data Portal (<https://envdata.boprc.govt.nz/>). We would like to thank all individuals who participated in the aforementioned projects, whose dedicated efforts in collecting and measuring such high-quality records are appreciated. Extended data access information is provided in Supporting Information S1.

- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings. Retrieved from <https://proceedings.mlr.press/v9/glorot10a.html>
- Gretton, A., Borgwardt, K. M., Rasch, M. J., & Sch, B. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13, 723–773.
- Gu, Y., Ge, Z., Bonnington, C. P., & Zhou, J. (2020). Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. *IEEE Journal of Biomedical and Health Informatics*, 24(5), 1379–1393. <https://doi.org/10.1109/JBHI.2019.2942429>
- He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J. R., & Parikh, N. A. (2020). A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Scientific Reports*, 10(1), 15072. <https://doi.org/10.1038/s41598-020-71914-x>
- Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., et al. (2019). A general lake model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geoscientific Model Development*, 12(1), 473–523. <https://doi.org/10.5194/gmd-12-473-2019>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. (2021). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5149–5169. <https://doi.org/10.1109/TPAMI.2021.3079209>
- Huang, A., Lazhu, Wang, J., Dai, Y., Yang, K., Wei, N., et al. (2019). Evaluating and improving the performance of three 1-D lake models in a large deep lake of the Central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 124(6), 3143–3167. <https://doi.org/10.1029/2018JD029610>
- Huisman, M., van Rijn, J. N., & Plaat, A. (2020). A survey of deep meta-learning. *Artificial Intelligence Review*, 54(6), 4483–4541. <https://doi.org/10.1007/s10462-021-10004-4>
- Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, 11(2), 40. <https://doi.org/10.3390/technologies11020040>
- Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. In *Proceedings of the 2019 SIAM International Conference on Data Mining* (pp. 558–566). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611975673.63>
- Jia, X., Willard, J., Karpatne, A., Read, J. S., Zwart, J. A., Steinbach, M., & Kumar, V. (2021). Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM/IMS Transactions on Data Science*, 2(3), 1–26. <https://doi.org/10.1145/3447814>
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6), 422–440. <https://doi.org/10.1038/s42254-021-00314-5>
- Karpatne, A., Jia, X., & Kumar, V. (2024). Knowledge-guided machine learning: Current trends and future prospects. *arXiv*. <https://doi.org/10.48550/arXiv.2403.15989>
- Korver, M. C., Lehner, B., Cardille, J. A., & Carrea, L. (2024). Surface water temperature observations and ice phenology estimations for 1.4 million lakes globally. *Remote Sensing of Environment*, 308(April), 114164. <https://doi.org/10.1016/j.rse.2024.114164>
- Lee, J. D., Simchowitz, M., Jordan, M. I., & Recht, B. (2016). Gradient descent only converges to minimizers. In *Conference on Learning Theory* (pp. 1246–1257). PMLR. Retrieved from <https://proceedings.mlr.press/v49/lee16.html>
- Lenters, J. D., Kratz, T. K., & Bowser, C. J. (2005). Effects of climate variability on lake evaporation: Results from a long-term energy budget study of Sparkling Lake, northern Wisconsin (USA). *Journal of Hydrology*, 308(1–4), 168–195. <https://doi.org/10.1016/j.jhydrol.2004.10.028>
- Li, Q., Wang, Z., Shangguan, W., Li, L., Yao, Y., & Yu, F. (2021). Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning. *Journal of Hydrology*, 600, 126698. <https://doi.org/10.1016/j.jhydrol.2021.126698>
- Li, Y., Swersky, K., & Zemel, R. (2015). Generative moment matching networks. *arXiv*. <https://doi.org/10.48550/arXiv.1502.02761>
- Liu, M., He, J., Huang, Y., Tang, T., Hu, J., & Xiao, X. (2022). Algal bloom forecasting with time-frequency analysis: A hybrid deep learning approach. *Water Research*, 219, 118591. <https://doi.org/10.1016/j.watres.2022.118591>
- Liu, M., Huang, Y., Hu, J., He, J., & Xiao, X. (2023). Algal community structure prediction by machine learning. *Environmental Science & Ecotechnology*, 14, 100233. <https://doi.org/10.1016/j.ese.2022.100233>
- Ma, K., Feng, D., Lawson, K., Tsai, W., Liang, C., Huang, X., et al. (2021). Transferring hydrologic data across continents—Leveraging data-rich regions to improve hydrologic prediction in data-sparse regions. *Water Resources Research*, 57(5), 1–11. <https://doi.org/10.1029/2020WR028600>
- Maberly, S. C., O'Donnell, R. A., Woolway, R. I., Cutler, M. E. J., Gong, M., Jones, I. D., et al. (2020). Global lake thermal regions shift under climate change. *Nature Communications*, 11(1), 1232. <https://doi.org/10.1038/s41467-020-15108-z>
- Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., et al. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv*. <https://doi.org/10.48550/arXiv.2304.06798>
- Markfort, C. D., Perez, A. L. S., Thill, J. W., Jaster, D. A., Porté-Agel, F., & Stefan, H. G. (2010). Wind sheltering of a lake by a tree canopy or bluff topography. *Water Resources Research*, 46(3), 1–13. <https://doi.org/10.1029/2009WR007759>
- Mendes, A., Julian, T., & Dos Santos Coelho, L. (2020). Multi-stage transfer learning with an application to selection process. In *Frontiers in artificial intelligence and applications*. IOS Press. <https://doi.org/10.3233/FAIA200291>
- Merz, E., Saberski, E., Gilarranz, L. J., Isles, P. D. F., Sugihara, G., Berger, C., & Pomati, F. (2023). Disruption of ecological networks in lakes by climate change and nutrient fluctuations. *Nature Climate Change*, 13(4), 389–396. <https://doi.org/10.1038/s41558-023-01615-6>
- Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1), 13603. <https://doi.org/10.1038/ncomms13603>
- Mironov, D. V. (2008). *Parameterization of lakes in numerical weather prediction. Description of a lake model* (p. 41). Offenbach am Main, Germany: Deutscher Wetterdienst. https://doi.org/10.5676/DWD_pub/nwv/cosmo-tr_1
- Moore, T. N., Mesman, J. P., Ladwig, R., Feldbauer, J., Olsson, F., Pilla, R. M., et al. (2021). LakeEnsemblR: An R package that facilitates ensemble modelling of lakes. *Environmental Modelling & Software*, 143, 105101. <https://doi.org/10.1016/j.envsoft.2021.105101>
- Muñoz-Sabater, J. (2019). ERA5-Land hourly data from 1950 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.e2161bac>
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- National Institute for Environmental Studies. (2016). National Institute for Environmental Studies (2016) Lake Kasumigaura Database [Dataset]. Retrieved from <https://db.cger.nies.go.jp/gem/moni-e/inter/GEMS/database/kasumi/index.html>
- O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., et al. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42(24), 1–9. <https://doi.org/10.1002/2015GL066235>

- Padarian, J., Minasny, B., & McBratney, A. B. (2019). Transfer learning to localise a continental soil vis-NIR calibration model. *Geoderma*, 340, 279–288. <https://doi.org/10.1016/j.geoderma.2019.01.009>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Paszke, A., Gross, S., Chintala, S., Chanan, G., & Yang, E. (2017). Automatic differentiation in PyTorch. In *31st Conference on Neural Information Processing Systems*. Long Beach, CA, USA. <https://doi.org/10.1145/3434309>
- Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, 11(1), 3357. <https://doi.org/10.1038/s41467-020-16970-7>
- Piccolroaz, S., Zhu, S., Ladwig, R., Carrea, L., Oliver, S., Piotrowski, A. P., et al. (2024). Lake water temperature modeling in an era of climate change: Data sources, models, and future prospects. *Reviews of Geophysics*, 62(1), e2023RG000816. <https://doi.org/10.1029/2023RG000816>
- Rasconi, S., Winter, K., & Kainz, M. J. (2017). Temperature increase and fluctuation induce phytoplankton biodiversity loss—Evidence from a multi-seasonal mesocosm experiment. *Ecology and Evolution*, 7(9), 2936–2946. <https://doi.org/10.1002/ece3.2889>
- R Core Team. (2013). *R: A language and environment for statistical computing*. Foundation for Statistical Computing, Vienna, Austria.
- Read, J. S., Winslow, L. A., Hansen, G. J. A., Van Den Hoek, J., Hanson, P. C., Bruce, L. C., & Markfort, C. D. (2014). Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecological Modelling*, 291, 142–150. <https://doi.org/10.1016/j.ecolmodel.2014.07.029>
- Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, 55(11), 9173–9190. <https://doi.org/10.1029/2019WR024922>
- Rogers, C. K., Lawrence, G. A., & Hamblin, P. F. (1995). Observations and numerical simulation of a shallow ice-covered midlatitude lake. *Limnology & Oceanography*, 40(2), 374–385. <https://doi.org/10.4319/lo.1995.40.2.0374>
- Rouso, B. Z., Bertone, E., Stewart, R., & Hamilton, D. P. (2020). A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*, 182, 115959. <https://doi.org/10.1016/j.watres.2020.115959>
- Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Richter, C. D., & Cha, K. H. (2019). Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Transactions on Medical Imaging*, 38(3), 686–696. <https://doi.org/10.1109/TMI.2018.2870343>
- Schmid, M., Hunziker, S., & Wüest, A. (2014). Lake surface temperatures in a changing climate: A global sensitivity analysis. *Climatic Change*, 124(1–2), 301–315. <https://doi.org/10.1007/s10584-014-1087-2>
- Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(11), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Sun, B., Feng, J., & Saenko, K. (2016). Correlation alignment for unsupervised domain adaptation. *arXiv*. <https://doi.org/10.48550/arXiv.1612.01939>
- Tayal, K., Jia, X., Ghosh, R., Willard, J., Read, J., & Kumar, V. (2022). Invertibility aware integration of static and time-series data: An application to lake temperature modeling. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)* (pp. 702–710). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611977172.79>
- Till, A., Rypel, A. L., Bray, A., & Fey, S. B. (2019). Fish die-offs are concurrent with thermal extremes in north temperate lakes. *Nature Climate Change*, 9(8), 637–641. <https://doi.org/10.1038/s41558-019-0520-y>
- Tong, Y., Feng, L., Wang, X., Pi, X., Xu, W., & Woolway, R. I. (2023). Global lakes are warming slower than surface air temperature due to accelerated evaporation. *Nature Water*, 1(11), 929–940. <https://doi.org/10.1038/s44221-023-00148-8>
- Tye, S. P., Siepielski, A. M., Bray, A., Rypel, A. L., Phelps, N. B. D., & Fey, S. B. (2022). Climate warming amplifies the frequency of fish mass mortality events across north temperate lakes. *Limnology and Oceanography Letters*, 7(6), 510–519. <https://doi.org/10.1002/lol2.10274>
- Wanders, N., van Vliet, M. T. H., Wada, Y., Bierkens, M. F. P., & van Beek, L. P. H. (2019). High-resolution global water temperature modeling. *Water Resources Research*, 55(4), 2760–2778. <https://doi.org/10.1029/2018WR023250>
- Wang, S., Teng, Y., & Perdikaris, P. (2021). Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5), A3055–A3081. <https://doi.org/10.1137/20M1318043>
- Wang, W., Shi, K., Wang, X., Zhang, Y., Qin, B., Zhang, Y., & Woolway, R. I. (2024). The impact of extreme heat on lake warming in China. *Nature Communications*, 15(1), 70. <https://doi.org/10.1038/s41467-023-44404-7>
- Wang, Y., Gupta, H. V., Zeng, X., & Niu, G. (2022). Exploring the potential of long short-term memory networks for improving understanding of continental- and regional-scale snowpack dynamics. *Water Resources Research*, 58(3), e2021WR031033. <https://doi.org/10.1029/2021WR031033>
- Willard, J. D., Read, J. S., Appling, A. P., Oliver, S. K., Jia, X., & Kumar, V. (2021). Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. *Water Resources Research*, 57(7), e2021WR029579. <https://doi.org/10.1029/2021WR029579>
- Willard, J. D., Read, J. S., Topp, S., Hansen, G. J. A., & Kumar, V. (2022). Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020). *Limnology and Oceanography Letters*, 7(4), 287–301. <https://doi.org/10.1002/lol2.10249>
- Willard, J. D., Varadharajan, C., Jia, X., & Kumar, V. (2024). Time series predictions in unmonitored sites: A survey of machine learning techniques in water resources. *arXiv Preprint*, 1–39.
- Woolway, R. I., Albergel, C., Frölicher, T. L., & Perroud, M. (2022). Severe lake heatwaves attributable to human-induced global warming. *Geophysical Research Letters*, 49(4), 1–10. <https://doi.org/10.1029/2021GL097031>
- Woolway, R. I., Jennings, E., Shatwell, T., Golub, M., Pierson, D. C., & Maberly, S. C. (2021). Lake heatwaves under climate change. *Nature*, 589(7842), 402–407. <https://doi.org/10.1038/s41586-020-03119-1>
- Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., & Sharma, S. (2020). Global lake responses to climate change. *Nature Reviews Earth & Environment*, 1(8), 388–403. <https://doi.org/10.1038/s43017-020-0067-5>
- Woolway, R. I., & Maberly, S. C. (2020). Climate velocity in inland standing waters. *Nature Climate Change*, 10(12), 1124–1129. <https://doi.org/10.1038/s41558-020-0889-7>
- Xie, Y., Wang, Z., Mai, G., Li, Y., Jia, X., Gao, S., & Wang, S. (2023). Geo-foundation models: Reality, gaps and opportunities. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems* (pp. 1–4). ACM. <https://doi.org/10.1145/3589132.3625616>
- Xu, C., Huang, Y., Xin, R., Wu, N., & Liu, M. (2025). Algal bloom forecasting leveraging signal processing: A novel perspective from ensemble learning. *Water Research*, 283, 123800. <https://doi.org/10.1016/j.watres.2025.123800>
- Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., et al. (2021). Machine learning: New ideas and tools in environmental science and engineering. *Environmental Science & Technology*, 55(19), 12741–12754. <https://doi.org/10.1021/acs.est.1c01339>