



## MODIS-based Daily Lake Ice Extent and Coverage dataset for Tibetan Plateau

Yubao Qiu, Pengfei Xie, Matti Leppäranta, Xingxing Wang, Juha Lemmetyinen, Hui Lin & Lijuan Shi

To cite this article: Yubao Qiu, Pengfei Xie, Matti Leppäranta, Xingxing Wang, Juha Lemmetyinen, Hui Lin & Lijuan Shi (2019) MODIS-based Daily Lake Ice Extent and Coverage dataset for Tibetan Plateau, Big Earth Data, 3:2, 170-185, DOI: [10.1080/20964471.2019.1631729](https://doi.org/10.1080/20964471.2019.1631729)

To link to this article: <https://doi.org/10.1080/20964471.2019.1631729>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.



Published online: 28 Jun 2019.



Submit your article to this journal [↗](#)




Article views: 229



View Crossmark data [↗](#)



## MODIS-based Daily Lake Ice Extent and Coverage dataset for Tibetan Plateau

Yubao Qiu <sup>a,b</sup>, Pengfei Xie<sup>a,c</sup>, Matti Leppäranta<sup>d</sup>, Xingxing Wang<sup>a,b</sup>,  
Juha Lemmetyinen <sup>b,e</sup>, Hui Lin<sup>f,g</sup> and Lijuan Shi<sup>a,b</sup>

<sup>a</sup>Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; <sup>b</sup>FMI-AIR Joint Research Center for Arctic Observations, Finnish Meteorological Institute and Aerospace Information Research Institute, Chinese Academy of Sciences, Sodankylä, Finland; <sup>c</sup>School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China; <sup>d</sup>Institute of Atmospheric and Earth Sciences, University of Helsinki, Helsinki, Finland; <sup>e</sup>Arctic Space Center, Finnish Meteorological Institute, Helsinki, Finland; <sup>f</sup>School of Geography and Environment, Jiangxi Normal University, Nanchang, China; <sup>g</sup>Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China

### ABSTRACT

The Tibetan Plateau houses numerous lakes, the phenology and duration of lake ice in this region are sensitive to regional and global climate change, and as such are used as key indicators in climate change research, particularly in environment change comparison studies for the Earth three poles. However, due to its harsh natural environment and sparse population, there is a lack of conventional *in situ* measurement on lake ice phenology. The Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Snow Index (NDSI) data, which can be traced back 20 years with a 500 m spatial resolution, were used to monitor lake ice for filling the observation gaps. Daily lake ice extent and coverage under clear-sky conditions was examined by employing the conventional SNOWMAP algorithm, and those under cloud cover conditions were re-determined using the temporal and spatial continuity of lake surface conditions through a series of steps. Through time series analysis of every single lake with size greater than 3 km<sup>2</sup> in size, 308 lakes within the Tibetan Plateau were identified as the effective records of lake ice extent and coverage to form the Daily Lake Ice Extent and Coverage dataset, including 216 lakes that can be further retrieved with four determinable lake ice parameters: Freeze-up Start (FUS), Freeze-up End (FUE), Break-up Start (BUS), and Break-up End (BUE), and 92 lakes with two parameters, FUS and BUE. Six lakes of different sizes and locations were selected for verification against the published datasets by passive microwave remote sensing. The lake ice phenology information obtained in this paper was highly consistent with that from passive microwave data at an average correlation coefficient of 0.91 and an RMSE value varying from 0.07 to 0.13. The present dataset is more effective at detecting lake ice parameters for smaller lakes than the coarse resolution passive microwave remote sensing observations. The published data are available in <https://data.4tu.nl/repository/uuid:fdfd8c76-6b7c-4bbf-aec8-98ab199d9093> and <http://www.sciencedb.cn/dataset/handle/744>.

### ARTICLE HISTORY

Received 9 April 2019  
Accepted 11 June 2019

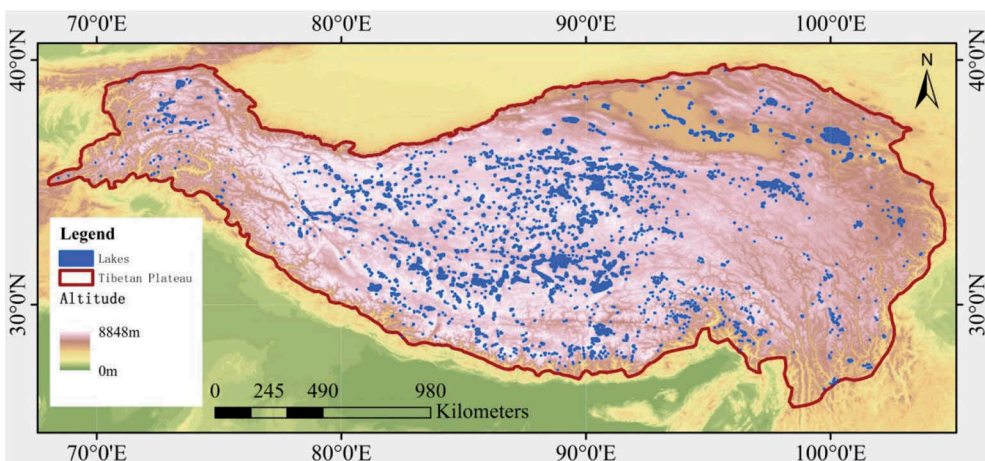
### KEYWORDS

Tibetan Plateau; Lake Ice Extent; Lake Ice Coverage; Daily; MODIS

## 1. Introduction

Known as the “Roof of the World”, the Tibetan Plateau has an average elevation of more than 2,000 m and an abundance of lakes. This region contains 2,612 lakes with a surface area greater than 0.25 km<sup>2</sup>, as estimated from the lake database (Messenger, Lehner, Grill, Nedeva, & Schmitt, 2016) (Figure 1). Lake ice is an indicator of changes in lake environments; moreover, it is also one of the key Environmental Climate Variables (ECVs) in the Global Climate Observing System (National centers for environmental information [EB/OL], 2017) (GCOS). Lake ice phenology, extent, and the thickness and type of ice are of great significance for climate and environmental studies (Liu, Chen, Wang, & Qiu, 2018). However, the existing *in situ* lake ice observation data for the Tibetan Plateau is severely limited due to the remoteness, harsh eco-environmental conditions, and thin atmosphere in this region. Therefore, a comprehensive lake ice monitoring dataset is urgently needed to compensate for this observation shortage.

At present, Many researchers have studied the lake ice phenology based on remote sensing data (Cai et al., 2019; Du, Kimball, Duguay, Kim, & Watts, 2017; Kropáček, Maussion, Chen, Hoerz, & Hochschild, 2013; Peng, Qinghua, & Qiufang, 2015; Qi et al., 2019), and released four data sets: (1) The Global Lake and River Ice Phenology Database (Benson & Magnuson, 2000), which contains freeze-up/breakup dates of 865 lakes and rivers in the Northern Hemisphere but includes only several large lakes in China; (2) The Canadian Ice Database (Canadian Cryospheric Information Network, 2018), which concentrates on North America, contains ice freeze-up and breakup data of 259 Canadian lakes, including lake ice thickness, snow depth, and freeze-up and breakup dates; (3) the Daily Lake Ice Phenology Time Series Derived from AMSR-E and AMSR2 (Du & Kimball, 2018) which describes daily lake-ice conditions (binary ice-on/ice-off) from 2002 to 2015 using a 5 km x 5 km grid (total 76,671 pixels) over the Northern Hemisphere and presents the ice coverage using downscaled coarse passive microwave observations; and, (4) The dataset of microwave brightness temperature and freeze-thaw status for



**Figure 1.** Distribution of lakes within the Tibetan Plateau (Zhang, Yao, Xie, Kang, & Lei, 2013). Lake boundaries derived from (Messenger et al., 2016). Background map derived from the global open digital elevation model data.

medium to large lakes over the High Asia (Yubao et al., 2018; Qiu, Guo, & Ruan et al., 2017), which includes lake-ice phenology for 51 lakes, of which 47 are located on the Tibetan Plateau, from 2002 to 2016. The dataset (4) was derived primarily from coarse resolution passive microwave observations, and it is not effective for monitoring the numerous small lakes in the Tibetan Plateau.

The dataset presented herein was prepared using the MODIS Normalized Difference Snow Index (NDSI) with a spatial resolution of 500 m as input for the NDSI snow cover algorithm SNOWMAP (George & Dorothy) in order to detect the lake ice coverage under daily clear sky conditions. Furthermore, lake ice coverage under cloud cover conditions was identified using the spatial and temporal continuity of the lake surface conditions. This method was used to calculate and classify the daily lake ice characteristics for the maximum number of lakes within the Tibetan Plateau from 2002 to 2018. Moreover, a time series analysis of lake ice coverage, which included lakes with a surface area greater than 3 km<sup>2</sup>, was carried out to provide a clear list of lakes where the lake ice phenology, i.e. Freeze-up Start (FUS), Freeze-up End (FUE), Break-up Start (BUS), or Break-up End (BUE), can be accurately estimated and retrieved.

## 2. Methodology and process steps

The MODIS Snow Cover Daily L3 Global 500 m Grid dataset was used as the input data, which was obtained from the US National Snow and Ice Data Center (Hall & Riggs, 2016), including MODIS/Terra morning acquisition data (MOD10A1) and MODIS/Aqua afternoon acquisition data (MYD10A1). The MODIS data pre-processing was carried out at the UTM/WGS84 system with steps of image splicing, projection, transformation, and image cropping. The NDSI snow cover algorithm SNOWMAP (George & Dorothy) was used to estimate the daily lake ice on/off at a single pixel under clear sky, and then a series of further processes was conducted under cloudy conditions based on spatial and temporal continuity of lake surface conditions.

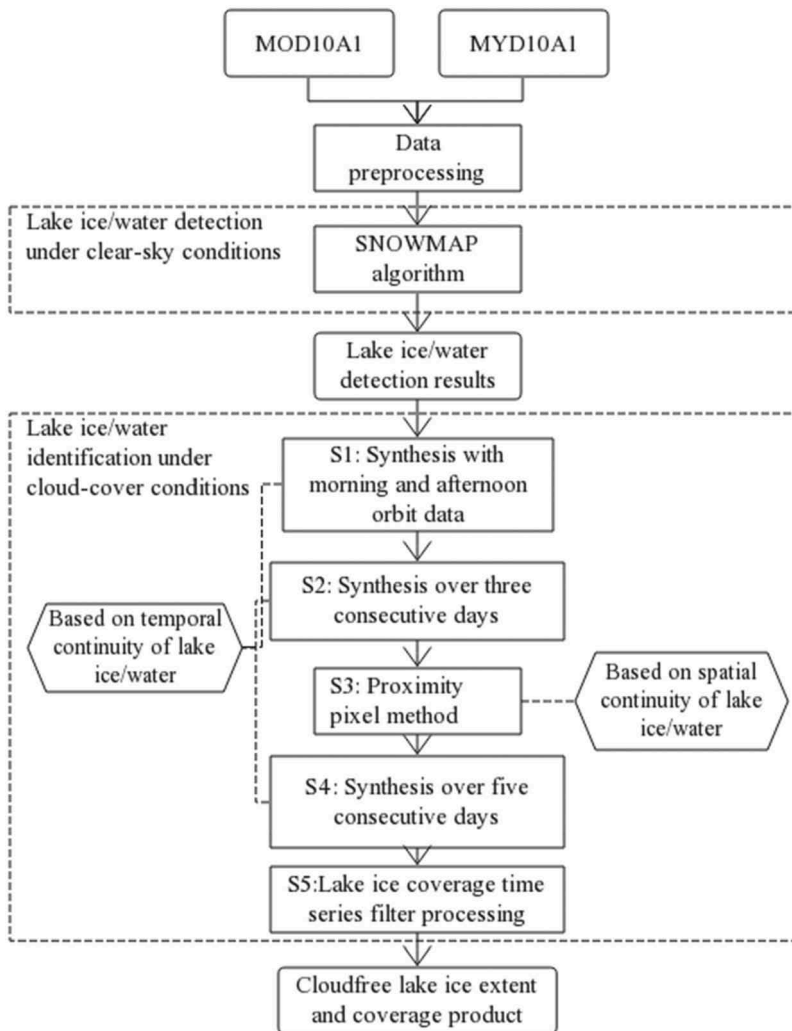
As a result, the lake ice extent data with partial cloud cover were obtained for 5,834 days, and lake ice, water and cloud coverage for each single lake, which was defined as the ratio of extent of lake ice and lake area, were obtained accordingly. The coverages for each lake were then processed with a time series filter to eliminate the contamination, thereby obtaining the final cloud-free lake ice coverage data product at cloud-free conditions.

The data-processing flow is shown in Figure 2 and described below.

### 2.1. Application of the SNOWMAP algorithm

Lake ice detection is based on the spectral reflectance characteristics of ice and water in the MODIS bands (Equation (1)). The reflectance of ice in the visible and near-infrared spectrum is high, whereas it is low for water. Therefore, the conventional SNOWMAP algorithm (George & Dorothy) was used to distinguish lake ice from lake water. For MODIS data, the NDSI is calculated as follows:

$$NDSI = \frac{band4 - band6}{band4 + band6} \quad (1)$$



**Figure 2.** Flow chart of the Tibetan Plateau daily lake ice extent and coverage algorithm used within this study.

where band4 and band6 are the reflectance at 545–565 nm and 1628–1652 nm bands, respectively. Under clear-sky conditions, when the NDSI is greater than 0.4 and the near-infrared reflectance of band 2 and band 4 are greater than 0.11 and 0.10, respectively, the pixels within lake areas are classified as lake ice.

## 2.2. Identification of lake ice/water under cloud area

First, a lake map containing the lake's vector boundary in the Tibetan Plateau was created based on the lake data obtained from the HydroLAKES database (Messenger et al., 2016) for lakes with an area larger than 1 km<sup>2</sup>. Next, the SNOWMAP algorithm was used to detect lake ice using MODIS data during the morning and afternoon orbit, respectively. However, the results still contain a large amount of cloud contamination.

To overcome this issue, the temporal and spatial continuity of lake ice for a single lake was used to re-identify lake ice/water under cloud cover conditions, the methods are described below by a sequence of steps depending on the spatial-temporal combination rule – the “more reliability, earlier process”, which is to say that the process steps with high correctness probabilities for cloud pixels were processed at a first priorities (Huan, Yubao, Zhaojun, Duo, & Yudan, 2016).

### **S1: Synthesis with morning and afternoon orbit data.**

Based on the complementary morning and afternoon satellite observations, daily lake ice/water data under cloud cover conditions was synthesized based on the following three rules: (a) If the same pixel was identified as lake water, either in the morning or afternoon, the cloudy pixel was designated as lake water, assuming that the water observed in a single pixel is more credible according to the algorithm; (b) If the same pixel was identified as lake ice both in the morning and afternoon, the pixel was designated as lake ice; and, (c) If the same pixel was identified as cloud-covered in either morning or afternoon, while being identified as lake ice during the other observation, the pixel was designated as lake ice.

### **S2: Temporal continuity synthesis over three consecutive days.**

Based on the temporal continuity of water/ice, the re-identification of cloud-covered pixels using a three consecutive day data interval were carried out based on the following rules: (a) If both the previous and following day observations for the same pixel was lake ice, the cloud-covered pixel was identified as lake ice; (b) If either the previous or following day observation for the same pixel was lake water, the cloud-covered pixel was identified as lake water; or, (c) If neither of the above rules applied, the original pixel value of the cloud was retained. This step was conducted before Step 3 and Step4 bellow (Huan et al., 2016).

### **S3: Spatially proximity pixel method.**

The cloud-covered pixels were re-identified using spatial adjacent relationships and its spatial continuity based on their proximity. Cloud-covered pixels adjacent to the shore of the lake were considered together with the eight adjacent pixels, after eliminating land pixels, if the remaining pixels were identified either as lake water or lake ice, then the cloud-covered pixel was identified as lake water or lake ice. On the other hand, any cloud pixel that was not close to the lake’s shore was identified by comparing it with the four adjacent pixels based on the following rules: (a) If at least three out of the four pixels adjacent to the cloud-covered pixel were identified as lake ice, then the center cloud pixel was designated as lake ice; (b) If at least three out of four of the adjacent pixels were identified as lake water, then the center cloud-covered pixel was designated as lake water; or, (c) If neither of the above rules applied, the pixel retained its cloud-cover value.

### **S4: Temporal continuity synthesis over five consecutive days.**

Based on the above method of processing, a five-day time period was used for the pixel identification in order to obtain lake ice surface data under partly cloudy coverage, based on the following rules: (a) If a pixel that was cloud-covered on day three was identified as lake ice on three out of the four remaining days, and no lake water was detected, then the pixel was designated as lake ice; (b) If a pixel that was cloud-covered on day three was identified as lake water on at least three out of the four remaining days, the pixel was designated as lake water; or, (c) If neither of the above rules applied, the pixel retained its cloud-cover value.



The daily map of lake ice or water has been eventually produced via the above steps, while some pixels are still marked as cloud.

#### **S5: Lake ice coverage time series filter processing.**

Aiming to obtain the inventory of lake ice/water change characteristics for every single lake, a further time series filter was conducted for each lake with ice, water, and cloud.

Firstly, a time series of filter with 10-days window was applied to these 5834-day lake ice and water coverages, and a secondary series of lake ice and water coverage record for each lake were produced. According to the MODIS-based lake ice phenology determination: the Freeze-Up End (FUE) is defined as the first day during the second half of the year when the lake ice coverage is greater than 0.9, whereas the Break-Up End (BUE) is the first day during the first half of the year when the lake ice coverage is less than 10% (Reed, Budde, Spencer, & Miller, 2009). As a result, when the lake ice coverage on a given day was greater than 90%, the cloud coverage was re-designated as lake ice part, and when the water coverage on a given day was greater than 90%, the cloud coverage was designated as the water part.

Secondly, when not that case (ice/water coverage <90%), these filters processed lake ice and water coverage was normalized into two new series of dataset regardless of cloud cover coverage, that means, the sum of coverage of lake ice and water is 100%. Then, the original cloud coverage for each lake was reassigned to the original lake ice and water coverage according to the normalization of ice and water time series data.

After the above processes, the final cloud-free lake ice coverage products at pixel-detail level have been produced.

### **2.3. Classification based on time series of lake ice coverage**

Of total 2612 examined lakes in Tibetan Plateau, the 5834 days' time series of cloud free time series of lake ice coverage for 625 lakes with areas larger than 3 km<sup>2</sup> were effectively recorded. Depending on the stability of ice coverage time series, the total 625 lakes are categorized into four types (Figure 3). Types I are those lakes evidenced to be not obviously thawing and freezing in winter time (November to April), and Type II are with obvious periodic changes of ice coverage larger than 90% for the whole freezing time span, Type III are more or less with periodic changes, not all examined 18 years are with ice coverage larger than 90%. Type IV are not convinced for its few pure pixels because of their small lake area and the complexity of lake shore.

## **3. Result and analysis**

In this paper, the ice coverage of 308 lakes for Type II and Type III and a table of records for 39 lake of Type I have been released openly as reference data for other applications. All the lakes with open data are shown in Figure 4 and Figure 5. The inland lakes in Tibetan Plateau were observed clearly, and most of the lakes with salty flats experience ice free seasons in our study. While the lakes of Type III are distributed in the southern part of Plateau mostly.

An examination on the behavior of the time series of lake ice coverage for each single lake has been conducted. The number of the pure pixel at MODIS 500 m resolution, lake

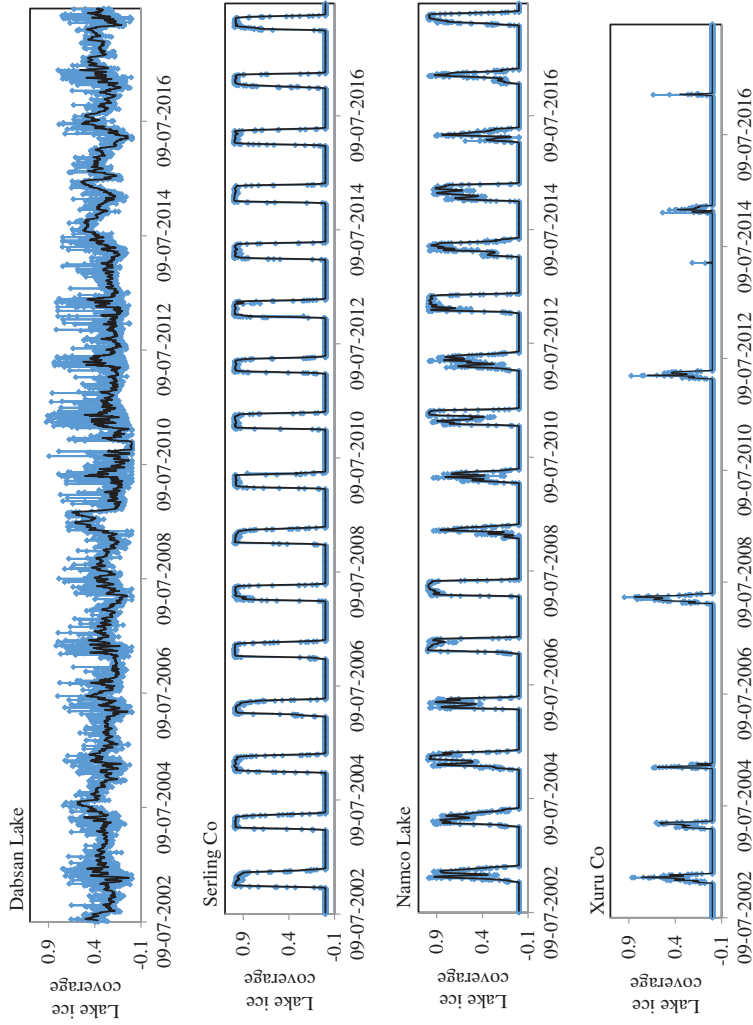
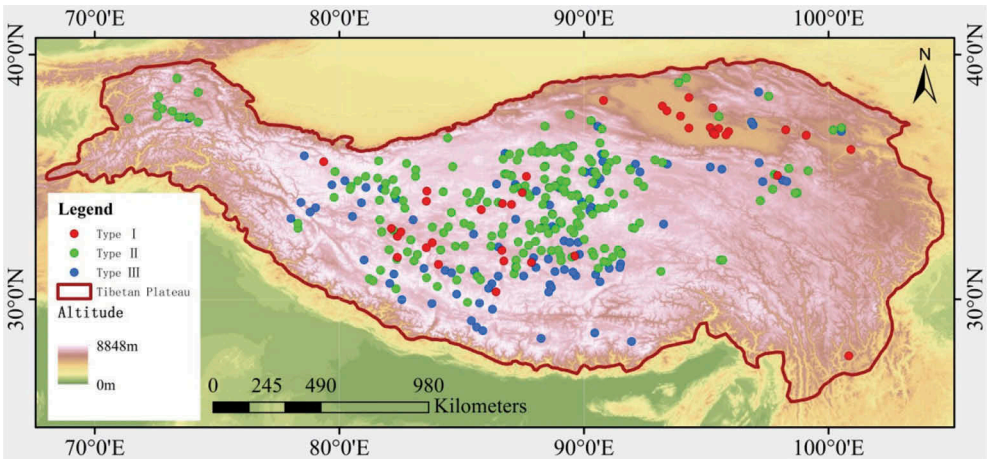
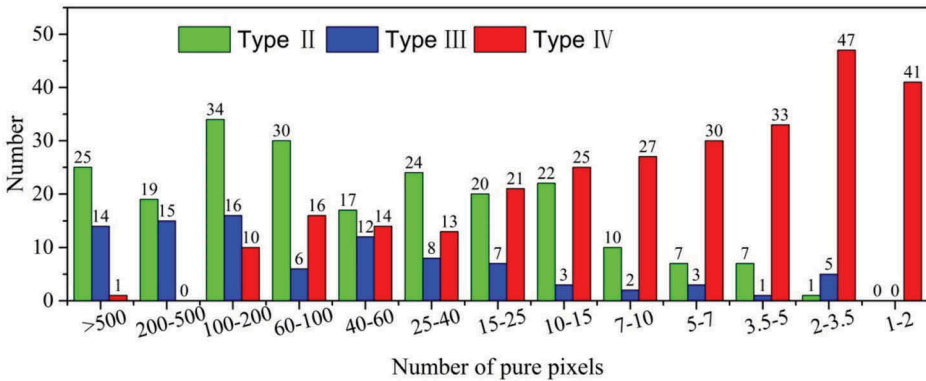


Figure 3. Examples of time series behaves for Type I (Dabsan Lake), Type II (Serling Co), Type III (Namco Lake) and Type IV (Xuru Co).





**Figure 4.** Lakes distribution for Type I, II, and III in Tibetan Plateau.



**Figure 5.** Lake numbers for Type II, III and IV published in this paper, with dependence to the number of pure pixels.

areas, salinity, and geographic locations influence characteristics of time series for ice coverage change. We find that three categories have been identified as follows.

#### (1) Lakes without periodic changes time series of lake ice coverage

Among 625 lakes of time series records of lake ice, there are 39 lakes that has been confirmed that there is no obvious periodic change signature throughout the whole 5834 days. These lakes are classified as Type I, for example, Dabsan lake showed in Figure 3.

A double-check with a comparison to the very recent high-resolution image through the Google Earth online image, and found that, 34 lakes and their shores are covered by salt flats clearly, and obviously the white salt leads to an abnormal signal of NDSI, and make the lakes are free of ice in the winter time. Another 3 lakes, named Xuru Co, Dangqiong Cohave, and Cam Co, have been confirmed that there are along with a fault in Tibetan Plateau, the geological reason, for example, the spring heat inside makes the

lakes ice-freely. The Longyangxia Reservoir and Lugu Lake are not freeze-lake due to their southern location and great water depth.

### **(2) Lakes with obvious periodic changes of time series lake ice coverage**

There are 216 lakes that are classified as lakes with obvious periodic changes with lake ice coverage larger than 90% in the freezing time, the four parameters of lake phenology, Freeze-up Start (FUS), Freeze-up End (FUE), Break-up Start (BUS), and Break-up End (BUE) can be easily estimated through the time series determination process. These are categorized into Type II, for example, the Serling Co lake showed in [Figure 3](#).

Another 92 lakes show a clear start and end of ice coverage, while the coverage is occasionally larger than 90% for some years, and normally less than 90% of the ice over the whole lake, these are of importance for understanding the partly freezing in the winter time, like Nam Co lake in [Figure 3](#), the information explicitly explained by the ice forming in different part of the lakes. This kind of lake is useful for the extraction of Freeze-up Start (FUS) and Break-up End (BUE) clearly, which are categorized into Type III.

### **(3) Lakes without periodic signal ascribe to the weak representative of time series of ice coverage**

Of the 625 lakes, 281 lakes with smaller area when compared to the MODIS spatial resolution, and lake shape complexity lead to a smaller number of pure pixels, and create much uncertainty for the time series analysis. These types of lake are categorized into Type IV, as shown in [Figure 3](#) with weak signal, which need to be a further analysis combined with other input.

An analysis has been done for the lake numbers of different types with the dependence to the MODIS pure pixels in [Figure 5](#), without consideration of the complexity of lakeshore, the results indicate clearly that the Types II and IV are quite relevant to the pure MODIS pixels, and the Type III exhibits a similar trend with the number of pixels. This result evidences that the remote sensing methodology for the lake ice monitoring is relevant to the pure pixel when the lake ice algorithm uncertainty fixed.

## **4. Validation and comparison**

### **4.1. Validation and comparison**

Two published datasets (Du & Kimball, 2018; Qiu et al. 2017) were selected as the reference dataset for the verification and inter-comparison. Both datasets were derived from AMSR-E and AMSR2 passive microwave observations, which are suitable for inter-comparison and verification of lake coverage as they are unaffected by cloud conditions; however, due to the intrinsic coarse spatial resolution of the sensors, only large lakes (above  $\sim 100 \text{ km}^2$ ) can be observed reliably. In the time series of lake ice coverage, four parameters of lake-ice phenology could only be extracted from Type II lakes ([Figure 3](#)). Therefore, the following six lakes with different locations, sizes and shapes were selected to inter-compare and validate the MODIS-based lake ice coverage: Qinghai Lake, Serling Co, Hala Lake, Dogze Co, Aksayqin, and Yaggain Co (detailed data are shown in [Table 1](#)).

In the Daily Lake Ice Phenology Time Series Derived from AMSR-E and AMSR2 data, passive microwave brightness temperatures are used to determine whether each pixel within a lake was completely frozen (Du & Kimball, 2018), with completely frozen pixels marked as “ice-on”. In the present study, for a given lake all “ice-on” pixels were

**Table 1.** Location and size of lakes of Type II chosen for comparison between AMSR-E/AMSR2 derived ice on/off proportion and MODIS-based lake ice coverage.

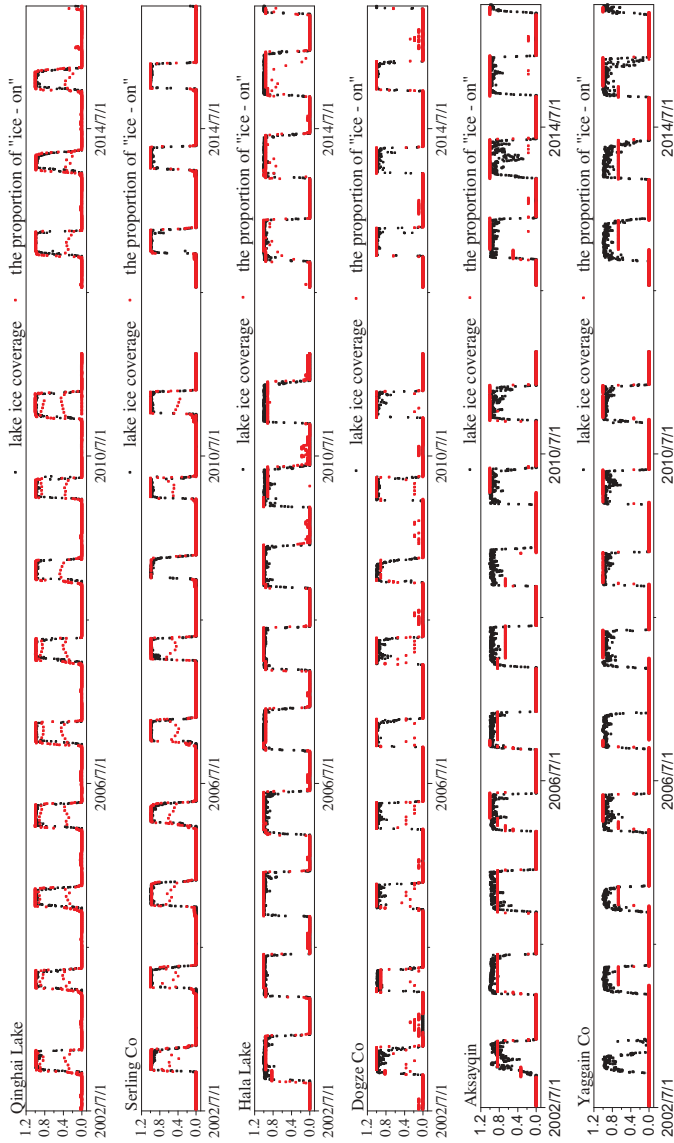
Lake No.	Lake Name	Lake-center longitude	Lake-center Latitude	Average depth (m)	Lake Area (km <sup>2</sup> )	Type
1	Qinghai Lake	100.1979	36.92292	16.8	4266.55	II
2	Serling Co	89.04792	31.77708	28	1749.53	II
3	Hala Lake	97.55625	38.30208	30	585.6	II
4	Dogze Co	87.52708	31.88542	8.5	251.56	II
5	Aksayqin	79.85750	35.20916	9.6	165.96	II
6	Yaggain Co	89.80850	33.00794	5.1	106.31	II

extracted and their proportions, i.e., the ice coverage of the lake, were calculated for the inter-comparison and verification with the MODIS-derived lake ice coverage, as shown in [Figure 6](#) for more than 12 years. [Figure 6](#) shows that MODIS ice coverage is consistent with the passive microwave records, especially for the lakes with size above 500 km<sup>2</sup>, which is comparable to the AMSR-E/2 resolution. But for the lakes with the size less than 200 km<sup>2</sup> (Aksayqin, and Yaggain Co), the agreement of ice coverage was weaker.

Finally, a linear regression analysis of the two time series was completed for all the six lakes. The results are shown in [Table 2](#). The time series of the four largest lakes (Qinghai Lake, Serling Co, Hala Lake, and Dogze Co) is consistent in the ice freeze-thaw cycle, with the coefficient of determination ( $R^2$ ) in the range 0.86–0.96, and the RMSE of the ice coverage ranging from 0.07 to 0.13. This result suggests that the MODIS lake ice coverage time series allowed for better identification of the ice phenology for smaller lakes than the passive microwave remote sensing observations.

Another inter-comparison between the MODIS-based lake ice coverage time series and microwave brightness temperature at 18.7 GHz and V polarization from AMSR-E was made for medium to large lakes over the High Asia region ([Figure 7](#)). The result showed that the two time-series had quite consistent thaw-freeze dates. Moreover, based on the MODIS definition of lake ice phenology, the AMSR-E lake ice phenology was extracted from the lake ice coverage for comparison for a single lake ([Figure 8](#)). The four parameters (Freeze-Up Start, Freeze-Up End, Break-Up Start, and Break-Up End) passed the validity test. The freeze-up start dates are quite consistent ( $R^2 = 0.96$ ) and close to each other (RMSE = 3.85 days).

Generally, the dates derived from the passive microwave data were earlier than those derived from MODIS ([Figure 8](#)). In addition, the spatial resolution of passive microwave data (27 km x 16 km) was far coarser than that of the MODIS data (500 m). Thus, some of the interferences were removed to obtain the overall brightness temperature at the initial stage of lake surface freezing. Additionally, compared to the optical–near IR data of MODIS, microwave data are more sensitive to phase changes between liquid water and ice. Therefore, when lake ice coverage is derived from MODIS data, the recurrent freezing and melting during the early stage increases the difficulty of decision on the parameters. As a result, broken ice is often neglected during the early stage, which causes delays in the MODIS freeze-up start date. Conversely, the linear relation of the freeze-up end date between passive microwave and MODIS data at 8.66 days has the lowest correlation ( $R^2 = 0.79$ ) and a relatively large RMSE. After the lake surface has begun to freeze, the passive microwave brightness temperature shows a ladder-peak



**Figure 6.** Visual comparison of time series of passive microwave-based proportion of ice on (Du & Kimball, 2018) and MODIS-based lake ice coverage, the gap of 2011 is the functional loss of AMSR-E sensor onboard.

**Table 2.** Parameters of linear fit between the “ice-on” proportion and lake ice coverage of this paper.

Lake number	Lake Name	Slope	Shift	R <sup>2</sup>	RMSE
1	Qinghai Lake	0.92,127	0.00871	0.86,335	0.13,808
2	Serling Co	0.9632	0.00077	0.9601	0.07438
3	Hala Lake	0.9763	0.00862	0.94,674	0.10,907
4	Dogze Co	0.93,485	0.00151	0.91,318	0.11,502
5	Aksayqin	0.90,571	0.0365	0.79,444	0.19,446
6	Yaggain Co	0.85,777	0.12,164	0.56,987	0.27,392

trend. It is generally considered that the first stable peak value corresponds to the complete freeze-up date. It is not possible to determine whether any subsequent peak values are caused by partial freezing, partial melting or weather phenomena, which might lead to errors.

The comparison of the breakup start dates was quite consistent, with an R<sup>2</sup> of 0.92 and an RMSE of 6.52 days. The main reason for some discrepancy is that, generally, only pixels in the center of the lake are considered when lake ice phenology is determined from passive microwaves, whereas ice freeze-thaw pixels close to the lake shore are neglected. This makes the passive microwave break-up start date slightly delayed compared to the date derived from the MODIS data, as the breakup typically begins at the shore (Jeffries, Morris, & Duguay). The results concerning the break-up end date have the highest consistency (R<sup>2</sup> = 0.98) and the minimum RMSE is 3.59 days.

The visual interpretation of the data determined that the Break-Up End date was the easiest to judge. Additionally, there is no clear standard for relating the lake ice phenology derived from passive microwave brightness temperature to the MODIS-derived lake ice phenology, which could also account for the difference between the two types of lake ice phenology definitions.

#### 4.2. Errors and missing data

Based on the above analysis, the present dataset still exhibits some deficiencies in lake ice detection. The main reason for this is that the SNOWMAP algorithm tends to incorrectly identify lake ice pixels under clear sky conditions, which indicates that an updated algorithm for the snow/ice recognition is needed. This problem could be caused by several factors that contribute to high NDSI values (see the Type I in the classification analysis), such as lake salinity, heavy snow cover on top of ice, or the uncertainty of the cloud-edge regions. Moreover, the optical characteristics of lakes – yellow substance, high sediment content and consequent high turbidity, and changes in aquatic vegetation and algae – also affect the reflectance characteristics of the lake during ice-free seasons, thereby leading to detection errors of lake-ice or river-ice coverage in the spring or summer (George & Dorothy). Some errors also occur in lake ice detection based on the temporal and spatial continuity of lake ice under cloud cover conditions.

In addition, there are some gaps in the 2002–2018 MODIS Terra and Aqua (day of year) dataset as applied in the present study (Table 3). When estimating lake ice extent under cloud cover conditions, data synthesis was not completed for days for which morning-pass or afternoon-pass data was missing. Instead, the available morning or afternoon-pass data were used, which could have resulted in the propagation of some errors.

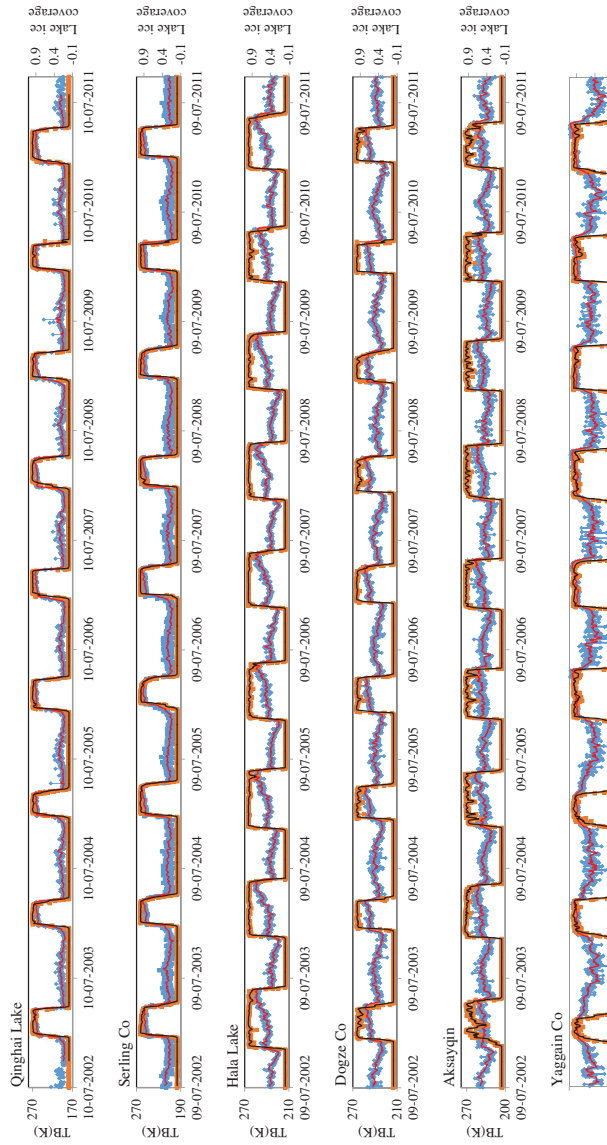


Figure 7. Visual comparison of passive microwave brightness temperature at 18.7 GHz, V polarization and lake ice coverage derived from MODIS in this paper.

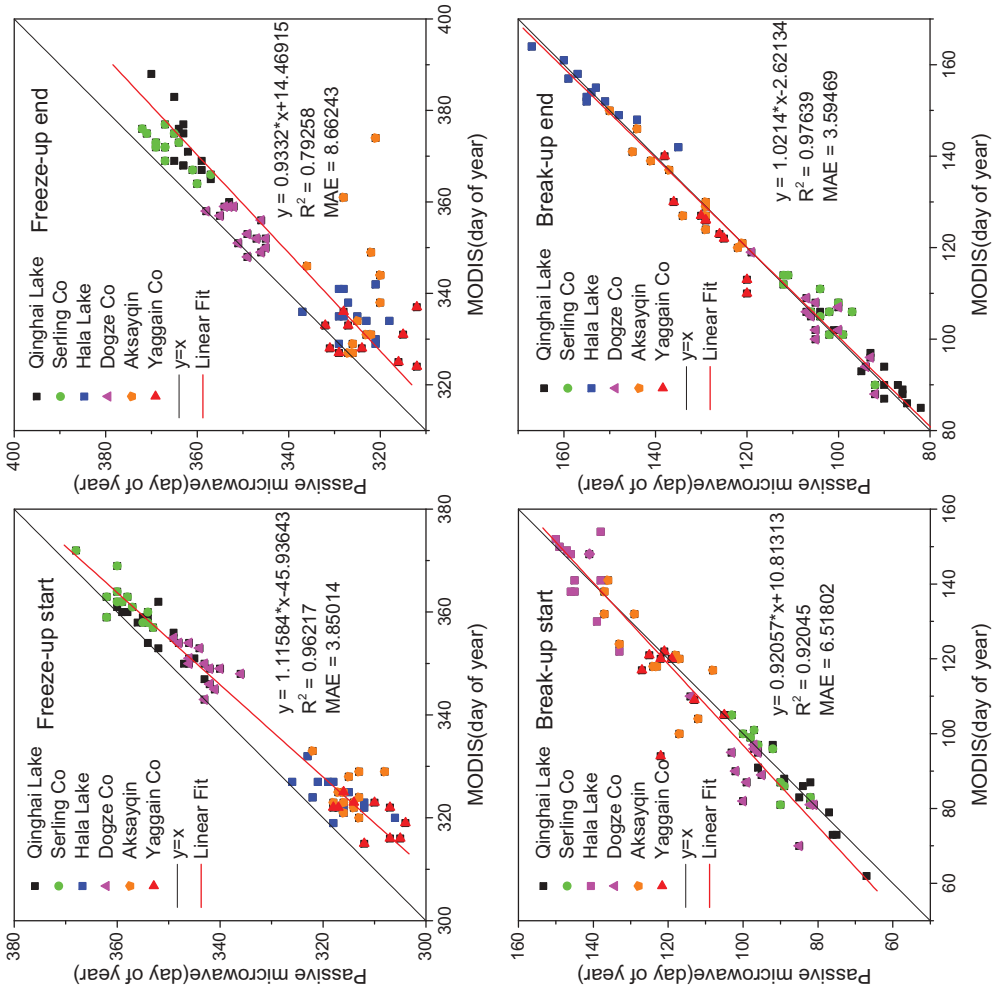


Figure 8. Comparison of lake ice phenology derived from MODIS and passive microwave brightness temperatures for six lakes.



**Table 3.** Missing data in the MODIS terra and aqua dataset (in Julian day) (“\” represents missing data).

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010
Morning pass	\	32,81,199,351–358	50	\	172,235	\	355–358	252	177
Afternoon pass	211–219,256,257,319	132	\	\	186	\	\	\	\
Year	2011	2012	2013	2014	2015	2016	2017	2018	
Morning pass	\	\	\	299	\	50–59	114	\	
Afternoon pass	339	\	\	\	\	\	\	\	

## Usage note

The data set contains 5,834 raster files, one vector file and 308 Excel files. The raster file is named Daily Lake Ice Extent. The vector file contains such information as the series number, name, location, surface area and classification number of the processed lake. The names of the excel files correspond to lake numbers. Each excel file contains six columns with the daily lake ice coverage information of its corresponding lake from July 2002 to June 2018. The attributes of each column are: succession, date, lake water coverage, lake ice coverage, cloud coverage, lake water coverage and lake ice coverage after cloud processing. Users can first use the vector file to determine the number, location and classification number of a given lake, and then obtain the corresponding daily lake ice coverage data for a given year from the Excel file to use for the monitoring of lake-ice freeze-thaw and research on climate change.

## Data Availability Statement

The data that support the findings of this study are openly available in 4TU.Centre for Research Data at <https://doi.org/10.4121/uuid:dfd8c76-6b7c-4bbf-aec8-98ab199d9093> and in Science Data Bank at <http://www.sciencedb.cn/dataSet/handle/744>.

## Acknowledgments

The research presented in this paper was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Number XDA19070201), Multi-Parameters Arctic Environmental Observations and Information Services (Grant number 2017YFE0111700) and “Big Earth Data – Based Environment Monitoring and Information Service for Arctic Sea Way of Polar Silk Road” (Grant Number 131211KYSB20170041).

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Chinese Academy of Sciences [XDA19070201]; Ministry of Science and Technology of the People’s Republic of China [2017YFE0111700]; The Chinese Academy of Sciences [131211KYSB20170041].

## ORCID

Yubao Qiu  <http://orcid.org/0000-0003-1313-6313>

Juha Lemmetyinen  <http://orcid.org/0000-0003-4434-9696>

## References

- Benson, B., & Magnuson, J. (2000). *updated 2012. global lake and river ice phenology database, Version 1. [Indicate subset used]*. Boulder, Colorado USA: NSIDC: National Snow and Ice Data Center. doi:10.7265/N5W66HP8
- Cai, Y., Ke, C. Q., Li, X., Zhang, G., Duan, Z., & Lee, H. (2019). Variations of lake ice phenology on the Tibetan Plateau from 2001 to 2017 based on MODIS data. *Journal of Geophysical Research: Atmospheres*, 124(2), 825–843.
- Canadian Cryospheric Information Network. Canadian Lake ice Database [EB/OL]. (2018). Retrieved from [https://www.polardata.ca/pdcsearch/PDCSearch.jsp?doi\\_id=1821](https://www.polardata.ca/pdcsearch/PDCSearch.jsp?doi_id=1821)
- Du, J., & Kimball, J. S. (2018). *Daily lake ice phenology time series derived from AMSR-E and AMSR2, Version 1. [Indicate subset used]*. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/HT4NQ07ZJF7M
- Du, J., Kimball, J. S., Duguay, C., Kim, Y., & Watts, J. D. (2017). Satellite microwave assessment of Northern Hemisphere lake ice phenology from 2002 to 2015. *The Cryosphere*, 11, 47–63.
- George, A., & Dorothy, K. MODIS Snow Products Collection 6 User Guide[EB/OL]. Retrieved from <http://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf>
- Hall, D. K., & Riggs, G. A. (2016). *MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6. [Indicate subset used]*. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/MODIS/MYD10A1.006
- Huan, Z. H. A. N. G., Yubao, Q. I. U., Zhaojun, Z. H. E. N. G., Duo, C. H. U., & Yudan, Y. A. N. G. (2016). Comparative study of the feasibility of cloud removal methods based on MODIS seasonal snow cover data over the Tibetan Plateau[J]. *Journal of Glaciology and Geocryology*, 38(3), 714–724.
- Jeffries, M., Morris, K., & Duguay, C. Floating ice: lake ice and river ice, In book: *Satellite Image Atlas of Glaciers of the World - State of the Earth's Cryosphere at the Beginning of the 21st Century: Glaciers, Global Snow Cover, Floating Ice, and Permafrost and Periglacial Environments Edition: Professional Paper 1386-AChapter: A-4-IIPublisher: U.S. Geological SurveyEditors: R.S. Williams, J. G. Ferrigno*
- Kropáček, J., Maussion, F., Chen, F., Hoerz, S., & Hochschild, V. (2013). Analysis of ice phenology of lakes on the Tibetan Plateau from MODIS data. *The Cryosphere*, 7(1), 287.
- Liu, Y., Chen, H., Wang, H., & Qiu, Y. (2018). The impact of the NAO on the delayed break-up date of lake ice over the Southern Tibetan Plateau. *Journal of Climate*, 31, 9073–9086.
- Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7, 13603.
- National centers for environmental information [EB/OL]. (2017). Retrieved from <https://www.ncdc.noaa.gov/gosic/gcos-essential-climate-variable-ecv-data-access-matrix/gcos-land-ecv-lakes>
- Peng, G., Qinghua, Y., & Qiufang, W. E. I. (2015). Lake ice change at the Nam Co Lake on the Tibetan Plateau during 2000–2013 and influencing factors. *Progress in Geography*, 34(10), 1241–1249.
- Qi, M., Yao, X., Li, X., Duan, H., Gao, Y., & Liu, J. (2019). Spatiotemporal characteristics of Qinghai Lake ice phenology between 2000 and 2016. *Journal of Geographical Sciences*, 29(1), 115–130.
- Qiu, Y., Guo, H., Ruan, Y., Fu, X., Shi, L., & Tian, B. (2017). *A dataset of microwave brightness temperature and freeze-thaw for medium-to-large lakes over the High Asia region 2002 – 2016*. Science Data Bank. doi: 10.11922/csdata.170.2017.0117
- REED, B., BUDDE, M., SPENCER, P., & Miller, A. E. (2009). Integration of MODIS-derived metrics to assess interannual variability in snowpack, lake ice, and NDVI in southwest Alaska[J]. *Remote Sensing of Environment*, 113(7), 1443–1452.
- Yubao, Q., Xingxing, W., Yongjian, R., Pengfei, X., Yu, Z., & Suping, Y. (2018). Passive microwave remote sensing of lake freeze-thawing over Qinghai-Tibet Plateau. *Journal of Lake Sciences*, 30(5), 1438–1449.
- Zhang, G., Yao, T., Xie, H., Kang, S., & Lei, Y. (2013). Increased mass over the Tibetan Plateau: From lakes or glaciers? *Geophysical Research Letters*, 40(10), 2125–2130.