

## Dataset Description

### Global annual lake ice phenological dataset 1861-2099

#### General Information

---

Title:	Global annual lake ice phenological dataset 1861-2099
Alternate Identifier:	10.1029/2022GL099022
Abstract:	<p>The global annual lake ice phenological dataset includes the freeze-up date, break-up date for 74,245 lakes in the northern hemisphere. The dataset is divided into three parts: 1) current data, obtained from MODIS productions through a DLRM model (with parameters provided), covering the period of 2001 to 2020; 2) historical and 3) future simulation data, obtained from the temperature-based lake-specific models, for the periods of 1861-2005 and 2006-2099, respectively. The historical and future simulations were only performed for 30,063 lakes that meet the model conditions and are presented in the dataset.</p>
Keywords:	Lake ice, Phenology, MODIS, Global warming, Climate change
Publication Date:	18 June 2022
Language:	English

#### People and Organizations

---

Contact & Creator:	<p>Xinchi Wang, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China Email: <a href="mailto:11930633@mail.sustech.edu.cn">11930633@mail.sustech.edu.cn</a></p>
Contact:	<p>Lian Feng, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China Email: <a href="mailto:fengl@sustech.edu.cn">fengl@sustech.edu.cn</a></p>

## Details

---

### Lake information in the study

File name: "Lake\_info\_74245lks.csv"

Size: 4.93 MB

Description: This file records information on 74,245 lakes in the study. Column A "Hylak\_id" corresponds to the ID in the HydroLAKES database<sup>1</sup>. Column B to H are some basic information in the lakes, such as continent, area (m<sup>2</sup>), shoreline length (m), shoreline development (-), volume (m<sup>3</sup>), average depth (m), and elevation (m). Column I "Cen\_lon" and J "Cen\_lat" are their central longitude and latitude. Column K "STD<2°C" is the comparison results of the minimum temperature STD in 20 years and 2°C (see more details in the paper). Here, "YES" means a temperature-based model can be used to simulate lake ice phenology in the historical and future periods. And the last column L "Array\_index(30,063lks)" shows a row index, that corresponds to historical and future data, with values from 0 to 30062.

### Data in the current period

File name: "Current\_74245lks\_2001-2020.zip"

Size: 19.61 MB

Description: Nine CSV files are compressed in this ZIP file, and each CSV table has the column "Hylak\_id" corresponds to HydroLAKES database. In addition, six files beginning with "fit\_" are related to DLRM model fitting. File "fit\_n\_obs.csv" records the annual number of observations from MODIS. File "fit\_para\_a1.csv", "fit\_para\_a2.csv", "fit\_para\_b1.csv", "fit\_para\_b2.csv" and "fit\_para\_c.csv" records five fitting parameters in DLRM:  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $c$ , respectively. The remaining three CSV files beginning with "ice\_", record the annual lake ice phenology from 2001-2020. That is, "ice\_on\_day.csv" for freeze-up date, "ice\_off\_day.csv" for break-up date, and "ice\_duration.csv" for ice duration. It is worth noting that August 1st of the previous year is the first day of DLRM, so the recorded freeze-up date and break-up date are calculated from August 1st of the previous year. We use "-1" to indicate unfrozen years.

Notice: For a few lakes, DLRM model may deviate from the actual situation. This situation generally occurs in lakes with less MODIS valid observations or low Snow Cover Index while freezing within a year (caused by the limitations of MODIS products). Therefore, while using this data, it is best to measure the accuracy of ice phenological metrics in combination with the

---

<sup>1</sup> HydroLAKES database: <https://www.hydrosheds.org/page/hydrolakes>

number of valid observations (in “fit\_n\_obs.csv”) and the peak value of model fitting (in “fit\_para\_c.csv”).

### **Data in the historical period**

File name: "Historical\_30063lks\_1861-2005.zip"

Size: 44.65 MB

Description: Twelve NPY files (NumPy array in Python) are compressed in this ZIP file. Each NPY file has a two-dimensional array with 30,063 rows and 144 columns, corresponding to 30,063 lakes and 144 years (1861-2005). The row index of these 30,063 lakes can be seen in the file “Lake\_info\_74245lks.csv” and corresponds to “Hylak\_id”. “Freezing\_”, “Breakup\_” and “Duration\_” in the file name refer to three phenological metrics, respectively. MIROC5, IPSL-CM5A-LR, GFDL-ESM2M and HadGEM2-ES are four global climate models from bias-corrected climate projections (Inter-Sectoral Impact Model Intercomparison Project, or ISIMIP2b), show the source of air temperature we used in the temperature-based model. August 1st of the previous year is the first day in the data, and we use “0” in “Freezing\_”, “Breakup\_” NPY files to represent unfrozen years.

### **Data in the future period**

File name: "Future\_30063lks\_2006\_2099.zip"

Size: 115.69 MB

Description: Forty-eight NPY files are compressed in this ZIP file. Each NPY file has a two-dimensional array with 30,063 rows and 93 columns, corresponding to 30,063 lakes and 93 years (2006-2099). Form of this data is similar to data in the historical period. But it should be noted that four representative concentration pathway (RCP) scenarios are investigated: RCP 2.6, 4.5, 6.0 and 8.5 (low to high), respectively. You can distinguish them by file name. August 1st of the previous year is the first day in the data, and we use “0” instead in “Freezing\_”, “Breakup\_” NPY files to represent unfrozen years.