# Response Letter

For

Manuscript ID: **egusphere-2025-1983**

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

**Reviewer #3:**

This manuscript presents HyLake v1.0, a hybrid lake–atmosphere model that embeds a Bayesian-optimized bidirectional LSTM surrogate within a process-based 1-D vertical transport framework to simulate lake surface temperature and surface fluxes. The work addresses a key challenge in environmental modeling: integrating data-driven surrogates with physical principles. The extensive validation on Lake Taihu (2012–2015) against FLake demonstrates clear performance gains, and the hybrid approach represents a meaningful methodological advance for lake modeling. While the methodology is sound and the Lake Taihu validation is comprehensive, the authors should more clearly discuss the requirements and limitations for applying this approach to other lake systems. The current multi-site validation within Lake Taihu provides good evidence of transferability, but broader applicability claims should be more cautiously framed.

**Response:** We sincerely thank Reviewer #3 for the constructive comments. In revision, we particularly discussed the requirements and limitations for HyLake v1.0 and presented an example using another morphologically distinct lake to show its transferability. All comments are accepted and **Relisted in black**, followed by our **Replies in blue** and **Revisions in red (highlighted revisions in bold)**. Before point-by-point response, we summarized major revisions followed by Reviewer #3's comments as:

| No. | Major Revisions | Important Messages |
|---|---|---|
| 1 | Presented a test case for applying HyLake v1.0 to another morphologically distinct lake. | The revised manuscript employed HyLake v1.0 to simulate lake-atmosphere interactions in another morphologically distinct lake, Lake Chaohu, and discussed the potential challenges in model application (**Materials and methodology; discussion**). |
| 2 | Discussed the limitations of deep-learning-based surrogates. | We discussed the cons and pros of computational requirements, BO-BLSTM-based surrogate, and the choice of lake surface temperature module (**Discussion**). |
| 3 | Provided future directions to improve HyLake v1.0. | We mainly discussed the uncertainty of Bayesian algorithms. Future improvements should focus on development of surrogates by using novel techniques. The employment of a Bayesian fully connected layer in surrogates could also provide probabilistic predictions by quantifying uncertainties in the future (**Discussion**). |

**Specific Comments**

The multi-site validation within Lake Taihu is convincing but add discussion of what adaptations would be needed for different lake types (e.g., deeper lakes, different climate zones, varying trophic states). Consider outlining a framework for applying the methodology to new lake systems.

**Response:** Good point! We agree that applying HyLake v1.0 to other lakes is essential. Therefore, we utilized it to another lake in the middle and lower reaches of the Yangtze River Plain—Lake Chaohu and discussed potential limitations for model application.

**(1) Applying HyLake v1.0 to another lake:** Lake Chaohu is the 5$^{th}$-largest shallow freshwater lake in China, with a deeper lake depth of 3.06 m and smaller lake area of 760 km$^2$ than Lake Taihu (Jiao et al., 2018), which has experienced heavy eutrophication and harmful algal blooms (Yang et al., 2020). Given the difficulty that Lake Chaohu does not have sufficient observations, unlike Taihu, we outlined a framework that utilized ERA5 datasets to force HyLake v1.0 and the MOD11A1 land surface temperature dataset for validating lake surface temperature changes. The results indicated that HyLake v1.0 performed well in Lake Chaohu, with an $R^2$ of 0.97, RMSE of 2.07 °C, and MAE of 1.57 °C, outperforming FLake compared to the MOD11A1 datasets (Figure S7-9). The successful attempt of HyLake v1.0 in Lake Chaohu demonstrated that HyLake v1.0 is promising to apply in ungauged lakes. The associated revisions can be found in Materials and methodology (Section 2.3.1, Lines 286-289, Lines 308-314) and Discussion (Section 4.1, Lines 560-567, Lines 609-611; Figure S7-S9).

**(2) Discussing potential challenges for model application:** Although HyLake v1.0 succeeded in estimating lake-atmosphere interactions in Lake Chaohu, it still has several limitations. Considering the diverse lake types worldwide, it remains challenging to validate the performance of HyLake v1.0 in every case due to the limited observations and simplified physical principles. The quantitative restriction on observations hampers our ability to improve the model's performance in regional cases by retraining or fine-tuning the LSTM-based surrogates for each lake type. Additionally, the inaccurate relationships between lake surface conditions (e.g. friction velocity, surface roughness length) and climate change pose a challenge to HyLake v1.0. Specifically, we found that there are biases in the surface roughness length ($z_0$) and friction velocity ($u^*$) between observations and predictions (Figure S6). These potential differences were hard to quantify due to data scarcity in the current process-based models, which impeded us to improve the understanding of lake-atmosphere interactions. Therefore, the physical principles between lake surface conditions and climate change should be focused in the future using novel process-based or data-driven techniques. The associated revisions are listed in Discussion (Section 4.2, Lines 647-654).

To sum up, HyLake v1.0 provided a novel method for improving the understanding of lake-atmosphere interactions on most lakes. However, the current limitations of data and physical principles restrict the generalization ability for all unknown lake types. We aim to expand the modules and functions of HyLake v1.0 and validate it in additional lakes in the future, to accurately predict lake-atmosphere interactions for a broader range of lake types.

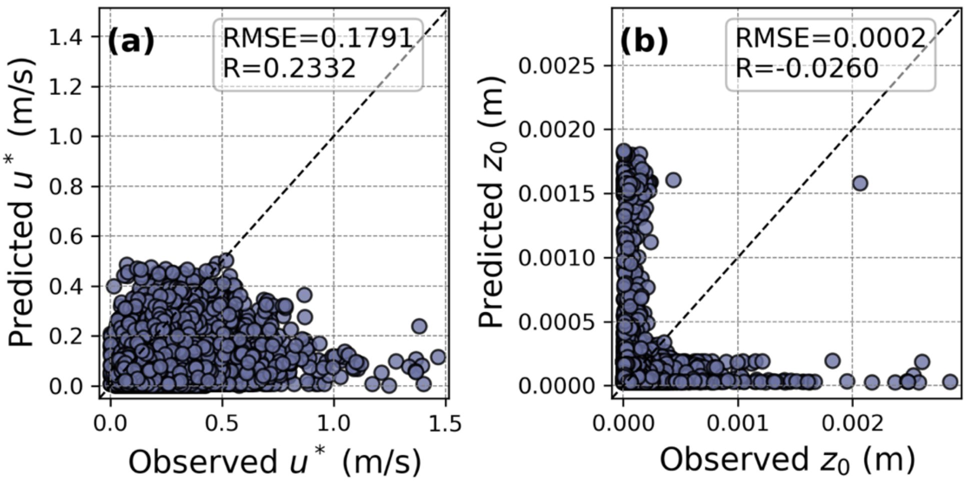



**Figure S6:** The comparison of friction velocity (u*) and surface roughness length ($z_{0m}$, m) in MLW lake site between simulation derived from PBBM and HyLake v1.0 and observations.

**References:**


Jiao, Y., Yang, C., He, W., Liu, W. X., and Xu, F. L.: The spatial distribution of phosphorus and their correlations in surface sediments and pore water in Lake Chaohu, China, *Environ. Sci. Pollut. Res.*, 25, 25906-25915, https://doi.org/10.1007/s11356-018-2606-x, 2018.

Yang, C., Yang, P., Geng, J., Yin, H., and Chen, K.: Sediment internal nutrient loading in the most polluted area of a shallow eutrophic lake (Lake Chaohu, China) and its contribution to lake eutrophication, *Environ. Pollut.*, 262, 114292, https://doi.org/10.1016/j.envpol.2020.114292, 2020.

**Revision:**

"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 **conducted** three numerical experiments**, including MLW experiment, Taihu-obs experiment, Taihu-ERA5 experiment, and Chaohu experiment,** using distinct **models** and forcing datasets (Table **2 and 3**), including FLake, **Baseline, and TaihuScene to intercompare.**" (Section 2.3.1, Lines 286-289)

"Furthermore, this study implemented the HyLake v1.0 into Lake Chaohu, the 5$^{th}$-largest shallow freshwater lake in China, which has experienced heavy eutrophication and harmful algal blooms (Yang et al., 2020), to assess its transferability to other lakes. A LST dataset in Lake Chaohu was obtained from MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061 imageries (MOD11A1, https://www.earthdata.nasa.gov/data/catalog/lpcloud-mod11a1-061), which were used to validate the performance of LST derived from HyLake v1.0. The computational efficiency for each 1-time prediction was recorded using a 16G 10-Core Apple M4 processor based on the established HyLake v1.0 model in this study. The training of the above-mentioned surrogates was run using a 24G NVIDIA GeForce RTX 4090 GPU." (Section 2.3.1, Lines 308-314)

"**Table 3: Intercomparison of model performance across different experiments conducted in diverse regions with different forcing datasets. Observations from all lake sites (MLW, DPK, BFG, XLS, and PTS) on Lake Taihu, were used to drive models in the Taihu-obs experiment. Bold values in the table present the best-performing model with each group of experiments. Computational efficiency is reported as the runtime for a single simulation.**

| **Exp** | **Model** | **Forcing** | **R** | | | **RMSE** | | | **MAE** | | | **Efficiency** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | LST | LE | HE | LST | LE | HE | LST | LE | HE | (s) |
| MLW | PBBM | MLW | 0.98 | 0.85 | 0.89 | 1.78 | 38.34 | 9.37 | 1.38 | 23.54 | 6.22 | 189.49 |
| | FLake | MLW | 0.98 | 0.82 | 0.84 | 1.76 | 42.73 | 7.24 | 1.35 | 24.76 | 5.01 | **16.40** |
| | Baseline | MLW | 0.96 | 0.74 | 0.75 | 2.71 | 51.77 | 14.63 | 2.11 | 33.52 | 9.30 | 151.46 |
| | **HyLake v1.0** | MLW | **0.99** | **0.94** | **0.93** | **1.08** | **24.65** | **7.15** | **0.85** | **15.18** | **4.73** | 270.21 |
| Taihu-obs | FLake | All sites | 0.97 | 0.61 | 0.74 | 2.24 | 15.46 | 69.11 | 1.69 | 41.95 | 10.44 | **89.00** |
| | TaihuScene | All sites | **0.99** | **0.82** | 0.89 | 1.52 | 14.93 | 43.49 | 1.23 | 29.53 | 10.63 | 6928.44 |
| | **HyLake v1.0** | All sites | **0.99** | 0.81 | **0.90** | **1.36** | **11.19** | **39.20** | **1.03** | **24.79** | **7.88** | 2693.23 |
| Taihu-ERA5 | FLake | ERA5 | 0.98 | 0.63 | 0.69 | 1.82 | 12.31 | 67.24 | 1.46 | 50.94 | 9.68 | **19.60** |
| | TaihuScene | ERA5 | **0.99** | 0.68 | 0.73 | 1.60 | 13.00 | 64.83 | 1.29 | 47.78 | 10.11 | 652.25 |
| | **HyLake v1.0** | ERA5 | **0.99** | **0.71** | **0.78** | **1.12** | **11.05** | **49.48** | **0.90** | **35.02** | **7.97** | 236.78 |
| Chaohu | FLake | ERA5 | **0.97** | \ | \ | 2.28 | \ | \ | 1.76 | \ | \ | **70.40** |
| | **HyLake v1.0** | ERA5 | **0.97** | \ | \ | **2.07** | \ | \ | **1.57** | \ | \ | 972.83 |

" (Table 3)

"To address issues related to model performance, generalization, and transferability in ungauged locations, three additional numerical experiments, including FLake, Baseline, and TaihuScene, were proposed for **intercomparison and a framework for applying HyLake v1.0 to another lake, such as Lake Chaohu, with a deeper depth of 3.06 m and area of 760 km$^2$ (Figure S7, Jiao et al., 2018), to validate the potential capacity of transferability further.** These experiments were compared using **observed meteorological datasets, and ERA5 datasets** and **then** validated for both spatial and temporal patterns at Lake Taihu **and Lake Chaohu (Tables 2-3). Similarly, ERA5 dataset-derived HyLake v1.0 outperformed FLake in estimating LST (R = 0.97, RMSE = 2.07 °C, MAE = 1.57 °C) in Lake Chaohu, compared to MOD11A1 datasets (Table 3 and Figures S7-9).**" (Section 4.1, Lines 560-567)

"HyLake v1.0, developed based on *in situ* observations from Lake Taihu, has been proven to be reliable and rigorously validated in Lake Chaohu (Table 3), demonstrating a faster and more accurate framework for enhancing the understanding of hybrid hydrological modeling." (Section 4.1, Lines 609-611)

"**HyLake v1.0 has been applied to Lake Chaohu and achieved superior performance in comparison to the MYD11A1 LST observations, showing a promising way for more applications.** 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). **However, in some exceptional cases, the lake may be influenced by regional inflows/outflows, or it may be covered by snow/ice for a long period, and the processes at the lake-air interface may differ from those in our experiments (Woolway et al., 2020). As a result, our model may not be quantifiable for these situations. Its surrogate will be required for more high-quality local datasets to retrain or finetune.**" (Section 4.2, Lines 647-654)

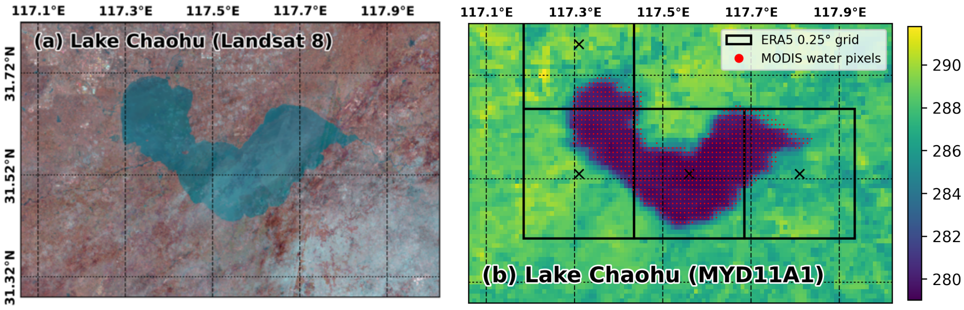
"



**Figure S7: The locations of Lake Chaohu overlaid on a true-color image from (a) Landsat 8 and daily land surface temperature from (b) MYD11A1 product.**" (Figure S7 in Supplementary materials)

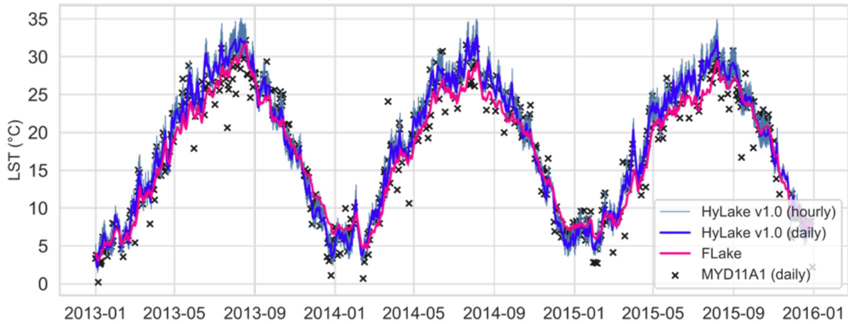



**Figure S8: Time series of daily grid-average LST on Lake Chaohu derived from MYD11A1, FLake simulation, and HyLake v1.0 from 2013 to 2015. HyLake v1.0 provides daily and hourly simulations.**

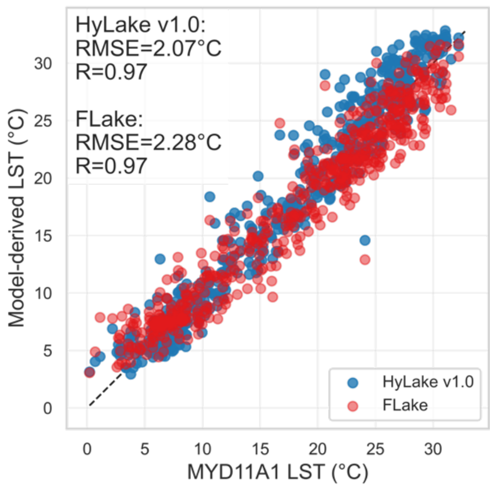



**Figure S9: The intercomparison of daily LST between model simulations (FLake and HyLake v1.0) and MYD11A1 observations on Lake Chaohu from 2013 to 2015.**" (Figures S7-S9 in Supplementary Material)

Better justify the choice of BO-BLSTM over simpler alternatives. provide clearer explanation of why Bayesian optimization and bidirectional LSTM architecture were chosen over deterministic alternatives.

**Response:** Thanks for the suggestions. Using the LSTM-based surrogate with the best group of hyperparameters based on Bayesian Optimization (BO), integrated the abilities of LSTM for time series forecasting and the high computational efficiency of Bayesian Optimization (BO) to represent the physical principles of lake surface temperature changes significantly.

**(1) The selection of BO-BLSTM over simpler alternatives:** LSTM is one of Recurrent Neural Networks (RNNs) that learn from past data by using several gates in their network architecture to remember the past data (Siami-Namini et al., 2019). It becomes feasible for long-term time series forecasting due to the ability to learn many-step dependencies and handle variable-length input sequences in fields such as hydrology (Liu et al., 2024). It outperformed traditional, process-based, and machine learning models in many cases, including predictions of soil moisture, streamflow, water temperature, and groundwater levels (Mao et al., 2021; Feng et al., 2020; Papacharalampous et al., 2018). Meanwhile, previous studies have shown that LSTM-based models outperform other traditional deep-learning models in auto-regressive predictions, supporting this study in predicting lake surface temperature changes robustly and reliably (Siami-Namini et al., 2019). Bayesian LSTM (BLSTM), an improved version of LSTM, adapts probability-distributed weight parameters, which reduce model overfitting and provide robust predictions in hydrology (Li et al., 2021; Lu et al., 2019). In comparison to these models in BO, we ultimately selected BLSTM-based surrogates to address the challenges in this autoregressive prediction task. Nevertheless, we agree that the surrogate should be improved due to its lower computational efficiency, which will be discussed in the future. Here we explained the advantages of LSTM and BLSTM in Materials and Methodology (Section 2.2.2, Lines 209-214) and discussed the limitations and potential improvements in Discussion (Section 4.1, Lines 614-624; Section 4.2, Lines 671-680).

**(2) Using Bayesian Optimization and Bayesian LSTM over deterministic alternatives:** In this study, we selected BO to search for the best group of hyperparameters in Bayesian LSTM (not Bidirectional LSTM) models. BO is a hyperparameter tuning algorithm based on the Bayesian theorem, which can significantly improve the performance and efficiency of deep learning models by building the relationships between model performance and their hyperparameters (Victoria et al., 2021; Wu et al., 2019). Previous studies have established that deep learning models often tune their hyperparameters using manual search or automatic search methods (Wu et al., 2019). Manual search methods depend on expert knowledge and are hard to reproduce and find the optimized hyperparameters. Traditional automatic search methods, such as grid search, train models with each combination of hyperparameters, which is exhaustive searching (Wu et al., 2019; Bergstra et al., 2012). BO adapted a random search technique to fit the data and update the posterior distribution of functions based on Gaussian processes and the Bayesian theorem (Victoria et al., 2021; Wu et al., 2019). Wu et al. (2019) compared the accuracy and costs between BO and grid search methods, finding that both methods performed almost equally well in the same case, while BO runs 12 times faster than grid search.

To summarize, given the large variability and complex relationships of the observations in this study, we would like to employ a more computationally efficient method to help users identify the most robust surrogate within a large hyperparameter space as soon as possible. Considering that the selection of optimization methods is not a focus of this study, the current manuscript provides detailed information about the hyperparameter space for each surrogate to help readers understand. The associated revisions are listed in Materials and Methodology (Section 2.2.3, Lines 276-279).

**References:**

Bergstra, J. and Bengio, Y.: Random search for hyper-parameter optimization, *J. Mach. Learn. Res.*, 13, 281–305, https://dl.acm.org/doi/10.5555/2188385.2188395, 2012.

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.

Feng, D., Fang, K., and Shen, C.: Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales, *Water Resour. Res.*, 56, e2019WR026793, https://doi.org/10.1029/2019WR026793, 2020.

Mao, G., Wang, M., Liu, J., Wang, Z., Wang, K., Meng, Y., et al.: Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall–runoff simulation, *Phys. Chem. Earth, A/B/C*, 123, 103026, https://doi.org/10.1016/j.pce.2021.103026, 2021.

Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D.: One-step ahead forecasting of geophysical processes within a purely statistical framework, *Geosci. Lett.*, 5, 12, https://doi.org/10.1186/s40562-018-0111-1, 2018.

Siami-Namini, S., Tavakoli, N., and Namin, A. S.: The performance of LSTM and BiLSTM in forecasting time series, in: 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019, 3285–3292, https://doi.org/10.1109/BigData47090.2019.9006190, 2019.

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., and Deng, S. H.: Hyperparameter optimization for machine learning models based on Bayesian optimization, *J. Electron. Sci. Technol.*, 17, 26–40, https://doi.org/10.11989/JEST.1674-862X.80904120, 2019.

Victoria, A. H. and Maragatham, G.: Automatic tuning of hyperparameters using Bayesian optimization, *Evol. Syst.*, 12, 217–223, https://doi.org/10.1007/s12530-020-09345-2, 2021.

**Revision:**

"It has been demonstrated that LSTM could capture historical time-step dependencies and handle variable-length input sequences using gradient optimization combined with backpropagation in hydrological applications (J. Liu et al., 2024). Bayesian LSTM (as an improved LSTM) adapts probability distributed weight parameters, which reduces the model overfitting, thereby providing robust predictions in hydrology (D. Li et al., 2021; Lu et al., 2019). The development of LSTM-based surrogates offers the possibility of accurate predictions in addressing the critical processes in lake-atmosphere modeling systems." (Section 2.2.2, Lines 209-214)

**References:**

Li, D., Marshall, L., Liang, Z., Sharma, A., and Zhou, Y.: Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process-based hydrological models. *Water Resour. Res.*, *57*(9), e2021WR029772, https://doi.org/10.1029/2021WR029772, 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.

Lu, D., Liu, S., and Ricciuto, D.: An efficient bayesian method for advancing the application of deep learning in earth science, in: *Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW)*, IEEE, November, 270-278, https://doi.org/ 10.1109/ICDMW.2019.00048, 2019.

"The hyperparameter space included the number of hidden layers (ranging from 1 to 8), neurons per layer (ranged from 16 to 1,024), optimizer (Adam, or RMSprop), batch size (ranging from 8 to 256), and learning rate (ranging from 1E-6 to 1E-2). The hyperparameters in BO-BLSTM-based surrogates were optimized using BO with a maximum of 100 iterations, 1000 epochs for each iteration, and 50 patience in a EarlyStopping strategy." (Section 2.2.3, Lines 276-279)

"**However, we found that HyLake v1.0 required slightly higher computational costs compared to process-based models, which depend on the hyperparameters of LSTM-based surrogates, despite achieving greater performance (Table 3). In an individual case of MLW prediction, HyLake v1.0 took about 9 times longer to run compared to FLake, with a cost of 151.46 seconds. To compare different experiments of hybrid lake models, Baseline, coupled to an LSTM-based surrogate with 1 layer and 256 neurons per layer, indicated the lowest cost. While TaihuScene, constructed by an LSTM-based surrogate with 7 layers and 836 neurons per layer, showed the most expensive in predictions. Given the sophisticated architecture of LSTM-based surrogates, which inevitably leads to higher costs in training and prediction, developing novel algorithms for approximating LSTMs is urgently needed. Furthermore, the recent research progress demonstrated that** LSTM-based surrogates are more suited for short-term predictions compared to **the prevalent Transformer-based family**, which **is** suited for long-term predictions and commonly used in global weather forecasting systems (K. F. Bi et al., 2023; L. Chen et al., 2023)." (Section 4.1, Lines 614-624)

"BO-BLSTM-based surrogate exhibits superior performance in estimating LST for HyLake v1.0. This study adapted BO and EarlyStopping strategies to ensure BLSTM provides accurate and reliable estimates in prediction but increases the computational demands for training due to its ability to converge from its more complex Bayesian architecture (Peng et al., 2025; Ferianc et al., 2021). In addition, the mere 1 Bayesian fully connected layer that was adapted in this surrogate only captures limited data uncertainty, which may lose several important aspects of probabilistic prediction (Klotz et al., 2022). Given the importance of uncertainty quantification for BLSTM, it is worth noting that HyLake v1.0 has the potential to assess the variance of predictions and probabilities of lake extreme events occurrence by developing its surrogate in future (Kar et al., 2024; Gawlikowski et al., 2023). Major limitations, including high computational demands and insufficient model performance, should be addressed by developing a novel deep-learning-based surrogate based on a more efficient architecture and larger datasets." (Section 4.2, Lines 671-680)

**References:**

Ferianc, M., Que, Z., Fan, H., Luk, W., and Rodrigues, M.: Optimizing Bayesian recurrent neural networks on an FPGA-based accelerator, in: 2021 International Conference on Field-Programmable Technology (ICFPT), IEEE, December, 1-10, 2021.

Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., and Nearing, G.: Uncertainty estimation with deep learning for rainfall–runoff modeling, *Hydrol. Earth Syst. Sci.*, 26, 1673–1693, https://doi.org/10.5194/hess-26-1673-2022, 2022.

Peng, Z., Mo, S., Sun, A. Y., Wu, J., Zeng, X., Lu, M., and Shi, X.: An explainable Bayesian TimesNet for probabilistic groundwater level prediction, *Water Resour. Res.*, 61, e2025WR040191, https://doi.org/10.1029/2025WR040191, 2025.

Discuss how the surrogate maintains physical consistency and whether energy balance is preserved through the hybrid coupling. Consider briefly addressing this in the discussion section.

**Response:** Good point. Indeed, we considered which processes in the lake model can be replaced by a deep-learning-based surrogate. Energy balance and lake water temperature approximations are two individual modules in process-based models, which are difficult to replace with deep-learning-based models simultaneously, while also ensuring numerical stability. There are two reasons to address this issue and listed in Discussion (Section 4.1, Lines 586-611):

**(1) Inadequate observations to build relationships between surface conditions and heat fluxes.** The energy balance equations are integrated modeling systems based on the bulk aerodynamic method from the Monin–Obukhov similarity theory, which covers the calculation of surface conditions (e.g., surface roughness length, friction velocity), as well as water and heat fluxes (e.g., latent heat, sensible heat, evaporation, precipitation-induced heat). Specifically, the latent and sensible heat fluxes are functions of transfer coefficients, which are iteratively updated using the Monin–Obukhov length, surface roughness length, and friction velocity, based on bulk flux algorithms (Verburg and Antenucci, 2010; Woolway et al., 2015). They performed well in estimating heat fluxes from the evidence in previous studies, which has been widely applied in process-based models (Woolway et al., 2020; Thiery et al., 2014). However, surface conditions in current research were always obtained from calculation instead of direct observations. Fewer studies have focused on monitoring surface conditions due to limited equipment, which hinders our ability to construct generalized lake models that reflect their potential relationships. Moreover, there is a large difference between observed surface conditions and predictions by Monin–Obukhov similarity theory, although the Lake Taihu Eddy Flux Network has monitored these conditions for a long time (Figure S6). In the future, high-quality observations and physical principles at the land-air interface should focus on addressing the significant discrepancies between observations and simulations.

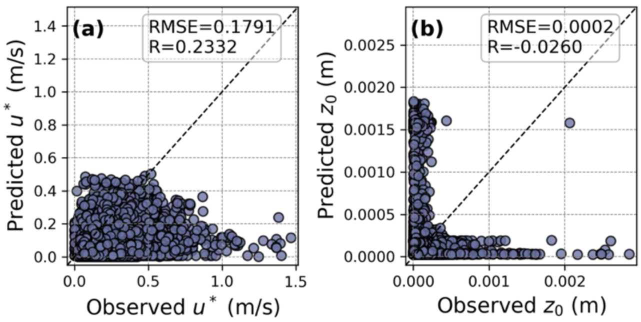



**Figure S6:** The comparison of friction velocity (u*) and surface roughness length ($z_{0m}$, m) in MLW lake site between simulation derived from PBBM and HyLake v1.0 and observations.

**(2) Lake surface temperature governing equations existed uncertainly.** The lake water temperature module is suitable for replacement by a deep-learning-based surrogate due to the rich and easily accessible observations and simplified schemes. Until now, accurately predicting lake water temperature using a generalized framework has remained a challenge due to the significant regional differences among lakes. Several researchers have attempted to approximate lake water temperature changes using complex integrated neural networks, such as physics-informed neural networks (PINNs), physics-guided neural networks (PGNNs), and modular networks (He et al., 2025; Ladwig et al., 2024; Read et al., 2019). These models may exhibit superior performance in specific tasks but require high computational power for pretraining or fine-tuning, and are challenging to predict untrained variables, such as latent heat, sensible heat fluxes, and evaporation. Choose this module to replace in this study, which hopes to propose a generalized integrated framework that combines physical principles and deep learning, and then improve the understanding of lake-atmosphere interactions in finer resolutions.

**References:**

He, Y., and Yang, X.: A physics-informed deep learning framework for estimating thermal stratification in a large deep reservoir, *Water Resour. Res.*, 61, e2025WR040592, https://doi.org/10.1029/2025WR040592, 2025.

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.

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.

Thiery, W. I. M., Stepanenko, V. M., Fang, X., Jöhnk, K. D., Li, Z., Martynov, A., et al.: LakeMIP Kivu: evaluating the representation of a large, deep tropical lake by a set of one-dimensional lake models, *Tellus A: Dyn. Meteorol. Oceanogr.*, 66(1), 21390, https://doi.org/10.3402/tellusa.v66.21390, 2014.

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.

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.

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.

**Revision:**

"Moreover, simplified parameterizations in traditional process-based lake models are commonly adopted (Golub et al., 2022; Mooij et al., 2010), which influence the coupling strategies in HyLake **v1.0. The two critical components, including energy balance equations and 1-D vertical lake water temperature transport equations, compose the physical principles of lake-atmosphere interaction modeling systems, which also possess simplification to some degrees.** 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; Woolway et al., 2015), but substantial discrepancies still exist between simulation results and observations (Figure S**6**), which in turn influence the physical principles between land surface conditions and LST. **The current approaches for solving energy balance equations uses bulk aerodynamic method based on the Monin–Obukhov similarity theory (Monin and Obukhov, 1954), and is the vital module in process-based lake models (e.g., FLake (Mironov et al., 2010), GLM (Hipsey et al., 2019), WRF-Lake (Gu et al., 2015)). However, it remains challenges to construct explainable approaches to quantify the relationships between surface conditions and fluxes and LST due to inadequate observations.** These potential differences in physical processes lead to uncertainties in training deep-learning-based surrogates, contributing to the **insufficient/limited** 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). **On top of the extensive observations of water temperature,** several hybrid models that integrate deep-learning-based and process-based models have been constructed in previous studies, achieving improved performance in model comparisons (**He et al., 2025**; Ladwig et al., 2024; Read et al., 2019). These models and their training strategies generally perform better on training and test datasets **due to their complex coupling strategies and higher computational requirements**, while their generalization and transferability need further validation. Lake Taihu, **as one of typical shallow**, eutrophic, and large **Chinese lakes** 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 potential ecological implications under thermodynamic changes. **HyLake v1.0, developed based on *in situ* observations from Lake Taihu, has been proven to be reliable and rigorously validated in Lake Chaohu (Table 3), demonstrating a faster and more accurate framework for enhancing the understanding of hybrid hydrological modeling.**" (Section 4.1, Lines 586-611)

**References added:**

He, Y., and Yang, X.: A physics-informed deep learning framework for estimating thermal stratification in a large deep reservoir, *Water Resour. Res.*, 61, e2025WR040592, https://doi.org/10.1029/2025WR040592, 2025.

Monin, A. S., and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, *Contrib. Geophys. Inst. Acad. Sci. USSR*, *151*(163), e187, 2954, 1954.

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.

Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., Hanson, P. C., Read, J. S., de Sousa, E., Weber, M., and Winslow, L. A.: A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON), *Geosci. Model Dev.*, 12, 473–523, https://doi.org/10.5194/gmd-12-473-2019, 2019.

Gu, H., Jin, J., Wu, Y., Ek, M. B., and Subin, Z. M.: Calibration and validation of lake surface temperature simulations with the coupled WRF-lake model. *Clim. Change*, *129*(3), 471-483, https://doi.org/10.1007/s10584-013-0978-y, 2015.

While full uncertainty quantification may be beyond the current scope, briefly discuss the uncertainty implications of the Bayesian surrogate and how this could be leveraged in future applications.

**Response:** The proposed Bayesian LSTM (BLSTM) in this study, an improved version of LSTM that replaces the last fully connected layer with a Bayesian fully connected layer, provides robust predictions by utilizing probability-distributed weight parameters in networks (D. Li et al., 2021; Lu et al., 2019). However, it inevitably causes uncertainties from challenging data sources and network architecture (Gawlikowski et al., 2023) and increases the computational requirements due to the complex architecture (Peng et al., 2025; Ferianc et al., 2021). The uncertainties caused by BLSTM's probability-distributed parameters, which have been widely used for assessing the variance of predictions and the probability of extreme events occurring when using out-of-bag samples, thereby improving the accuracy of decision-making for users (Kar et al., 2024; Gawlikowski et al., 2023). We are expected to improve the surrogate in HyLake v1.0 and quantify its uncertainties to further enhance our understanding of the occurrence and frequency of lake extreme events in the future. The associated revisions can be found in Materials and Methodology (Section 2.2.2, Lines 209-214), and Discussion (Section 4.2, Lines 671-680).

**References:**

Ferianc, M., Que, Z., Fan, H., Luk, W., and Rodrigues, M.: Optimizing Bayesian recurrent neural networks on an FPGA-based accelerator, in: 2021 International Conference on Field-Programmable Technology (ICFPT), IEEE, December, 1-10, 2021.

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., et al.: A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.*, *56*, 1513-1589, https://doi.org/10.1007/s10462-023-10562-9, 2023.

Kar, S., McKenna, J. R., Sunkara, V., Coniglione, R., Stanic, S., and Bernard, L.: XWaveNet: enabling uncertainty quantification in short-term ocean wave height forecasts and extreme event prediction. *Appl. Ocean Res.*, *148*, 103994, https://doi.org/10.1016/j.apor.2024.103994, 2024.

Li, D., Marshall, L., Liang, Z., Sharma, A., and Zhou, Y.: Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process-based hydrological models. *Water Resour. Res.*, *57*(9), e2021WR029772, https://doi.org/10.1029/2021WR029772, 2021.

Lu, D., Liu, S., and Ricciuto, D.: An efficient bayesian method for advancing the application of deep learning in earth science, in: *Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW)*, IEEE, November, 270-278, https://doi.org/ 10.1109/ICDMW.2019.00048, 2019.

Peng, Z., Mo, S., Sun, A. Y., Wu, J., Zeng, X., Lu, M., and Shi, X.: An explainable Bayesian TimesNet for probabilistic groundwater level prediction, *Water Resour. Res.*, 61, e2025WR040191, https://doi.org/10.1029/2025WR040191, 2025.

**Revision:**

"It has been demonstrated that LSTM could capture historical time-step dependencies and handle variable-length input sequences using gradient optimization combined with backpropagation in hydrological applications (J. Liu et al., 2024). Bayesian LSTM (as an improved LSTM) adapts probability distributed weight parameters, which reduce the model overfitting, thereby providing robust predictions in hydrology (D. Li et al., 2021; Lu et al., 2019). The development of LSTM-based surrogates offers the possibility of accurate predictions in addressing the critical processes in lake-atmosphere modeling systems." (Section 2.2.2, Lines 209-214)

"BO-BLSTM-based surrogate exhibits superior performance in estimating LST for HyLake v1.0. This study adapted BO and EarlyStopping strategies to ensure BLSTM provides accurate and reliable estimates in prediction but increases the computational demands for training due to its ability to converge from its more complex Bayesian architecture (Peng et al., 2025; Feriance et al., 2021). In addition, the mere 1 Bayesian fully connected layer that was adapted in this surrogate only captures limited data uncertainty, which may lose several important aspects of probabilistic prediction (Klotz et al., 2022). Given the importance of uncertainty quantification for BLSTM, it is worth noting that HyLake v1.0 has the potential to assess the variance of predictions and probabilities of lake extreme events occurrence by developing its surrogate in future (Kar et al., 2024; Gawlikowski et al., 2023). Major limitations, including high computational demands and insufficient model performance, should be addressed by developing a novel deep-learning-based surrogate based on a more efficient architecture and larger datasets." (Section 4.2, Lines 671-680)

**References added:**

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., et al.: A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.*, *56*, 1513-1589, https://doi.org/10.1007/s10462-023-10562-9, 2023.

Kar, S., McKenna, J. R., Sunkara, V., Coniglione, R., Stanic, S., and Bernard, L.: XWaveNet: enabling uncertainty quantification in short-term ocean wave height forecasts and extreme event prediction. *Appl. Ocean Res.*, *148*, 103994, https://doi.org/10.1016/j.apor.2024.103994, 2024.

**Minor Comments**

Terminology: Define LE (latent heat) and HE (sensible heat) at first mention.

**Response:** We have defined LE and HE in the Introduction.

**Revision:** "Lake-atmosphere interactions represent a tightly coupled system (B. B. Wang et al., 2019), where process-based models traditionally approximate the interdependence between LST, **latent heat (LE) and sensible heat (HE) fluxes**." (Section 1, Lines 77-79)

References: Standardize citation formats (e.g., "Hersbach et al. (2020)" vs. "Hersbach et al., 2020").

**Response:** We have checked the manuscript. We adopted "Hersbach et al. (2020)" when discussing their contributions and used "Hersbach et al., 2020" to cite their conclusions.

Section Organization: Consider moving deep implementation details (e.g., GUI remarks) into a Supplement or Code &

Data Availability section.

**Response:** We now provided example bash scripts to run HyLake v1.0 and other models (e.g., Baseline, TaihuScene) in Taihu and Chaohu experiments. The example script for run these models was given by Figure R1. The information of data and code availability was given in Lines 728-732.

```
# --- Example 1: Run PBBM in MLW experiment -----------------------
python HyLake.py \
  --data_source MLW \
  --model "PB"
# --- Example 2: Run Baseline in MLW experiment -------------------
python HyLake.py \
  --data_source MLW \
  --model "Baseline"
# --- Example 3: Run HyLake v1.0 in MLW experiment ----------------
python HyLake.py \
  --data_source MLW \
  --model "LSTM"
# --- Example 4: Run TaihuScene in MLW experiment -----------------
python HyLake.py \
  --data_source MLW \
  --model "Taihu_LSTM"
# --- Example 5: Run HyLake v1.0 in Taihu-ERA5 experiment----------
python HyLake.py \
  --data_source ERA5 \
  --model "LSTM"
# --- Example 5: Run TaihuScene in Taihu-ERA5 experiment-----------
python HyLake.py \
  --data_source ERA5 \
  --model "Taihu_LSTM"
# --- Example 6: Run HyLake v1.0 in 1 grid of Chaohu experiment ----
python HyLake.py \
  --Lake_Name Chaohu \
  --Lake_Lat 31.53 \
  --Lake_Lon 117.31 \
  --Lake_altitude -4 \
  --Lake_depth 4 \
  --SimLength 35040 \
  --initial_temp 5.0 \
  --model "LSTM" \
  --data_source custom \
  --csv_path "./data/chaohu/ERA5_forcings/forcing_csv/Chaohu_forcing_lat31.53_lon117.31.csv" \
  --exp "Chaohu_Lake_lat31.53_lon117.31" \
  --col_indices 1,2,3,4,5,6,7
```

**Figure R1: The example scripts for run HyLake v1.0 for MLW, Taihu-Obs and Taihu-ERA5 experiments.**

**Revision:** "***Code and data availability.*** The datasets, codes **and scripts** of HyLake v1.0 and other models **(e.g., Baseline, TaihuScene)** used in this study are available at https://doi.org/10.5281/zenodo.15289113 (He et al., 2025). **FLake model was run via LakeEmsemblR tool (https://aemon-j.github.io/LakeEnsemblR/).** 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)." (Code and data availability, Lines 728-732)

Caption Detail: Enhance figure captions to specify whether plotted values are observed or simulated and note dataset origins (real vs. semi-synthetic).

**Response:** We have improved the captions to clearly describe the datasets in Figures 5 to 11.