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 #1:

The manuscript presents HyLake v1.0, a hard-coupled hybrid lake model in which an LSTM surrogate replaces the implicit-Euler surface-temperature solver embedded within an in-house one-dimensional physical backbone. The surrogate is trained at the MLW site on Lake Taihu and then applied to five other sites that differ in both biological characteristics and meteorological forcing. Although the hybrid framework outperforms several process-based and deep-learning-based benchmarks, its validation strategy and treatment of uncertainty require further refinement. Overall, the paper is clearly written and could be suitable for publication after moderate revision.

Response: We thank Reviewer #1 for his/her positive feedback and constructive comments. The comments are all accepted and **Relisted in black**, followed by our **Replies in blue** and **Revisions in red (highlighted revisions in bold)**. According to the comments, we modified the manuscript, particularly on the capacity of HyLake v1.0 from the aspects of model transferability, computational efficiency, and future improvements. Major changes are summarized as follows:

No.	Major Revisions	Important Messages
1	Applied HyLake v1.0 for Lake Chaohu and validated based on MYD11A1 imagery.	HyLake v1.0 outperformed FLake using ERA5 forcing dataset in Lake Chaohu **(Discussion)**.
2	Intercompared computational efficiency between FLake, Baseline, TaihuScene, and HyLake v1.0.	Computational efficiency depends on surrogate architecture. HyLake v1.0 costed fewer than Baseline and TaihuScene but much than FLake in all experiments **(Discussion)**.
3	Cross-validated different BO-BLSTM-based surrogates that were individually trained with five lake site observations.	MLW observations are the most suitable site for training BO-BLSTM-based surrogate, which achieved the best performance in cross-validation **(Discussion)**.

Major comments

1. To address the model's generality, the authors should apply HyLake to at least one morphologically distinct lake or extend the simulation period to include additional years.

Response: Good point. We agree that the capacity of HyLake v1.0's transferability is important. Therefore, we conducted one additional experiment via FLake and HyLake v1.0, and then validated the lake surface temperature (LST) on a morphologically distinct lake - Lake Chaohu, a large, shallow lake in southeastern China (Fig. R1). The average depth and area of Lake Chaohu are 2.7 m and 768 km^2, which is one of several large shallow lakes in the middle and lower reaches of the Yangtze River, and is densely populated and strongly influenced by human activities (Wei et al., 2022; Zhang et al., 2022). The observation was derived by MYD11A1 MODIS/aqua daily products with 1 km spatial resolution. The ERA5 dataset was used as the forcing dataset to drive the Chaohu experiment. ERA5 forcing dataset used 4 black grids to cover Lake Chaohu, while MODIS observations covered Lake Chaohu by red points in Figure S7.

Our assessments indicated that HyLake v1.0 performed better than the FLake model, showing the promise of applying it to other lakes (Table 3, Figure S8-9). Changes can be found in Materials and Methodology (Section 2.3.1, Lines 286-289, Lines 308-314), Discussion (Section 4.1, Lines 560-567, Lines 609-611; Section 4.2, Lines 647-648), Table 3, and Figure S7-9.

References:

Wei, Z., Yu, Y., and Yi, Y.: Spatial distribution of nutrient loads and thresholds in large shallow lakes: the case of Chaohu Lake, China, *J. Hydrol.*, *613*, 128466, https://doi.org/10.1016/j.jhydrol.2022.128466, 2022.

Zhang, J., Gao, J., Zhu, Q., Qian, R., Zhang, Q., and Huang, J.: Coupling mountain and lowland watershed models to characterize nutrient loading: An eight-year investigation in Lake Chaohu Basin, *J. Hydrol.*, *612*, 128258, https://doi.org/10.1016/j.jhydrol.2022.128258, 2022.

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 for intercomparison." (Section 2.3.1, Lines 286-289)

"Furthermore, this study implemented the HyLake v1.0 into Lake Chaohu, the 5th-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 (MYD11A1, 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)

References added:

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.

"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 lake depth of 3.06 m and lake area of 760 km^2 (Figure S6, Jiao et al., 2018), to validate the potential capacity of model application.** These experiments were compared using **observed meteorological datasets and ERA5 datasets, then** validated for both spatiotemporal patterns of LST 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 MYD11A1 datasets (Table 3 and Figures S7-9).**" (Section 4.1, Lines 560-567)

References added:

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.

"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." (Section 4.2, Lines 647-648)

"**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)

“

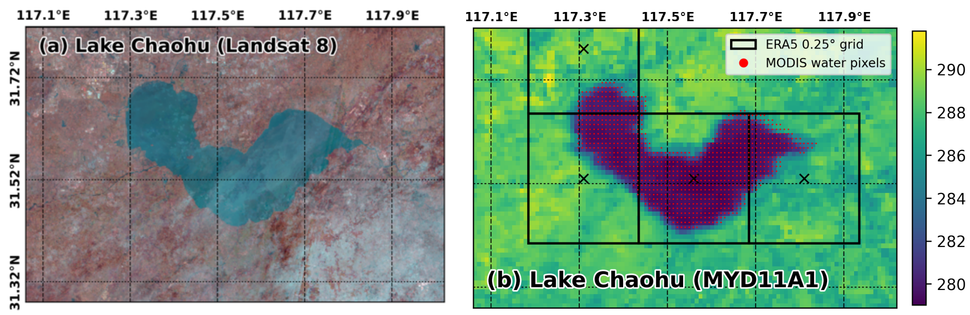


Figure S7: The locations of Lake Chaohu overlaid on a true-color image from (a) Landsat 8 and daily land surface temperature from (b) MYD11A1 product.” (Figure S7 in Supplementary materials)

“

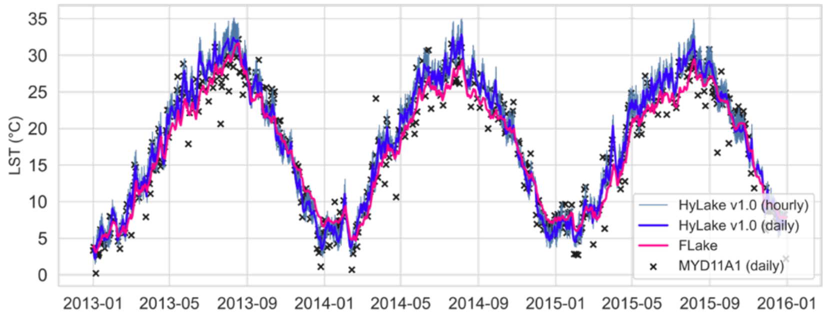


Figure S8: Time series of daily grid-average LST on Lake Chaohu derived from MYD11A1, FLake simulation, and HyLake v1.0 from 2013 to 2015. HyLake v1.0 provides daily and hourly simulations.” (Figure S8 in Supplementary materials)

“

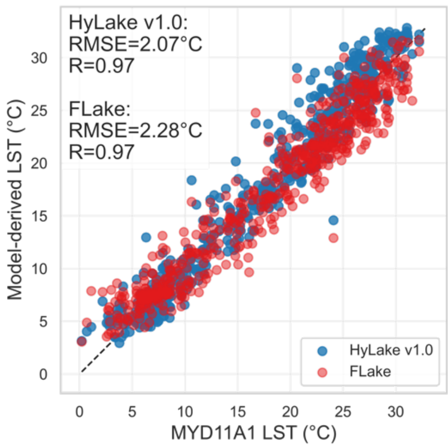


Figure S9: The intercomparison of daily LST between model simulations (FLake and HyLake v1.0) and MYD11A1 observations on Lake Chaohu from 2013 to 2015.” (Figure S9 in Supplementary materials)

2. The study employs Bayesian optimization to optimize network depth, width, optimizer, and learning rate, but ignore the critical information. Please provide the search ranges of hypermeters, objective function, stopping criterion, and computational cost.

Response: We apologize for the missing hypermeters, which are being searched for in the space of Bayesian Optimization and computational cost. In the revised manuscript, we have added information about surrogate training and compared the computational efficiency of each model in all experiments. We also employed an EarlyStopping strategy to optimize the best set of hyperparameters using Bayesian Optimization. The comparison results of computational efficiency for each model indicated that the computational costs depended on the surrogate's architecture, suggesting that the BO-BLSTM-based surrogate in TaihuScene, which has a larger network, required more computational resources than HyLake v1.0 and the Baseline. The associated revisions can be found in Materials and Methodology (Section 2.2.3, Lines 276-279) and Table 3.

Revision:

"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 an EarlyStopping strategy." (Section 2.2.3, Lines 276-279)

"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)

3. HyLake performed well at MLW, PTS, and XLS, whereas TaihuScene outperforms it at BFG and DPK. Discuss possible causes and advise when multi-site versus single-site training is preferable.

Response: Good point. We were surprised to find that the BO-BLSTM-based surrogate, which used a larger training dataset encompassing all lake sites in Lake Taihu, performed worse than one trained solely with MLW observations. The references found two hypothesis that might explain this issue: (1) The scaling laws for deep learning models determined that model performance will not continue to increase indefinitely with stacking of neurons and layers (Hestness et al., 2017); (2) Training data with high representation samples helps improve model performance, while using all over datasets would neglect heterogeneous functional samples and hinder the model's performance gains (Wang et al., 2025). We agreed that the surrogate in HyLake v1.0 meets the hypotheses. However, a comprehensive assessment of how many samples should be adapted at different scales (e.g., individual-lake, regional, or global scale) needs to be discussed in the future. The current manuscript provides additional explanations regarding these two hypotheses, although it does not reach a definitive conclusion. The associated revisions are listed in Discussion (Section 4.1, Lines 654-670).

References:

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

Wang, S., Yu, L., Gao, C., Zheng, C., Liu, S., Lu, R., et al.: Beyond the 80/20 rule: High-entropy minority tokens drive effective reinforcement learning for LLM reasoning, arXiv [preprint], arXiv:2506.01939, 2025.

Revision:

"Future development of HyLake v1.0 **needs to** collect more observations, including heat fluxes and water temperature, and searching for more variables in datasets to train LSTM-based surrogates and acquire more general models at a larger scale. However, it is important to note that **the** performance of HyLake may not always improve with an increase in the training data size. **The training datasets with higher representation of physical principles help improve the model performance. Similar phenomenon has already been observed in many deep-learning-based large models, demonstrating that directly training models using all datasets would neglect heterogeneous functional samples and thereby hinder performance gains (Wang et al., 2025). Therefore, this study assumed that an individual-site-trained LSTM-based surrogate would have better capacity in representing lake-atmosphere interactions, which was collectively matched to the above-mentioned hypotheses. Due to insufficient observations at other lake sites (DPK, PTS, and XLS), to some degree, the surrogates trained on their datasets performed closely in estimating ΔLST except for XLW (Table S1). For the relatively complete observed datasets in BFG (although its biological characteristics cannot represent the whole Lake Taihu), the surrogate performed poorer than the proposed BO-BLSTM-based surrogate in terms of diurnal patterns of LST of HyLake v1.0 (Figure S10).** As HyLake is extended to larger scales or more lakes, the computational architecture will need to accommodate large training datasets, which may limit performance for specific lakes. Specifically, the scaling laws for deep-learning models indicate that model performance does not continue to increase indefinitely with the stacking of neurons and layers (Hestness et al., 2017). Adapting more powerful deep-learning-based surrogates will further improve HyLake v1.0 performance, leading to a better representation of lake-atmosphere interactions in ungauged lakes." (Section 4.1, Lines 654-670)

References added:

Wang, S., Yu, L., Gao, C., Zheng, C., Liu, S., Lu, R., et al.: Beyond the 80/20 rule: High-entropy minority tokens drive effective reinforcement learning for LLM reasoning, arXiv [preprint], arXiv:2506.01939, 2025.

4. The discussion regarding computational efficiency of HyLake is inadequate. Provide a detailed table comparing training time and wall-clock simulation speed for HyLake, FLake, and any other relevant models.

Response: We have provided the computational efficiency of models in all numerical experiments, including PBBM, Baseline, FLake, TaihuScene, and HyLake v1.0. As for the training time of LSTM-based surrogates, we do not compare each other because it is evident that the BO-BLSTM-based surrogate in TaihuScene, which used larger datasets to train, costs more than that in HyLake v1.0. Therefore, we discussed the computational efficiency of process-based models and hybrid lake models in response to this comment. The results indicated that HyLake v1.0 required higher resources compared to the traditional process-based models, including PBBM and FLake, in some cases, but cost less than TaihuScene. We found that their computational efficiency depends on the architecture of their surrogates. It is undeniable that the surrogate with deeper and broader networks requires more resources to train and predict. Therefore, we need to develop more deep-learning-based approaches to simulate results accurately and rapidly in the future. The associated revision can be found in Table 3 and Discussion (Section 4.1, Lines 615-624; Section 4.2, Lines 671-680).

Revision:

"**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 615-624)

“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 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:

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.

“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)

Minor comments

Line 164: Define the acronym LWT on first use.

Response: Corrected.

Revision: "T_s (°C) accounts for LST solved by 1-D vertical **lake water temperature (LWT)** transport equation; …" (Section 2.2.1, Lines 176-177)

Figure 3: Text are crowded in d-f. Consider summarizing the model accuracy in a table.

Response: We have added Table 3 to summarize the model's performance. The citations to Table 3 have already been added to the manuscript.

Revision:

"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)

Figure 4-6: Add the relevant validation statistics directly to the LST, LE, and HE plots for clarity.

Response: We have added the Pearson coefficient in Figure 4-6 (now Figure 6-8 in the revised manuscript) due to the aesthetic appeal of the figures. The detailed information about model validation can be found in Table 3.

Revision:

“

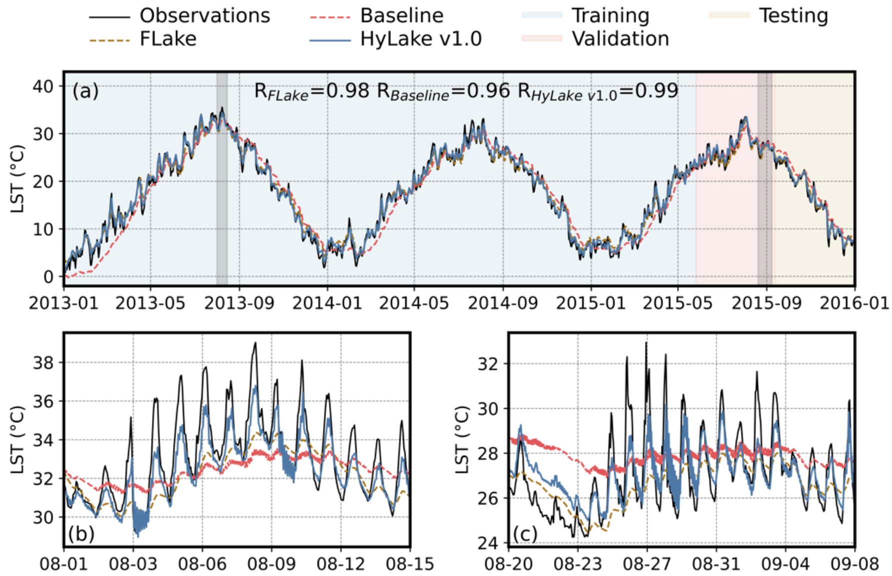


Figure 1: Comparison of observations and predictions by FLake, Baseline, and HyLake v1.0 in temporal trends of LST. Comparison of (a) the full time series and (b-c) partial time series of models derived LST and observations from 2013 to 2015. All results in (a) were presented at a daily-average scale by resampling. Blue, red, and yellow regions represent the period for the train, validation, and test datasets, respectively. Black solid, brown dashed, red dashed, and blue solid lines represent LST from observations, FLake, Baseline, and HyLake v1.0, respectively.” (Figure 6)

“

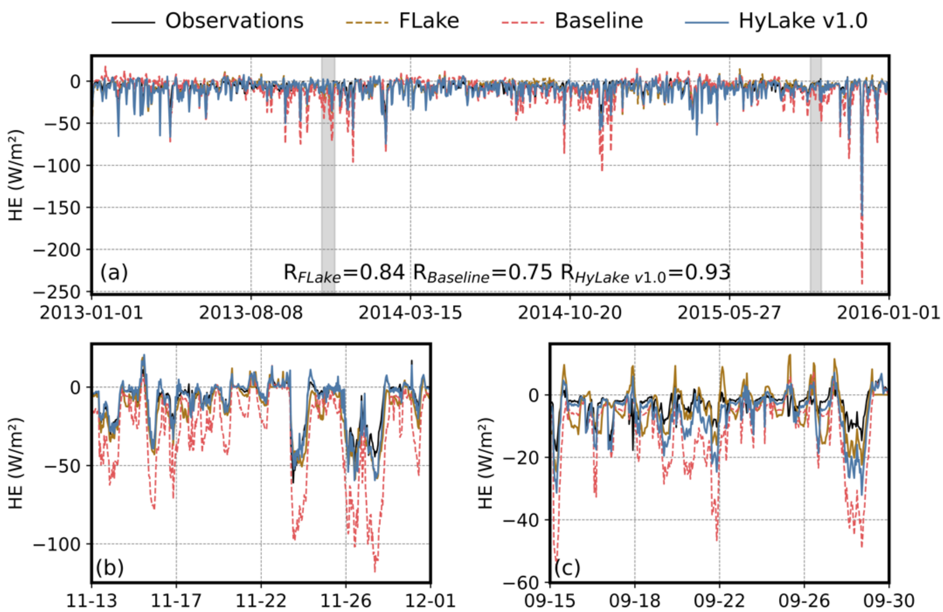


Figure 2: Comparisons of observations and predictions by FLake, Baseline, and HyLake v1.0 in temporal trends for LE. Comparison of (a) full and (b-c) partial time series of model derived LE and observations from 2013 to 2015.” (Figure 7)

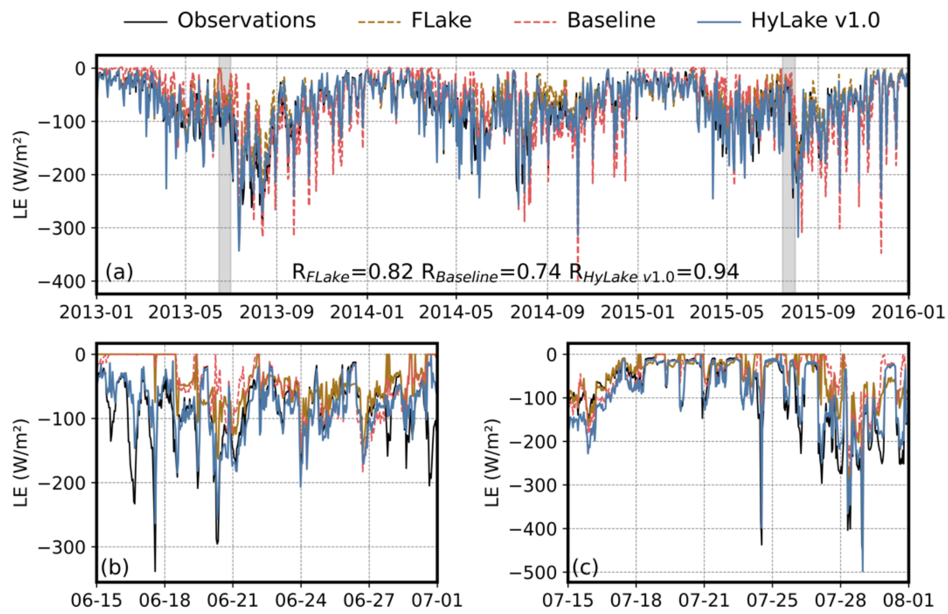
“


Figure 3: Comparisons of observations and predictions by FLake, Baseline, and HyLake v1.0 in temporal trends for HE. Comparison of (a) full and (b-c) partial time series of model derived HE and observations from 2013 to 2015.” (Figure 8)

Figure 9: a-c and d-i represent distinct data types and should not share a single figure label.

Response: Corrected. Figure 9 was now spitted into Figure 10 and 11.

Revision:

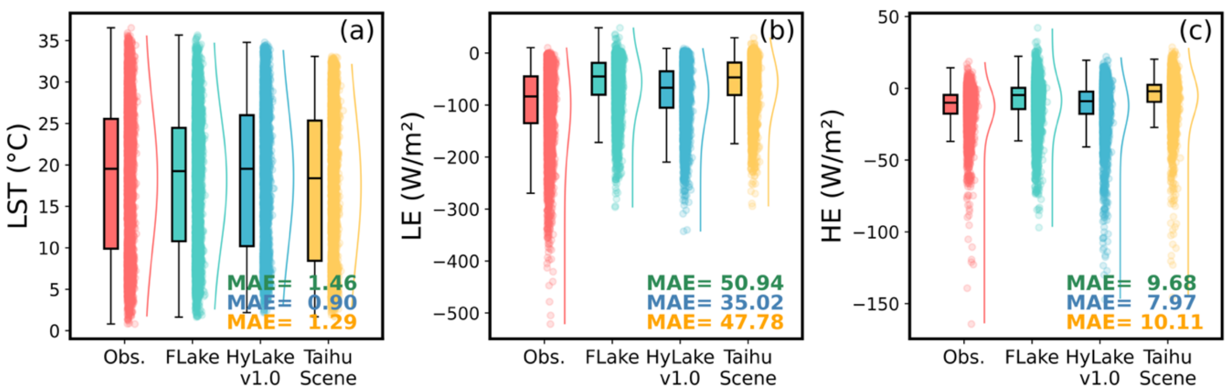
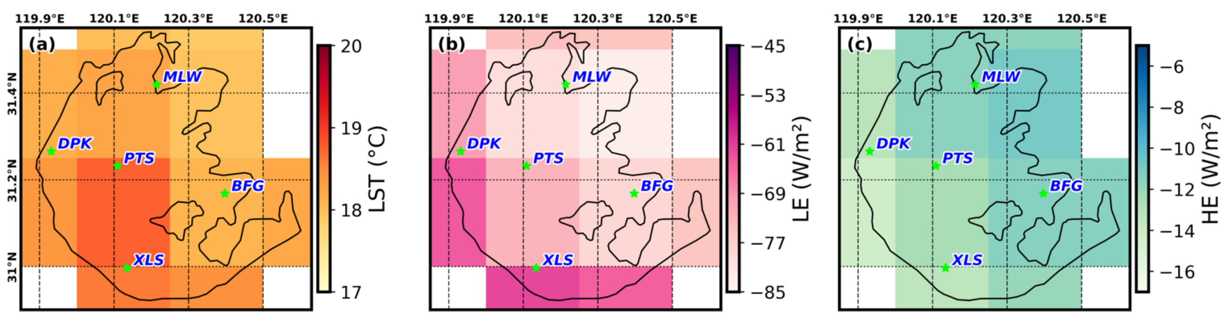
“


Figure 10: The statistical characteristics and spatial average of LST, LE and HE for observations, FLake, HyLake v1.0 and TaihuScene in all sites using ERA5 forcing datasets. Green, blue and yellow texts in figures represent the MAEs of LST, LE and HE for FLake, HyLake v1.0 and TaihuScene, respectively.” (Figure 10)

“


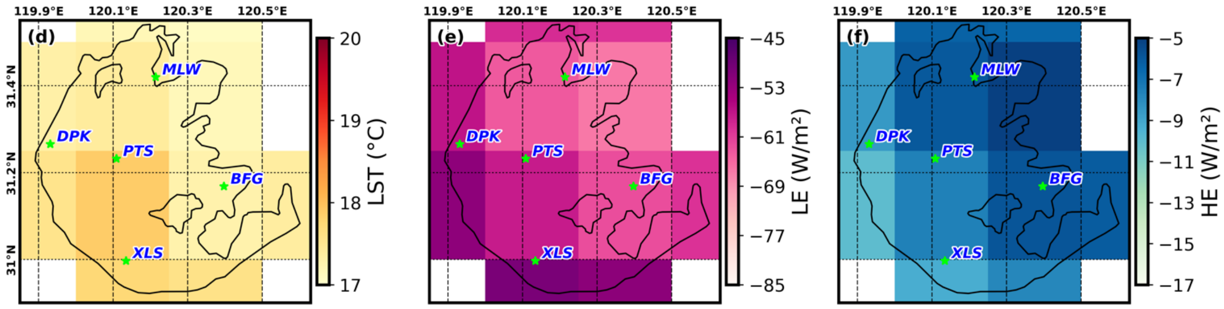


Figure 4: The statistical characteristics and spatial average of predicted LST, LE and HE for HyLake v1.0 and TaihuScene drove by ERA5 forcing datasets. (a-c) represent LST, LE and HE for HyLake v1.0, respectively; (d-f) represent LST, LE and HE for TaihuScene, respectively. The green stars noted in all figures are lake sites in Lake Taihu." (Figure 11)

Line 847: correct the citation format for code repositories with the rest of the reference list.

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