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 #2:

The manuscript entitled "Hybrid Lake Model (HyLake) v1.0: unifying deep learning and physical principles for simulating lake-atmosphere interactions" written by Yuan He and Xiaofan Yang (egusphere-2025-1983) presented the HyLake v1.0 hybrid model, which performed better than other models. The manuscript is generally well-written, which will be within the scope of GMD. Please clarify the following points before the possible publication.

Response: We thank Reviewer #2 for the positive and constructive comments. All the comments have been accepted and **Relisted in black**, followed by our **Replies in blue** and **Revisions in red (highlighted revisions in bold)**. According to the comments, we particularly discussed the reasons of using individual-site observations to train LSTM-based surrogates and explained the details of model development. Major changes are summarized in the following table:

No.	Major Revisions	Key Messages
1	Discussed the selection of MLW observations to train LSTM surrogate in HyLake v1.0.	As one of the long-term monitored lake sites in Lake Taihu, MLW has high-quality of observations and highly represents the eutrophic status of Lake Taihu. We cross-validated the performance using different observations from lake sites to train LSTM surrogate and confirmed that observations of the MLW are reliable **(Materials and methodology; Discussion)**.
2	Described how to fill the data gaps using ERA5 dataset	We used meteorological variables in ERA5 dataset to fill the missing data in the lake site that existed missing in their time series. These observations were used to force lake models to predict lake surface temperature and heat fluxes **(Materials and methodology)**.

Major comments:

Line 117-119 (and Table 1): This might be the trial and error in the authors and was not presented explicitly within the manuscript, but why was only the MLW site used for training and other sites used for validation? I missed the information, but why was the cross-validation not attempted in the process? I am wondering about the robustness of the developed model based on the training data from one site.

Response: We apologize for the confusion regarding our use of MLW observations to train an LSTM-based surrogate. Specific reasons are listed as follows:

(1) Reason for choosing MLW observations to train LSTM-based surrogate. There are five sites in Lake Taihu where hydrometeorological variables are observed via the lake Eddy Flux Network, including air temperature, rainfall rate, net longwave and shortwave radiation, wind speed, surface pressure, relative humidity, lake surface temperature, and latent and sensible heat fluxes. Air temperature, rainfall rate, net longwave and shortwave radiation, wind speed, surface pressure, and relative humidity were adapted to force lake models, while lake surface temperature, latent and sensible heat fluxes were used to validate the model performance. Among these observations, we found there are data gaps to different degrees in these sites. For example, about 475-time steps (~1.36%) of observed surface pressure were found to be lacked in DPK site during 2012 and 2015; 7,959 time steps (~22.71%) of all observed variables were missing in XLS site; 12,539 time steps (~35.78%) of all observed variables were missing in PTS site during 2013-2015. We think these lake sites were not suitable for training the LSTM-based surrogates. Given the MLW and BFG sites, citations have evidenced that Lake Taihu and its MLW site are quintessential examples of severe eutrophication in China (Yan et al., 2024; Wang et al., 2019), which differs from BFG's biological characteristics. The association descriptions can be found in Material and methodology (Section 2.1, Lines 114-139).

(2) Cross-validation between each lake site in training LSTM-based surrogates: The MLW observations are the most reliable among the 5 lake sites after our rigorous validation. We also used observations from 5 lake sites to individually search for optimized BO-BSTM-based surrogates, respectively. The validation results are given in Figure 4, S10, R1 and Table S1. The results indicated that, theoretically, the surrogates trained with MLW, BFG, DPK, and PTS performed well in validation, while those trained with XLS performed poorer than the other surrogates (Figure 4, R1 and Table 1). Considering the absence of DPK and PTS observations, we only choose the surrogate trained with BFG to couple to the PBBM backbone model (namely HyLake-BFG in this comment) and then compare its performance to HyLake v1.0 in MLW and BFG site. Results indicated that HyLake v1.0 outperformed HyLake-BFG in both MLW and BFG site (Figure S10), indicating that using MLW observations to train surrogate helps hybrid lake models learn physical knowledge and improve their accuracy. Therefore, we decided to developed HyLake v1.0 based on MLW observations to according to the comprehensive validation. The associated revisions were listed in Discussion (Section 4.2, Lines 660-666).

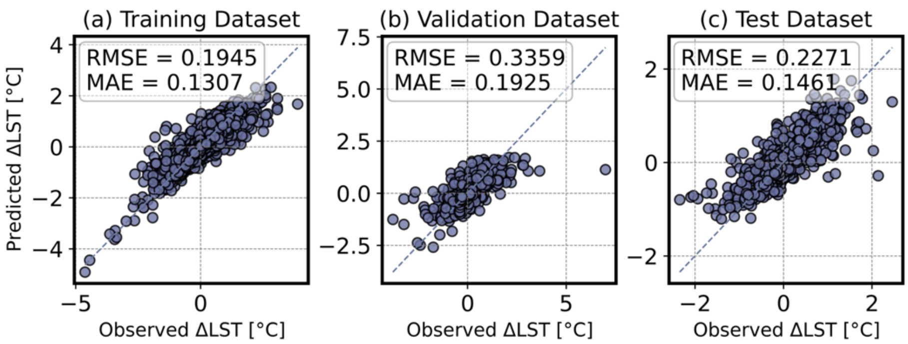
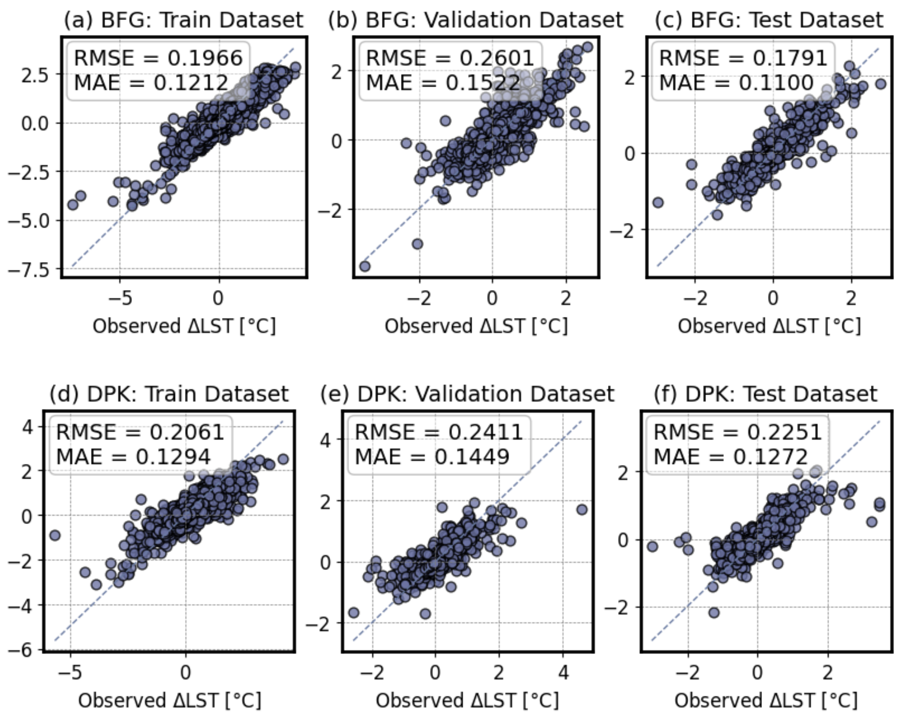


Figure 1: Validation of BO-BLSTM-based surrogate trained with MLW observations in HyLake v1.0 for (a) training, (b) validation and (c) test datasets.



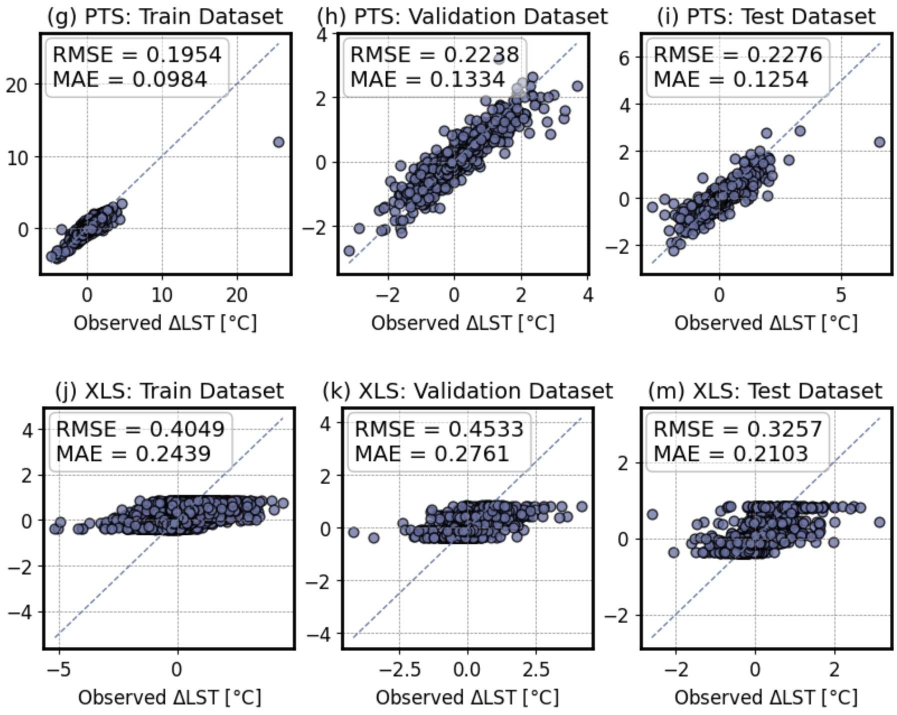


Figure R1: Validation of BO-BLSTM-based surrogates trained with (a-c) BFG, (d-f) DPK, (g-i) PTS, and (j-m) XLS observations in train, validation and test datasets, respectively.

Table S1: Model specifications of BO-BLSTM-based surrogates that trained with BFG, DPK, PTS, and XLS observations and performance in training sets, validation sets, and test sets of MLW. The RMSE for each surrogate was calculated from the difference between their training datasets.

NO.	Training dataset	Model specifications				RMSE (℃)		
		Number of layers	Neurons per layer	Batch size	Learning rate	Train	Validation	Test
1	MLW	4	467	64	9.6E-4	0.19	0.34	0.23
2	BFG	5	30	94	2.5E-3	0.20	0.26	0.18
3	DPK	5	94	124	3.0E-3	0.21	0.24	0.23
4	PTS	6	143	124	7.5E-4	0.20	0.22	0.23
5	XLS	5	170	29	1.0E-2	0.40	0.45	0.33
6	Whole	7	836	145	2.5E-2	0.24	0.33	0.23

References:

Wang, J., Fu, Z., Qiao, H., and Liu, F.: Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.*, *650*, 1392-1402, https://doi.org/10.1016/j.scitotenv.2018.09.137, 2019.

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

Revision:

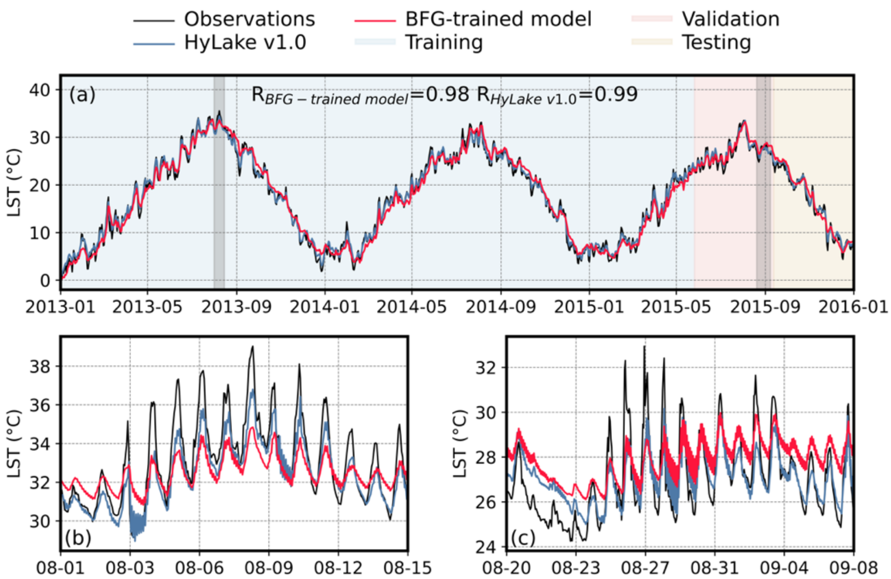
"**The datasets included two parts: (1) hydrometeorological variables observed from the Taihu Lake Eddy Flux Network to force and validate the models, and (2) meteorological variables from ERA5 datasets to fill the gaps of observations and force the models. Within the network,** each site is equipped with an eddy covariance system that continuously monitors **LE and HE** using sonic anemometers and thermometers (Model CSAT3A; Campbell Scientific, Logan, UT, USA) positioned 3.5 to 9.4 m above the lake surface. Hydrometeorological variables, including air humidity and temperature (Model HMP45D/HMP155A; Vaisala, Helsinki, Finland), wind speed (Model 03002; R.M. Young Co., Traverse City, MI, USA), and net radiation components (Model CNR4; Kipp & Zonen, Delft, the

Netherlands), are also measured. **These meteorological variables were used to force lake models while LE, HE and LST from observations were used to validate the results of each numerical experiment, on top of which, the** inferred radiative LST were collected at 30-minute intervals that are publicly accessible via Harvard DataVerse (Lee, 2004; Zhang et al., 2020; https://doi.org/10.7910/DVN/HEWCWM). The dataset spans from 2012 to 2015 **and contains several data gaps across these lake sites. Specifically, 475 time steps (~1.36%) of observed surface pressure were found missing at the DPK site during 2012 and 2015; 7,959 time steps (~22.71%) of all observed variables were missing at the XLS site; 12,539 time steps (~35.78%) of all observed variables were missing at the PTS site. Observations at the MLW and BFG sites were complete during the entire study periods. For the model evaluation of Taihu-obs experiment, the data gaps of observed variables in these lake sites were directly filled by ERA5 datasets at the corresponding time steps, which were used to predict lake-atmosphere interactions.** In this study, **observed meteorological** variables from the MLW site**, an eutrophic lake site that presents the trophic status of Lake Taihu (Table 1, Wang et al., 2019),** are used to train the Long Short-Term Memory (LSTM)-based surrogates (Sect. 2.2); while data from the remaining sites serve to evaluate the generalization of HyLake v1.0 and train the LSTM-based surrogates. To **further** address the generalization **and transferability** of HyLake v1.0 across different forcing datasets, this study utilized **8 meteorological variables that obtained from** hourly ERA5 datasets from 2012 to 2015, with a spatial resolution of 0.25° at a single level to **force** HyLake v1.0. These datasets, available from the Climate Data Store (Hersbach et al., 2020; http://cds.climate.copernicus.eu), include variables such as air temperature, dew point temperature, surface pressure, wind speed, and surface net longwave and shortwave radiation**, which has similar probability distribution to observations across Lake Taihu (Figure S1).**The ERA5 datasets are **also individually** used to **force FLake and TaihuScene for comparison and predict lake-atmosphere interactions** in Lake Taihu, providing insights into the model's generalization, transferability and performance **using** different climatic forcing datasets.” (Section 2.1, Lines 114-139)

References added:

Wang, J., Fu, Z., Qiao, H., and Liu, F.: Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.*, *650*, 1392-1402, https://doi.org/10.1016/j.scitotenv.2018.09.137, 2019.

“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).” (Section 4.2, Lines 660-666)



“

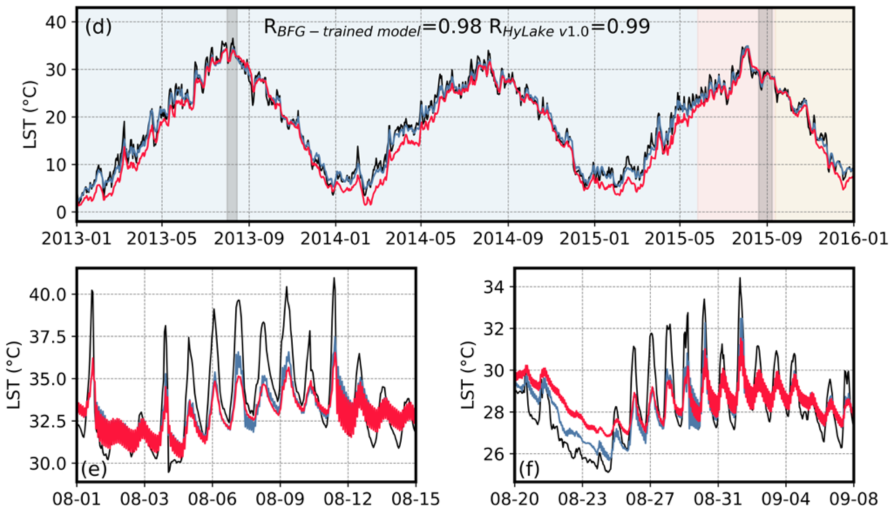


Figure S10: Comparison between HyLake v1.0 used MLW-train surrogate and BFG-trained surrogate in temporal trends of LST. (a-c) and (d-f) present the time series comparison at MLW and BFG site, respectively. Comparison of (a, and d) the full time series and (b-c, and e-f) partial time series of models derived LST and observations from 2013 to 2015. Blue, red, and yellow regions represent the period for the training, validation, and test datasets, respectively." (Figure S10 in Supplementary Materials)

"**Table S1: Model specifications of BO-BLSTM-based surrogates that trained with BFG, DPK, PTS, and XLS observations and performance in training sets, validation sets, and test sets of MLW. The RMSE for each surrogate was calculated from the difference between their training datasets.**

NO.	Training dataset	Model specifications				RMSE (°C)		
		Number of layers	Neurons per layer	Batch size	Learning rate	Train	Validation	Test
1	MLW	4	467	64	9.6E-4	0.19	0.34	0.23
2	BFG	5	30	94	2.5E-3	0.20	0.26	0.18
3	DPK	5	94	124	3.0E-3	0.21	0.24	0.23
4	PTS	6	143	124	7.5E-4	0.20	0.22	0.23
5	XLS	5	170	29	1.0E-2	0.40	0.45	0.33
6	Whole	7	836	145	2.5E-2	0.24	0.33	0.23

" (Table S1 in Supplementary Materials)

Specific comments:

Line 40-42: The reference for each process-based model will be better.

Response: Corrected.

Revision: "Process-based lake **thermodynamics** models, such as the Freshwater Lake model (FLake) **(Mironov et al., 2010)**, the General Lake Model (GLM) **(Hipsey et al., 2019)**, and **the lake thermodynamics model in Weather Research & Forecasting Model (WRF-Lake) (Gu et al., 2015)**, are built on relationships between climate variables and LST, often employing simplified assumptions based on empirical physical principles (Mironov et al., 2010; Piccolroaz et al., 2024; L. J. Xu et al., 2016)." (Section 1, Lines 42-46)

References added:

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.

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.

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.

Line 116-117 and 120-125: How to fill the gap by the ERA5 reanalysis dataset was ambiguous. For example, what was the deficit rate from 2012 to 2015? Please rewrite this explanation.

Response: Sorry for missing this information. We first provided a conceptual figure to describe the way of using the ERA5 dataset to fill the observations (Figure R2). We used the most straightforward method, which involved checking and replacing missing data in observations with ERA5 datasets for each variable, as the two datasets share a similar probability distribution in their meteorological variables (Figure S1). Then, we calculated the deficit rate (missing length/length of time series) for observations at each lake site from 2012 to 2015. Specifically, 475 time steps (~1.36%) of observed surface pressure were found to be lacking in the DPK site during 2012 and 2015; 7959 time steps (~22.71%) of all observed variables were missing in the XLS site; 12539 time steps (~35.78%) of all observed variables were missing in the PTS site; Observations at the MLW and BFG sites were complete during the study periods. More details about using ERA5 datasets in this study were provided in Materials and methodology (Section 2.1, Lines 114-139) and Figure S1.

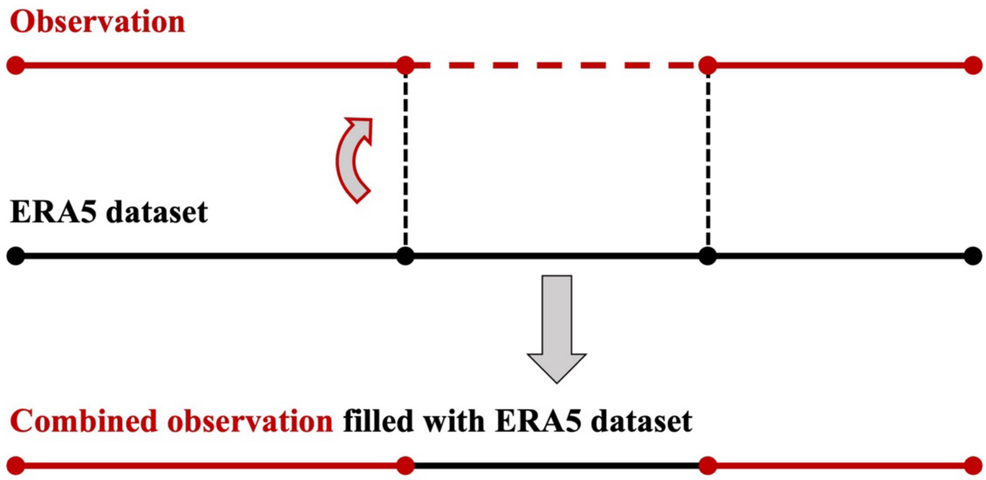


Figure R2: Conceptual diagram for gap filling of the observations by using ERA5 dataset.

Revision: "The datasets included two parts: (1) hydrometeorological variables observed from the Taihu Lake Eddy Flux Network to force and validate the models, and (2) meteorological variables from ERA5 datasets to fill the gaps of observations and force the models. Within the network, each site is equipped with an eddy covariance system that continuously monitors **LE and HE** using sonic anemometers and thermometers (Model CSAT3A; Campbell Scientific, Logan, UT, USA) positioned 3.5 to 9.4 m above the lake surface. Hydrometeorological variables, including air humidity and temperature (Model HMP45D/HMP155A; Vaisala, Helsinki, Finland), wind speed (Model 03002; R.M. Young Co., Traverse City, MI, USA), and net radiation components (Model CNR4; Kipp & Zonen, Delft, the Netherlands), are also measured. **These meteorological variables were used to force lake models while LE, HE and LST from observations were used to validate the results of each numerical experiment, on top of which, the** inferred radiative LST were collected at 30-minute intervals that are publicly accessible via Harvard DataVerse (Lee, 2004; Zhang et al., 2020; https://doi.org/10.7910/DVN/HEWCWM). The dataset spans from 2012 to 2015 **and contains several data gaps across these lake sites. Specifically, 475 time steps (~1.36%) of observed surface pressure were found missing at the DPK site during 2012 and 2015; 7,959 time steps (~22.71%) of all observed variables were missing at the XLS site; 12,539 time steps (~35.78%) of all observed variables were missing at the PTS site. Observations at the MLW and BFG sites were complete during the entire study periods. For the model evaluation of Taihu-obs experiment, the data gaps of observed variables in these lake sites were directly filled by ERA5 datasets at the corresponding time steps, which were used to predict lake-atmosphere interactions.** In this study,

observed meteorological variables from the MLW site**, an eutrophic lake site that presents the trophic status of Lake Taihu (Table 1, Wang et al., 2019),** are used to train the Long Short-Term Memory (LSTM)-based surrogates (Sect. 2.2); while data from the remaining sites serve to evaluate the generalization of HyLake v1.0 and train the LSTM-based surrogates. To **further** address the generalization **and transferability** of HyLake v1.0 across different forcing datasets, this study utilized **8 meteorological variables that obtained from** hourly ERA5 datasets from 2012 to 2015, with a spatial resolution of 0.25° at a single level to **force** HyLake v1.0. These datasets, available from the Climate Data Store (Hersbach et al., 2020; http://cds.climate.copernicus.eu), include variables such as air temperature, dew point temperature, surface pressure, wind speed, and surface net longwave and shortwave radiation**, which has similar probability distribution to observations across Lake Taihu (Figure S1).**The ERA5 datasets are **also individually** used to **force FLake and TaihuScene for comparison and predict lake-atmosphere interactions** in Lake Taihu, providing insights into the model's generalization, transferability and performance **using** different climatic forcing datasets.” (Section 2.1, Lines 114-139)

“

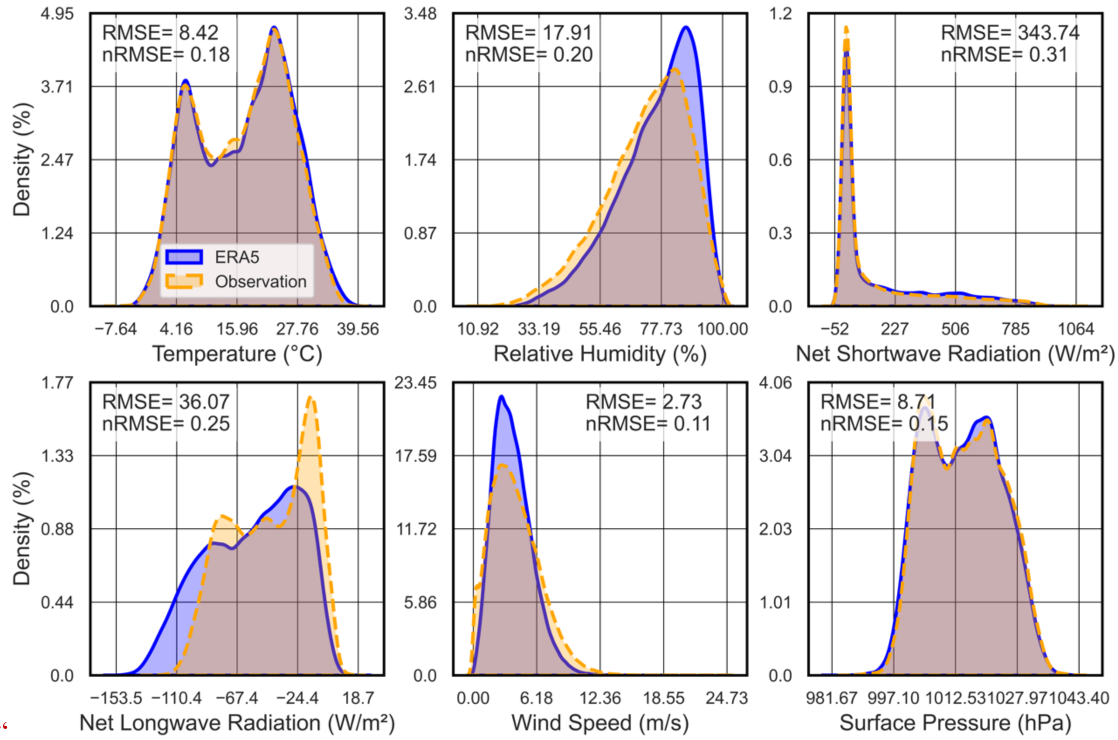


Figure S1: The probability density distribution of meteorological variables from observation and ERA5 reanalysis datasets in MLW, BFG, DPK, PTS, and XLS site during 2012 to 2015. A normalized RMSE (nRMSE) was assigned to assess the error between observation and ERA5 reanalysis datasets.” (Figure S1 in Supplementary Materials)

Line 120-125: In addition to the above comment, when the ERA5 reanalysis gaps the data at the MLW site, is this the self-validation? Please clarify.

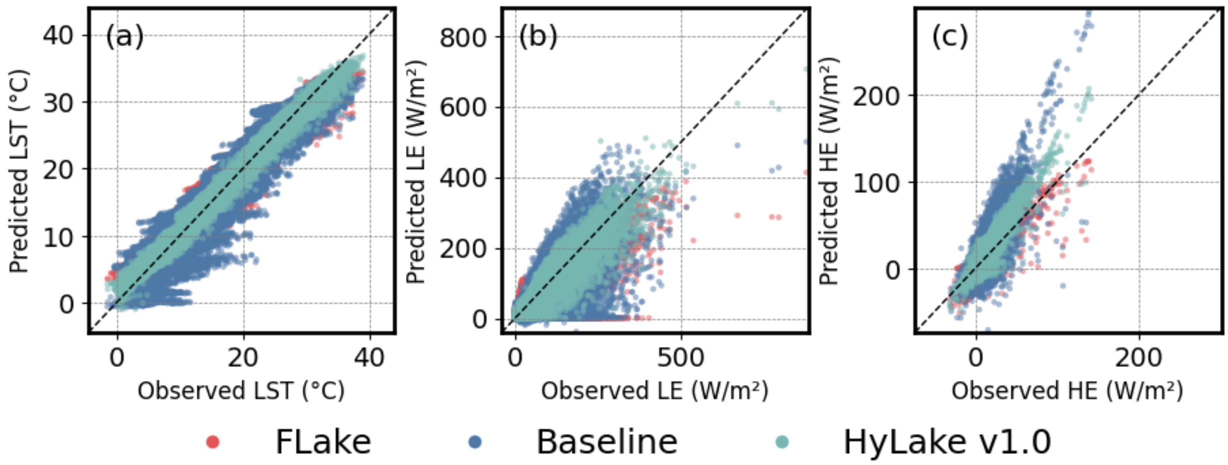
Response: The MLW site has complete observations for 2013 and 2015, which DOESN’T require any gap filling with ERA5 datasets. We only checked and filled the meteorological variables from observations, including air temperature, relative humidity, surface pressure, wind speed, and surface net longwave and shortwave radiation, with ERA5 datasets to force HyLake v1.0 and other lake models in DPK, PTS, and XLS sites during the studied period if needed.

Revision: “In the evaluation of all observations-forced experiments, the data gaps of observed variables in these lake sites were directly filled by ERA5 datasets at the corresponding time steps to predict lake-atmosphere interactions.” (Section 2.1, Lines 127-129)

Line 345: The legend “HyLake-baseline” will be confusing. I would like to recommend expressing “Baseline”.

Response: Corrected.

Revision:



"

Figure 2: Comparison of predicted (a) LST, (b) LE and (c) HE by using FLake (red points), Baseline (blue points) and HyLake v1.0 (green points) in MLW experiments." (Figure 5)

Technical comments:

Line 29: "surface water temperature" will not match the abbreviation of "LST". Is this "lake surface temperature"? Please confirm.

Response: Thanks. Corrected to "lake surface temperature (LST)".

Line 110: No need to repeat these abbreviations.

Response: Corrected.

Revision: "**Within the network, e**ach site is equipped with an eddy covariance system that continuously monitors **LE and HE** using sonic anemometers and thermometers (Model CSAT3A; Campbell Scientific, Logan, UT, USA) positioned 3.5 to 9.4 m above the lake surface." (Lines 116-118)