



LI-CCR: Dataset of daily lake ice evolution (2002-2024) across global cold climate regions based on gap-filled MODIS observations

Zhengxin Jiang^{1,2,3}, Yubao Qiu^{1,2,3}, Matti Leppäranta^{1,4}, Xiaoting Li^{1,2}, Peng Yao^{1,2,3}, Guoqiang Jia^{1,2}, Jiancheng Shi⁵

5 ¹ Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

² International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China

³ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

⁴ University of Helsinki, Helsinki 00014, Finland

10 ⁵ National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

Correspondence to: Yubao Qiu (qiuyb@aircas.ac.cn)

Abstract. Lake ice is an essential component of the terrestrial cryosphere and plays an important role in the socioeconomic and ecological systems of cold regions. However, existing lake ice datasets generally suffer from poor temporal continuity, limited spatial coverage, and a lack of observations for small and medium-sized lakes. Consequently, the global spatial patterns and long-term trends of lake ice have remained insufficiently understood. In this study, using MODIS observations, we developed global datasets of daily lake ice coverage (LIC), annual lake ice-cover status, annual lake ice phenology (LIP), and the probability of complete ice-cover occurrence (PCIO), including 32,800 lakes across global cold climate regions from 2002 through 2024. Validation against multiple remote sensing datasets demonstrated high accuracy and confirmed that our Lake Ice - Cold Climate Regions dataset (LI-CCR) effectively captures the spatiotemporal evolution of the lake ice zone. Quality assessments were conducted for both LIC and LIP, and the information was included as an integral component of the LI-CCR. The results show that most cold regions lakes in the Northern (Southern) Hemisphere freeze before November (May) and melt from May (November) onward, with an average ice cover duration of about 200 days. Presently 89 % of the lakes experience completely frozen annually, and the number of intermittent or even ice-free lakes is increasing under the warming climate. The LI-CCR dataset provides comprehensive records of lake ice evolution in cold climate regions and offers valuable information for studies on lake-climate interactions, cryosphere changes, and ecosystem responses. This LI-CCR dataset v2 is freely available from the Zenodo platform at <https://doi.org/10.5281/zenodo.17687698> (Jiang et al., 2025b).

1 Introduction

Lakes occupy about 3.7 % of the global glacier-free land surface with more than half (over 50 million) distributed in the mid-to-high-latitude cold climate regions which experience seasonal ice cover, either annually or intermittently (Verpoorter et al., 2014; Sharma et al., 2019). Lake ice provides series of ecosystem services for local communities, including



transportation, fishing, and cultural practices (Culpepper et al., 2024; Hampton et al., 2024). Ice cover also profoundly regulates lake ecosystems by limiting lake-air interactions and modifying key physical and biogeochemical processes, including heat and light transfer, mixing, and oxygen dynamics (Li et al., 2022b; Woolway and Merchant, 2019; Xie et al., 2023; Jansen et al., 2025; Zhang Yixiao and Matti, 2024). For instance, methane emission season largely overlaps the ice-free season (Wik et al., 2016; Johnson et al., 2022; Song et al., 2024), while in winter, gas bubbles become trapped beneath the ice (Li et al., 2011; Engram et al., 2020; Mu et al., 2025). Additionally, earlier ice melting in spring leads to accelerated lake warming, with effects lasting up to three months, thereby influencing aquatic biological activity and water quality (Li et al., 2022b).

Lake ice phenology (LIP) refers to dates of recurring ice events, encompassing the freeze-up start (FUS), freeze-up end (FUE), break-up start (BUS), and break-up end (BUE). The ice cover duration (ICD), and complete ice cover duration (CICD) are subsequently derived from LIP. LIP is primarily driven by atmospheric variables, mainly the air temperature, wind, and precipitation (Xu et al., 2024; Cai et al., 2022b; Lepparanta, 2023), and responds rapidly to climate change. Recognized as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS), LIP serves as a key indicator of regional and global climate change (Woolway et al., 2020). In recent decades, cold climate regions including the Arctic and High Mountain Asia have warmed at rates exceeding the global average due to polar amplification (Rantanen et al., 2022; Xie et al., 2022; You et al., 2021). Consequently, lake ice has undergone unprecedented changes with delayed freeze-up and earlier break-up (Huang et al., 2022; Filazzola et al., 2020; Sharma et al., 2021), and a northward shift of the boundary of ice-covered lakes has taken place (Sharma et al., 2019; Wang et al., 2025).

Nearly half (49.8 %) of the world's lakes are located north of 60 °N, representing 23.6 % of the total lake area, and most lakes in cold climate regions are small to medium size (<100 km²) (Pi et al., 2022). Although the significance of small lakes in the global lake system is particularly present in lake expansion and greenhouse gas emission (Pi et al., 2022; Wik et al., 2016; Holgerson and Raymond, 2016; Woolway et al., 2020), current research is still primarily focused on large lakes (Messenger et al., 2016; Jiang et al., 2025a). Compared to other cryosphere elements, lake ice has been underexplored (Williams and Ferrigno, 2012) and faces lack of continuity in time and space. Consequently, long-term, high frequency dataset of ice cover in small lakes is essential for understanding and modelling Earth system processes under climate change. In-situ observations have long history (Sharma et al., 2019; Benson et al., 2012; Magnuson et al., 2000), but the data is limited by uneven spatial distribution, decreasing number of stations, and inconsistencies in the definitions of ice phenology variables (Sharma et al., 2022; Hodgkins et al., 2002). Process-based physical models can be used to estimate lake ice evolution where observations are unavailable (Mackay et al., 2017; Li et al., 2022a; Brooks et al., 2013; Cheng et al., 2014; Zhang Yixiao and Matti, 2024) but they are limited by simplified geometry, coarse forcing data, and low portability between different lakes (Sharma et al., 2020).

Remote sensing observations have been effective in monitoring large-scale lake ice variations but still exhibit notable limitations. The passive microwave method offers weather-independent and high-frequency brightness temperature data, but its low-resolution limits applications to small lakes and spatial details of large lakes (Wang et al., 2022a; Cai et al., 2022a;



Kang et al., 2014; Shi et al., 2025). Active microwave method provides high-resolution backscatter coefficient data (Li et al., 2022a; Mugunthan et al., 2023; Mangilli et al., 2022; Tom et al., 2020) but is limited by long revisit periods and sensitivity to surface roughness and ice microstructure (Yang et al., 2024; Murfitt et al., 2024). Thermal infrared method provides the radiative surface temperature but is limited by weather dependence and resolutions (Wu et al., 2022; Korver et al., 2024; Weber et al., 2016).

Optical remote sensing with appropriate spatial and temporal resolution has good performance in lake ice monitoring based on the different reflectance of ice and water. Although high-resolution Landsat imagery enables ice cover mapping in small lakes, its 16-day revisit period constrains the monitoring of detailed ice evolution dynamics (Wang et al., 2021; Yang et al., 2022). The Moderate Resolution Imaging Spectroradiometer (MODIS) has strong potential for capturing ice evolution details through its moderate resolution and high revisit frequency (twice daily). However, current studies typically focus on large lakes where cloud cover does not much affect the overall results (Wang et al., 2022b; Xu et al., 2024; Weber et al., 2016). Although it is impossible to directly obtain surface information, several studies have applied advanced cloud removal algorithms to enhance the data availability in optical remote sensing (Cai et al., 2019; Gafurov and Bardossy, 2009; Qiu et al., 2019). By such approach, gaps in MODIS data can be largely filled in ice monitoring in small and medium-sized lakes.

To address the limited availability of remote sensing data on lake ice across the Earth's cold climate regions, the primary objective of this study is to develop a long-term lake ice dataset (2002-2024) of daily lake ice coverage (LIC), annual lake ice-cover status, annual lake ice phenology (LIP) and the probability of complete ice-cover occurrence (PCIO) for 32,800 lakes suitable for MODIS observations. These lakes include 14,369 lakes in Eurasia, 17,914 in North America, 336 in South America, 141 in Oceania, and 40 in Africa. The LI-CCR (Lake Ice - Cold Climate Regions) dataset was generated through four key steps: (1) Developing a cloud-gap-filled algorithm for MODIS observations to produce continuous daily LIC; (2) Mapping the annual distribution of completely frozen, partially frozen, and unfrozen lakes; (3) Calculating the PCIO, and (4) Extracting annual LIP for each lake. The products were systematically validated against multiple datasets to ensure the data accuracy, while quality assessments were documented as parts of the dataset. This dataset enables a comprehensive characterization of the spatial and temporal variations in lake ice cover during the past two decades, providing a robust data foundation for evaluating the impacts of climate change and supporting Earth system modelling in cold regions.

2 Data

2.1 MODIS daily snow cover products

Since 2002, the MODIS instruments onboard the polar-orbiting Terra and Aqua satellites have provided twice-daily, synchronized observations of the Earth's surface, collecting radiances in both visible, near-infrared, short-wave infrared, and thermal infrared bands. The MODIS daily snow products (MOD10A1/MYD10A1) provide the global Normalized Difference Snow Index (NDSI) at 500 m resolution, along with classification for land, lakes, and cloud cover (Hall and



Riggs, 2021). In this study, the data from September 2002 through August 2024 were utilized to monitor lake ice cover dynamics across cold climate regions worldwide.

2.2 Lake mask from the HydroLAKES

100 The HydroLAKES database (Messenger et al., 2016) integrates multiple global and regional lake datasets, offering high-resolution shoreline polygons for over 1.43 million lakes and reservoirs larger than 0.1 km². Additionally, the database provides key lake attributes, such as area, shoreline length, shoreline development, total volume, and average depth. To align with the resolution of MODIS data, lake masks from HydroLAKES were converted to grid format and resampled accordingly. Lakes with the geometry and spatial extent appropriate for the MODIS monitoring were selected for this study.

105 2.3 Köppen-Geiger climate classification

The 1-km gridded Köppen-Geiger climate classification map (Beck et al., 2018) provides five major climate groups: tropical (A), arid (B), temperate (C), cold (D), and polar (E). These groups are further subdivided into 30 distinct climate types based on the seasonal temperature and precipitation. The climate classification of each lake was used to determine whether it potentially experiences ice-covered winters.

110 2.4 CPC Global Unified Temperature

To reduce noise in the daily LIC, we utilized the CPC Global Unified Temperature dataset from the National Centers for Environmental Prediction (<https://psl.noaa.gov>) offering daily maximum and minimum surface air temperature at a spatial resolution of 0.5° × 0.5°. This information enables reliable assessments of temperature conditions influencing lake ice dynamics at regional scales.

115 2.5 Lake ice cover dataset

The lake ice extent for the Northern Hemisphere (LIE-NH) dataset (Heinila et al., 2021), derived from Sentinel-3 with a spatial resolution of 0.005° (0.556 km in latitude degrees), provides daily observations that are closely aligned with MODIS overpasses. This dataset enables the assessment of MODIS-based ice cover under clear sky, and owing to different cloud distributions between the sensors, serves an independent benchmark for evaluating the performance of our cloud-gap-filled
 120 algorithms.

2.6 Lake ice phenology datasets

Satellite-based and model-based LIP datasets were compiled to validate our results. the passive microwave-based dataset provided long-term records covering 56 lakes in the Northern Hemisphere from 1979 to 2019, with a focus on large lakes(Cai et al., 2022a). At the regional scale, the ice phenology for 132 lakes on the Tibetan Plateau from 1978 to 2016 was



125 reconstructed using Air2Water numerical model to simulate the surface temperature for small lakes (Wu et al., 2022). Most
 of the lakes in these datasets were also monitored in our study, providing an independent perspective for validating the
 retrieved results.

3 Methods

3.1 Lake observability assessment with MODIS

130 Due to moderate spatial resolution of MODIS, lakes with small surface area or irregular shoreline show a high proportion of
 mixed pixels caused by contamination from adjacent land. Although mixed pixels may partially capture lake ice presence,
 they introduce considerable uncertainty and reduce data accuracy and reliability (Carrea et al., 2023). To ensure the scientific
 reliability and usability of the extracted results, the maximum distance to land was adopted as the key criterion for assessing
 the lake observability in the MODIS data. Only lakes with proper observation conditions were selected for further analysis.
 135 First, lakes from the HydroLAKES database were filtered according to their morphology and the resolution limitations of
 MODIS. At a resolution of 500 m, certain lake shapes and sizes possess high likelihood that pixels classified as 'lake' include
 adjacent land. To ensure accurate MODIS monitoring and minimize uncertainties due to mixed pixels, only lakes larger than
 2 km² with at least two pure lake pixels were included. This threshold enhances the accuracy and robustness of lake ice
 observations but limits the coverage of small water bodies. Second, to focus on lakes with potential to freeze during winter,
 140 those located in cold and polar climate zones as well as in arid regions with cold winter in the Köppen-Geiger climate
 classification (Table 1) were selected for further analysis.

Table 1: Description of selected Köppen-Geiger climate classifications.

Main climate	Seasonal precipitation type	Seasonal temperature type
B (Arid)	W (Desert)	k (Cold)
	S (Steppe)	
D (Cold)	f (Without dry season)	a (Hot summer)
	s (Dry summer)	b (Warm summer)
	w (Dry winter)	c (Cold summer)
		d (Very cold winter)
E (Polar)	T (Tundra)	
	F (Frost)	

3.2 Daily LIC algorithm

To obtain detailed information on lake ice evolution, daily LIC data were derived as follows. First, MODIS pixels within
 145 lake boundaries defined by the HydroLAKES database were extracted, and non-informative pixels, such as night or saturated



detector conditions, were reclassified as 'cloud'. Second, the classical NDSI lake ice detection algorithm was applied to classify ice and water among cloud-free pixels, while cloud-gap-filled algorithm leveraging spatial and temporal continuity, was employed to estimate lake surface conditions obscured by clouds (Qiu et al., 2019). Third, the remaining cloudy pixels were allocated according to the cloud-free pixels of the same lake. The obtained ice fraction, defined as LIC, was further processed to address missing values and remove outliers. For consistency in the temporal analysis, the long-term LIC (2002 to 2024) was segmented into hydrological years: from August 1st to July 31st of the following year for lakes in the Northern Hemisphere, and from February 1 to January 31 of the following year for lakes in the Southern Hemisphere. This time frame captures the complete freeze-thaw cycle of seasonally freezing lakes, including the autumn and winter ice growth and the subsequent spring and summer melting, providing continuous records for analysing spatiotemporal evolution of lake ice.

3.2.1 Lake ice detection for clear-sky pixels

Snow and ice exhibit distinct spectral reflectance characteristics - high in the visible band while low in the shortwave infrared (SWIR) and near-infrared (NIR) bands - making this window well-suited for ice identification in bare land or open water environments (Hall et al., 2002). Therefore, the reflectance from MODIS band 4 (green, 545-565 nm) and band 6 (SWIR, 1628-1652 nm) were used to calculate the NDSI index, along with the reflectance of band 4 and band 2 (NIR, 841-876 nm) applied to identify pixels covered by snow or ice together:

$$NDSI = \frac{\text{band 4} - \text{band 6}}{\text{band 4} + \text{band 6}}, \quad (1)$$

A pixel is classified as ice when the NDSI is ≥ 0.40 , and the reflectance of band 2 and band 4 exceed 0.11 and 0.10, respectively.

3.2.2 Cloud-gap-filled for cloudy pixels

To address the limitations caused by extensive cloud cover in MODIS observations, a four-step cloud-gap-filled algorithm was implemented, leveraging the spatial and temporal continuity of freeze-thaw conditions. Each step used the output from the previous stage to progressively reduce cloud-related gaps. Temporal interpolations were performed sequentially with varying time window lengths, where shorter windows applying stricter continuity constraints (Figure 1).

Step 1 is based on short-term temporal continuity within a single day. Terra and Aqua satellites pass over the same location within approximately 3 hours, possibly under different cloud conditions. If their cloud condition differed, it was assumed that ice cover was unchanged: if one pixel was cloudy while the other was clear, the clear observation was assigned. Otherwise, if one pixel was ice while the other was open water, it was assigned as an open water pixel.

Step 2 is based on medium-term temporal continuity between the preceding and following days. Cloudy pixels were filled based on ice conditions of the adjacent days: (a) If both days indicated ice, or if one day was cloudy while the other indicated ice, it was assigned as ice; and (b) If either day was open water, it was assigned as open water regardless of the other day's



condition. Although this may overlook potential changes during freeze-up and break-up phases, it was effective for most cases.

Step 3 is based on the spatial continuity among neighbouring pixels. Assuming that the ice condition beneath clouds is consistent with majority of neighbouring pixels, the algorithm proceeds as: if at least 3 out of 4 adjacent pixels were identified as either ice or open water, the same label was assigned to the cloud pixel. In nearshore areas, where spatial consistency tends to be weaker, a stricter criterion was applied: if all lake pixels in the 8 surrounding pixels were identified as either ice or open water, the same classification was assigned.

Step 4 is based on the long-term temporal continuity, incorporating observations from the two days before and after the target date. Due to the extended time window, the criterion is more stringent than in Step 2: (a) If at least 3 days were identified as open water, it was assigned as open water; and (b) If at least 3 days were identified as ice and none as open water, it was assigned as ice.

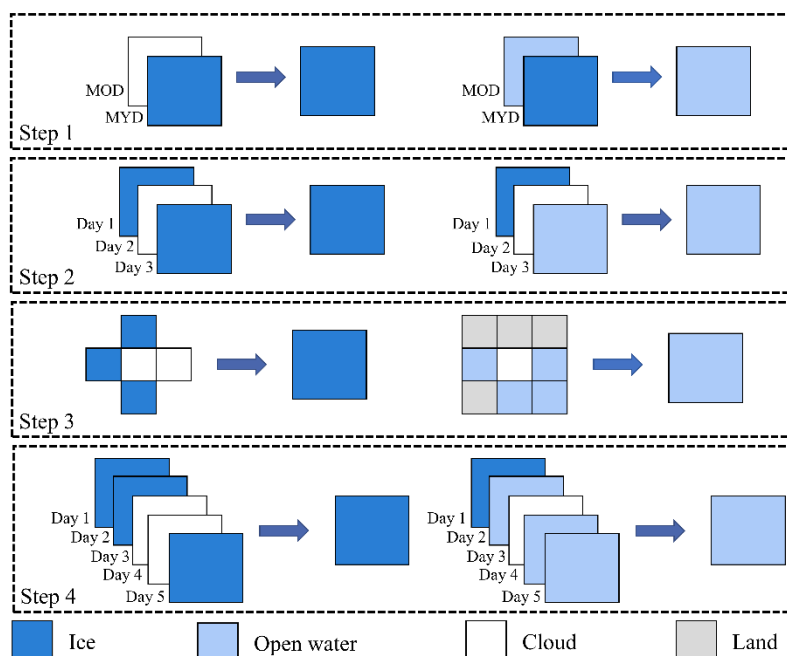


Figure 1: Flowchart illustrating of the cloud-gap-filled algorithm.

3.2.3 LIC reprocessing and noise correction

To calculate the lake ice extent, the numbers of pixels identified as lake ice, open water, and cloud were recorded as N_{ice} , N_{water} , and N_{cloud} , respectively. The raw LIC was defined as:

$$Raw LIC = \frac{N_{ice}}{N_{ice} + N_{water} + N_{cloud}}, \quad (2)$$



195 The raw lake water coverage (LWC) and cloud coverage (CC) were calculated the same way. A LIC threshold of 10 % was used to indicate the onset of ice, whereas 90 % indicated complete freezing. However, random fluctuations caused by remaining cloud pixels and missing observations introduced noise into the data that limited its ability to accurately represent the actual lake ice evolution.

200 Despite the implementation of the gap-filled procedure, some cloud pixels still remained. To address this, it was assumed that the ice conditions in the remaining cloudy areas and cloud-free areas were consistent in the same lake, and the following rules were established: If at least 90 % of the lake was open water, cloud coverage was assigned to LWC; and if at least 90 % was ice, it was assigned to LIC. Otherwise, it was allocated according to the observed ice-water ratio.

205 Low illumination conditions in high-latitude regions during autumn and winter, due to high solar zenith angles (SZA), limits the acquisition of surface reflectance from optical sensors (Hall and Riggs, 2021). When $SZA \geq 85^\circ$, MODIS data was marked as "night", may leading to missing ice observations of lakes above 60°N . To address this issue, the lake ice data above 60°N from October to March of the following year, when both LIC and LWC were missing, was reprocessed: If the lake had completely frozen before this period, it was assumed to remain frozen during the dark season-a high-probability event-and the missing values was filled accordingly. Similarly, if the lake was not fully frozen before the period but was completely frozen afterward, missing values were filled by the interpolation using the nearest valid observations. While this approach may not capture the precise dynamics of ice formation, it provides a reasonable estimate for the dark season.

210 In this study, LIC values below 10 % were considered negligible and filtered out as minor noise. However, high-reflectance pixels caused by undetected clouds, turbid waters, or algal blooms could be misclassified as lake ice, introducing outliers during ice-free periods. Assuming that lake ice dynamics are air-temperature-driven, we divided the ice evolution progress into four phases according to the daily temperature cycle: (1) Minimum temperature first drops below 0°C , (2) Maximum temperature remains consistently below 0°C , (3) Maximum temperature first rises above 0°C , and (4) Minimum temperature remains consistently above 0°C . Anomalous LIC values occurring during phase 2 and phase 4 were identified and corrected to maintain the physical consistency of the LIC dataset.

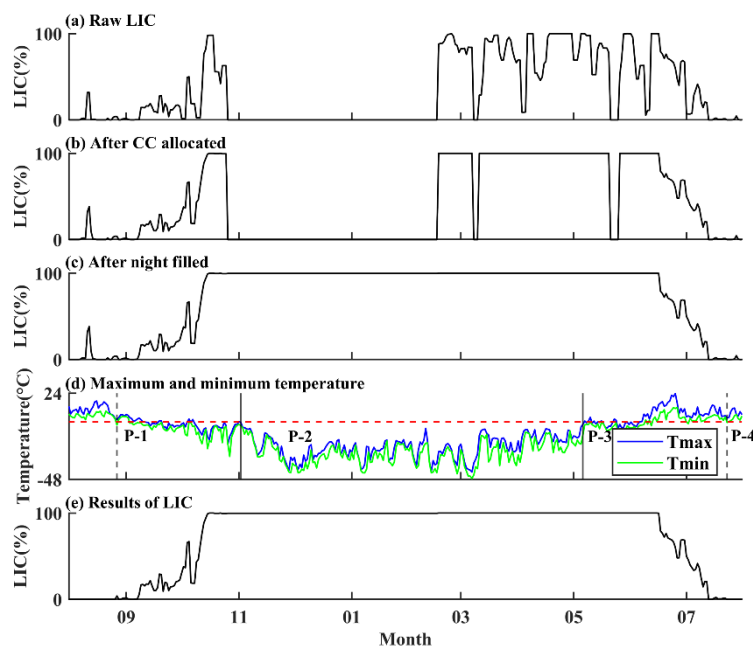


Figure 2: An example of calculating and reprocessing of LIC for Ozero Nadudoturku Lake in northern Siberia, Russia (2002-2003). (a) Raw LIC, LWC and CC. (b) LIC after the allocation of the remaining CC. (c) LIC after filling the gap caused by low illumination conditions. (d) Four phases of ice evolution divided by daily maximum and minimum temperature; and (e) LIC after the removal of outlines based on air temperature criteria.

3.3 Extraction of lake ice phenology

The phenology considered in this study included the start and end of freeze-up and break-up. Due to the wide latitudinal range and diverse climate conditions, some cold region lakes remained partially or completely ice-free in mild winters. To ensure a valid LIP extraction, lakes were classified each year into three types according to their ice-cover status: (1) Unfrozen lakes, less than 10 % ice cover for 7 consecutive days, were not assigned any ice phenology; (2) Completely frozen lakes, at least 90 % ice cover for 3 consecutive days, were assigned full ice phenology; and (3) Partially frozen lakes which were ice-covered but did not reach 90 % were assigned only FUS and BUE. These 3-day and 7-day thresholds helped to exclude unstable short-term ice cover cases and minimize the impact of anomalies. In addition, we determined the completely frozen years for each lake throughout the study period and quantified the PCIO.

LIP was defined based on daily LIC as follows: In autumn or winter, FUS was the first day when LIC remained ≥ 10 % for 3 consecutive days, and FUE was the first day when LIC remained ≥ 90 % for 3 consecutive days. In spring or summer, BUS was the first day when LIC remained ≤ 90 % for 3 consecutive days, and BUE was the first day when LIC remained ≤ 10 % for 3 consecutive days. For each lake LIP was annually extracted based on its ice-cover status. The ICD was calculated from FUS to BUE, while CICD (FUE to BUS) was additionally calculated when well defined.



235 3.4 Validations of lake ice cover and phenology

3.4.1 Evaluation of the cloud-gap-filled algorithms with MODIS observations

Although in situ ice observations provide direct information even under cloud cover, their sparse distribution and limited accessibility restrict their use for large-scale validation of remote sensing. Given the brief time interval, Step 1 of the cloud gap algorithm is generally considered reliable and was therefore excluded from explicit validation. Taking advantage of the short time interval between Terra and Aqua overpasses, when lake ice was assumed stable, we employed a simulation-based approach to develop a paired-observation framework for quantitative evaluation the cloud-gap-filled algorithm (Gafurov and Bardossy, 2009).

Lake ice cover from Terra's cloud-free areas were used as a reference, while cloud-covered areas in Aqua were designated as the test data. To assess both individual processing steps and the overall algorithm, the test data underwent the same procedures to estimate freeze/thaw status. These estimates were then compared pixel-by-pixel with the Terra-based reference data from the same day. Overall Accuracy (OA) was used to evaluate the algorithm, defined as:

$$OA = \frac{N_{ice}^{ice} + N_{water}^{water}}{N_{remove}}, \quad (3)$$

Here, $N_{removed}$ represents the number of pixels removed in the test data that correspond to clear pixels in the reference, and N_{ice}^{ice} and N_{water}^{water} represent the number of pixels where the cloud-gap-filled results match the reference data for ice and water, respectively.

3.4.2 Comparison with reference ice cover extent dataset

To further validate the lake ice time series, lake ice extent derived from Sentinel-3 imagery was used as a reference (Heinila et al., 2021). Two scenarios were considered: (1) Clear-sky conditions, validating direct observations from Terra and Aqua; and (2) Cloudy conditions, assessing MODIS results processed using the cloud-gap-filled algorithm. The same metrics described in Section 3.4.1 were employed to quantify the MODIS detection performance and the effectiveness of the cloud-gap-filled algorithm.

3.4.3 Comparison with reference lake ice phenology dataset

To validate our retrieval method, the lakes that spatially and temporally coincided between our dataset and existing satellite-based and model-based LIP products (Cai et al., 2022a; Wu et al., 2022) were selected. The four key ice phenology events—FUS, FUS, BUS and BUE—were analysed. The accuracy and consistency of these events were quantified using the coefficient of determination (R^2), mean error (ME), and Root Mean Squared Error (RMSE).

$$R^2 = 1 - \frac{\sum (V_j - M_j)^2}{\sum (V_j - \bar{M}_j)^2}, \quad (4)$$



$$ME = \frac{1}{n} \sum_{j=1}^n (M_j - V_j), \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (M_j - V_j)^2}, \quad (6)$$

265 where n denotes the total number of lakes for the validation, M_j is the MODIS-based phenology for lake j , and V_j refers to the corresponding date from the validation dataset.

3.5 Quality assessment

3.5.1 Quality assessment for LIC

270 According to the daily LIC algorithm described in Section 3.2, the final LIC for each lake on each day is constructed as the sum of six individual components, each representing distinct sources of original MODIS observation or LIC algorithm-derived increments. These component-wise structure provides a framework for quality assessment, allowing the relative contribution of different data sources to be identified and evaluated.

$$Final\ LIC = LIC_1 + LIC_2 + LIC_3 + LIC_4 + LIC_5 + LIC_6, \quad (7)$$

275 where LIC_1 represents the original MODIS observations combined with Terra and Aqua sensors, LIC_2 to LIC_4 denote gap-filled increments derived from three-day, neighbouring-pixel, and five-day scales combination, respectively, LIC_5 represents the increments applied to remain cloud allocation and LIC_6 corresponds to the reprocessing adjustment based on the physically consistent evolution of lake ice.

The reliability of daily LIC for each lake was estimated using an error propagation approach as follows:

$$Q = \sqrt{LIC_1 * (1 - A_1)^2 + LIC_2 * (1 - A_2)^2 + LIC_3 * (1 - A_3)^2 + LIC_4 * (1 - A_4)^2 + LIC_5 * (1 - A_5)^2}, \quad (8)$$

280 where Q represents the quality of the final LIC, and A_1, A_2, A_3, A_4 and A_5 denote the accuracies of the first five components, respectively.

The first four components can be evaluated using the validation methods described in Section 3.4, allowing their associated accuracies to be characterized. The accuracy of LIC_5 is constrained by and propagated from the accuracies of the first four components. LIC_6 cannot be directly assessed due to the lack of independent validation data; however, it primarily addresses 285 data gaps during the low illumination periods or noise in ice-free conditions, when most lakes have already reached a stable ice-cover state. Under such conditions, temporal variability in lake ice coverage is limited, and the likelihood of large errors introduced by this step is therefore low. Based on the final LIC and its quality Q , we further provide the estimated daily range of LIC, expressed as:

$$LIC\ Range = [max(0, Final\ LIC - Q), min(1, Final\ LIC + Q)], \quad (9)$$



290 3.5.2 Quality assessment for LIP

In Section 3.2.3, missing ice observations during low-illumination periods were reconstructed by interpolation using the nearest valid observations. However, for lakes that have not reached completely frozen state before these periods, this approach may influence the representation of LIC, and may consequently affect the identification of the FUE date. To support quality assessments, a quality-control flag was provided for the FUE date of all lakes. An FUE records is flagged
 295 when all LIC values within the three-day window used for FUE extraction are derived from the reprocessing step, indicating that the estimate is based solely on reprocessed LIC rather than original observations or cloud-gap-filled increments. The continuity of LIC records strongly influences the extraction of LIP. To further assess the quality of all ice events, we calculated, for each lake, the annual winter (define by $LIC > 10\%$) proportion of LIC values that are entirely derived from the reprocessing step, rather than the original observations and cloud-gap-filled increments. The statistics are provided with
 300 the dataset for users informed the use of the extracted phenology metrics, particularly for applications sensitive to data source composition.

4 Results

The lakes with reliable monitoring conditions in cold climate regions are widely distributed across both high-latitude and high-altitude regions, including 32,800 lakes worldwide, 14,369 in Eurasia, 17,914 in North America, 336 in South America,
 305 141 in Oceania, and 40 in Africa (Figure 3). Over half of these lakes lie north of 60°N , and 608 lakes are in the High Mountain Asia region. Compared with previous studies, our dataset substantially improves the spatial representativeness by incorporating a large number of small to medium-sized lakes that were previously underrepresented. This broad spatial coverage and environmental diversity provide an excellent basis for investigating lake ice cover dynamics and its response to climate change.

310 The LI-CCR dataset provides the daily LIC, annual lake ice-cover status, annual LIP and PICO information. Daily records of LIC, ranging from 0 to 100 %, quantitatively capture the ice cover variations. In addition to the final LIC, it also includes the six individual LIC components, the quality, and the estimated daily minimum and maximum LIC ranges. And for all lakes, the annual lake ice-cover status (completely frozen, partially frozen, or unfrozen) is provided. For completely frozen years, the phenology records key freeze and melt events (FUS, FUE, BUS, BUE) and the derived parameters ICD and CICD. For
 315 partially frozen lakes, FUS, BUE and ICD are included. Finally, the PCIO (0 to 100 %) is given. The data base spans from 1 August 2002 through 31 July 2024, covering 22 complete freeze-thaw cycles for the Northern Hemisphere and 21 cycles for the Southern Hemisphere.

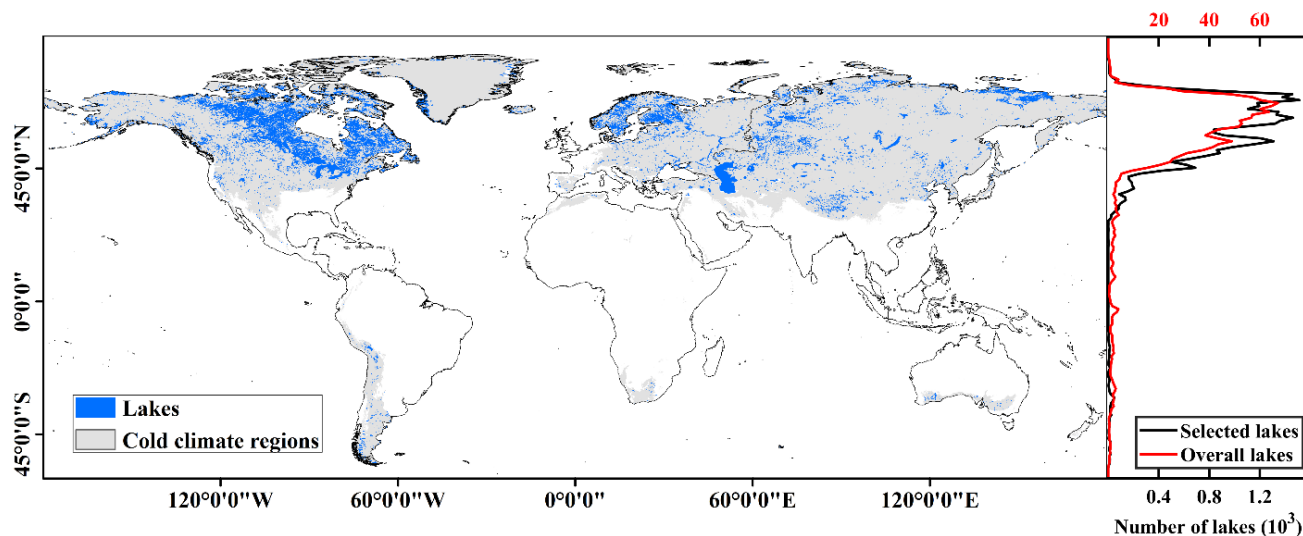


Figure 3: Selected lakes in the cold climate regions.

320 4.1 Accuracy evaluation

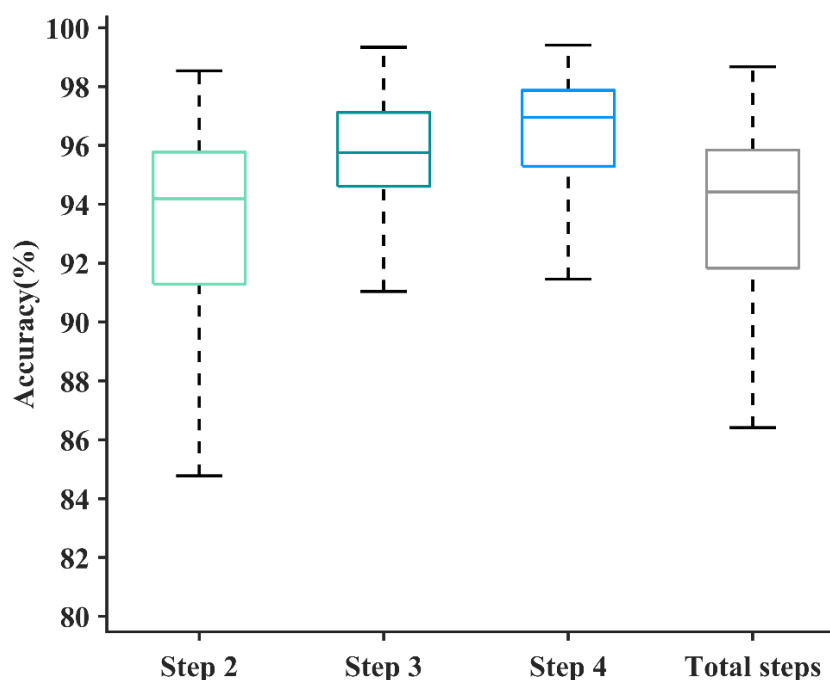
4.1.1 Performance and accuracy of cloud-gap-filled algorithms

To evaluate the performance of cloud-gap-filled algorithms, we selected several representative lakes in Asia. For each lake, we computed the annual mean cloud coverage from raw MODIS observations in 2015 and compared it with the remaining cloud coverage after each step (Table 2). The initial cloud cover typically ranged from 40 % to 60 %, substantially limiting the ability to monitor lake ice evolution with temporal and spatial continuity. Progressive application of algorithms significantly reduced cloud contamination after each step. In the final output, the remaining cloud cover was generally reduced to below 10 %, greatly improving data availability and continuity.

Table 2: A comparison of cloud cover (%) between original MODIS Aqua/Terra data and the outcome after each step in 2015.

Lake name	Terra	Aqua	Step 1	Step 2	Step 3	Step 4
Qinghai	39.8	41.2	27.1	4.1	4.0	2.3
Hulun	42.3	43.1	31.9	5.9	5.7	4.1
Namco	46.6	56.5	35.7	14.2	14.0	12.0
Xingkai	49.0	48.5	37.6	7.7	7.5	3.5
Baikal	54.6	55.3	45.0	14.6	14.3	9.7
Hala	46.1	56.3	34.3	6.2	6.0	2.3

To evaluate the accuracy of the cloud-gap-filled algorithm, we conducted cross-sensor validation for individual steps and the overall algorithm using daily MODIS data from the Terra and Aqua imagery in 2008. These twin satellites provide near-synchronous observations over the same lakes, offering a robust basis for evaluating the reconstruction accuracy under similar atmospheric and surface conditions. Due to the brief time interval between the observations, Step 1 was considered



340 **Figure 4: Accuracy of cloud-gap-filled algorithm steps in 2015.**

Overall, the cloud-gap-filled algorithm reliably reconstructs lake ice cover evolution obscured by cloud cover while maintaining high accuracy. By progressively reducing cloud-contaminated pixels, the spatially and temporally continuous MODIS-based lake ice record is substantially improved, thereby enhancing its suitability for long-term monitoring of LIP.

4.1.2 Comparison with reference lake ice cover dataset

345 To assess the accuracy of lake ice cover under both clear-sky and cloudy conditions, we compared our dataset with LIE-NH products (Table 3). Under clear-sky conditions, pixel-level cross-validation was conducted for over 220 million valid pixels. The MODIS-derived results from both Terra and Aqua showed strong consistency with LIE-NH across different seasons, achieving an overall classification accuracy of 98.6 %. The Terra results were slightly more accurate, possibly due to errors in the reconstructed Band 6 reflectance from Aqua. While generally reliable, MODIS-based results showed a slight tendency
 350 to overestimate ice extent.



To further evaluate the performance under cloud cover, we examined approximately 31 million pixels where MODIS observed clouds but LIE-NH remained cloud-free. The results indicated that our method successfully estimated lake ice cover across seasons, effectively addressing the data gaps caused by cloud contamination, improving the spatial and temporal continuity of the dataset.

Table 3: Accuracy of MODIS-derived lake ice cover under clear sky and cloudy conditions across seasons.

Observation conditions		Accuracy (%)				
		Annual	Summer	Autumn	Winter	Spring
Clear sky	Terra	98.7	99.7	99.0	96.3	98.4
	Aqua	98.6	99.5	98.3	97.3	98.0
	Total	98.6	99.6	98.6	96.8	98.3
Cloudy	Total	98.1	99.2	98.4	96.3	97.8

4.1.3 Comparison with reference lake ice phenology dataset

To assess the accuracy of MODIS-derived LIP, we compared it against two representative reference datasets: one based on passive microwave remote sensing and the other derived from physically based model simulations.

As shown in Figure 5a, our results show strong consistency with passive microwave data for both FUE and BUE, with R2 exceeding 0.94, RMSE of 6-8 days, and ME within 3 days. Notably, 60 % of the observations differed by less than 3 days, and about 10 % matched exactly. In contrast, although FUS and BUS also showed good correlation, our events tend to occur earlier, with the ME ranging from -10 to -7 days. This systematic bias may be attributed to spatial resolution differences: passive microwave sensors mainly capture the ice conditions in central lake area-where freezing and melting usually occur last-whereas MODIS is more sensitive in shallow or nearshore waters, enabling earlier detections of these events. We also compared our result with a physical model simulation (Figure 5b). Although the ME and RMSE increased for the key events, about 27 % of them deviated by less than 3 days. The R² remained above 0.76, indicating consistent seasonal patterns and reasonable consistency between satellite observations and model simulations.

In summary, our results showed high accuracy and robustness in capturing lake ice dynamics. Our dataset provides a reliable foundation for future studies on lake-climate interactions and environmental changes in cold climate regions.

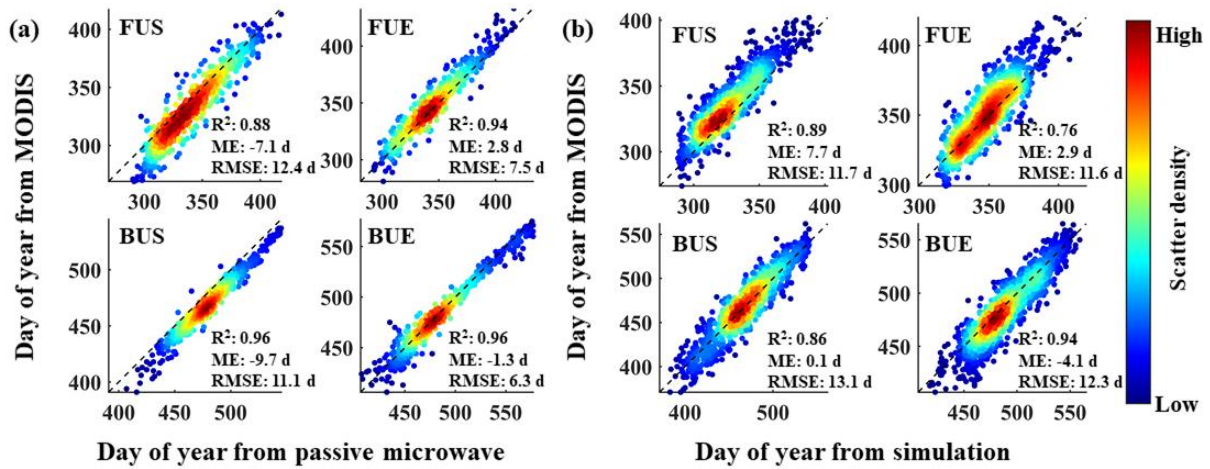


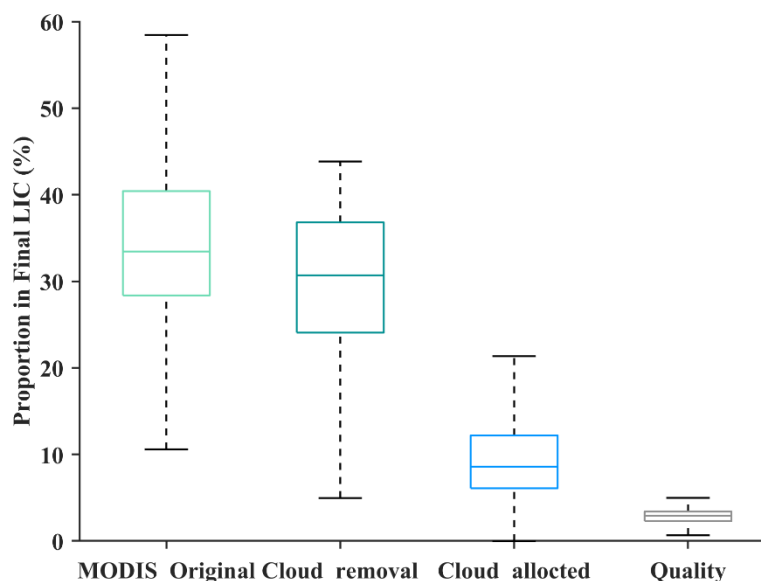
Figure 5: Comparison of MODIS-derived ice phenology with passive microwave data (a) and model simulations (b).

4.2 Quality assessment

4.2.1 Quality assessment results of LIC

In Section 3.5.1, we describe the method used to assess the quality of daily LIC. For A_1 , representing the accuracy of the original MODIS observations, we adopted the overall accuracy under clear-sky conditions derived from comparison with the LIE-NH dataset (98.6 %). For A_2 , A_3 , and A_4 , accuracies were obtained from the validation results presented in Section 4.1.1, and the 75th percentile of each accuracy distribution was used as input for the error propagation. Compared with the median or mean, this approach better captures a broader range of most conditions. The accuracy of A_5 was not independently evaluated but was derived through accuracies of the first four components. Using this framework, we derived, for each lake and each day, the final LIC together with the values of its individual components, the quality, and the corresponding LIC range.

To further examine the quality characteristics of LIC across different lakes, we quantified, for each lake, the mean contribution of each component to the final LIC and the quality during the winter season (defined as LIC > 10%). In this analysis, LIC₂, LIC₃, and LIC₄ were grouped as the cloud-removal contribution. Figure 6 summarizes the mean contributions and qualities across all lakes. Overall, the median contribution of direct MODIS observations accounts for approximately 33.4 % of the final LIC, while the cloud-removal procedures (Steps 2-4) contribute about 30.7 %, and remaining cloud allocation contributes roughly 8.6 %. The median relative quality of the final LIC remains low, at approximately 2.9 %. The component-wise contribution ratios for individual lake are also included in the dataset for user reference.



390 **Figure 6 Mean contribution of different component to the final lake ice coverage (LIC) and the quality during the winter season (LIC > 10%), averaged across all lakes.**

4.2.2 Quality assessment results of LIP

The following presents the quality-control flag results for the FUE dates. Approximately 29.8 % of the FUE records are flagged under this criterion, suggesting that the extracted FUE may be earlier than the true event. It should be noted, however, that this flag did not reflect a confirmed error. Across all lakes, the median number of flagged years is 6, and about 17.8 % of the lakes are flagged in more than 10 years. These lakes are mainly located at high latitudes or correspond to small lakes in mid- to low-latitude regions, where data gaps are more frequent. Users are therefore encouraged to consider this quality flag when higher temporal precision is required.

The following presents the quality assessment results for all ice events. It shows that, across all lakes and all winter seasons, the median proportion of LIC values obtained entirely through reprocessing is approximately 23.9%. When considering individual lakes across all years, about 18.1% of lakes have a median annual proportion below 10%, while the overall median is 22.7%. for the majority of lakes and years, phenology extraction is primarily based on original and cloud-gap-filled data. Users are advised to consider the proportion of reprocessed values when using the dataset for specific applications.

405 4.3 Patterns of lake ice phenology

We analysed the spatial patterns of median annual LIP across global cold-region lakes from 2002 through 2024 (Figure 7), focusing on the timing of FUS, FUE, BUS, and BUE. In the Northern Hemisphere, the median FUS and FUE occurred on days 293 and 298, respectively, while BUS and BUE occurred on days 498 and 508. The median ICD and CICD were 213



and 194 days, respectively. In contrast, there were just a few freezing lakes in the Southern Hemisphere, with mean FUS,
410 BUE, and ICD values of 170, 261, and 98 days, respectively.

The results show that during autumn and winter, over half of the lakes became ice-covered before November in the Northern Hemisphere or May in the Southern Hemisphere, mainly in high-latitude regions. Only about 5 % of the lakes froze after December or June, mostly in mid-latitude and low-altitude regions with mild winters. During spring and summer, only a small number of lakes began to melt before April in the Northern Hemisphere or October in the Southern Hemisphere,
415 mainly in mid-latitude and low-altitude areas. About 35 % of lakes started melting in May or November, whereas the latest melting occurred in high-latitude regions, where ice typically disappeared late in the season. To complement these analyses, we further mapped the ICD and CICD (Figure 8). The results show that around 70 % of the lakes remained ice-covered for six months or even longer annually, mainly in high-latitude regions and partly on the Tibetan Plateau.

To further explore the distribution of LIP, we analysed the phenological characteristics of lakes across different latitude
420 bands. Because only a few lakes ($n = 504$) were monitored in the Southern Hemisphere and their spatial distribution is sparse, the latitudinal analysis was therefore limited to lakes in the Northern Hemisphere ($n = 32,296$). The results (Figure 9) reveal strong linear relationships between the latitude and all phenological metrics: lakes at higher latitudes tend to freeze earlier, break up later, and maintain ice-cover for longer periods. For every 1° increase in latitude in the Northern Hemisphere, FUS and FUE advance by 2.5 days ($R = 0.85$ and 0.83), BUS and BUE are delayed by 3 days ($R = 0.79$ and 0.77), and ICD and
425 CICD increase by 5.5 days ($R = 0.83$ and 0.84). All these trends are statistically significant ($p < 0.05$), confirming a robust latitudinal control on lake ice dynamics.

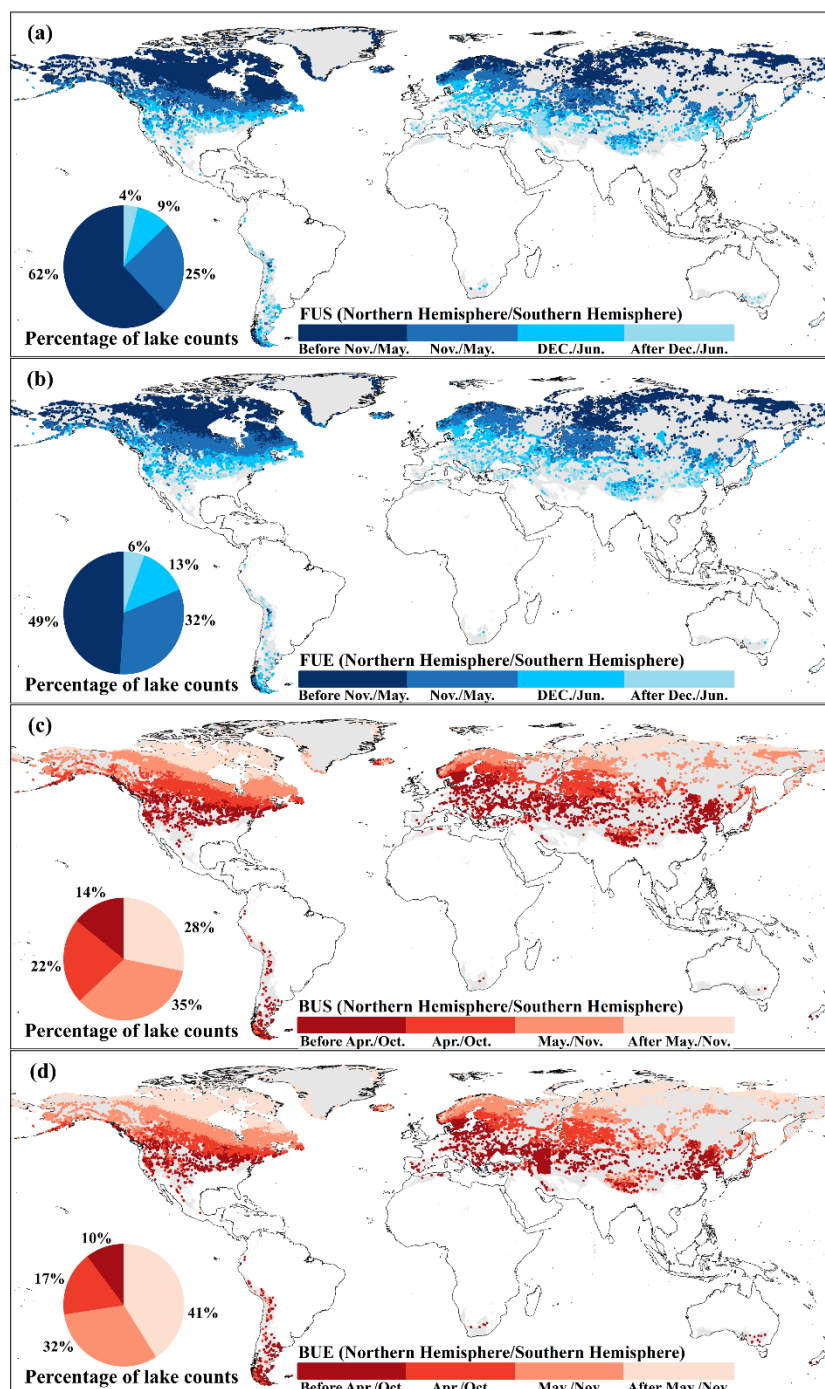


Figure 7: Spatial patterns of the median annual lake ice phenology for lakes in global cold regions during 2002-2024. The (a) mean FUS, (b) mean FUE, (c) mean BUS and (d) mean BUE were mapped.

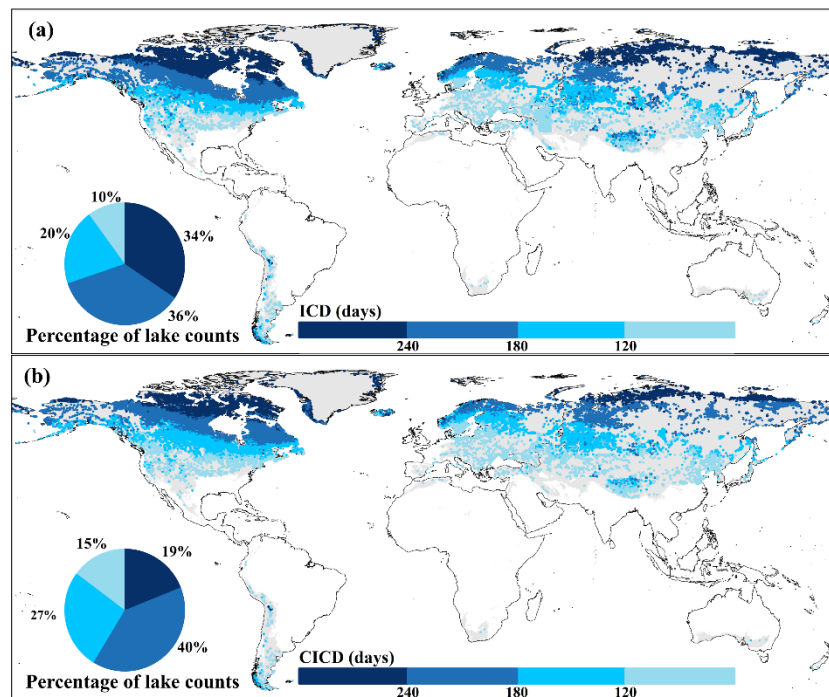


Figure 8: Spatial patterns of the median annual (a) ICD and (b) CICD for lakes in global cold regions during 2002-2024.

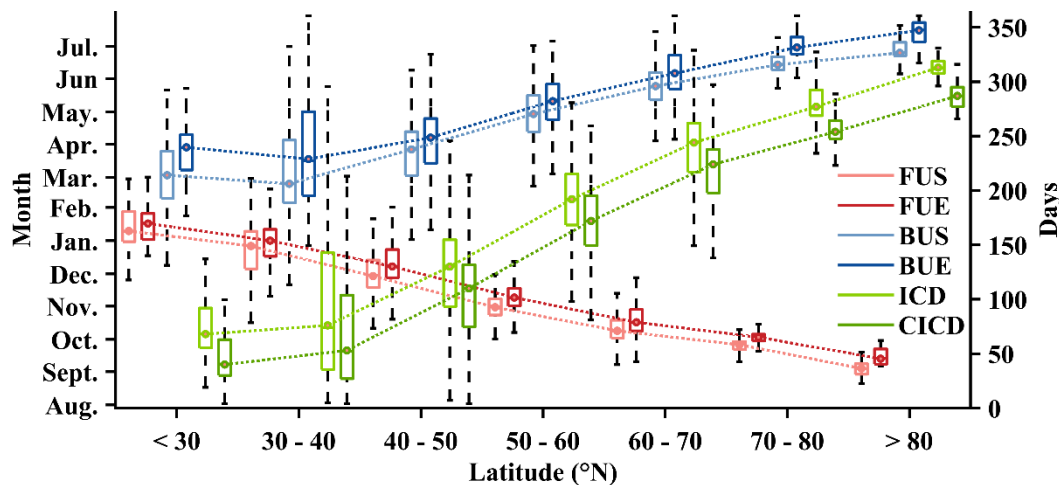


Figure 9: Geographical distribution of lake ice phenology for 32,296 lakes across latitudes in the Northern Hemisphere.

4.4 Patterns of lake ice-cover status

During the preparation of the LIP dataset, each lake was classified for each hydrological year as completely frozen, partially frozen, or unfrozen. Building on this information, the predominant ice-cover status of each lake was defined as the state in which it most frequently occurred from 2002 through 2024 (Figure 10). The results indicate that about 96 % of lakes were predominantly completely frozen during winter, primarily north of 30°N. About 2 % of lakes were partially frozen, mostly



large lakes in mid-latitude regions, while the remaining 2 % remained unfrozen, occurring mainly in areas with mild winters.

Overall, the proportion of completely frozen lakes was higher in North America (99 %) than in Eurasia (95 %).

The PCIO was further subsequently calculated for each lake over the study period, and lakes were classified into three categories: annually completely frozen (PCIO = 100 %), intermittently completely frozen ($0 < \text{PCIO} < 100$ %), and never completely frozen (PCIO = 0) (Figure 11). The results show that 89 % of the lakes are annually completely frozen, mainly north of 30°N. About 10 % reach complete ice cover only in some years with the mean areas of 114 km², locating primarily in mid-latitude regions.

The remaining 1 % have never been completely frozen, occurring mostly in low latitudes. The northernmost is Chilko Lake in Canada (51°19'N, 124°06'W), where only the northern section occasionally freezes due to the large area, great depth, and strong water flow (Brown et al., 2019). Overall, the proportion of annually completely frozen lakes is higher in North America (97 %) than in Eurasia (82 %), whereas intermittent completely frozen lakes account for 2.7 % and 17.3 % in North America and Eurasia, respectively. These spatial patterns reflect the greater stability of complete ice cover in high-latitude lakes, while mid-latitude lakes exhibit higher interannual variability in freezing conditions. In colder climate, ice becomes thick fast and forms a stable lid in small and medium-size lakes.

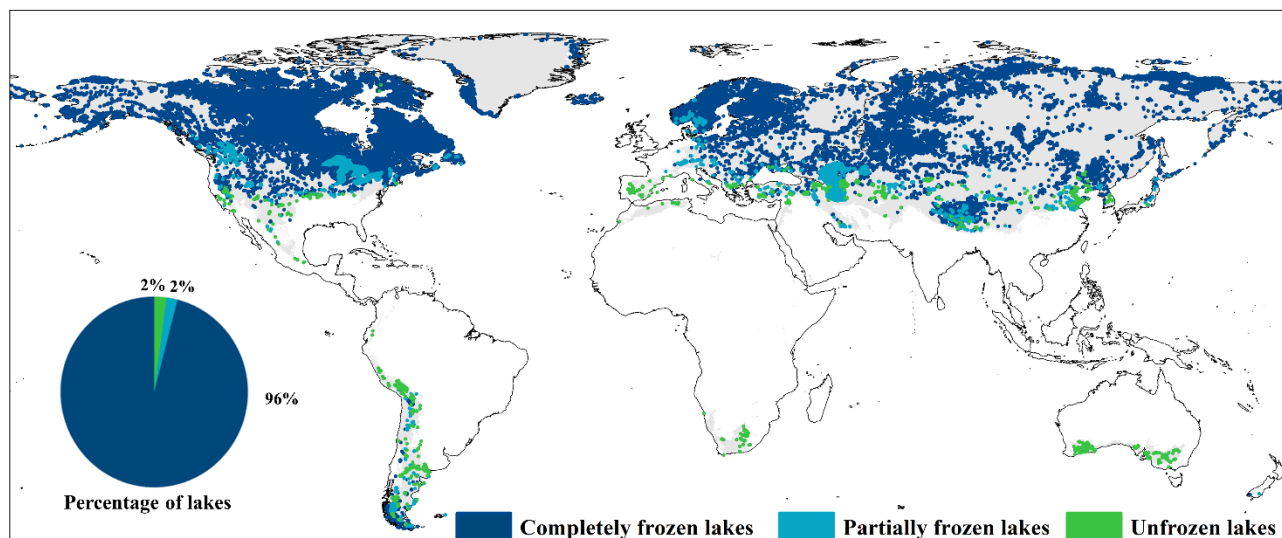


Figure 10: Predominant ice-cover status of lakes in global cold regions from 2002 through 2024, classified as completely frozen, partially frozen, or unfrozen.

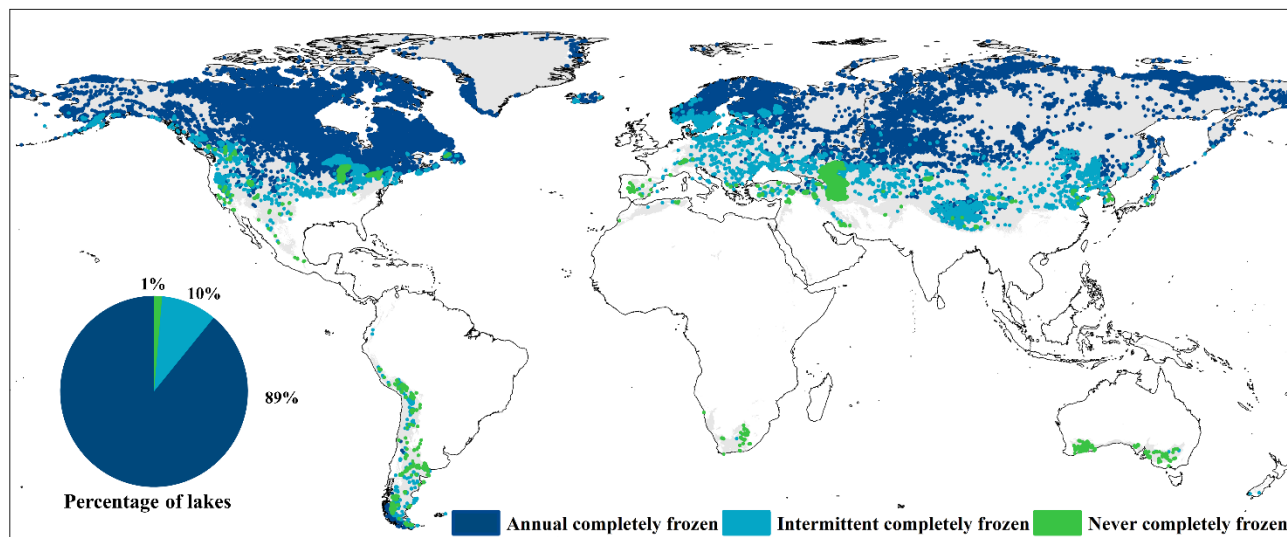


Figure 11: Classification of lakes in global cold regions based on the PCIO from 2002 through 2024, showing annually completely frozen, intermittently completely frozen, and never completely frozen lakes.

The dataset also provided the annual counts of lakes in each ice-cover status category for each year since 2002., allowing users to further explore lake ice conditions. Due to the absence of ice cover data for 2002-2003 in the Southern Hemisphere, the predominant ice-cover status during 2003-2008 was used as a substitute for 2002. The results show a gradual decline in the number of completely frozen lakes, on average decreasing by approximately 19 lakes per year (Figure 12). In contrast, the numbers of partially frozen and unfrozen lakes increased by about 10.6 and 8.4 lakes per year, respectively, indicating that an increasing proportion of lakes were experiencing intermittent or entirely absent ice-cover periods. These findings underscore the suppressive effect of climate warming on lake ice formation and provide a quantitative basis for long-term studies of LIP.

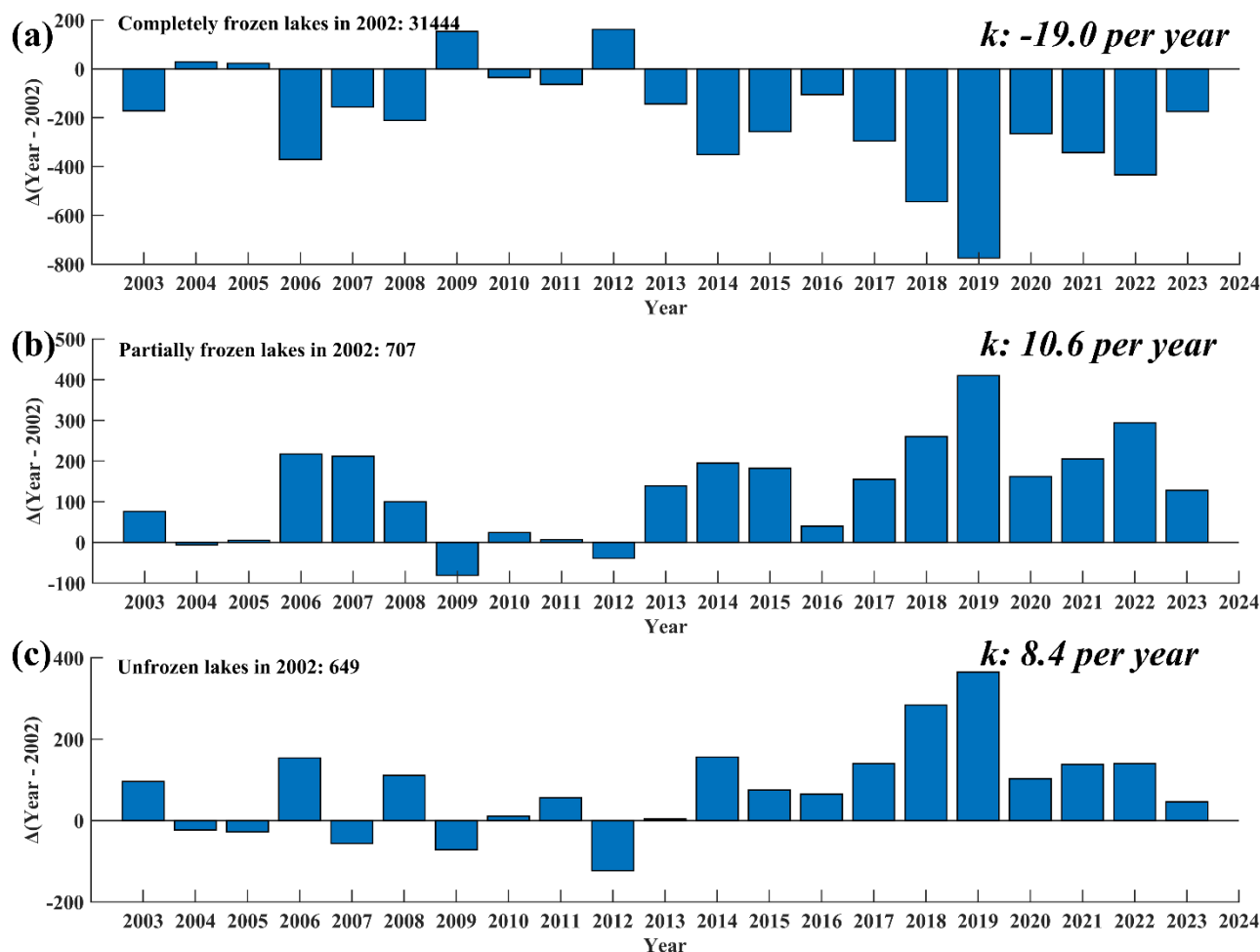


Figure 12: Temporal changes in the number of lakes within each ice-cover category from 2003 onward, relative to 2002 baseline counts, showing trends in completely frozen, partially frozen, and unfrozen lakes in global cold regions.

5 Discussions

470 Based on applications of lake observability and series of cloud-gap-filled algorithms, the LI-CCR dataset significantly improves upon existing lake ice climatology products by enhancing daily monitoring capabilities for small lakes (down to 2 km²). Covering the period from 2002 through 2024, it currently stands as the most spatially comprehensive and lake-inclusive satellite-mapped lake ice dataset. Nevertheless, it still contains uncertainties and limitations in both retrieval and analysis stages.



5.1 Temporal variability in the cloud-gap-filled performance and accuracy

The performance of the cloud-gap-filled algorithm depends on both the original cloud coverage and the stability of lake freeze-thaw conditions during the selected time windows. To evaluate the performance under varying cloud coverage, Qinghai Lake was chosen as a representative example. MODIS data from one day in each month of 2015 were used to compare the original cloud coverage with the outcomes after each processing step, as summarized in Table 4.

The results indicated that the short-window-based step has effectively removed most cloud pixels in most instances (Table 4). However, persistent cloud cover limits the performance, requiring an extended temporal window to include additional cloud-free observations. For example, on January 26, 2015, heavy cloud cover rendered Step 1 almost ineffective. Conversely, Step 2 utilized the lower cloud cover on the preceding and following days (24.9 % and 29.1 %), introducing more cloud-free pixels to fill the observation gaps. Each step aims to extend the spatiotemporal window to introduce more cloud-free information, further reducing the number of cloud pixels. In instances of consecutive cloud cover days, expanding the window can further reduce cloud cover effects but may also introduce uncertainties due to the lower spatiotemporal consistency. Therefore, this study only applied a five-day window and an eight-neighbours spatial window to enhance cloud removal performance.

Table 4: Cloud coverage of MODIS original data and after each cloud-gap-filled step for Qinghai Lake (%).

Date	Terra	Aqua	Step 1	Step 2	Step 3	Step 4
26 Jan	98.7	99.3	98.4	9.1	8.3	2.0
25 Feb	53.1	85.6	41.2	18.7	18.2	10.3
2 Mar	52.5	41.1	30.3	0.1	0.0	0.0
18 Apr	94.4	100.0	94.4	12.8	12.5	7.9
8 May	65.7	28.4	26.1	0.4	0.3	0.0
11 Jun	32.9	25.5	5.6	0.0	0.0	0.0
10 Jul	20.9	4.1	0.7	0.0	0.0	0.0
16 Aug	88.1	79.1	70.7	29.7	28.9	4.7
24 Sep	65.2	52.0	34.7	0.6	0.5	0.5
23 Oct	40.6	15.4	4.4	0.1	0.1	0.0
4 Nov	80.0	55.4	51.1	0.4	0.4	0.1
28 Dec	14.6	97.5	13.4	4.2	3.6	3.2

To assess the impact of unstable ice conditions on the accuracy of the algorithm, we analysed the performance across different freeze-thaw periods at Qinghai Lake using the method in Section 3.4.1. The melting period spanned over March 3–29, 2015, with freezing occurring as late as December 25. The results are summarized in Table 5. The algorithm maintains high accuracy during fully frozen and ice-free periods; however, rapid ice changes during unstable periods markedly reduced the spatiotemporal continuity, decreasing the accuracy. For example, during melting, the accuracy first declined and then



recovered due to relatively stable ice conditions in early and late phases, in contrast with rapid mid-phase changes. A similar pattern was observed during freezing, with accuracy showing clear variation over time.

Table 5: Cloud-filled algorithm accuracy during the melting process in Qinghai Lake.

Lake ice condition	Date	Accuracy (%)
Fully frozen	22 Jan	98.5
	16 Feb	98.0
Melting	11 Mar	94.9
	14 Mar	93.3
	19 Mar	84.4
	24 Mar	77.9
	28 Mar	90.3
	8 Apr	99.8
Ice-free	12 May	100.0

5.2 Potential improvements of the LIC algorithms

To address data gaps caused by polar night during autumn and winter at high latitudes, this study developed an interpolation method that performs well in most situations, especially for lakes that are completely frozen before the onset of the low-illumination period. However, for lakes that remain partially ice-free before polar night, a slight overestimation of ice coverage may be introduced. Future improvements could involve multi-source observations to enhance the accuracy in the gap filling. High-frequency passive microwave data is suitable for monitoring medium and large lakes during the polar night, and for small lakes, thermal infrared or radar data-despite their lower temporal resolution and potential to miss rapid ice dynamics-may still provide valuable information to improve interpolation performance.

Lake surface areas change over time due to long-term climate trends, interannual variability of water balance, and seasonal cycles (Li et al., 2025; Pekel et al., 2016). The lake mask used here comes from the HydroLAKES dataset released in 2016, which does not capture changes in lake extent. This limitation may reduce the accuracy and timeliness of lake ice datasets. Over the past two decades, many cold-region lakes have expanded, with some extending beyond the HydroLAKES boundaries. Because freeze-thaw processes usually begin near shorelines, the newly formed areas were excluded from monitoring, potentially causing ice phenology estimates to lag behind the actual conditions. HydroLAKES defines boundaries using the summer extent, which does not account for winter shrinkage due to reduced inflow. Snow-covered exposed shorelines can be misclassified as ice, leading to overestimation of ice cover, and adding uncertainty. Using time-varying lake masks would help to mitigate these effects and improve the accuracy of future lake ice datasets.



515 5.3 Limitations and considerations for dataset use

The LI-CCR dataset offers freeze-thaw observations at the lake scale, reflecting the overall seasonal ice dynamics of each lake. Lake thermodynamics are influenced by large scale weather patterns and lake characteristics which bring spatial heterogeneity, such as variations water depth, lake size, salinity, and water exchange, which may lead to substantial differences in freeze-thaw timing even across different parts of a lake. For localized analyses, modelling, or validation of LIP, users need to incorporate information on the lake for its spatial structure, in situ observations, or high-resolution datasets to improve the accuracy of the results and their interpretation.

Some lakes remain ice-free throughout the year due to factors such as deep water, high salinity, geothermal activity, or groundwater inflow. If they are excluded from spatial analyses of ice phenology or assessments of interannual variability, the results may be biased, leading to misinterpretation of regional lake ice dynamics. For instance, in certain low-latitude plateaus, a local change could be mistakenly interpreted as a shift from partially frozen to ice-free conditions driven by climate warming, thereby exaggerating the spatial extent and rate of lake ice loss. Excluding these lakes from dynamic monitoring and trend analyses, would enhance the dataset's accuracy and reliability for climate response studies.

6 Data availability

Daily Lake ice coverage (LIC) with quality assessments, annual lake ice-cover status, annual lake ice phenology (LIP) with quality assessments and the probability of complete ice-cover occurrence (PCIO) for 32,800 lakes from 2002 through 2024 in cold climate regions worldwide and the codes used to generate the LI-CCR dataset are available at <https://doi.org/10.5281/zenodo.17687698> (Jiang et al., 2025b).

7 Conclusions

Using MODIS observations, our study derived a comprehensive global lake ice dataset LI-CCR, including daily LIC, annual lake ice-cover status, annual LIP and the PCIO records for 32,800 lakes (size >2 km) in cold climate regions from 2002 through 2024. The classical SNOWMAP algorithm, combined with a spatiotemporal cloud-gap-filled algorithm, was used to obtain the ice cover status. A series of noise-removal approaches were then applied to obtain the daily LIC. Based on these results, annual lake ice cover status was assessed and then LIP was extracted using thresholds of 10 % and 90 % of the total lake area for the ice presence and for the full ice cover, respectively. Validation against Sentinel-3 products confirmed that the MODIS-based lake ice cover maintains high accuracy under both clear-sky (98.6 %) and cloudy (98.1 %) conditions. Compared with existing satellite-based and model-based phenology records, our dataset expands LIP monitoring to include a significantly larger number of small lakes, while maintaining high accuracy and reliability.

From 2002 through 2024, lakes experiencing ice-covered winters had an average ICD about 200 days. The dataset revealed a pronounced latitudinal gradient, with higher-latitude lakes tending to freeze earlier, melt later, and maintain ice-cover for



545 longer periods. Over 95 % of the lakes in cold regions are completely frozen during winter, reaching 99 % in North America; however, under climate warming, an increasing number of lakes are transitioning from fully frozen to partially frozen or even unfrozen state. Collectively, these patterns indicate that climate warming is leading to shorter and more intermittent ice-cover periods.

This dataset, encompassing daily LIC, annual LIP, annual lake ice-cover status and the PCIO records from 2002 through
550 2024, serves as a robust indicator of cryosphere changes under global warming. It provides a valuable foundation for investigating the impacts of altered lake ice changes on lake ecosystems, the response mechanisms of ice phenology to atmospheric teleconnections, and the socio-economic implications of changing ice regimes in cold regions.

Author contributions.

JZ conducted the study and wrote the manuscript. QY conceived the study. LM supervised the development and revision of
555 the manuscript. LX, YP, JG and SJ contributed to algorithm development and data processing. All the authors contributed to review and editing.

Competing interests.

The contact author has declared that none of the authors has any competing interests.

Acknowledgements.

560 We sincerely thank the data providers for their support to this study. The MODIS daily snow products were provided by the NASA. The HydroLAKES database was obtained from the HydroSHEDS. The CPC Global Unified Temperature data was downloaded from the NOAA Physical Sciences Laboratory. We also acknowledge the providers of the Köppen-Geiger climate classification map, the LIE-NH dataset and the lake ice phenology dataset.

Financial support.

565 This research has been supported by the National Key R&D Program of China (No:2022YFF0711700), the National Science and Technology Major Project (No. 2022ZD0117202), Xinjiang Key Laboratory of Water Cycle and Utilization in Arid Zone, Xinjiang Institute of Ecology and Geography, Chinese Academy of Science (No.XJYS0907-2023-03), the Innovative Research Program of the International Research Center of Big Data for Sustainable Development Goals (No. CBAS2022IRP08), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19070201 and
570 No. XDA19070102), and the Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (No.2021VTA0007).



References

- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Koppen-Geiger climate classification maps at 1-km resolution, *Scientific Data*, 5, 10.1038/sdata.2018.214, 2018.
- 575 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), *Climatic Change*, 112, 299-323, 10.1007/s10584-011-0212-8, 2012.
- Brooks, R. N., Prowse, T. D., and O'Connell, I. J.: Quantifying Northern Hemisphere freshwater ice, *Geophysical Research Letters*, 40, 1128-1131, 10.1002/grl.50238, 2013.
- 580 Brown, L. N., Borstad, G. A., Ersahin, K., Loos, E., Selbie, D., Costa, M., and Irvine, J. R.: Satellite-Based Time Series of Chlorophyll in Chilko Lake, British Columbia, Canada, *Canadian Journal of Remote Sensing*, 45, 368-385, 10.1080/07038992.2019.1632699, 2019.
- Cai, Y., Duguay, C. R., and Ke, C.-Q.: A 41-year (1979-2019) passive-microwave-derived lake ice phenology data record of the Northern Hemisphere, *Earth System Science Data*, 14, 3329-3347, 10.5194/essd-14-3329-2022, 2022a.
- 585 Cai, Y., Ke, C. Q., Xiao, Y., and Wu, J.: What caused the spatial heterogeneity of lake ice phenology changes on the Tibetan Plateau?, *Science of the Total Environment*, 836, 10.1016/j.scitotenv.2022.155517, 2022b.
- 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, *Journal of Geophysical Research-Atmospheres*, 124, 825-843, 10.1029/2018jd028993, 2019.
- 590 Carrea, L., Cretaux, J.-F., Liu, X., Wu, Y., Calmettes, B., Duguay, C. R., Merchant, C. J., Selmes, N., Simis, S. G. H., Warren, M., Yesou, H., Mueller, D., Jiang, D., Embury, O., Berge-Nguyen, M., and Albergel, C.: Satellite-derived multivariate world-wide lake physical variable timeseries for climate studies, *Scientific Data*, 10, 10.1038/s41597-022-01889-z, 2023.
- Cheng, B., Vihma, T., Rontu, L., Kontu, A., Pour, H. K., Duguay, C., and Pulliainen, J.: Evolution of snow and ice temperature, thickness and energy balance in Lake Orjarvi, northern Finland, *Tellus Series a-Dynamic Meteorology and Oceanography*, 66, 10.3402/tellusa.v66.21564, 2014.
- 595 Culpepper, J., Jakobsson, E., Weyhenmeyer, G. A., Hampton, S. E., Obertegger, U., Shchapov, K., Woolway, R. I., and Sharma, S.: Lake ice quality in a warming world, *Nature Reviews Earth & Environment*, 5, 671-685, 10.1038/s43017-024-00590-6, 2024.
- 600 Engram, M., Anthony, K. M. W., Sachs, T., Kohnert, K., Serafimovich, A., Grosse, G., and Meyer, F. J.: Remote sensing northern lake methane ebullition, *Nature Climate Change*, 10, 511-+, 10.1038/s41558-020-0762-8, 2020.
- Filazzola, A., Blagrove, K., Imrit, M. A., and Sharma, S.: Climate Change Drives Increases in Extreme Events for Lake Ice in the Northern Hemisphere, *Geophysical Research Letters*, 47, 10.1029/2020gl089608, 2020.
- Gafurov, A. and Bardossy, A.: Cloud removal methodology from MODIS snow cover product, *Hydrology and Earth System Sciences*, 13, 1361-1373, 10.5194/hess-13-1361-2009, 2009.
- 605 Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid (61) [dataset], <https://doi.org/10.5067/MODIS/MOD10A1.061>, 2021.
- Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, *Remote Sensing of Environment*, 83, 181-194, 10.1016/s0034-4257(02)00095-0, 2002.
- 610 Hampton, S. E., Powers, S. M., Dugan, H. A., Knoll, L. B., McMeans, B. C., Meyer, M. F., O'Reilly, C. M., Ozersky, T., Sharma, S., Barrett, D. C., Chandra, S., Jansen, J., McClure, R. P., Rautio, M., Weyhenmeyer, G. A., and Yang, X.: Environmental and societal consequences of winter ice loss from lakes, *Science*, 386, 10.1126/science.adl3211, 2024.
- Heinila, K., Mattila, O.-P., Metsamäki, S., Vakeva, S., Luojus, K., Schwaizer, G., and Koponen, S.: A novel method for detecting lake ice cover using optical satellite data, *International Journal of Applied Earth Observation and Geoinformation*, 104, 10.1016/j.jag.2021.102566, 2021.
- 615 Hodgkins, G. A., James, I. C., and Huntington, T. G.: Historical changes in lake ice-out dates as indicators of climate change in New England, 1850-2000, *International Journal of Climatology*, 22, 1819-1827, 10.1002/joc.857, 2002.
- Holgerson, M. A. and Raymond, P. A.: Large contribution to inland water CO₂ and CH₄ emissions from very small ponds, *Nature Geoscience*, 9, 222-U150, 10.1038/ngeo2654, 2016.



- 620 Huang, L., Timmermann, A., Lee, S.-S., Rodgers, K. B., Yamaguchi, R., and Chung, E.-S.: Emerging unprecedented lake ice loss in climate change projections, *Nature Communications*, 13, 10.1038/s41467-022-33495-3, 2022.
- Jansen, J., Weyhenmeyer, G. A., Harkonen, L. H., Paterson, A. M., del Giorgio, P. A., and Prairie, Y. T.: Divergent oxygen trends in ice-covered lakes driven by ice-cover decline and ecological memory, *Proceedings of the National Academy of Sciences of the United States of America*, 122, 10.1073/pnas.2426140122, 2025.
- 625 Jiang, Q., Sun, Y., Jeppesen, E., Smol, J. P., Scavia, D., Hecky, R. E., Mehner, T., Qin, Y., Tong, Y., Qin, B., Hambright, K. D., Jin, X., Li, J., Cai, K., Wu, Z., and Liu, Y.: Persistent inequities in global lake science, *Nature Reviews Earth & Environment*, 6, 629-631, 10.1038/s43017-025-00722-6, 2025a.
- Jiang, Z., Qiu, Y., Leppäranta, M., Li, X., Yao, P., Jia, G., and Shi, J.: LI-CCR: Dataset of daily lake ice evolution (2002-2024) across global cold climate regions based on gap-filled MODIS observations [dataset], <https://doi.org/10.5281/zenodo.17687698>, 2025b.
- 630 Johnson, M. S., Matthews, E., Du, J., Genovese, V., and Bastviken, D.: Methane Emission From Global Lakes: New Spatiotemporal Data and Observation-Driven Modeling of Methane Dynamics Indicates Lower Emissions, *Journal of Geophysical Research-Biogeosciences*, 127, 10.1029/2022jg006793, 2022.
- Kang, K. K., Duguay, C. R., Lemmetyinen, J., and Gel, Y.: Estimation of ice thickness on large northern lakes from AMSR-E brightness temperature measurements, *Remote Sensing of Environment*, 150, 1-19, 10.1016/j.rse.2014.04.016, 2014.
- 635 Korver, M. C., Lehner, B., Cardille, J. A., and Carrea, L.: Surface water temperature observations and ice phenology estimations for 1.4 million lakes globally, *Remote Sensing of Environment*, 308, 10.1016/j.rse.2024.114164, 2024.
- Leppäranta, M.: Ice Phenology and Thickness Modelling for Lake Ice Climatology, *Water*, 15, 10.3390/w15162951, 2023.
- Li, L., Long, D., Wang, Y., and Woolway, R. I.: Global dominance of seasonality in shaping lake-surface-extent dynamics, *Nature*, 642, 10.1038/s41586-025-09046-3, 2025.
- 640 Li, X., Long, D., Huang, Q., and Zhao, F.: The state and fate of lake ice thickness in the Northern Hemisphere, *Science Bulletin*, 67, 537-546, 10.1016/j.scib.2021.10.015, 2022a.
- Li, X., Peng, S., Xi, Y., Woolway, R. I., and Liu, G.: Earlier ice loss accelerates lake warming in the Northern Hemisphere, *Nature Communications*, 13, 10.1038/s41467-022-32830-y, 2022b.
- 645 Li, Z.-j., Huang, W.-f., Jia, Q., and Leppäranta, M.: Distributions of crystals and gas bubbles in reservoir ice during growth period, *Water Science and Engineering*, 4, 204-211, 10.3882/j.issn.1674-2370.2011.02.008, 2011.
- Mackay, M. D., Versegny, D. L., Fortin, V., and Rennie, M. D.: Wintertime Simulations of a Boreal Lake with the Canadian Small Lake Model, *Journal of Hydrometeorology*, 18, 2143-2160, 10.1175/jhm-d-16-0268.1, 2017.
- Magnuson, J. J., Robertson, D. M., Benson, B. J., Wynne, R. H., Livingstone, D. M., Arai, T., Assel, R. A., Barry, R. G., Card, V., Kuusisto, E., Granin, N. G., Prowse, T. D., Stewart, K. M., and Vuglinski, V. S.: Historical trends in lake and river ice cover in the Northern Hemisphere, *Science*, 289, 1743-1746, 10.1126/science.289.5485.1743, 2000.
- 650 Mangilli, A., Thibaut, P., Duguay, C. R., and Murfitt, J.: A New Approach for the Estimation of Lake Ice Thickness From Conventional Radar Altimetry, *Ieee Transactions on Geoscience and Remote Sensing*, 60, 10.1109/tgrs.2022.3186253, 2022.
- Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, *Nature Communications*, 7, 10.1038/ncomms13603, 2016.
- 655 Mu, C., Lei, P., Mu, M., Zhang, C., Zhou, Z., Song, J., Jia, Y., Fan, C., Peng, X., Zhang, G., Yang, Y., Wang, L., Li, D., Song, C., Wang, G., and Zhang, Z.: Methane emissions from thermokarst lakes must emphasize the ice-melting impact on the Tibetan Plateau, *Nature Communications*, 16, 10.1038/s41467-025-57745-2, 2025.
- Mugunthan, J. S., Duguay, C. R., and Zakharova, E.: Machine learning based classification of lake ice and open water from Sentinel-3 SAR altimetry waveforms, *Remote Sensing of Environment*, 299, 10.1016/j.rse.2023.113891, 2023.
- 660 Murfitt, J., Duguay, C., Picard, G., and Lemmetyinen, J.: Forward modelling of synthetic-aperture radar (SAR) backscatter during lake ice melt conditions using the Snow Microwave Radiative Transfer (SMRT) model, *Cryosphere*, 18, 869-888, 10.5194/tc-18-869-2024, 2024.
- Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, *Nature*, 540, 418-+, 10.1038/nature20584, 2016.
- 665 Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., Ma, E., Cheng, R., Fensholt, R., Brandt, M., Cai, X., Gibson, L., Liu, J., Zheng, C., Li, W., and Bryan, B. A.: Mapping global lake dynamics reveals the emerging roles of small lakes, *Nature Communications*, 13, 10.1038/s41467-022-33239-3, 2022.
- Qiu, Y., Xie, P., Leppäranta, M., Wang, X., Lemmetyinen, J., Lin, H., and Shi, L., - 3, - 185, 2019.



- 670 Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvarinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, *Communications Earth & Environment*, 3, 10.1038/s43247-022-00498-3, 2022.
- Sharma, S., Blagrove, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., Magee, M. R., Straile, D., Weyhenmeyer, G. A., Winslow, L., and Woolway, R. I.: Widespread loss of lake ice around the Northern Hemisphere in a warming world, *Nature Climate Change*, 9, 227–+, 10.1038/s41558-018-0393-5, 2019.
- 675 Sharma, S., Meyer, M. F., Culpepper, J., Yang, X., Hampton, S., Berger, S. A., Brousil, M. R., Fradkin, S. C., Higgins, S. N., Jankowski, K. J., Kirillin, G., Smits, A. P., Whitaker, E. C., Yousef, F., and Zhang, S.: Integrating Perspectives to Understand Lake Ice Dynamics in a Changing World, *Journal of Geophysical Research-Biogeosciences*, 125, 10.1029/2020jg005799, 2020.
- 680 Sharma, S., Richardson, D. C., Woolway, R. I., Imrit, M. A., Bouffard, D., Blagrove, K., Daly, J., Filazzola, A., Granin, N., Korhonen, J., Magnuson, J., Marszelewski, W., Matsuzaki, S.-I. S., Perry, W., Robertson, D. M., Rudstam, L. G., Weyhenmeyer, G. A., and Yao, H.: Loss of Ice Cover, Shifting Phenology, and More Extreme Events in Northern Hemisphere Lakes, *Journal of Geophysical Research-Biogeosciences*, 126, 10.1029/2021jg006348, 2021.
- Sharma, S., Filazzola, A., Nguyen, T., Imrit, M. A., Blagrove, K., Bouffard, D., Daly, J., Feldman, H., Felsine, N., Hendricks-Franssen, H.-J., Granin, N., Hecock, R., L'Abée-Lund, J. H., Hopkins, E., Howk, N., Iacono, M., Knoll, L. B., Korhonen, J., Malmquist, H. J., Marszelewski, W., Matsuzaki, S.-I. S., Miyabara, Y., Miyasaka, K., Mills, A., Olson, L., Peters, T. W., Richardson, D. C., Robertson, D. M., Rudstam, L., Wain, D., Waterfield, H., Weyhenmeyer, G. A., Wiltse, B., Yao, H., Zhdanov, A., and Magnuson, J. J.: Long-term ice phenology records spanning up to 578 years for 78 lakes around the Northern Hemisphere, *Scientific Data*, 9, 10.1038/s41597-022-01391-6, 2022.
- 685 Shi, L., Peng, C., Qiu, Y., Zhang, Y., Lemmetyinen, J., Li, L., Lepparanta, M., Cheng, B., Kontu, A., and Shi, J.: Lake Ice Thickness Estimation Using Calibrated Enhanced-Resolution Passive Microwave Data, *Ieee Transactions on Geoscience and Remote Sensing*, 63, 10.1109/tgrs.2025.3586818, 2025.
- Song, C., Liu, S., Wang, G., Zhang, L., Rosentreter, J. A., Zhao, G., Sun, X., Yao, Y., Mu, C., Sun, S., Hu, Z., Lin, S., Sun, J., Li, Y., Wang, Y., Li, Y., Raymond, P. A., and Karlsson, J.: Inland water greenhouse gas emissions offset the terrestrial carbon sink in the northern cryosphere, *Science Advances*, 10, 10.1126/sciadv.adp0024, 2024.
- 690 Tom, M., Aguilar, R., Imhof, P., Leinss, S., Baltsavias, E., and Schindler, K.: Lake Ice Detection from Sentinel-1 SAR with Deep Learning, *Arxiv*, arXiv:2002.07040, 2020.
- Verpoorter, C., Kutser, T., Seekell, D. A., and Tranvik, L. J.: A global inventory of lakes based on high-resolution satellite imagery, *Geophysical Research Letters*, 41, 6396–6402, 10.1002/2014gl060641, 2014.
- 700 Wang, W., Shi, K., Woolway, R. I., and Zhang, Y.: Chinese ice-lake line shifts under climate change, *Science Bulletin*, 70, 804–807, 10.1016/j.scib.2024.12.013, 2025.
- Wang, X., Qiu, Y., Zhang, Y., Lemmetyinen, J., Cheng, B., Liang, W., and Lepparanta, M.: A lake ice phenology dataset for the Northern Hemisphere based on passive microwave remote sensing, *Big Earth Data*, 6, 401–419, 10.1080/20964471.2021.1992916, 2022a.
- 705 Wang, X., Feng, L., Gibson, L., Qi, W., Liu, J., Zheng, Y., Tang, J., Zeng, Z., and Zheng, C.: High-Resolution Mapping of Ice Cover Changes in Over 33,000 Lakes Across the North Temperate Zone, *Geophysical Research Letters*, 48, 10.1029/2021gl095614, 2021.
- Wang, X., Feng, L., Qi, W., Cai, X., Zheng, Y., Gibson, L., Tang, J., Song, X.-p., Liu, J., Zheng, C., and Bryan, B. A.: Continuous Loss of Global Lake Ice Across Two Centuries Revealed by Satellite Observations and Numerical Modeling, *Geophysical Research Letters*, 49, 10.1029/2022gl099022, 2022b.
- 710 Weber, H., Riffler, M., Noges, T., and Wunderle, S.: Lake ice phenology from AVHRR data for European lakes: An automated two-step extraction method, *Remote Sensing of Environment*, 174, 329–340, 10.1016/j.rse.2015.12.014, 2016.
- Wik, M., Varner, R. K., Anthony, K. W., MacIntyre, S., and Bastviken, D.: Climate-sensitive northern lakes and ponds are critical components of methane release, *Nature Geoscience*, 9, 99–+, 10.1038/ngeo2578, 2016.
- 715 Williams, R. S. and Ferrigno, J. G.: State of the Earth's cryosphere at the beginning of the 21st century : glaciers, global snow cover, floating ice, and permafrost and periglacial environments: Chapter A in *Satellite image atlas of glaciers of the world*, Woolway, R. I. and Merchant, C. J.: Worldwide alteration of lake mixing regimes in response to climate change, *Nature Geoscience*, 12, 271–+, 10.1038/s41561-019-0322-x, 2019.



- 720 Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., and Sharma, S.: Global lake responses to climate change, *Nature Reviews Earth & Environment*, 1, 388-403, 10.1038/s43017-020-0067-5, 2020.
- Wu, Y., Guo, L., Zhang, B., Zheng, H., Fan, L., Chi, H., Li, J., and Wang, S.: Ice phenology dataset reconstructed from remote sensing and modelling for lakes over the Tibetan Plateau, *Scientific Data*, 9, 10.1038/s41597-022-01863-9, 2022.
- 725 Xie, A., Zhu, J., Kang, S., Qin, X., Xu, B., and Wang, Y.: Polar amplification comparison among Earth's three poles under different socioeconomic scenarios from CMIP6 surface air temperature, *Scientific Reports*, 12, 10.1038/s41598-022-21060-3, 2022.
- Xie, F., Lu, P., Lepparanta, M., Cheng, B., Li, Z., Zhang, Y., Zhang, H., and Zhou, J.: Heat budget of lake ice during a complete seasonal cycle in lake Hanzhang, northeast China, *Journal of Hydrology*, 620, 10.1016/j.jhydrol.2023.129461, 2023.
- 730 Xu, Y., Long, D., Li, X., Wang, Y., Zhao, F., and Cui, Y.: Unveiling lake ice phenology in Central Asia under climate change with MODIS data and a two-step classification approach, *Remote Sensing of Environment*, 301, 10.1016/j.rse.2023.113955, 2024.
- Yang, X., Pavelsky, T. M., Bendezu, L. P., and Zhang, S.: Simple Method to Extract Lake Ice Condition From Landsat Images, *Ieee Transactions on Geoscience and Remote Sensing*, 60, 10.1109/tgrs.2021.3088144, 2022.
- 735 Yang, Z., Wang, J., Liu, L., Miao, H., Miao, X., and Zhang, Q.: Estimating effects of wind and waves on the Doppler centroid frequency shift for the SAR retrieval of ocean currents, *Remote Sensing of Environment*, 311, 10.1016/j.rse.2024.114312, 2024.
- You, Q., Cai, Z., Pepin, N., Chen, D., Ahrens, B., Jiang, Z., Wu, F., Kang, S., Zhang, R., Wu, T., Wang, P., Li, M., Zuo, Z., Gao, Y., Zhai, P., and Zhang, Y.: Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences, *Earth-Science Reviews*, 217, 10.1016/j.earscirev.2021.103625, 2021.
- 740 Zhang Yixiao, Q. I. U. Y. P. C. B. C. S. H. I. L. and Matti, L.: - Brightness temperature simulation of subarctic lake-ice evolution by HIGHTSI and SMRT model, - 35, - 323, 2024.