Reply to RC2

Note: the comments and authors' replies are in black and blue font, respectively. All changes in the revised manuscript are highlighted in yellow.

Lake ice phenology is a sensitive indictor to climate change.

The paper proposed a new double-threshold moving t test (DMTT) algorithm applied to CETB dataset, and reconstructed the ice phenology of Lake Ulansu from 1941 to 2023 using Random Forest based on the ERA5 data, which have been proved relatively good performance. The method is novel in identifying lake ice phenology and the influencing factors, but there still need some improvement.

(1) The introduction is not well organized, and need to improve.

(2) The logic of methodology is not clear. The overview of Data makes the reader confusing.

(3) The caption of figures should be concise, and the author need to check and improve them.

(4) The correlation between solar radiation and lake ice phenology is lower than we expected. The authors analyzed this in terms of month. Freeze-up and break-up date are one value in a winter, and other parameters varies each month, how to connect them together?

In conclusion, the subject of the study is suitable for "The Cryosphere" and can potentially be of interest for the journal's wide audience, and we suggest a major revision.

Reply: Thank you for your constructive comments on our manuscript, which are highly valuable to us during the revision. We have made major revisions to the manuscript accordingly, and the main points are listed as follows:

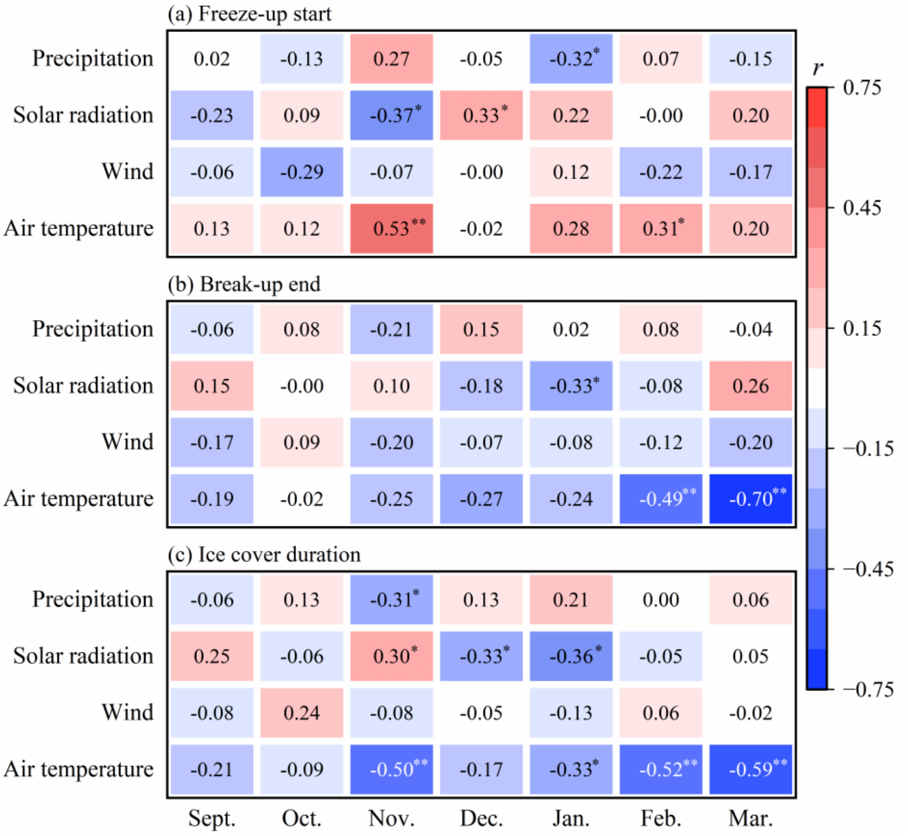
(1) We have restructured the Introduction to clarify the importance of studying shallow, vegetated lakes such as Lake Ulansu for ice phenology research. The revised sections now better highlight the novelty of our study and the progress of current methodologies and clearly state our research objectives. These changes can be found on lines 23-75 of the revised manuscript.

(2) We have revised Sections 3.1 and 3.2 to provide a clearer overview of the data sources and their specific uses within our study, as well as detailed descriptions of the methodologies employed. Figure 2 has been updated to visualize represent the flow from the raw data to the ice phenology results more clearly. Additional comparisons between our algorithm and existing methods, including MTT and field observations, are detailed in a new subsection, 4.1 Algorithm evaluation, to underline the enhancements our approach offers.

(3) The figure captions have been revised to be more concise and clear, enhancing the readability and

directness of the presented data. Specific changes have been made to Figures 1, 2, 3, 5, 6, B1, and D1.

(4) We elaborate on the unique characteristics of Lake Ulansu as a eutrophic lake. The presence of algae and suspended particles in Lake Ulansu can absorb and scatter solar radiation, mitigating the effects of warming on the water body. These details have been added in lines 344-349. Furthermore, we adjusted our analysis to focus on the relationships between meteorological factors from September to March and ice phenology, as illustrated in Figure 7.



(a) Freeze-up start

	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
Precipitation	0.02	-0.13	0.27	-0.05	-0.32^{*}	0.07	-0.15
Solar radiation	-0.23	0.09	-0.37^{*}	0.33^{*}	0.22	-0.00	0.20
Wind	-0.06	-0.29	-0.07	-0.00	0.12	-0.22	-0.17
Air temperature	0.13	0.12	0.53^{**}	-0.02	0.28	0.31^{*}	0.20

(b) Break-up end

	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
Precipitation	-0.06	0.08	-0.21	0.15	0.02	0.08	-0.04
Solar radiation	0.15	-0.00	0.10	-0.18	-0.33^{*}	-0.08	0.26
Wind	-0.17	0.09	-0.20	-0.07	-0.08	-0.12	-0.20
Air temperature	-0.19	-0.02	-0.25	-0.27	-0.24	-0.49^{**}	-0.70^{**}

(c) Ice cover duration

	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
Precipitation	-0.06	0.13	-0.31^{*}	0.13	0.21	0.00	0.06
Solar radiation	0.25	-0.06	0.30^{*}	-0.33^{*}	-0.36^{*}	-0.05	0.05
Wind	-0.08	0.24	-0.08	-0.05	-0.13	0.06	-0.02
Air temperature	-0.21	-0.09	-0.50^{**}	-0.17	-0.33^{*}	-0.52^{**}	-0.59^{**}

Figure 7: Correlation coefficients between meteorological factors and ice phenology in different months. Significance levels for correlation coefficients are denoted by * ($p < 0.05$) and ** ($p < 0.01$).

Specific comments:

1. The Introduction is not clear and attractive. The meaning of long-term tracking of climate records such as lake ice phenology, is not well expressed. The research meanings in the first paragraph of Introduction are not closely related with the content, and the author should rewrite. The scientific problem and study goals are note clear.

Reply: We have thoroughly revised the first paragraph to better articulate the significance of long-term lake ice phenology tracking in the context of climate change and its ecological implications. We have also clarified the scientific problems and objectives of this study by emphasizing how the research fills critical knowledge gaps in understanding complex lake systems under changing climatic conditions. The Introduction now effectively sets the stage for the methods and findings presented, underscoring the importance of ice phenology data for enhancing climate models and weather forecasting. See details on lines 23-35.

2. Line 38 the best quality is not appropriate.

Reply: The revised text in the manuscript "*In situ observations provide a wealth of information on ice phenology (Benson et al., 2011; Yang et al., 2020).*"

3. Line 41 MODIS just mentioned 8-day product, and the author should update recent work.

Reply: The text in the manuscript has been revised: "*For example, the MODIS Terra and Aqua products, specifically MOD10A1 and MYD10A1, have been utilized to determine lake ice phenology across the Tibetan Plateau from 2001 to 2017 (Cai et al., 2019).*"

4. Line 46 The active microwave is not suitable for the large lakes. This sentence should be improved, and supplement more contents.

Reply: We added this information to the Introduction: "*Active microwave remote sensing is utilized primarily for extracting ice phenology data from large water bodies because of its relatively low (> 10 days) temporal resolution (Antonova et al., 2016; Howell et al., 2009).*"

5. P47 One advantage of passive microwave is frequent revisit, which is not mentioned.

Reply: The original text has been modified to "*Passive microwave remote sensing provides data with long temporal coverage and frequent revisit times, although it offers coarse spatial resolution.*"

6. The second paragraph is too long, and should be divided. The core of this work belongs to passive microwave, and could be discussed alone.

Reply: Revised accordingly.

7. The fitting methods of lake ice phenology based on climate records (such as air temperature) are note well discussed, just listed some previous works.

Reply: We have revised the relevant text to better illustrate the practicality of our methodology. The updated section now emphasizes the importance of integrating diverse meteorological factors with ice phenology data to increase the effectiveness of machine-learning models. See details on lines 59-66.

8. The author explained how the study organized, and the goals of this study is not clear. The last paragraph should be rewritten.

Reply: To clarify and specify, we have revised the last paragraph of the Introduction to better articulate these specifics: "*Specifically, our research strategy was composed of the following steps: (1) A new algorithm was developed to classify the ice and water states on the lake surface in brightness temperature data from the SMMR and SSM/I-SSMIS sensors for the period 1979–2023. (2) A random forest model was trained using the results in step (1) to reconstruct the ice phenology from 1941 to 1978. (3) The meteorological impact on the ice phenology of Lake Ulansu from 1941 to 2023 was analyzed to explore the key drivers of its variations.*"

9. The general description of ice regime should be added in Study are, like ice thickness, regular winter recreation.

Reply: We have amended the manuscript to include a detailed description of the ice thickness and winter recreation. See details on lines 91-92.

10. Line 75 300 need to check, or provide the citation.

Reply: The original text has been modified to "*The lake covers an area of 306 km*^{2}."

11. Line 80 how to exclude the influence of vegetation?

Reply: Given the lake's narrow and irregular shoreline, we implemented several strategies to ensure the accuracy and reliability of our phenology assessments, which are detailed in Section 3.2 of the manuscript.

(1) To address the complex mixed pixels in Lake Ulansu, we developed and applied the DMTT algorithm. We adapted the detection of abrupt change points (ACPs) in the Tb time series by considering seasonal variations and air temperature constraints, ensuring that freezing and melting transitions are accurately identified even in the presence of mixed pixels. The algorithm calculates distinct thresholds for freezing and melting based on the mean T_b values before and after each detected ACP, tailored specifically to accommodate the variability introduced by mixed pixels.

(2) We selected five T_b grids with a water proportion greater than 0.70 to represent the lake surface status. This threshold selection is crucial for reducing the influence of mixed pixels and is specifically tailored to the unique geographical characteristics of Lake Ulansu.

These details are further elaborated in Section 3.2.1, providing a comprehensive view of how we address the challenges posed by mixed pixels at Lake Ulansu.

12. Line 83-84 remove the date.

Reply: Revised accordingly.

13. Line 85 The definition of hydrological year is not accurate. For example, HY 2022 lasted from August in 2021 to July in 2022.

Reply: To clarify and specify, we have revised the sentence to better articulate these specifics: "*In this study, each hydrological year is defined as beginning on August 1st and ending on July 31st of the following year. For example, HY2022 spans August 1, 2021, to July 31, 2022.*"

14. Line 90 The title of figure 1need to revise.

Reply: The original text has been modified to "*Figure 1: (a) Geographical context and elevation profile of Lake Ulansu within the Hetao Basin, Inner Mongolia. (b) CETB data grids with shaded areas representing brightness temperature pixels selected for Lake Ulansu surface identification. (c) Photographic depiction of aquatic reeds within Lake Ulansu. (d) On-ice instrumentation for field observations (Cao et al., 2021).*"

15. Line 95-105 The logic of flowchart is not clear, and Figure 2 should be revised. The input and output is not clear. The different colors have certain meanings? The method is too short.

Reply: We have thoroughly revised the flowchart to better illustrate the logical progression from data inputs through methodologies to outputs. In addition to revising the Fig. 2, we have also enhanced the corresponding textual descriptions within the manuscript. See details on lines 102-109.

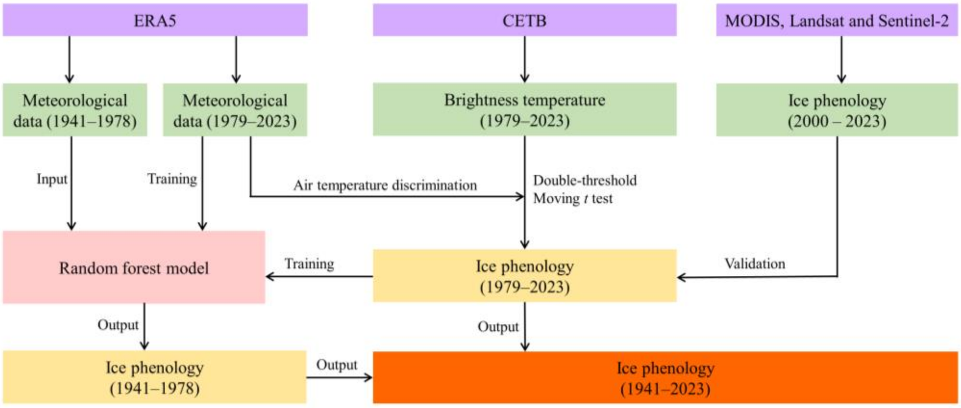


Figure 2: Reconstruction of the ice phenology of Lake Ulansu (1941–2023) based on ERA5, CETB, and optical satellite data via the double-threshold moving *t* test algorithm and random forest model.

16. Line 119 high spatial resolution Tb data are essential, gramma error

Reply: Revised accordingly.

17. Line 124 These data were sourced from the SMMR on the Nimbus 7 satellite; the SSM/I on the F08, F10, F11, F13, F14 and F16 satellites; and the SSMIS on the F16 satellite. Move this to the above content.

Reply: We have carefully revised this section to clarify the selection and usage of CETB data for our study. See details on lines 122-123.

18. Figure 3 have two choices: (1) make the content short and concise, and add the figure in the supplementary materials; (2) move this part to Results

Reply: We agree that the best approach is to make the content short and concise and to move the detailed validation section, along with Figure 3, to Appendix A. This adjustment will streamline the main manuscript while still providing access to the detailed validation results for interested readers. We believe that this modification maintains the narrative flow of the paper and allows us to focus more directly on the results in the main text.

19. Figure 4 “between 1 August and 31 July from 1979 to 2023 for Lake Ulansu”: check this. “The solid lines represent the interannual average brightness temperature and air temperature time series, respectively. “ delete this sentence. Lack the description of Figure 1 (a), just (b) appears.”

Reply: We appreciate your attention to detail and have made the following revisions to address your concerns:

(1) We verified the date range "between 1 August and 31 July from 1979 to 2023" for Lake Ulansu and confirmed its accuracy. To clarify, each series represents data from a hydrological year, defined as 1 August to 31 July of the following year.

(2) We have removed the sentence "The solid lines represent the interannual average brightness temperature and air temperature time series, respectively." to streamline the figure caption."

(3) We have added a description for part (a) of the figure.

Revised Figure 3 Caption:

"*(a) Time series of brightness temperatures (blue dashed lines) for each hydrological year from 1979 to 2023. A hydrological year (HY) was defined from 1 August to 31 July of the following year for Lake Ulansu. (b) The brightness temperature (blue solid line) and air temperature (red solid line) for HY2001.*

The ice and water statuses were determined via the double-threshold moving t test algorithm, where the red and blue shaded regions represent the water and ice states, respectively."

20. Line 174 “proportion lake water greater than 0.70” how the thresholds are determined? Why do not use 0.8/0.2?

Reply: In our study, we opted for a threshold of 0.70 for the proportion of lake water in the T_b grids to ensure a significant representation of water surfaces while still capturing sufficient spatial coverage across Lake Ulansu. The choice of this threshold was driven by the need to balance the inclusion of enough grid cells for robust statistical analysis and minimize the influence of adjacent land and vegetation within each pixel.

Reference thresholds of 0.80 or 0.20, which might be familiar from studies utilizing optical satellite data such as MODIS for ice phenology, were considered. However, these thresholds typically cater to different analysis objectives and sensor characteristics. For passive microwave data, especially in the context of lakes with complex shorelines such as Lake Ulansu, a threshold of 0.70 provides a pragmatic compromise between data availability and the accuracy of representing lake surface conditions.

21. Line 236 “Seventy percent” and “30 %”, keep the same expression.

Reply: Revised accordingly.

22. Line 258 ERA5 have various types of climate data, which one you used? It is not clear. The input and output of RF is confusing.

Reply: We have clarified the description of the ERA5 meteorological data used as input for our random forest model from 1941 to 1978. The revised sentence now reads: "*After the RF model was established and validated, we used ERA5 meteorological data from 1941 to 1978 as input features to obtain the historical ice phenology. These data included average air temperature, wind speed, solar radiation, and cumulative precipitation for the months of September to November and January to March each year.*" We have also revised Figure 2 to more clearly illustrate the input and output processes within our modeling workflow.

23. Figure 6 Please add the math equation and basic index of linear regression. Please add the lengend of different lines, it would be better to remover the description of dashed lines. (a)-(f) explained separately.

Reply: Thank you for your valuable suggestions regarding Figure 6. We have revised the figure to include the mathematical equations and basic indices of the linear regression for each segment. We have also updated the legend to clearly differentiate between the various lines depicted in the graphs and have removed the description of the dashed lines from the legend for clarity. The explanations for parts (a) through (f) have been distinctly outlined to ensure that each component is clearly understood.

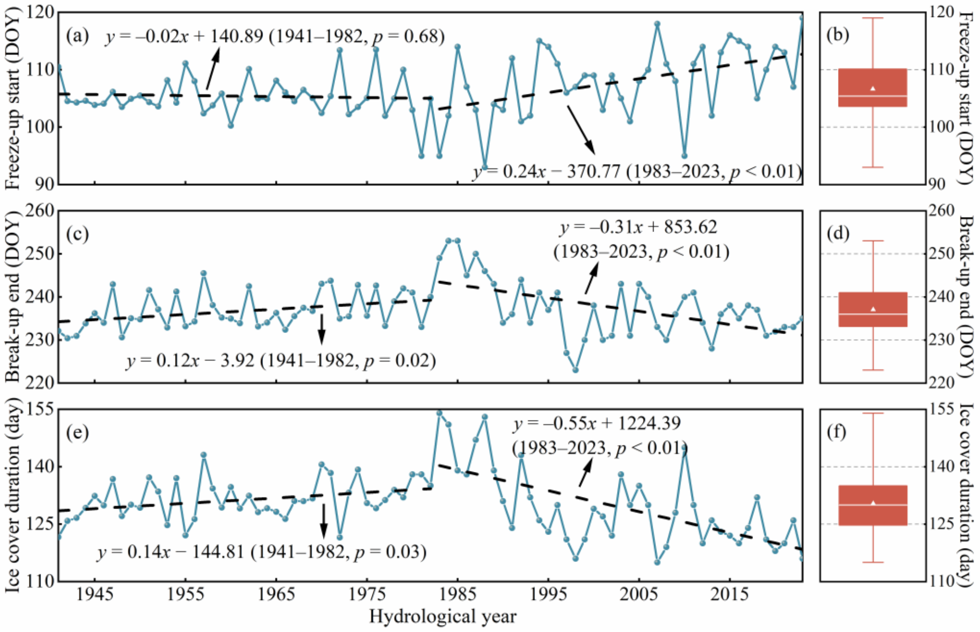


Figure 6: Ice phenology results for Lake Ulansu from 1941 to 2023. (a) Freeze-up start with trend lines. (b) Freeze-up start date distribution. (c) Break-up end with trend lines. (d) Break-up end date distribution. (e) Ice cover duration with trend lines. (f) Ice cover duration distribution.

24. Line 280 add the yearly changing rates.

Reply: We have updated the manuscript to include the yearly changing rates for ice phenology trends in Lake Ulansu. The revised text now reads: "*Overall, from 1941 to 2023, the ice phenology in Lake Ulansu exhibited several notable features. The FUS occurred between the 93rd and 119th days, with an average date of approximately the 107th day (November 15), showing a slight delaying trend of 0.07 d yr*$^{-1}$*. The BUE ranged from the 223rd to the 253rd day, typically occurring around the 237th day (March 25), with an advancing trend of 0.01 d yr*$^{-1}$*. The ICD spans from 115 to 154 days, with an average of approximately 130 days, decreasing by 0.08 d yr*$^{-1}$ *over the study period.*"

."

25. Line 282 The title of 4.2 is not proper.

Reply: We have revised the title of this section to "*Impact of meteorological factors on ice phenology*"

26. Table 3 Linear tread is yearly changing rates? Please add the unit. The linear tread of wind speed is 0. Check this?

Reply: Thank you for your attention to the details in Table 3. We have now added the appropriate units to indicate that the linear trends represent yearly changing rates, expressed in the respective units of each ERA5 parameter per year. With respect to the wind speed trend, our analysis reveals a minimal change, indeed nearing zero, which reflects the stable wind conditions over the study period at Lake Ulansu.

27. Figure 8 The correlation between solar radiation and lake ice phenology is lower than we expected.

Reply: In eutrophic lakes, due to the high concentrations of algae and suspended particles, the absorption and penetration of solar radiation differ significantly from those in clear lakes. Eutrophic lakes are rich in nutrients (such as nitrogen and phosphorus), promoting the proliferation of algae and phytoplankton (Li et al., 2024). These algae and suspended particles absorb and scatter solar radiation, especially blue light (450–495 nm). After being absorbed, this light energy is converted into heat or used for photosynthesis (Lin et al., 2024).

Because blue light is strongly absorbed, the penetration depth of light in water is significantly reduced. In clear lakes, light can penetrate deeper into the water column, but in eutrophic lakes, light is mainly absorbed and scattered by algae and suspended particles near the surface, reducing the penetration depth of light. Consequently, the heating effect of solar radiation on water is diminished.

The revised text in the manuscript "*For Lake Ulansu, the correlation may be due to the high attenuation coefficient caused by the presence of numerous suspended particles and algae, which are common in eutrophic lakes (Yang et al., 2020). These particles and algae absorb and scatter solar radiation, especially blue light (450–495 nm), reducing the penetration depth of light in water (Lin et al., 2024). As a result, the heating effect of solar radiation is confined primarily to the surface layers, leading to a diminished overall heating effect on the water.*"

28. As for random forest, we have following questions: (1) The abbreviation of correlation coefficients is r, not R. Need to check the whole manuscript. (2) The evaluate the performance of RF, the determination coefficients R^2 is more usually used, rather than R. (3) How the author avoid overfitting in the work? Need to explained more clearly. (4) Why the author chose random forest rather than other methods?

Reply: We have carefully reviewed and addressed each point as follows:

(1) We have revised the notation for correlation coefficients throughout the manuscript from "R" to "r" to align with standard statistical notation practices.

(2) We have updated the manuscript to use the determination coefficient "R^2" instead of "r" when discussing the performance of the random forest model.

(3) We determined that using 20 trees in the random forest model provided a balanced complexity that effectively captured the underlying patterns without overfitting. This decision was based on observing diminishing returns on model performance metrics with more than 20 trees (Appendix C). We implemented 3-fold cross-validation during the training phase to ensure that the model did not learn the noise and outliers of the training data, thereby enhancing its ability to generalize to unseen data. We continuously monitored the model's performance on both the training and validation sets to detect signs of overfitting, which was characterized by high performance on the training data but poor performance on the validation data.

(4) The random forest is robust in handling nonlinear relationships between variables, which is often the case in ecological and climatological studies. This method allows for an effective assessment of feature importance, which is crucial for understanding the influence of different meteorological factors on ice phenology. Compared with other models, random forest models are generally more resilient to overfitting, especially when the correct number of trees and other hyperparameters are chosen. Previous studies have successfully used random forest models in similar contexts (Anilkumar et al., 2023; Ruan et al., 2020), providing a tested framework for our analysis.

29. The two paragraphs in Author contribution are repeatedly expressed.

Reply: Revised accordingly.

References:

Anilkumar, R., Bharti, R., Chutia, D., and Aggarwal, S. P.: Modelling point mass balance for the glaciers of the Central European Alps using machine learning techniques, Cryosphere, 17, 2811–2828, doi:10.5194/tc-17-2811-2023, 2023.

Antonova, S., Duguay, C., Kääb, A., Heim, B., Langer, M., Westermann, S., and Boike, J.: Monitoring Bedfast Ice and Ice Phenology in Lakes of the Lena River Delta Using TerraSAR-X Backscatter and Coherence Time Series, Remote Sens., 8, doi:10.3390/rs8110903, 2016.

Benson, B. J., Magnuson, J. J., Jensen, O. P., Card, V. M., Hodgkins, G., Korhonen, J., Livingstone, D. M., Stewart, K. M., Weyhenmeyer, G. A., and Granin, N. G.: Extreme events, trends, and variability in Northern Hemisphere lake-ice phenology (1855–2005), Clim. Change, 112, 299–323, doi:10.1007/s10584-011-0212-8, 2011.

Cai, Y., Ke, C.-Q., Li, X., Zhang, G., Duan, Z., and Lee, H.: Variations of lake ice phenology on the Tibetan Plateau from 2001 to 2017 based on MODIS data. J. Geophys. Res.-Atmos., 124, doi:10.1029/2018JD028993, 2019.

Howell, S. E. L., Brown, L. C., Kang, K.-K., and Duguay, C. R.: Variability in ice phenology on Great Bear Lake and Great Slave Lake, *Northwest Territories, Canada, from SeaWinds/QuikSCAT: 2000–2006, Remote Sens. Environ., 113, 816–834, doi:10.1016/j.rse.2008.12.007, 2009.*

Li, H., Song, C., Yang, L., Qin, H., Cao, X., and Zhou, Y.: Phosphorus supply pathways and mechanisms in shallow lakes with different regime, Water Resour., 193, 116886, doi:10.1016/j.watres.2021.116886, 2021.

Lin, X., Wu, X., Chao, J., Ge, X., Tan, L., Liu, W., Sun, Z., and Hou, J.: Effects of combined ecological restoration measures on water quality and underwater light environment of Qingshan Lake, an urban eutrophic lake in China, Ecological Indicators, 163, doi:10.1016/j.ecolind.2024.112107, 2024.

Ruan, Y., Zhang, X., Xin, Q., Qiu, Y., and Sun, Y.: Prediction and Analysis of Lake Ice Phenology Dynamics Under Future Climate Scenarios Across the Inner Tibetan Plateau, J. Geophys. Res.-Atmos., 125, doi:10.1029/2020jd033082, 2020.

Yang, Q., Song, K., Hao, X., Wen, Z., Tan, Y., and Li, W.: Investigation of spatial and temporal variability of river ice phenology and thickness across Songhua River Basin, northeast China, Cryosphere, 14, 3581–3593, doi:10.5194/tc-14-3581-2020, 2020.

Yang, W., Xu, M., Li, R., Zhang, L., and Deng, Q.: Estimating the ecological water levels of shallow lakes: a case study in Tangxun Lake, China, Sci Rep, 10, doi:10.1038/s41598-020-62454-5, 2020.