Prediction of Surface Water Temperature and Its Spatial-Temporal Variation Characteristics of 11 Main Lakes in Yunnan–Guizhou Plateau

Shanshan Xu, Kun Yang, Xiaolu Zhou, Jie Zhang, *Student Member, IEEE*, Chenghao Yang, and Yi Luo

***Abstract*—In the context of global warming, the lake surface water temperature (LSWT) exhibits a general upward trend. As an indicator of climate change, the surface water temperature of plateau lakes is particularly sensitive to climate warming. In this study, the machine learning model and physical process model are combined to predict water temperature in lake systems: The Air2water model and the long short term memory (LSTM) model are combined to form a prediction model for the inversion and prediction of 11 typical plateau lakes in the Yunnan–Guizhou Plateau. The modeling results show that the Air2water model performs best, followed by the LSTM model. Overall, the Air2water model and the LSTM model effectively reproduce the monthly, seasonal, and interannual variations of LSWT dynamics in the 11 lakes. The research results show that the LSWT dataset from 2021 to 2025 is constructed based on the forecast result, and it is found that six lakes, including Caohai, Dianchi, Erhai, Qionghai, Xingyun Lake, and Yangzonghai, showed significant warming in the next five years (the maximum warming rate was less than +0.2 °C/year), while the remaining five lakes showed no significant changes, and the size classification of LSWT data in different spaces on the interannual scale was consistent with the classification of altitudes. By conducting extensive exploration and research on the inversion of LSWT using physical process models and machine learning models, this study offers novel solutions for LSWT prediction and inversion. The study also provides a solid theoretical and practical foundation for future LSWT research, thereby holding practical significance and research value.**

***Index Terms*—Air2water, climate change, global warming, lake surface water temperature (LSWT), long-short term memory (LSTM).**

Manuscript received 13 August 2023; revised 16 October 2023; accepted 6 November 2023. Date of publication 29 November 2023; date of current version 3 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42271441, in part by The Ten Thousand Talent Plans for Young Topnotch Talents of Yunnan Province under Grant YNWR-QNBJ-2019-200, and in part by Yunnan Province Innovation Team Project under Grant 202305AS350003. *(Corresponding author: Yi Luo.)*

Shanshan Xu is with the Engineering Research Center of GIS Technology in Western China, Ministry of Education of China, Yunnan Normal University, Kunming 650500, China, and also with the School of Geographic Sciences, East China Normal University, Shanghai 200062, China (e-mail: xss_5233@163.com).

Kun Yang, Chenghao Yang, and Yi Luo are with the Engineering Research Center of GIS Technology in Western China, Ministry of Education of China, Yunnan Normal University, Kunming 650500, China (e-mail: yangkun@ynnu.edu.cn; mlych847497053@163.com; lysist@ynnu.edu.cn).

Xiaolu Zhou is with the Department of Geography, Texas Christian University, Fort Worth, TX 76109 USA (e-mail: xiaolufile@gmail.com).

Jie Zhang is with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: jiezhang@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2023.3335291

I. Introduction

WATER temperature is one of the most important indicators of lake ecosystems, controlling physical processes (e.g., thermal stratification, mixing processes), chemical processes (e.g., chemical reaction rate, oxygen solubility), and ecological processes (e.g., biological metabolism, growth, and reproduction) in lakes and have significant effects on physical, biological, and chemical processes throughout the lake [1], [2]. The lake surface water temperature (LSWT) refers to the water temperature within the uppermost layer (0–1 m) of a lake, which is one of the parameters that characterize the water temperature of the lake, serving as an important physical parameter in the field of lake ecosystem research [3], [4]. LSWT plays a main role in controlling the physical, chemical, and ecological processes of lakes, with significant implications on lake water quality and ecosystem function [5]. LSWT plays a critical role in influencing the ecological environment of lakes and serves as a vital indicator of the impact of global climate change [6], [7], [8] and its key driver is temperature [9], [10] influenced by other objective factors such as solar radiation [11], relative humidity [12], ice sheet [13] etc, also influenced by local factors such as cloud cover and lake morphology [14], [15], human activities [16] and surface runoff [17], etc. Matulla et al. [16] used Alpine lakes to show that the extent to which anthropogenic climate (atmospheric variables based on different data related to the hydrosphere) changes in LSWT did not begin to decline until mid-1980s. Oswald et al. [18] showed that shallow lakes have a tendency to warm up more rapidly compared to deep lakes due to their lower thermal storage capacity, especially at higher surface water temperatures. However, Sharma et al. [19] proposed that climatic factors have a greater effect on surface water temperature than lake morphology on a wide spatial scale. Woolway et al. [15] showed that the response of lakes to climate change is closely related to their location and scale, and lakes with different locations and scales have different responses to climate change. Studies have shown that the global lake LSWT is rising at a rate of about 0.3 °C per decade [14], [20], [21]. Significant transformations in Earth's freshwater resources and their associated processes are already in progress. Along with this trend, significant changes in water temperature may have serious consequences for lake ecosystems [22]. Due to the limitations of inverting existing remote sensing data, which restricts the ability to capture historical changes in LSWT and forecast future trends, as well as the lack of adequate spatiotemporal coverage for in situ measurements, an increasing number of researchers are now focusing on the development of LSWT-oriented models. Piccolroaz et al. [23] developed an Air2Water model that only needs to input air temperature for LSWT models that are plagued by multiple uncontrollable factors, which is simple and the output LSWT has high accuracy, is widely used by researchers for LSWT prediction and can easily be used to predict the lake's response to climate change. Indeed, as highlighted by Piccolroaz et al. [24], making reliable predictions regarding the evolution of climate change is challenging. Therefore, O'Beirne et al. [25] studied LSWT response to climate change from the perspective of climate multifactor trends, and Yankova et al. [26] studied the response of lakes to climate change from a thermodynamic perspective. Bruce et al. [27] investigated the response of LSWT to complex thermodynamic fluxes. Their approach involved utilizing process-based numerical models to accurately quantify these responses, which necessitated the utilization of detailed on-lake meteorological data (such as wind speed, humidity, and cloud cover) as input parameters. The current complex input parameters make neural networks widely favored in model building. Zhu et al. [28] designed a multilayer perceptron neural network (MLPNN) model, along with a wavelet transform and MLPNN ensemble model (WT_MLPNN), to predict LSWT and compared the results with the Air2Water model based on physical statistics and the nonlinear regression model, and the results showed that the Air2Water model was optimal. On the basis of these studies, Zhu et al. [29] estimated the LSWT of the 25 Polish lowland lakes based on the Air2Water model. Fabio et al. [30] designed a novel machine learning algorithm called multilayer perceptron and random forest (MLP-RF) to predict LSWT, where its predictive ability is much better than existing models such as Air2Water. Yu et al. [31] constructed prediction model based on support vector regression, analytic hierarchy and backpropagation artificial neural networks, combining meteorological data, human factors and lake intrinsic data to predict lake LSWT changes. LSWT inversion and prediction face challenges due to the complex interactions among multiple factors. These complexities make the task of inverting and predicting LSWT more challenging. However, it is important to note that changes in LSWT have significant implications for lake stratification and seasonal deep convection. They impact various aspects such as mixing time, duration, and intensity, which are crucial factors in LSWT inversion and prediction.

In this study, we mainly apply the long short-term memory (LSTM) model and the Air2Water model to predict the LSWT of 11 major lakes in the Yunnan–Guizhou Plateau: First, long-term observed monthly average air and LSWT data for 11 major lakes in the Yunnan–Guizhou Plateau in China were collected and established. Then, performance of the two models was compared and analyzed under monthly, quarterly, and yearly spatial and temporal dimensions for the 11 major lakes in the Yunnan–Guizhou Plateau, the effects of natural and anthropogenic factors on the LSWTs of these plateau lakes are elucidated. Finally, a high-precision dataset of LSWTs of 11 major lakes in the Yunnan–Guizhou Plateau from 2021 to 2025 is constructed by combining the LSTM and Air2Water models, the expected trends of the LSWT were analyzed during this period.

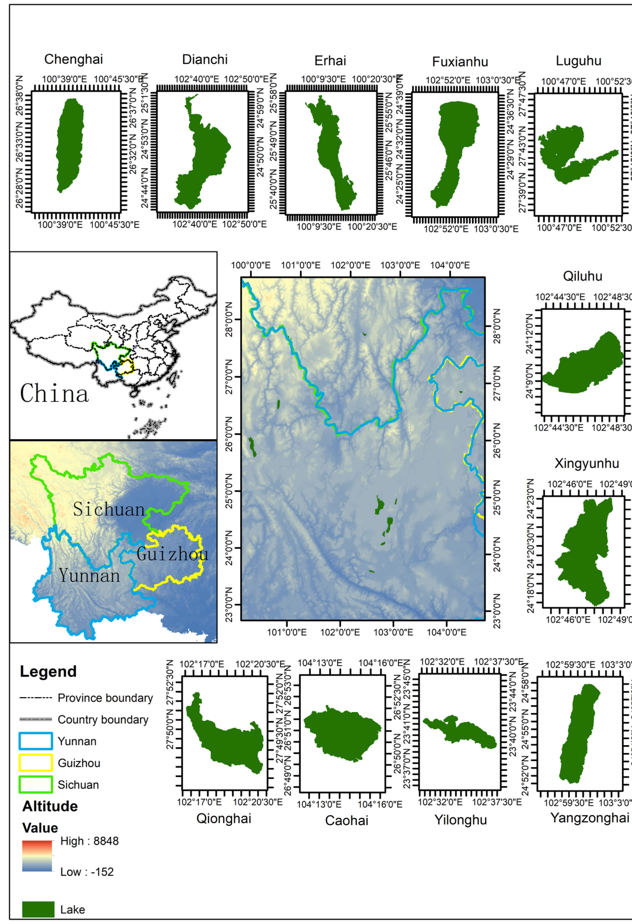


Fig. 1. Study area map.

II. Materials and Methods

A. Study Area and Data

This study focuses on the 11 primary lakes located in the Yunnan–Guizhou Plateau, as illustrated in Fig. 1. The Yunnan–Guizhou Plateau, characterized by a substantial distribution of freshwater resources, encompasses nine lakes within Yunnan Province, namely Dianchi, Erhai, Fuxian Lake, Chenghai, Lugu Lake, Qilu Lake, Xingyun Lake, Yilong Lake, and Yangzonghai. In addition, Qionghai Lake in Sichuan Province and Caohai Lake in Guizhou are included. Together, these 11 lakes, covering central and northwestern Yunnan, account for approximately 85% of the total lake area. Dianchi Lake, located southwest of Kunming City, represents a significant lake in Yunnan Province, while Erhai Lake, situated on the outskirts of Dali, ranks as the province's second-largest freshwater lake. Fuxian Lake, the largest deep-water freshwater lake in China, holds a prominent position as a natural lake. Chenghai Lake, located in Yongsheng County, Lijiang City, is the second-largest freshwater lake in western Yunnan. Lugu Lake, situated at a higher altitude in the southwest area of Sichuan Province, exhibits a subtropical character. Qilu Lake falls into the category of rifted lakes, while Xingyun Lake, Yilong Lake, and Yangzonghai Lake are classified as plateau rifted lakes. Caohai Lake, the largest natural lake in Guizhou, ranks among the three major plateau freshwater lakes in China. Similarly, Qionghai Lake in Sichuan Province represents the second-largest freshwater lake and an early rifted lake. The 11 main lakes in the Yunnan–Guizhou Plateau have played a vital role in nurturing the area's biodiversity, climatic diversity, and cultural richness. Unfortunately, the discharge of industrial wastewater and domestic sewage has resulted in the deterioration of lake water quality. Consequently, it is of utmost significance to investigate the influence of natural environmental distribution and human activities on LSWT for effective lake environment preservation.

This study covers a time span from January 2001 to December 2020 and focuses on various datasets related to the 11 main lakes in the Yunnan–Guizhou Plateau. The key data sources include information on lake boundary and area, LSWT, near-surface air temperature, and other relevant variables. The lake boundary data were obtained by utilizing Landsat Series TM, ETM+, and OLI remote sensing image data from the geospatial data cloud. LSWT data were extracted from MOD11A2 images, which are 1 km resolution 8-day synthetic level three products obtained through NASA's Earth Observation System (EOSDIS). In addition, near-surface air temperature data were acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF), providing 2 m atmospheric temperature data at a resolution of 0.125°× 0.125°.

B. Air2Water Model

The model is a simple tool for predicting LSWT only when air temperature is available, can correctly simulate hysteresis loops of air and water temperature in shallow and deep lakes, and accurately captures seasonal and interannual fluctuations in LSWT. The model is based on the volumetric integrated thermal balance equation that simulate stratification dynamics in lakes without the need for complex descriptions of air–water interface processes based on detailed quantification of individual heat flux components. Since the lakes in the Yungui Plateau were not frozen, a 6-parameter model was selected for the entire study. Simple linearization of heat flux terms by using Taylor unfolding and using air temperature instead of the combined effects of related processes and fluxes [23], [32]. Six-parameter version [23] (a_1–a_6) can be expressed as follows:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left\{a_1 \cdot \cos 2\pi(\frac{t}{t_y} - a_2) + a_3 + a_4(T_a - T_w) + a_5 T_w\right\} \tag{1}$$

$$\text{where } \delta = \begin{cases} \exp \frac{T_r - T_w}{a_6} & , T_w \leq T_r \\ 1 & , \text{else} \end{cases} \tag{2}$$

where T_w is LSWT, T_r is the deep lake temperature, T_a is air temperature, t is time, and t_y is the duration of the year in suitable time units. σ is a dimensionless number representing the ratio between the volume of the surface lake layer and a reference volume, and a_1–a_6 are parameters that can be estimated by model calibration using the observed LSWT data.

C. LSTM Model

The LSTM model is a type of recurrent neural network (RNN) that is specifically designed for learning and predicting based on long-term sequential data. Unlike traditional RNNs, LSTM incorporates a memory mechanism that allows it to capture and utilize historical information for predicting current output conditions, even over long distances in the sequence. This is achieved by introducing a cell state, also known as the "memory cell," to the hidden layer of the original RNN. The LSTM model takes three inputs: 1) the current input value, 2) the output value from the previous time step, and 3) the cell state from the previous time step. It produces two outputs: 1) the current output value and 2) the current cell state. The model utilizes "gates" in its fully connected layers to control the flow of information. Each cell in the LSTM is governed by three types of gates: 1) the forgetting gate, 2) the input gate, and 3) the output gate. The forgetting gate determines which parts of the previous cell state to retain and which to discard, enabling the model to handle long-range dependencies and effectively address the limitations of traditional RNNs

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4}$$

$$\mathfrak{c}_t = tanh\left(W_{\mathfrak{c}} \cdot [h_{t-1}, x_t] + b_{\mathfrak{c}}\right) \tag{5}$$

$$c_t = f_t \bigodot c_{t-1} + i_t \bigodot \mathfrak{c}_t \tag{6}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{7}$$

$$h_t = o_t \bigodot tanh(c_t) \tag{8}$$

W_f is the weight matrix of the forgetting gate, W_i is the weight matrix of the input gate, W_c is the weight matrix of the calculation unit, and W_o is the weight matrix of the output gate; b_f is the bias term of the forgetting gate, b_i is the bias term of the input gate, b_c is the bias term of the calculation unit state, and b_o is the bias term of the output gate; h_{t-1} is the output and cell state of the previous moment, x_t is the input of the current moment, $\bigodot$ refers to the dot product of the matrix elements; f_t is the current memory of the forgetting gate, i_t is the current memory of the input gate, c_t is the current memory of the state of the computing unit, o_t is the current memory of the output gate, and h_t is the current output value.

D. Model Performance Evaluation

In this study, the models were utilized to classify 11 lakes based on the error between the estimated and observed values. The dataset was divided into two subsets, with 70% of the data used for model training and the remaining 30% for model validation. The accuracy of the prediction was assessed using key metrics. The root mean square error (RMSE) served as a reliable indicator of prediction accuracy. The mean absolute error (MAE) quantified the average difference between the predicted and true values. In addition, the standard deviation (SD) provided insights into the dispersion of values relative to the mean. These measures, namely RMSE, MAE, and SD, were calculated using the equations as shown in the accompanying diagrams

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{\mathrm{obs},i} - X_{\mathrm{model},i}\right)^2}{n}} \tag{9}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|X_{\mathrm{model},i} - X_{\mathrm{obs},i}\right| \tag{10}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}} \tag{11}$$

where $i = 1, \ldots, n$, n is the number of measurements; $X_{\mathrm{obs},i}$ refers to the i time observation value; $X_{\mathrm{model},i}$ refers to the i time simulated value; σ reflects the degree of dispersion of a dataset; X_i represents the real numbers in the dataset; $\bar{X}$ represents the average value of the dataset.

III. Results and Discussion

The objective of this study is to conduct experimental research on 11 lakes situated in the Yunnan–Guizhou Plateau. The objective of the study is to analyze outcomes and assess the effectiveness of the applied models in the context of lake environments. First, the classification results of the models employed in the Yunnan–Guizhou Plateau lakes are summarized, and a comprehensive analysis of the newly constructed LSWT dataset is conducted. Next, a detailed assessment of the Air2Water model and the LSTM model are performed, considering their performance on various temporal scales, including monthly, seasonal, and annual variations. Further, by comparing the simulated and observed data from both models, the strengths and weaknesses of simulating LSWT using physical models and machine learning algorithms are determined. This comparative analysis aims to evaluate the accuracy of the models. By conducting experiments and simulations on lakes located in the Yunnan–Guizhou Plateau, this study evaluates the predictive powers of various models in capturing lake characteristics. The ultimate goal is to determine the most reliable model for simulating the distribution of LSWT.

A. Model Classification on Spatiotemporal Scales

This study assesses the suitability of both models in predicting lake water temperature by examining the trends in error between simulated and observed data. The aim is to identify the most suitable model for different time periods based on these findings. By analyzing the discrepancies between the simulated and observed data, the study aims to determine the model that exhibits the highest accuracy in predicting lake water temperature.

The presented Tables I and II provide a comprehensive overview of the model performance for the 11 lakes, considering monthly, quarterly, and annual averages. It is evident from the tables that the Air2water model demonstrates better suitability for lakes with a higher sensitivity to temperature, while the LSTM model shows superior performance during May to June and November. Furthermore, the seasonal average analysis reveals that during the first quarter, the Air2water model outperforms the LSTM model for all lakes, suggesting that these lakes are primarily influenced by air temperature in spring and less affected by other factors. In the second and third quarters, the majority of lakes exhibit better results with the LSTM model, indicating that during the warm seasons of summer and autumn, the LSWT is influenced by various factors, including internal lake dynamics. In the fourth quarter, a mix of lakes show better performance with either the Air2water or LSTM model. This study analyzes the RMSE, MAE, and SD error values between the estimated and observed data of the two models for different time periods. Based on this analysis, a new dataset is constructed, and its predictive ability for future data variations is examined.

TABLE I
Model Selection Results on the Monthly Mean Scale

Month	1	2	3	4	5	6	7	8	9	10	11	12
Caohai	A	A	A	A	A	L	A	A	L	L	A	A
Chenghai	A	L	L	A	A	A	L	L	A	A	L	A
Dianchi	A	A	A	A	L	L	A	A	A	A	L	A
Erhai	L	L	A	A	L	A	L	A	A	L	L	A
Fuxianhu	A	A	A	A	L	A	A	L	A	L	L	A
Luguhu	L	A	A	A	A	L	L	A	L	A	A	L
Qiluhu	L	A	A	A	A	A	A	A	A	L	L	A
Qionghai	L	L	A	A	L	L	A	L	A	A	L	A
Xingyunhu	A	A	A	A	L	L	A	L	A	A	L	A
Yangzonghai	L	A	A	A	L	L	A	L	A	A	L	A
Yilonghu	L	A	L	L	L	A	A	A	A	L	L	A

“A” represents the Air2water model, “L” represents the LSTM model.

TABLE II
Model Selection Results on the Seasonal and Annual Scales

	1Q	2Q	3Q	4Q	Yearly
Caohai	A	L	L	A	A
Chenghai	A	L	A	L	A
Dianchi	A	A	L	A	A
Erhai	A	L	L	A	A
Fuxianhu	A	A	L	A	A
Luguhu	A	L	L	L	A
Qiluhu	A	A	A	L	A
Qionghai	A	L	L	A	A
Xingyunhu	A	A	A	A	A
Yangzonghai	A	L	A	L	A
Yilonghu	A	A	A	L	L

“A” represents the Air2water model, “L” represents the LSTM model.

B. Trend Analysis of New Datasets

Through the analysis of classification data based on the models, it is evident that both the LSTM and Air2water models exhibit the poorest performance on Lugu Lake. During the

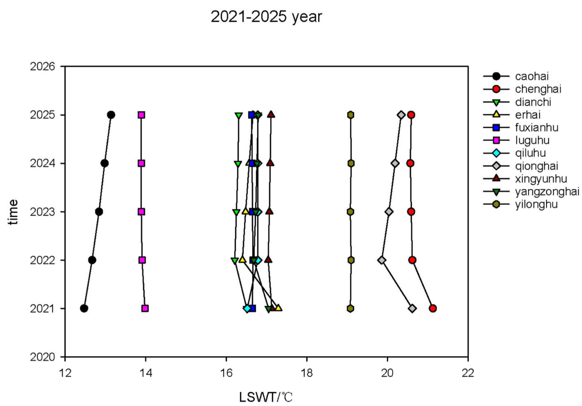


Fig. 2. Variation trend of the annual LSWT of 11 lakes in the next five years.

experiment, systematic errors were observed in both models. As the monthly average temperature data from January 2021 to December 2025 is predicted by the LSTM model, the Air2water model utilizes these predicted air temperatures to estimate the monthly average LSWT data within the same timeframe. Therefore, the estimated data will reflect the combined systematic error of both the LSTM and Air2water models. However, due to the intricate interannual thermal behavior and the considerable lake depth of Lugu Lake, the interannual variation of LSWT deviates significantly from a sinusoidal curve. The error values of the two models in capturing the interannual change curve of LSWT are substantial. As a result, the LSTM model demonstrates superior performance in predicting future data, reducing the systematic error associated with the Air2water model and mitigating the reproducibility error caused by the lake's thermal behavior during estimation. Consequently, the future dataset predicted by the LSTM model is solely adopted based on the data for Lugu Lake.

Fig. 2 shows the annual average scatter plot of LSWT in 11 lakes from 2021 to 2025. We use the newly built dataset to analyze the future trend changes in the LSWT.

Fig. 2 reveals notable warming trends in Caohai, Dianchi, Erhai, Qionghai, Xingyun Lake, and Yangzonghai over the next five years. Caohai and Qionghai exhibit a change rate of approximately 0.2 °C/year, while Dianchi Lake and Xingyun Lake show a change rate of around 0.02 °C/year. Erhai and Yangzonghai demonstrate a change rate of approximately 0.1 °C/year. Conversely, Qilu Lake's overall LSWT shows no significant change, indicating a relatively stable interannual variation trend. Yilong Lake also exhibits a minimal overall change trend in the future, with an estimated change rate of approximately 0.02 °C/year. In addition, Fig. 2 illustrates that Caohai and Lugu Lake have annual temperatures below 15 °C, while Chenghai, Qionghai, and Qilu Lake experience average temperatures exceeding 18 °C annually. The other lakes fall within the temperature range of 16–18 °C. Regarding altitude, lakes such as Dianchi, Erhai, Fuxian Lake, Qilu Lake, Xingyun Lake, and Yangzonghai are situated between 1700 and 2000 m above sea level. Chenghai, Yilong Lake, and Qionghai are located between 1400 and 1600 m above sea level. Caohai and Yilong Lake are positioned above 2100 m, indicating a decrease in LSWT with increasing altitude.

C. Explore the Performance of Air2water Model on the Spatial-Temporal Scales

In this study, the Air2water model, which functions as the mechanism model, is employed. Before training the model, we incorporate the average depth of each lake as a preprocessing input. This assists in determining the six parameters that are required for the model. Fig. 3 illustrates the satisfactory performance of the Air2water model on the monthly averaged data for the 11 lakes, as indicated by low RMSE, MAE, and SD values. The RMSE ranges from 0.2 to 2.65 °C (mean: 1.04 °C), the MAE ranges from 0.17 to 2.36 °C (mean: 0.89 °C), and the SD ranges from 0.38 to 2.07 °C (mean: 0.93 °C).

Fig. 3 reveals notable variations in model performance among the lakes. Fuxian Lake demonstrates superior performance, while Lugu Lake, Qionghai, Chenghai, Erhai, and Caohai exhibit relatively poorer performance. The model's performance in Yilong Lake, on the other hand, is comparatively better, suggesting its effectiveness in shallow lakes. Nonetheless, the model's performance in other lakes varies. Notably, adaptability is not solely determined by lake depth and water storage capacity, as evidenced by the model's favorable performance in both Fuxian Lake and Yilong Lake. The red areas in the figure indicate regions of poor model performance, highlighting overall deficiencies during June to August. Conversely, the darker blue areas represent improved model performance, particularly during January to April and December.

Fig. 4 presents the performance indicators of the seasonally averaged Air2water model for the 11 lakes. The graphs clearly illustrate the changing trends of the three error values on a seasonal basis, indicating an overall declining trend. From a seasonal perspective, all lakes exhibited improved performance during the model's validation period. The RMSE ranged from 0.16 to 1.63 °C (mean: 0.76 °C), the MAE ranged from 0.13 to 1.41 °C (mean: 0.66 °C), and the SD ranged from 0.22 to 1.07 °C (mean: 0.61 °C).

When comparing Fuxian Lake and Yilong Lake in Fig. 4, it can be observed that the model performance of Fuxian Lake is better in the first and fourth quarters, while the model performance of Yilong Lake is better in the second and third quarters. This difference can be attributed to the stratification phenomenon in Fuxian Lake from June to November, which exacerbates the thermal behavior of the lake and reduces the model's estimation performance. Other lakes generally showed poor performance in the second quarter, whereas Lugu Lake performed poorly in the fourth quarter. Overall, all lakes exhibited poorer performance in the second and third quarters, while performing better in the first and fourth quarters. Among them, Fuxian Lake performed exceptionally well in the first and fourth quarters. This can be attributed to Fuxian Lake's subtropical monsoon climate, which exhibits stable weather conditions and less pronounced interannual stratification phenomena. Therefore, the model is highly suitable for predicting and estimating the LSWT of lakes similar to Fuxian Lake.

Fig. 5 illustrates the performance indicators of the Air2water model for the 11 lakes on an annual basis. The graph indicates that the average annual RMSE for all the studied lakes ranges

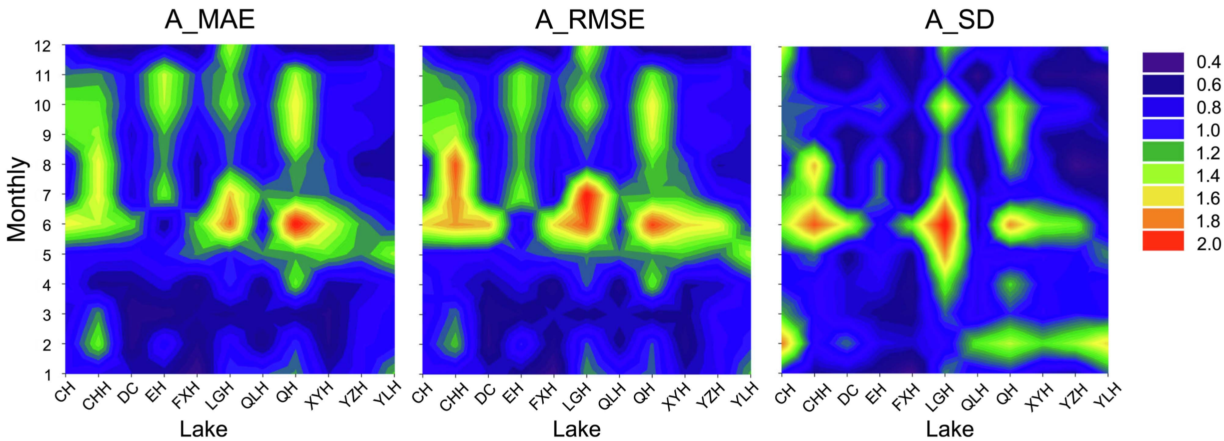


Fig. 3. RMSE, MAE, and SD of the monthly average of the Air2water model.

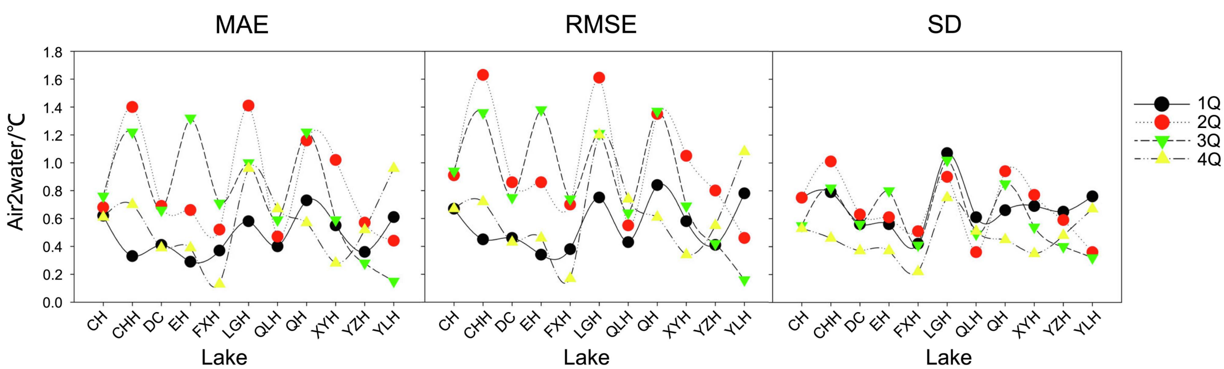


Fig. 4. RMSE, MAE, and SD for the seasonally of the Air2water model.

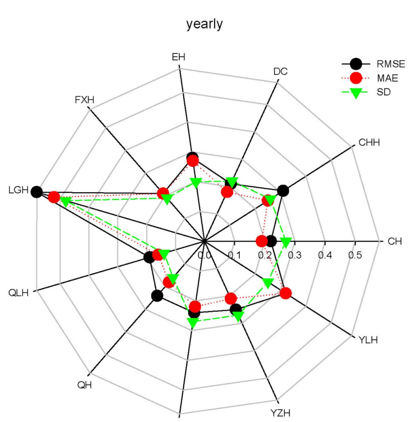


Fig. 5. RMSE, MAE, and SD for the annual average of the Air2water model.

from 0.19 to 0.58 °C (mean: 0.28 °C), while the MAE ranges from 0.16 to 0.52 °C (mean: 0.25 °C). In addition, the SD varies between 0.14 and 0.48 °C (mean: 0.25 °C). These results provide strong evidence of the effectiveness and reliability of the Air2water model in accurately estimating LSWT.

Fig. 5 depicts a decreasing trend in the error values of the model's annual average data compared to the error values of the monthly average data. Among the lakes, Lugu Lake exhibits the poorest model performance compared to other lakes, while the performance is relatively good in the remaining lakes. The data presented in Figs. 3–5 provide insights into the model's performance across different lakes. Fig. 3 indicates that, on a monthly time scale, the model performs well for lakes such as Dianchi, Erhai, Fuxian Lake, and Yangzonghai, while demonstrating poorer performance for Chenghai, Qionghai, and Lugu Lake. The performance for other lakes falls within a moderate range. Fig. 4, illustrating the quarterly average time scale, shows that the model performs better for Fuxian Lake and Yilong Lake, but exhibits weaker performance for Chenghai, Lugu Lake, and Qionghai Lake. The model's performance for other lakes generally falls within an acceptable range. Examining the annual average data in Fig. 5, the model performs better for Qilu Lake but shows poorer performance for Lugu Lake. In summary, the model's performance in simulating LSWT follows the order of interannual > quarterly average > monthly average.

D. Explore the Performance of the LSTM Model on the Spatial-Temporal Scales

After conducting tests and evaluations using the LSTM model, we have identified the range of parameters that are

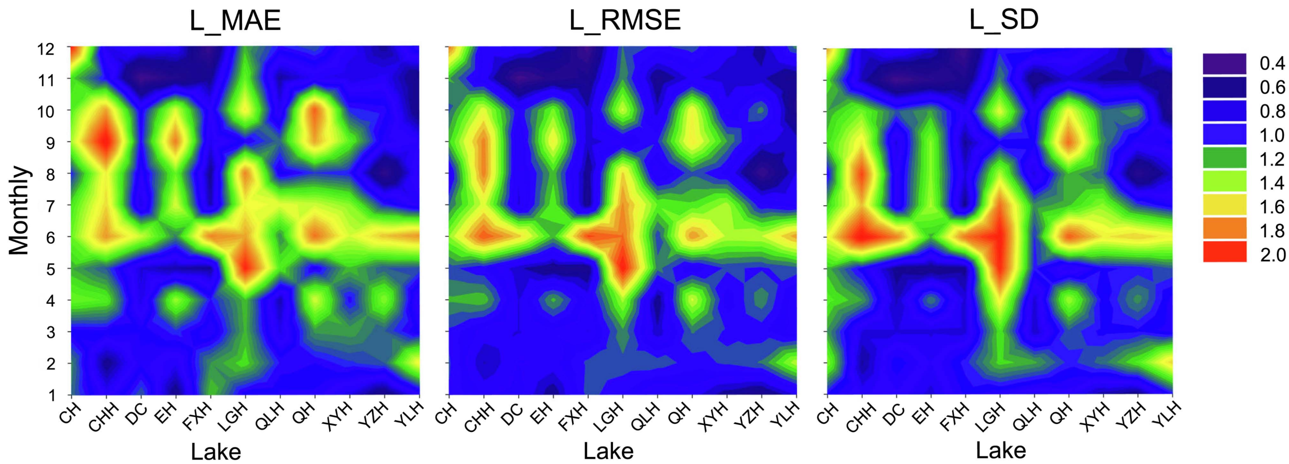


Fig. 6. RMSE, MAE, and SD of the monthly average of the LSTM model.

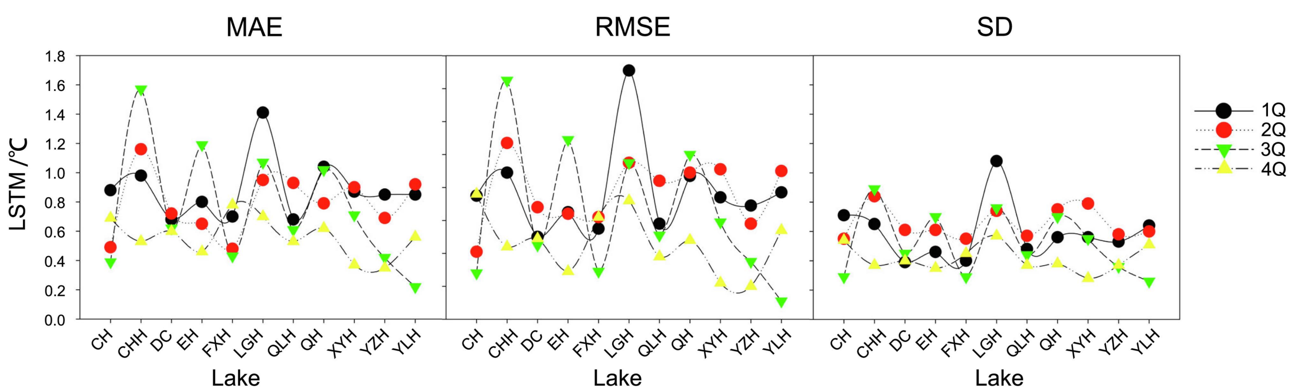


Fig. 7. RMSE, MAE, and SD of the seasonal average of the LSTM model.

suitable for this research in the solver. Taking into consideration the findings from our investigations and previous experiments, it has been determined that the optimal number of hidden neurons for each lake should be set at 200. The effectiveness of the estimates for the 11 lakes can be observed in the figure below, demonstrating that the model's performance matches with that of the underlying model used for data prediction. To assess the model's performance and usability, an error analysis is conducted using data from the validation period. Fig. 6 illustrates the error values for the 11 lakes when the LSTM model is employed for data prediction, showcasing the model's performance across different time periods for each lake. The RMSE for these lakes during the validation phase ranged from 0.33 to 2.42 °C (mean: 1.23 °C). The MAE varied between 0.25 and 2.11 °C (mean: 1.03 °C), while the SD ranged from 0.22 to 1.67°C (mean: 0.79 °C).

From Fig. 6, it is evident that the model exhibits better performance for Fuxian Lake, while Lugu Lake and Chenghai demonstrate poor performance. The data variations in Lugu Lake deviate from the sinusoidal curve, making it challenging for the model to capture the interannual LSWT changes accurately. This finding highlights the similarity between the LSTM model and the Air2water model, both of which rely on LSWT exhibiting sinusoidal curve regularity over the years to enable more precise

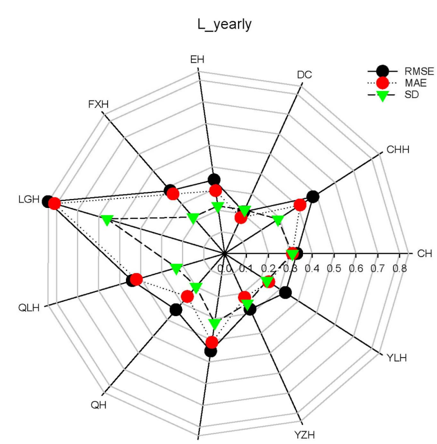


Fig. 8. RMSE, MAE, and SD of the quarter average of the LSTM model.

estimation. As depicted in the figure, the model's performance is subpar from June to October but improves during the time period of January to April and November. The poor performance observed in Chenghai, Lugu Lake, and Qionghai during this period indicates that the trend of LSWT in these three lakes is unstable and may be influenced to varying degrees by the thermal behavior of the lake and the intensity of human activities.

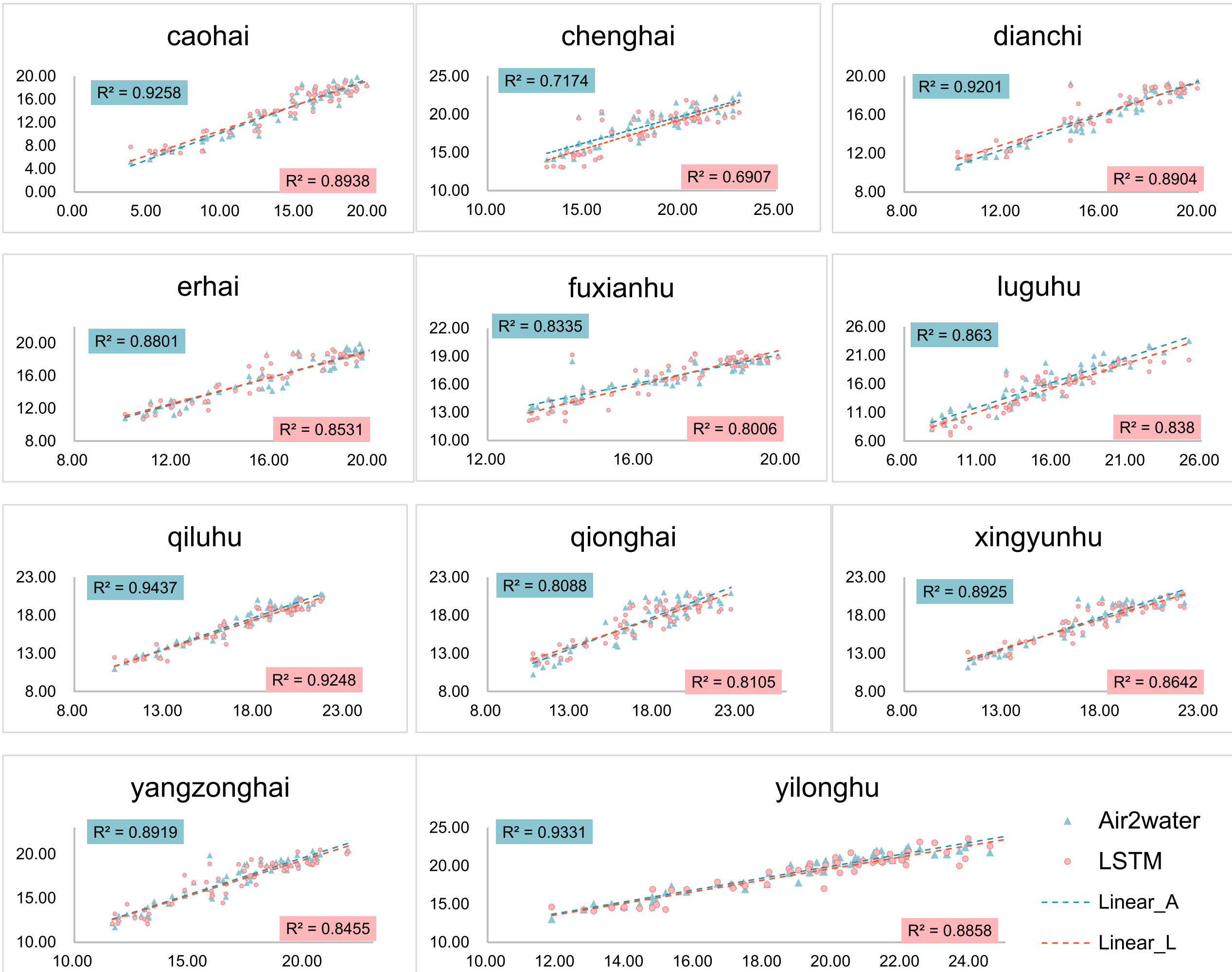


Fig. 9. Fits between the estimated data and the observed data of the Air2water model and the LSTM model.

In addition, it suggests that neither model is suitable for accurately estimating the occurrence of extreme weather events.

Fig. 7 illustrates the performance index of the LSTM model for the seasonal average, demonstrating its effectiveness in estimating data on this scale for the 11 lakes. The evaluation of performance is based on the three error values: RMSE, MAE, and SD, aiming to assess the consistency between the model's estimated quarterly average data and the observed data. During the validation period, the RMSE for all lakes ranged from 0.31 to 1.71 °C (mean: 0.86 °C), while the MAE varied between 0.22 and 1.57 °C (mean: 0.75 °C), and the SD ranged from 0.26 to 1.08 °C (mean: 0.54 °C). These findings highlight the model's ability to effectively estimate the quarterly average LSWT data for the studied lakes.

Fig. 7 provides evidence of the model's reliability in predicting LSWT, particularly in lakes such as Caohai, Dianchi, Fuxian Lake, Alder Lake, Yilong Lake, and Yangzonghai, where it demonstrates better performance on a quarterly scale. However, Lugu Lake exhibited poor performance in the first and fourth quarters, while Chenghai performed poorly in the second and third quarters. Taking into account the overall picture of the lakes, it is observed that all lakes performed better in the fourth quarter but performed worse in the second and third quarters. These observations suggest the seasonal variations in LSWT and highlight the varying performance of the model across different quarters and lakes.

Fig. 8 presents the performance indicators of the LSTM model on the average annual time scale for the 11 lakes. These data highlight the higher accuracy of the model in estimating LSWT during the validation period. Furthermore, the figure illustrates the variation in average annual error values. The RMSE of the 11 lakes ranged from 0.21 to 0.84 °C (mean: 0.40 °C), while the MAE varied between 0.18 and 0.81 °C (mean: 0.36 °C). In addition, the SD ranged from 0.20 to 0.56 °C (mean: 0.28 °C). These results emphasize the model's ability to provide reliable estimates of LSWT on an annual average basis for the studied lakes.

Fig. 8 reveals a gradient decline in data error values on the annual average scale compared to the monthly average scale. Among the three-error indicators, RMSE exhibits the highest values overall, while SD shows the smallest values, indicating that the model's estimated values are more volatile and less dispersed. By considering Figs. 6–8 collectively, the performance of the LSTM model in predicting LSWT can be assessed.

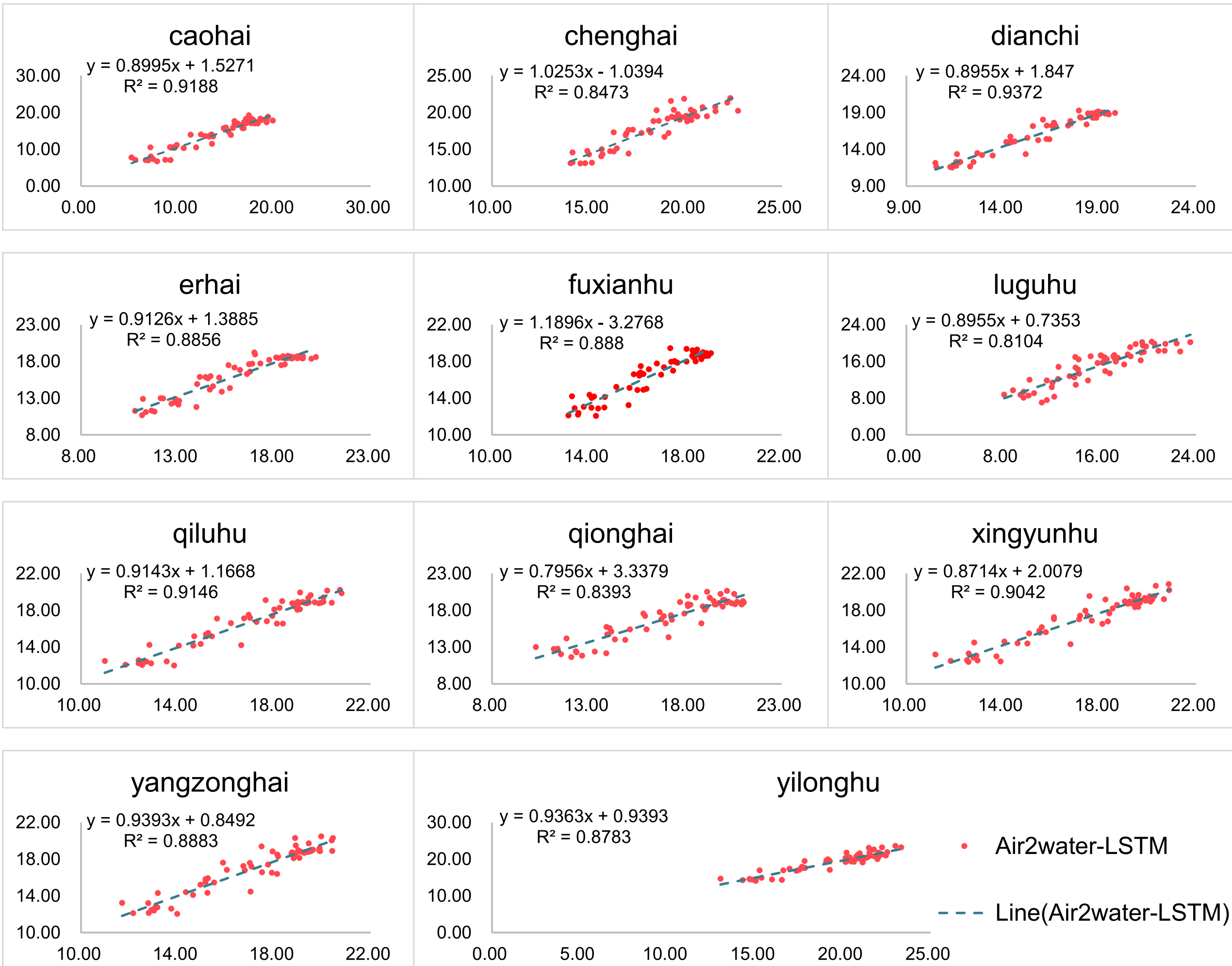


Fig. 10. Fit of the estimated data of the two models.

Fig. 6 demonstrates the model's superior performance on lakes such as Fuxian Lake, Qilu Lake, and Yangzonghai. Fig. 7 highlights the model's better performance on lakes including Caohai, Dianchi, Yangzonghai, and Yilong Lake. Finally, Fig. 8 indicates that the model performs well on Dianchi Lake. These findings provide evidence of the LSTM model's effectiveness in predicting LSWT for the selected lakes.

E. Accuracy Analysis of Estimation Data From the Two Models

Based on the findings presented in Fig. 9, the suitability of the two models in predicting LSWT was evaluated by assessing the goodness of fit between the models and the observed data. The study revealed a high degree of consistency between the two models and the observed LSWT data. The difference in R2 values between the model based on 11 lakes and the observed data ranged from 0.002 to 0.047, indicating a close match between the predicted and observed values. However, for Chenghai Lake, both models exhibited relatively lower fit, with an R2 of 0.72 for the Air2water model and 0.69 for the LSTM model. In contrast, the R2 values for the remaining lakes were greater than 0.80 for both models. These results demonstrate that the two models effectively capture the interannual variation of LSWT in the time series data of these 11 high-precision lakes. Furthermore, when considering the performance based on the 11 lakes, the Air2water model (RMSE: 0.28 °C, MAE: 0.25 °C, SD: 0.25 °C) outperformed the LSTM model (RMSE: 0.40 °C, MAE: 0.36 °C, SD: 0.28 °C).

Based on the analysis presented in Fig. 9, it can be observed that the LSTM model demonstrates a good fit in Qionghai, while both models exhibit the best fit in Qilu Lake. This suggests that the two models are well suited for predicting ecological environments such as Qilu Lake. Overall, the Air2water model outperforms the LSTM model in the 11 lakes, indicating that air temperature has a stronger influence on LSWT compared to other comprehensive factors. Therefore, the Air2water model is more suitable for lake inversion. The results reveal that the effect of air temperature on LSWT in Yangzonghai and Yilong Lake is 0.046 and 0.047 higher than that of the comprehensive factors, respectively. In Chenghai, the effect of air temperature on LSWT is 0.027 higher than the comprehensive factors, while in Qionghai, it is 0.002 lower than the combined factors. In summary, Qionghai is significantly influenced by factors other than air temperature that contribute to positive feedback on LSWT. On the other hand, Caohai and other lakes are greatly affected by factors other than air temperature that contribute to negative feedback on LSWT.

F. Goodness-of-Fit of the Two Models is Estimated Data

Based on the analysis presented in Fig. 10, it can be observed that the simulated data of both the Air2water model and the LSTM model exhibit a high level of goodness of fit. The difference between the simulated data of the two models on the 11 lakes is minimal, further affirming their reliability. Furthermore, the study results indicate that factors other than air temperature have a relatively small influence on the surface water temperature of the lakes, accounting for roughly 20% or less.

Fig. 10 illustrates the strong performance of the estimated data from both models, disregarding the systematic errors associated with the LSWT estimation. The figure also presents the interaction coefficient R2 between air temperature and LSWT, as well as the interaction coefficient of other factors, which can be calculated as 1 - R2. This analysis provides insights into the degree of interaction between various factors and LSWT. The results show that the determination coefficient R2 of the influence of the temperature of Caohai, Dianchi, Qilu Lake and Xingyun Lake on the LSWT is greater than 0.9, while the other lakes are around 0.85 except Lugu Lake which is lower at 0.81, indicating that the air temperature of the study area has a dominates influence on the LSWT, and other factors account for a relatively weak proportion. Lugu Lake has the highest latitude than other lakes, and similarly Qionghai has the second highest latitude, and the determination coefficient shows that the LSWT at high latitudes may be affected by factors other than air temperature higher than that of lakes at low latitudes. Therefore, the two models are well used in the spatial distribution of 11 lakes in the Yunnan–Guizhou–Sichuan Plateau. Through the above analysis, it is calculated that the effect of other factors of Caohai, Dianchi, Qilu Lake, and Xingyun Lake is less than 10%, while the other factors of other lakes are greater than 10% and less than 20%. It also shows that the combined role of other factors except air temperature is far less than the relative effect of air temperature on the LSWT, but it also shows that the role of other comprehensive factors in some lakes cannot be ignored. In order to further quantify the impact of climate change on LSWT, the data on the monthly time scale were selected for analysis to correctly describe the complex feedback system of LSWT. This study uses different models to simulation data in the prediction process; it shows that the Air2water model is more suitable for the response of LSWT to climate change [14], [15]. The machine learning model is more suitable for the learning of empirical processes, and deep learning also has good results in the trend of interannual changes, which expands the research method of factors in time series [31].

IV. Conclusion

This study focuses on 11 lakes in the Yunnan–Guizhou Plateau. The Air2water model was selected to estimate the LSWT, and the LSTM model was also used to construct the future LSWT. We first employed LSTM to forecast the monthly average temperature data spanning from 2021 to 2025. Subsequently, we utilized these predictions as meteorological indicators to estimate the LSWT during the same time frame. We then compared the estimated LSWT data with the monthly average LSWT data predicted by LSTM for the years 2021 to 2025.By analyzing the accuracy of data during the verification period, we can evaluate the performance of both models in terms of spatial and temporal aspects. This evaluation helps determine the model that performs better in each space-time scale, providing a solid foundation for establishing a new dataset. Reestablishing a new dataset for 2021–2025, the analysis of the results showed that the annual change of the surface water temperature of each lake was less than 0.2 °C (Fig. 2). The study results present a comparative analysis of the performance between two models utilized for the estimation and prediction of water temperature. The Air2water model based on physical process (monthly average RMSE = 1.04 °C, MAE = 0.89 °C, and SD = 0.93 °C) outperforms the machine learning LSTM model (monthly mean RMSE = 1.23 °C, MAE = 1.03 °C, and SD = 0.79 °C). By the year 2025, several lakes, including Caohai, Dianchi Lake, Erhai Lake, Qionghai Lake, Xingyun Lake, and Yangzong Lake, are projected to experience a significant warming trend with a rate of 0.2 °C. Notably, Caohai Lake exhibits the highest warming rate among these lakes. In addition, the spatial distribution reveals that Caohai Lake, located in the northeast of the other lakes, experiences a more pronounced warming trend compared to other regions within the study area. These findings contribute to our understanding of the regional dynamics and temperature changes in the studied lakes, furthering the field of geography. Finally, this research discusses the application of the models to estimate and predict LSWT in different temporal and spatial contexts, highlighting the significance of accurate lake physical processes and model construction methods in achieving high accuracy. The study emphasizes the dependence of inversion and prediction accuracy of LSWT on the energy influencing the lake's physical processes and the inherent systematic errors of the models. To improve the simulation and forecasting of future LSWT changes, it is crucial to refine the understanding of lake physical processes and enhance model construction techniques. The findings demonstrate the utilization of air temperature and LSWT as input and output variables in the models to assess the influence of atmospheric factors on LSWT predictions and investigate future trend patterns. While the machine learning approach may be comparatively less effective than the physical process-based method, this research paves the way for exploring hybrid approaches that combine physical processes with machine learning to enhance both runtime efficiency and prediction accuracy of LSWT. Importantly, this study highlights the importance of integrating various hydrological time series research methods into the framework of machine learning to enhance the prediction of LSWT. These findings make a valuable contribution to the field of geography by enhancing our understanding and predictive capabilities of LSWT dynamics.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization, designed and performed this research, K.Y.; analyzed the data and performed the experiments, Y.L.; drafted the manuscript, S.X. and Y.L.; edited the manuscript, S.X.; Software, S.X. and J.Z.; data collection, S.X. and C.Y.; revised the article, X.Z. All authors have read and agreed to the published version of the manuscript.

References

[1] A. Choiński, M. Ptak, and A. Strzelczak, "Changeability of accumulated heat content in alpine-type lakes," *Polish J. Environ. Stud.*, vol. 24, no. 6, pp. 2363–2369, 2015, doi: 10.15244/pjoes/58871.

[2] N. Gallina, N. Salmaso, G. Morabito, and M. Beniston, "Phytoplankton configuration in six deep lakes in the peri-alpine region: Are the key drivers related to eutrophication and climate?," *Aquatic Ecol.*, vol. 47, no. 2, pp. 177–193, 2013.

[3] M. Hondzo and H. G. Stefan, "Long-term lake water quality predictors," *Water Res.*, vol. 30, no. 12, pp. 2835–2852, 1996, doi: 10.1016/0043-1354(95)00286-3.

[4] S. Wang, X. Qian, B.-P. Han, L.-C. Luo, and D. P. Hamilton, "Effects of local climate and hydrological conditions on the thermal regime of a reservoir at tropic of cancer, in southern China," *Water Res.*, vol. 46, no. 8, pp. 2591–2604, 2012, doi: 10.1016/j.watres.2012.02.014.

[5] S. Piccolroaz, R. I. Woolway, and C. J. Merchant, "Global reconstruction of twentieth century lake surface water temperature reveals different warming trends depending on the climatic zone," *Climatic Change*, vol. 160, no. 3, pp. 427–442, 2020.

[6] T. R. Karl et al., "Possible artifacts of data biases in the recent global surface warming hiatus," *Science*, vol. 348, no. 6242, pp. 1469–1472, 2015, doi: 10.1126/science.aaa5632.

[7] S. Sima, A. Ahmadalipour, and M. Tajrishy, "Mapping surface temperature in a hyper-saline lake and investigating the effect of temperature distribution on the lake evaporation," *Remote Sens. Environ.*, vol. 136, pp. 374–385, 2013, doi: 10.1016/j.rse.2013.05.014.

[8] W. Xu et al., "A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900," *Climate Dyn.*, vol. 50, pp. 2513–2536, Apr. 2018, doi: 10.1007/s00382-017-3755-1.

[9] J. Edinger, D. Duttweiler, and J. Geyer, "The response of water temperature to meteorological conditions," *Water Resour. Res.*, vol. 4, pp. 1137–1143, 1968.

[10] M. Schmid, S. Hunziker, and A. Wüest, "Lake surface temperatures in a changing climate: A global sensitivity analysis," *Climatic Change*, vol. 124, pp. 301–315, 2014.

[11] G. Fink, M. Schmid, B. Wahl, T. Wolf, and A. Wüest, "Heat flux modifications related to climate-induced warming of large European lakes," *Water Resour. Res.*, vol. 50, no. 3, pp. 2072–2085, 2014.

[12] D. Livingstone, "Impact of secular climate change on the thermal structure of a large temperate central European lake," *Climatic Change*, vol. 57, pp. 205–225, 2003.

[13] J. Magnuson et al., "Historical trends in lake and river ice cover in the northern hemisphere," *Science*, vol. 289, pp. 1743–1746, 2000.

[14] C. M. O'Reilly et al., "Rapid and highly variable warming of lake surface waters around the globe," *Geophysical Res. Lett.*, vol. 42, no. 24, pp. 10773–10781, 2015, doi: 10.1002/2015GL066235.

[15] R. I. Woolway and C. J. Merchant, "Worldwide alteration of lake mixing regimes in response to climate change," *Nature Geosci.*, vol. 12, no. 4, pp. 271–276, 2019.

[16] C. Matulla et al., "Establishment of a long-term lake-surface temperature dataset within the European alps extending back to 1880," *Climate Dyn.*, vol. 52, no. 9/10, pp. 5673–5689, 2018.

[17] Y. Luo, Y. Zhang, K.-Z. Yang, X. Zhou, and Z. Peng, "Urban surface thermal runoff generation mechanism and scenario simulation," *Water Resour. Res.*, vol. 59, no. 4, 2023, Art. no. e2022WR033881, doi: 10.1029/2022WR033881.

[18] C. Oswald and W. Rouse, "Thermal characteristics and energy balance of various-size canadian shield lakes in the mackenzie river basin," *J. Hydrometeorol.*, vol. 5, pp. 129–144, 2004.

[19] S. Sharma, S. C. Walker, and D. A. Jackson, "Empirical modelling of lake water-temperature relationships: A comparison of approaches," *Freshwater Biol.*, vol. 53, no. 5, pp. 897–911, 2008.

[20] A. Hirst, K. Giri, D. Ball, and R. Lee, "Determination of the physical drivers of zostera seagrass distribution using a spatial autoregressive lag model," *Mar. Freshwater Res.*, vol. 68, no. 9, pp. 1752–1763, 2017.

[21] A. Witze, "Lakes warm worldwide," *Nature*, Dec. 2015, doi: 10.1038/nature.2015.19034.

[22] D. Beauregard, E. Enders, D. Boisclair, and K. Kidd, "Consequences of circadian fluctuations in water temperature on the standard metabolic rate of atlantic salmon parr (salmo salar)," *Can. J. Fisheries Aquatic Sci.*, vol. 70, no. 7, pp. 1072–1081, 2013.

[23] S. Piccolroaz, M. Toffolon, and B. Majone, "A simple lumped model to convert air temperature into surface water temperature in lakes," *Hydrol. Earth System Sci.*, vol. 17, no. 8, pp. 3323–3338, 2013.

[24] S. Piccolroaz, M. Toffolon, and M. Bruno, "The role of stratification on lakes' thermal response: The case of lake superior," *Water Resour. Res.*, vol. 51, pp. 7878–7894, 2015.

[25] M. D. O'Beirne, J. P. Werne, R. E. Hecky, T. C. Johnson, S. Katsev, and E. D. Reavie, "Anthropogenic climate change has altered primary productivity in lake superior," *Nature Commun.*, vol. 8, no. 1, Jun. 2017, Art. no. 15713, doi: 10.1038/ncomms15713

[26] Y. Yankova, S. Neuenschwander, O. Köster, and T. Posch, "Abrupt stop of deep water turnover with lake warming: Drastic consequences for algal primary producers," *Sci. Rep.*, vol. 7, no. 1, Oct. 2017, Art. no. 13770, doi: 10.1038/s41598-017-13159-9

[27] L. C. Bruce et al., "A multi-lake comparative analysis of the general lake model (GLM): Stress-testing across a global observatory network," *Environ. Modelling Softw.*, vol. 102, pp. 274–291, 2018, doi: 10.1016/j.envsoft.2017.11.016.

[28] S. Zhu, M. Ptak, Z. M. Yaseen, J. Dai, and B. Sivakumar, "Forecasting surface water temperature in lakes: A comparison of approaches," *J. Hydrol.*, vol. 585, 2020, Art. no. 124809, doi: 10.1016/j.jhydrol.2020.124809.

[29] S. Zhu, M. Ptak, M. Sojka, A. P. Piotrowski, and W. Luo, "A simple approach to estimate lake surface water temperatures in polish lowland lakes," *J. Hydrol.: Regional Stud.*, vol. 48, 2023, Art. no. 101468, doi: 10.1016/j.ejrh.2023.101468.

[30] F. Di Nunno, S. Zhu, M. Ptak, M. Sojka, and F. Granata, "A stacked machine learning model for multi-step ahead prediction of lake surface water temperature," *Sci. Total Environ.*, vol. 890, 2023, Art. no. 164323, doi: 10.1016/j.scitotenv.2023.164323.

[31] Z. Yu, K. Yang, Y. Luo, C. Shang, and Y. Zhu, "Lake surface water temperature prediction and changing characteristics analysis - a case study of 11 natural lakes in yunnan-guizhou plateau," *J. Cleaner Prod.*, vol. 276, 2020, Art. no. 122689, doi: 10.1016/j.jclepro.2020.122689.

[32] D. M. Livingstone and J. Padisák, "Large-scale coherence in the response of lake surface-water temperatures to synoptic-scale climate forcing during summer," *Limnology Oceanogr.*, vol. 52, no. 2, pp. 896–902, 2007, doi: 10.4319/lo.2007.52.2.0896.

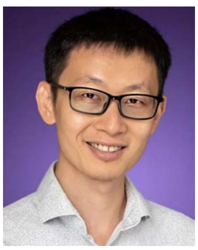
Shanshan Xu received the B.S. degree in computer science and technology from the School of Information Science and Technology, Zhengzhou Normal University, Henan, China, in 2019, and the M.S. degree in computer application technology from the School of Information Science and Technology, Yunnan Normal University, Kunming, China, in 2022. She is currently working toward the Ph.D. degree in physical geography with the School of Geographic Sciences, East China Normal University, Shanghai, China.

Her research interests include ecological security and ecosystems.

Kun Yang received the master's degree in engineering from the University of Geographical Information Engineering, New South Wales, Sydney, Kensington, Australia, in 1998.

He is currently the Dean with the Faculty of Geography, Yunnan Normal University, Kunming, China, the Director with Engineering Research Center, GIS Technology, Western China of Ministry of Education of China, and a Doctoral Supervisor in cartography and geographic information engineering specialty with Wuhan University, Wuhan, China.

His research interests include GIS and data fusion.



Xiaolu Zhou received the bachelor's degree from Wuhan University, Wuhan, China, in 2008, the master's degree in geography from the National University of Singapore, Singapore, in 2010, and the Doctor of Philosophy degree from the University of Illinois Urbana-Champaign, Champaign, IL, USA, in 2014.

He is currently an Associate Professor with Texas Christian University, Fort Worth, TX, USA. His research interests include geographic information systems, spatial big data, and urban informatics

Jie Zhang (Student Member, IEEE) received the B.S. degree in information countermeasure technology from the Southwest University of Science and Technology, Sichuan, China, in 2019, and the M.S. degree in computer application technology from the Yunnan Normal University, Kunming, China, in 2022. He is currently working toward the Ph.D. degree in computer science and technology with the Tianjin University, Tianjin, China.

His research interests include blockchain and network security.

Chenghao Yang received the bachelor's degree in geographic information science from the School of Geospatial Information, Yunnan Normal University, Kunming, China, in 2024.

His research interests include urban sustainable development and geographic information science.

Yi Luo received the bachelor's, master's, and doctoral degrees in test and measurement technology and instrumentation from the Harbin University of Science and Technology, Heilongjiang, China, in 2009, 2012, and 2014, respectively.

He is currently the Deputy Director with Engineering Center, Ministry of Education for Geographic Information Technology of Western Resources and Environment and the Dean with the School of Mapping and Geoinformation, Department of Geography, Yunnan Normal University, Kunming, China. His research interests include electrical engineering, remote sensing, and geoinformatics.