Fine-Resolution Mapping of Pan-Arctic Lake Ice-Off Phenology Based on Dense Sentinel-2 Time Series Data

Chong Liu 1,2, Huabing Huang 1,2, Fengming Hui 1,2, Ziqian Zhang 1 and Xiao Cheng 1,2, *  

1 School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou 510275, China; liuc@mail.sysu.edu.cn (C.L.); huanghb55@mail.sysu.edu.cn (H.H.); huifm@mail.sysu.edu.cn (F.H.); zhangzq55@mail2.sysu.edu.cn (Z.Z.)  
2 Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China  
* Correspondence: chengxiao9@mail.sysu.edu.cn

Abstract: The timing of lake ice-off regulates biotic and abiotic processes in Arctic ecosystems. Due to the coarse spatial and temporal resolution of available satellite data, previous studies mainly focused on lake-scale investigations of melting/freezing, hindering the detection of subtle patterns within heterogeneous landscapes. To fill this knowledge gap, we developed a new approach for fine-resolution mapping of Pan-Arctic lake ice-off phenology. Using the Scene Classification Layer data derived from dense Sentinel-2 time series images, we estimated the pixel-by-pixel ice break-up end date information by seeking the transition time point when the pixel is completely free of ice. Applying this approach on the Google Earth Engine platform, we mapped the spatial distribution of the break-up end date for 45,532 lakes across the entire Arctic (except for Greenland) for the year 2019. The evaluation results suggested that our estimations matched well with both in situ measurements and an existing lake ice phenology product. Based on the generated map, we estimated that the average break-up end time of Pan-Arctic lakes is $172 \pm 13.4$ (measured in day of year) for the year 2019. The mapped lake ice-off phenology exhibits a latitudinal gradient, with a linear slope of 1.02 days per degree from 55$^\circ$N onward. We also demonstrated the importance of lake and landscape characteristics in affecting spring lake ice melting. The proposed approach offers new possibilities for monitoring the seasonal Arctic lake ice freeze–thaw cycle, benefiting the ongoing efforts of combating and adapting to climate change.

Keywords: arctic lake; ice-off phenology; dense time series; Sentinel-2

1. Introduction

Lakes, through their seasonal ice cover extent and duration, are a major landscape in the Arctic [1–5]. Governed by both geographical and morphological settings, the spring lake ice phenomenology (i.e., timing of ice-off) is a primary driver of Arctic hydrological processes [6], and has been identified as a sensitive metric to measure climate change [7–9]. Regional coherence or synchrony of lake break-up timing also regulates greenhouse gas emissions from fresh water into the atmosphere [9–12], affecting the land surface energy balance [13], which in turn generates feedbacks into the climate system. Given their wide-ranging importance, accurately estimating the spatial distribution of lake ice-off phenomenology is essential to understand Arctic ecosystem functions and project their responses to global environmental dynamics under various climate change scenarios [14,15].

Before the advent of remote sensing, lake ice-off phenomenology data could only be collected through in-field observations, which are nevertheless laborious and spatially sparse [16]. Satellite remote sensing has revolutionized our ability to monitor fresh water freeze–thaw cycling processes, especially at the continental to global scales [17]. Using high temporal frequency satellite image time series, many attempts have been made to extract lake ice-off timing information [18]. The microwave brightness temperature, with long available archives and the all-weather imaging capability, is perhaps the most widely used...
data source for lake ice phenology parameter identification [15,19–21]. However, most of the passive microwave sensors have coarse spatial resolutions (typically ranging from 6.25 km to 25 km), hence limiting their effectiveness within heterogeneous landscapes [15]. Therefore, other remote sensing data types with higher spatial resolutions, including synthetic aperture radar (SAR) [7,22], optical imagery [8,9,23,24], and UAV [25], have also been used for mapping lake ice-off phenology.

Despite previous efforts at estimating fresh water freeze–thaw cycling in many regions of the world, the extent and distribution of lake ice-off phenology have not yet been quantified across the entire Arctic, which vary with respect to ground thermal regime, topography, surface geology, and land cover [3]. Moreover, most of the current studies have been conducted by treating a lake as the smallest unit of interest, concealing the internal variability of lake ice-off timing distribution [26]. Technical challenges exist in almost every aspect of lake ice phenology mapping, including data acquisition, lake extent extraction, time series modeling, and ice melting/freezing identification. These challenges are even more critical for small lakes (typically less than 1000 ha), which account for approximately 99% of the total Arctic lake population [1], and which are more sensitive to climate change and land disturbance events [9,27].

This study aims to develop a spatially continuous map that first estimates Pan-Arctic lake ice-off phenology at a fine spatial resolution (i.e., 20 m). To achieve this goal, we design a new approach that identifies the per-pixel ice break-up date during the spring melting season based on dense time series stacks of Sentinel-2 data. We implement the algorithm on the Google Earth Engine (GEE) platform and apply it to lakes across the Pan-Arctic region (except for Greenland). To evaluate the mapping performance, we compare our results with those derived from daily freeze–thaw satellite products as well as in situ data for validation. In the remainder of this paper, we introduce the materials and methods (Section 2), report on the validation and mapping results of the Arctic lake ice-off phenology (Section 3), present discussions (Section 4), and draw some conclusions from the study (Section 5).

2. Materials and Methods

2.1. Study Area Boundaries and Lake Selection

The study area covers the entire Arctic, which is defined as the northernmost part of the Earth featured by tundra vegetation, an arctic climate, and arctic flora, with the tree line and the coastline joint-shaping its extent borders (Figure 1a) [27–29]. The lake selection in this study was based on HydroLAKES [30], which provides the shoreline polygons of global lakes with a surface area of no less than 10 ha. We filtered the HydroLAKES dataset by the extent of the study area, and excluded lakes that have a surface area less than 1 km\(^2\) to avoid noise introduced by mixing pixels at the shore line [8]. We also masked out supraglacial lakes in Greenland because of their inter-annual fluctuations in extent and distribution [31]. After the abovementioned selection procedure, there remained 47,230 lakes, ranging from 1 km\(^2\) to 4909 km\(^2\), with their spatial concentration displayed in Figure 1b.
2.2. Datasets

Our primary data source used was Sentinel-2 A/B satellite imagery, accessed from the European Space Agency (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi, accessed on 23 November 2020) via the GEE platform [32]. Among the various Sentinel-2 products, we selected the category of Level 2A because it meets the formal quality of orthorectified Bottom-of-Atmosphere (BOA) reflectance, and thus multi-temporal images can be directly used to construct time series for further analysis [33]. Within the Sentinel-2 Level 2A data, we used mainly the Scene Classification Layer (termed SCL hereafter) band for lake ice-off phenology mapping at a 20 m resolution. For each Sentinel-2 image, the SCL band is a pixel classification map derived from four technical steps: cloud/snow detection, cirrus detection, cloud shadow detection, and classification map generation. A specific land cover scheme is designed for classifying each pixel into 11 different categories, including cloud, cloud shadows, vegetation, soils/deserts, water, snow/ice, etc. Readers can refer to the Sentinel-2 Level-2A Algorithm Overview (https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm, accessed on 28 November 2020)) for more technical details of SCL generation. We processed all Sentinel-2 Level 2A data with less than 85% cloud obscuration acquired between 1 February and 1 September 2019, a collection of 1621 images with their average revisit interval pattern displayed in Figure 1c. In addition to satellite imagery, we also obtained ERA5 air temperature (ERA5 Ta hereafter), which is a reanalysis product that combines modeled data with observations into a globally gridded dataset, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). Compared with its predecessors, ERA5 Ta benefits from an enhanced spatial resolution and is becoming increasingly popular in lake/river ice applications [15,17]. Similar to the Sentinel-2 data collection, we accessed ERA5 Ta from the GEE’s public data archive, and utilized the daily average air temperature at 2 m height for better identification of the ice-melting timing information.

Figure 1. Overview of the study area. (a) Spatial coverage of the study area; (b) distribution of lake concentration, calculated as the percent area covered by lakes within a 25 km × 25 km grid; (c) distribution of the Sentinel-2 average revisit interval (day) within the study period (i.e., from 1 February to 1 September 2019).
2.3. Lake Ice-Off Phenology Mapping Algorithm

Many ice phenology variables have been used for characterizing the spring melting process at the level of a pixel or lake. Here, we mainly focus on break-up end (BUE), which is defined as the day of the year (DOY) when the pixel becomes totally ice free [20,24]. Figure 2 displays the proposed approach that aims to utilize pixel-level SCL time series information for BUE mapping on the GEE platform. Particularly, the framework includes three sequentially integrated parts. In the first part, we refined the lake extent derived from HydroLAKES. This procedure was implemented by referring to Sentinel-2 SCL class information (Section 2.3.1). In the second part, we again used time series of Sentinel-2 SCL and ERA5 Ta to generate a seamless, binary phenophase time series at the 20 m resolution and 5-day interval. Then, such a dataset was utilized as the input of the “max-seg-difference” algorithm, which identifies the per-pixel BUE by seeking the transition time point when the pixel is completely free of ice (Section 2.3.2). Finally, we performed an accuracy assessment (Section 2.3.3) and analyzed the spatial pattern of our estimated lake ice-off phenology result across the entire Arctic (Section 2.3.4).

Figure 2. Flow chart of the lake ice BUE mapping proposed in this study.

2.3.1. Lake Extent Refinement

Despite overall high accuracy, the HydroLAKES shoreline layer still bears errors for lake ice-off phenology mapping. Some waterbodies are frozen or clear of ice all year round, neither of which should be included in our analysis. Moreover, several Arctic lakes may have changed in their extents [34] or seasonal fluctuation patterns [35] since the release of HydroLAKES. Therefore, a lake extent refinement is required in this study. We rasterized the original HydroLAKES shoreline layer into a binary image consisting of two types: lake and non-lake. Then, a frequency threshold-based method [36,37] was applied only to the lake pixels for selecting waterbodies with seasonal ice duration based on Sentinel-2 SCL information (Section 2.3.1). In the second part, we again used time series of Sentinel-2 SCL and ERA5 Ta to generate a seamless, binary phenophase time series at the 20 m resolution and 5-day interval. Then, such a dataset was utilized as the input of the “max-seg-difference” algorithm, which identifies the per-pixel BUE by seeking the transition time point when the pixel is completely free of ice (Section 2.3.2). Finally, we performed an accuracy assessment (Section 2.3.3) and analyzed the spatial pattern of our estimated lake ice-off phenology result across the entire Arctic (Section 2.3.4).

2.3.1. Lake Extent Refinement

Despite overall high accuracy, the HydroLAKES shoreline layer still bears errors for lake ice-off phenology mapping. Some waterbodies are frozen or clear of ice all year round, neither of which should be included in our analysis. Moreover, several Arctic lakes may have changed in their extents [34] or seasonal fluctuation patterns [35] since the release of HydroLAKES. Therefore, a lake extent refinement is required in this study. We rasterized the original HydroLAKES shoreline layer into a binary image consisting of two types: lake and non-lake. Then, a frequency threshold-based method [36,37] was applied only to the lake pixels for selecting waterbodies with seasonal ice duration based on Sentinel-2 SCL data. In particular, for pixel \( i \), only by satisfying the following criterion can it be identified as valid and used for lake ice BUE mapping:

\[
P_i = \begin{cases} 
1, & F_{\text{bare}} \leq 10\% \text{ and } F_{\text{veg}} \leq 10\% \text{ and } F_{\text{ice}} \geq 10\% \text{ and } F_{\text{water}} \geq 10\% \\
0, & \text{otherwise}
\end{cases}
\]

where \( P_i \) is the final lake extent refinement result, with “1” denoting valid and “0” invalid. \( F_{\text{bare}}, F_{\text{veg}}, F_{\text{ice}}, \) and \( F_{\text{water}} \) are percentages of the SCL time series classified as bare soil, vegetation, water, and ice, respectively, within the study period (from 1 February 2019 to 1 September 2019). The threshold of 10% was adopted due to the overall classification...
uncertainty of the SCL data [38]. To evaluate the refined lake extent result, 800 stratified random sample points were selected, with one half collected within the refined lake extent while the other half from the masked regions. After removing points that were unable to be identified, there remained 324 and 325 sample points, respectively. Three metrics, including commission error, omission error, and F1 score, were selected for refinement assessment [39]. Here, commission and omission represent the percentage of valid lake samples that were falsely masked, and the percentage of invalid lake samples that were not detected by the refinement procedure, respectively. The F1 score quantifies the level of balance between precision (commission) and recall (omission).

2.3.2. BUE Identification

• Creation of phenophase time series

At the pixel level, this study used Sentinel-2 SCL data to create a binary (ice/water) phenophase time series. Within the refined lake extent, we first masked poor quality or irrelevant observations, leaving only observations classified as water or ice by the SCL (labeled as 6 and 11, respectively). Second, an equitemporal phenophase composite was obtained by deriving the mode value of all valid observations during each temporal interval, which was set to 5 days as a balance of multiple factors, including the sensor revisit cycle, swath (the area imaged on Earth surface) overlap, and the temporal uncertainty control of the final BUE map. It should be noted that an extreme case may exist; that is, in a 5-day interval, there might be no valid data from any Sentinel-2 observations. In this case, a filling value was determined according to temporally adjacent (within 15 days) valid observations.

• Outlier removal

Several outliers remain in the created phenophase data cube because of the misclassification errors of the SCL. These errors on the temporal dimension could undermine or even lead to the failure of the BUE identification. In this study, we used ERA5 Ta data for outlier removal because the lake ice melting is highly correlated to the warming process of air temperature [9,25,40]. Due to the spatial resolution mismatch, the bicubic interpolation algorithm [41] was used for resizing the original ERA5 Ta data to 20 m, and the 28-day average air temperature ($T_a^{28}$) was adopted as the metric to correct unreasonable ice or water presence [9]. Two threshold values, $-5 \degree C$ and $5 \degree C$, were selected, corresponding to the upper and lower boundaries of $T_a^{28}$ when the outlier removal takes effect. Specifically, two situations are involved. In the first situation, we assume no ice should melt if $T_a^{28}$ is equal or lower than $-5 \degree C$. By contrast, in the second situation, we assume no ice should exist when $T_a^{28}$ is equal or higher than $5 \degree C$. Such a threshold filter approach has been widely applied in previous studies [9,42], and proven effective for reducing the spurious detection results of ice melting.

• The “max-seg-difference” algorithm

Traditionally, the lake ice BUE is identified from satellite sensors by detecting a sharp change in surface reflectance or brightness temperature trajectory during the spring melting period [8,43]. This method is straightforward, but its usefulness could be affected by a number of factors, including sun illumination [8], ice thickness [25], and microwave dielectric properties [22]. As a consequence, additional intervention from the user is usually needed, hampering the application of BUE mapping over large areas. To deal with this issue, here we proposed a new BUE identification algorithm, aiming to be more self-adaptive at the per-pixel level through segmentation of the phenophase time series. Figure 3 graphically displays how the BUE date is determined for one pixel. First, we used the SCL trajectory to create a new Boolean time series profile, indicating the phenophase (water as 1, ice as 0) at each temporal interval. Then, for a given temporal interval $DOY_i$ (highlighted with the blue circle), the entire phenophase profile can be divided into two parts: $Seg_{prior}^{i}$ and $Seg_{post}^{i}$ with $DOY_i$ as the boundary point. Since BUE represents the ending of ice cover and the beginning of open water, our core target here is to find a
temporal interval $DOY_{max}$ when the difference between $Seg^i_{prior}$ (ice dominated) and $Seg^i_{post}$ (water dominated) can be maximized. The objective function is thus written as

$$\argmax |\overline{Seg}^i_{prior} - \overline{Seg}^i_{post}| \quad i \in [1, N]$$

where $\overline{Seg}^i_{prior}$ and $\overline{Seg}^i_{post}$ are the average values of $Seg^i_{prior}$ and $Seg^i_{post}$, respectively, and $N$ is the length of the temporal domain (the total number of temporal intervals). In practice, we created a loop for each temporal interval and generated a “seg-diff” profile, through which BUE can be identified according to the maximum principle. A primary advantage of the “max-seg-difference” algorithm is that it can make full use of the time series information so that the impacts of “false” alarms (e.g., Case 1 and Case 3 in Figure 3) would be greatly reduced or even eliminated. We implemented the “max-seg-difference” algorithm within the refined lake extent using GEE, and the final Pan-Arctic BUE map were produced and exported at a 20 m spatial resolution.

Figure 3. Illustration of the proposed “max-seg-difference” algorithm for BUE identification.

2.3.3. Evaluation of Lake Ice-Off Phenology Mapping

The performance evaluation of the Pan-Arctic lake ice BUE map was conducted in two ways. In the first evaluation method, we randomly created 400 sample points within the refined lake extent, and their BUE dates for the year 2019 were visually identified based on multi-temporal satellite images, including Sentinel-1, Sentinel-2, Landsat-8, and Planet, for the purpose of interpretation bias minimization. We kept only well-interpreted points with a high level of confidence, which eventually resulted in a total of 234 samples (termed RDsite hereafter). We also collected in situ ice phenology observations of 29 lakes provided by the Finnish Environmental Institute. It should be noted that these lakes are geographically beyond the Arctic extent. To make use of these valuable measurements (termed RDsite hereafter) for algorithm validation, we created a 20 m buffer for each lake ice observation site, and implemented our approach of lake ice BUE identification within each buffer. Previous studies suggested that using only a pixel-level evaluation may be insufficient to reflect the quality of a spatially continuous map [44]. Therefore, we implemented a second evaluation method by comparing our result with an independent lake ice BUE map derived from the 1 km Interactive Multisensor Snow and Ice Mapping System (IMS) snow and ice product for the year 2019 (termed RDims hereafter) [15]. We aggregated our 20 m × 20 m lake ice BUE map to 1 km × 1 km using the majority rule, and kept only those pixels whose surrounding area within the 1 km buffer is all lake, to reduce the uncertainty caused by sub-pixel heterogeneity. The evaluation metrics include
mean error (ME), mean absolute error (MAE), and root mean square error (RMSE), where \( \hat{f}_i \) and \( f_i \) are the estimated and reference BUE results for sample point \( i \), respectively. \( M \) represents the sample number.

\[
ME = \frac{1}{M} \sum_{i=1}^{M} (\hat{f}_i - f_i)
\]

(3)

\[
MAE = \frac{1}{M} \sum_{i=1}^{M} |\hat{f}_i - f_i|
\]

(4)

\[
RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (\hat{f}_i - f_i)^2}
\]

(5)

2.3.4. Analysis Method

To investigate the distribution patterns of lake ice-off phenology across the entire Arctic, we conducted our analysis through three levels of spatial aggregation. Based on our 20 m lake ice-off phenology mapping result, we first calculated the mean BUE value for each 5 km \( \times \) 5 km grid. In this way, a coarse resolution map was derived, which was used for visual interpretation of the Pan-Arctic lake ice BUE distribution. We also divided the Arctic landmass into zones defined by administrative boundaries and calculated the country-level lake ice BUE estimates. To investigate the sensitivity of the lake ice BUE to the geographical environment, we examined the latitudinal trend of our estimated lake ice BUE map. For each country (except for Norway and Iceland, due to their limited lake numbers), a latitudinal profile of BUE was created with an aggregation interval of 500 m, and the trend was calculated as the slope of the linear regression outputs. The \( t \)-test was used to examine the significance level of the trend.

We further analyzed the lake-level relationship between the ice BUE and four morphology/landscape parameters: lake area, lake depth, residence time of the lake waters, and shoreline development, measured as the ratio between the shoreline length and the circumference of a circle with the same area. All these parameter records were derived from the HydroLAKES dataset. Similar to the latitudinal analysis, for each parameter–BUE pair, we derived the linear trend as well as its significance level.

3. Results

3.1. Assessment of the Refined Lake Extent

Table 1 shows the evaluation results of the lake extent refinement procedure. For the entire Pan-Arctic region, 45,532 lakes of 138,733 km\(^2\) were identified as the valid lake extent within the study area. In general, we observed the refined Pan-Arctic lake extent exhibits high accuracies, with a commission error, omission error, and F1 score of 6.7%, 6.8%, and 93.2%, respectively. In other words, the refinement procedure is effective for masking out irrelevant land covers (bare soil, vegetation, stable ice, and stable water) that were included in the original HydroLAKES dataset. We also found substantial accuracy variations across countries. Referring to the F1 score, the best performances are observed in Iceland (97.3%) and Norway (95.3%), followed by the USA (93.9%), Russia (91.2%), and Canada (88.3%). Moreover, in the USA and Canada, the accuracies measured by commission are slightly lower than those by omission, while in other countries the opposite trend (higher omission than commission) is observed. Referring to the lake area, we found that the majority of Arctic lakes are located in three countries, including Canada (30,325 lakes, 100,938 km\(^2\)), Russia (11,495 lakes, 27,841 km\(^2\)) and the USA (3647 lakes, 9905 km\(^2\)), where the permafrost landscape is widely distributed. On the other hand, Norway and Iceland have tiny shares of the total lake area extent, primarily due to the smaller landmass of these two countries.
Table 1. Evaluation of the refined lake extent.

| Region    | Commission (%) | Omission (%) | F1 Score (%) | Refined Lake Number | Refined Lake Area (km²) |
|-----------|----------------|--------------|--------------|---------------------|-------------------------|
| Canada    | 17.3           | 5.3          | 88.3         | 30,325              | 100,938                 |
| Iceland   | 1.3            | 4.0          | 97.3         | 8                   | 6                       |
| Norway    | 2.7            | 6.7          | 95.3         | 57                  | 43                      |
| Russia    | 5.3            | 12.0         | 91.2         | 11,495              | 27,841                  |
| USA       | 8.0            | 4.2          | 93.9         | 3647                | 9905                    |
| Pan-Arctic| 6.7            | 6.8          | 93.2         | 45,532              | 138,733                 |

The discrepancy between the original HydroLAKES product and the refined lake extents can be attributed to two primary reasons: misclassification errors of the HydroLAKES dataset and changes in lake extents. Figure 4 displays three typical subsets (200 × 200 pixels for each), representing the diversity of Arctic lake and landscape characteristics. The lake shoreline polygons before and after the refinement were compared and analyzed. Subset 1 is located in Siberia, Russia, while Subsets 2 and 3 are located in Alaska, USA. In Case 1, HydroLAKES mistakenly identifies a cluster of pixels as a lake, which is actually a shaded region covered by shrubs. Such an omission error, however, is not observed in our refined output. In Case 2, both the results with and without the refinement procedure capture the overall pattern of the lake boundaries. However, differences still existed following a further look. Specifically, we observed a slight shrinking trend in the largest lake in the middle, which is correctly reflected in the refined shoreline result. Case 3 further shows a more extreme situation in which the majority of lake areas has been lost and replaced by new plants. Therefore, the refined lake extent used in this study is more reasonable than the original HydroLAKES product for lake ice BUE mapping.

Figure 4. Three typical subsets showing lake shorelines before and after the refinement procedure. The background maps are Sentinel-2 images (shortwave infrared, near infrared, and red bands are displayed in as red, green, and blue layers) acquired on 18 July 2019 for Subset 1 and on 24 July 2019 for Subsets 2–3, respectively.
3.2. Evaluation of Lake Ice-Off Phenology Mapping

We evaluated the estimated lake ice BUE map based on the RDsite and RDsat datasets. We found that our predictions and the actual BUE dates are overall in good agreement (Figure 5). Table 2 further displays the quantitative accuracy metrics. We found that the predicted outputs are more subject to positive bias errors, with $ME$ values of 0.34 and 6.14 days for RDsite and RDsat, respectively, meaning a higher probability of a delay in the lake ice BUE date detection. Comparatively, our BUE predictions show a slightly better performance in matching RDsite than RDsat by all measures.

![Figure 5. Box plots of the observed and predicted lake ice BUE dates using in situ observations (RDsite) and sample points from satellite visual interpretation (RDsat).](image)

Table 2. Error metrics of the predicted BUE results using RDsite and RDsat (unit: day).

| Number of Observations | ME | MAE | RMSE |
|------------------------|----|-----|------|
| RDsite                 | 29 | 0.34| 6.55 | 7.40 |
| RDsat                  | 234| 6.14| 7.79 | 9.92 |

We further evaluated the performance of our proposed algorithm in estimating lake ice BUE using the RDims dataset. These results were summarized in the density plots using pixel and lake as the statistical unit, respectively (Figure 6). In general, we found that our predictions are generally consistent with those derived from the IMS time series, although the BUE difference is inherently variable within and among lakes. By forcing the intercept of the linear fit to zero [44], we found that our predictions show systematic underestimations (i.e., slope $> 1.00$ and $ME < 0$) with the IMS-based BUE. Specifically, at the pixel level our predictions explain 65% of the variations of RDims, while at the lake level 57% of the variation can be explained by the predicted BUE. Regarding $MAE$ and $RMSE$, similar results were observed, indicating consistent performance across different spatial levels.
Figure 6. Density plots showing the relationships between the predicted and RDims BUE dates using pixel (a) and lake (b) as the statistical unit. The red line represents the linear fit with the intercept forced to 0.

Figure 7 further compares lake ice BUE maps from RDims and our predictions by selecting five Arctic lakes, each of which represents one typical landscape. In general, we found that both mapping results correctly reflect the BUE variation within and among lakes, with BUE dates ranging from 110 to 220. Spatially, our BUE maps can provide more details over space and exhibit more consistent BUE patterns. Taking lake Taymyr (A) and lake Hall (D) as examples, the BUE difference between the lake center and lake shore is better captured by our predictions than those of RDims. Such a mapping performance enhancement is primarily due to the improved spatial resolution of Sentinel-2, which allows us to reveal the hidden patterns of BUE heterogeneity that would otherwise not be detected using coarse resolution products (e.g., IMS). Apart from the improvement in mapping accuracy, our 20 m BUE map is also associated with reduced uncertainty caused by the noise introduced by mixing pixels at the shoreline (e.g., lake Ozero Yarato (C) and lake Nanvaranak (E) in Figure 7), and is capable of offering more spatially continuous lake ice BUE estimates across the entire Arctic.

Figure 7. Comparison of BUE maps for five Arctic lakes. (A–E) represent lake Taymyr (Russia), lake Teshekpuk (U.S.), lake Ozero Yarato (Russia), lake Hall (Canada), and lake Nanvaranak (U.S.), respectively.

3.3. Spatial Pattern of Pan-Arctic Lake Ice BUE

Using the proposed approach, we calculated the Pan-Arctic lake ice BUE and mapped its distribution (Figure 8). Based on this map, we estimated that the average BUE of Pan-Arctic lakes for the year 2019 is 21 June (DOY = 172 ± 13.4 days, one standard deviation). Spatially, we found that the BUE distribution is strongly associated with latitudinal gradi-
ents. Earlier lake ice-off is observed in Iceland, Southwest Alaska, the Southampton Island of Canada, and some parts of Southern Siberia. Conversely, later clearing of lake ice is detected over relatively high latitudes, such as the Svalbard Archipelago of Norway, the Ellesmere Island of Canada, and the Severny Island of Russia. We also displayed finer-scale lake ice BUE mapping results by selecting four hydrological basins. In general, we found that small lakes have earlier ice BUE dates than large ones (see Basin 1 and Basin 3), though the magnitude of difference varies across basins. Similar with Figure 7, for some big lakes there exist discrepancies between the nearshore pixels and those located in the central lake areas.

Figure 8. (a) Pan-Arctic lake ice BUE map for the year 2019 at a 25 km × 25 km tile scale. Four hydrological basins are highlighted for visual display at a 20 m spatial resolution, including (b) North Slope, (c) Victoria Island, (d) Southampton Island, and (e) Yamal Peninsula.

We further divided the Pan-Arctic lake ice BUE map into countries, and the results are shown in Figure 9a. Regarding the average value of the BUE, we found that Norway exhibits the latest BUE date (181), followed by Canada (175), Russia (169), and the USA (150), respectively. In contrast, an earlier BUE estimation by our output is observed in Iceland (112). Overall, countries with greater lake concentrations are less subject to internal variability, as reflected by the lower standard deviations (error bar in Figure 9a). In particular, for Russia, Canada, Iceland, the USA, and Norway, the BUE standard deviations are ±10.2, ±10.7, ±17.9, ±22.1, and ±27.6 days, respectively. Along latitude, all countries (except for Norway and Iceland) exhibit significant trends, together resulting in a linear slope of 1.02 days per degree for the Pan-Arctic (Figure 9b). Moreover, the latitudinal profile of Pan-Arctic lake ice BUE can be roughly divided into four phases by visual interpretation. There is a slight negative trend from 55°N to 62°N, followed by a dramatic increase from 62°N to 65°N, and a relatively stable sequence from 65°N to 78°N. Finally, the BUE is likely
to increase from 78°N onward. At the country level, the greatest positive trend is detected in the USA, although its latitudinal range is limited (approximately from 59°N to 72°N). For Russia, relatively obvious increases are found in its southern and northern parts. The latitudinal profile of Canada exhibits the strongest agreement with the entire study area, indicating its prominent role in shaping the Pan-Arctic lake ice-off phenology.

To examine the sensitivity of the BUE to landscape parameters, we conducted an analysis of the influence of parameter change on BUE distribution at the lake level. These results are displayed and summarized in Figure 10. Except for shoreline development, we found that lake ice BUE is significantly impacted by the selected parameters, including lake area (\( p = 0.00 \)), lake depth (\( p = 0.05 \)), and residence time of the lake waters (\( p = 0.00 \)). As suggested by previous studies [25,45], we demonstrated that bigger lakes are more likely to have later BUE dates and less internal variability than smaller lakes (Figure 10a). On the other hand, an overall later trend is also detected with lake depth increasing, with the most striking BUE change occurring at a lake depth of approximately 1 m (Figure 10b). Additionally, we found the BUE response to lake depth gradient is highly nonlinear, displaying a “saturation” phenomenon for deep lakes. Our analysis further indicates that the relationship between lake ice BUE and shoreline development is insignificant (\( p = 0.82 \), Figure 10c). Such a result, however, is different from that of several regional research...
studies, highlighting the impact of lake morphology on lake ice phenology [46,47]. Last but not least, we found a consistent and linear increase along the residence time profile (Figure 10d), suggesting that Arctic lakes with higher hydrologic connectivity tend to have earlier ice melting dates [48].

![Figure 10](image-url)

**Figure 10.** Characteristics of the Pan-Arctic lake ice BUE given (a) lake area (km²); (b) lake depth (m); (c) shoreline development, measured as the ratio between the shoreline length and the circumference of a circle with the same area; (d) residence time of the lake waters (day). The linear trend was calculated using the group-aggregated parameter-BUE pair, and the red point denotes the average value of each data group along the x-axis.

4. Discussion

Regulating many aspects of Arctic ecosystems, the spatial distribution of lake ice-off phenology is important for us to understand the earliest detected impacts of climatic warming [40,43,49,50]. However, the diversity of freshwater lakes, landscape heterogeneity, and varying weather conditions make fine-resolution Pan-Arctic lake ice BUE mapping a challenging task. In this study, we proposed a novel approach for per-pixel BUE identification based on Sentinel-2 constellation data. We examined the lake ice BUE mapping performance and found that the estimated results matched well with in situ measurements and existing microwave satellite records (Table 2, Figures 5–7). The main advantages of the new approach are threefold. First, we confirmed the effectiveness of the SCL band of the Sentinel-2 Level 2A product for seasonal ice/water monitoring and incorporated it into our BUE identification approach. We showed that the refined lake extent can better
reflect the shoreline boundaries than the original HydroLAKES dataset (Table 1, Figure 4); therefore, the errors caused by lake/land confusion can be greatly reduced or even eliminated. Second, we designed a “max-seg-difference” algorithm, which detects the abrupt change of the binary phenophase time series through temporal segmentation at the pixel level (Figure 3). Compared with commonly used threshold-based methods (e.g., [9,43]), the “max-seg-difference” algorithm is more robust and self-adaptive. As a result, it can be applied to various geographical environments without user intervention. Moreover, the use of phenophase information rather than initial satellite reflectance data has the potential of integrating more data types (e.g., SAR) for fine-resolution lake ice BUE mapping. Finally, we employed the GEE platform to run the full process under the proposed approach, which enables us to apply BUE mapping over a large spatial extent in an effective and efficient manner.

To the best of our knowledge, this study provides the first spatially continuous map of Pan-Arctic lake ice BUE at a 20 m resolution, which can facilitate ongoing efforts to understand the changing Arctic ecosystems. Finer spatial resolution BUE information is not only statistically beneficial for more accurate estimation, but also allows us to reveal the hidden patterns of lake ice melting that would otherwise not be detected using coarse resolution products [15,20]. For example, we found that the nearshore lake ice is likely to melt earlier than that distributed in the central lake area (Figures 7 and 8). Consistent with some previous studies [15,24], our lake ice BUE estimates exhibited an overall later trend with latitudinal gradients across the Pan-Arctic region. However, the magnitude of the latitudinal BUE shift is unevenly distributed (Figure 9), highlighting the complex suite of climatic and ecological forces that can play roles in spring lake ice melting. With an improved resolution, the new BUE map is particularly valuable for monitoring spring ice melting of small lakes, which constitutes the majority of total Arctic lake populations that are most vulnerable to climate change [27]. Theoretically, a small lake is fundamentally different from a big one in thermodynamic processes [25]. Our results confirmed this discrepancy by showing that the BUE date tends to occur later with increasing lake areas and lake depths (Figure 10a,b). Since small lakes are commonly ignored in previous Arctic lake ice phenology studies, reanalysis may be needed using finer-resolution data to clarify the existing findings of Arctic lake ice-off timing dynamics over space and time. We noted an insignificant change in lake ice BUE along the shoreline development profile (Figure 10c), reflecting the internal heterogeneity that obscures the expected effects of lake morphology. For example, within the lake extent, its shoreline topography can be very different. This may influence the wind speed and direction, both of which are drivers of initial ice breaking in early spring [25]. Additionally, our study also provides observational evidence of the hydrological drivers in joint-shaping BUE diversity through connectivity (Figure 10d). Water bodies with high hydrologic connectivity and short residence times are more subject to heat exchange, thus taking fewer days to become ice free in spring [25].

While presenting substantial confidence regarding the Pan-Arctic mapping of fine-resolution lake ice BUE, the proposed approach has limitations and uncertainties. In this study, we selected twin Sentinel-2 satellites as the base data source because of its relatively high spatial and temporal resolutions. With a temporal interval of 5 days, the reconstructed Sentinel-2 composite performed well in general for extracting Pan-Arctic BUE information. Nevertheless, cloud obscuration will lower the actual revisit frequency, making our approach less effective in tracking sub-seasonal variations in lake ice extent. In this sense, the increasing availability of miniature satellites in low Earth orbit, known as CubeSats (e.g., Planet and RapidEye), may offer added value for the timely and accurate estimation of lake ice phenology [5,49]. The derivation of accurate lake extent has always been a critical constraint on ice BUE mapping. In this study, the depicting of lake extent mainly relies on the HydroLAKES dataset, even after a refinement procedure was conducted. The errors of incorrect lake shoreline boundaries will be eventually propagated to the final BUE outputs, and make our estimates less informative. The “max-seg-difference” algorithm depends heavily on the accuracy of SCL, which inevitably contains classification
errors. A recent study showed that there is room for improvement in the current scene classification algorithm by ESA [38]. Hence, a better BUE mapping performance of our approach can be expected once the new version of the SCL product is ready for use.

This study has both methodological and practical implications for future endeavors on Arctic-wide lake ice monitoring. By further including SAR time series data, the proposed approach can be extended to track the ice freeze-up process, which is equally important in understanding the Arctic lake system. Using this study as the baseline, the spatial-temporal dynamics of Arctic lake ice phenology can be quantified, making it possible to examine their drivers and project future changes. Moreover, the fine-resolution lake ice BUE map will benefit a variety of scientific research on Arctic ecosystems. For example, lakes are considered a major source of atmospheric methane (CH4) [11,12], and the timing of ice-off can affect the annual quantity of methane emitted from water to the atmosphere [17]. Finally, we emphasize the necessity of more comprehensive, spatially representative databases that can provide in situ ice phenology records to support model development and validation. So far, there are many lakes where the measurements are discontinued and outdated [8]. We expect this data gap can be filled by new and collaborative programs in the near future.

5. Conclusions

Accurately mapping lake ice-off phenology is essential for us to understand a changing Arctic. In this study, we developed a novel approach for fine-resolution BUE mapping with the use of dense Sentinel-2 time series data via the GEE platform. Experimental results in 45,532 lakes across the Arctic reveal the feasibility of this approach. From the estimated Pan-Arctic lake ice BUE map, four major conclusions can be drawn. First, the Sentinel-2 SCL time series dataset shows great potential in monitoring the sub-seasonal dynamics of ice/water coverage with high spatial (20 m) and temporal (5 day) resolutions, which can facilitate the fine-scale detection of lake ice melting. Second, the utilization of a binary phenophase profile rather than original spectral signatures can reduce data uncertainties, making the “max-seg-difference” algorithm self-adaptive for per-pixel BUE identification and efficient for large area application. Third, the mapped lake ice-off phenology exhibits gradients along the latitude profile and lake characteristics, indicating the joint control of climate and local landscapes in determining Arctic lake ice melting. Finally, we concluded that the new approach provided an improved estimate of Pan-Arctic lake ice BUE, which is expected to enlighten the continuing endeavor for achieving sustainable development goals. Future studies can consider environmental characteristics that were not explored in this study, such as lake ice quality, human footprint, and their impacts on Arctic lake ice BUE. Additionally, SAR satellite data hold potential in capturing ice melting/freezing information without the influence of unfavorable weather conditions. They can be thus combined with optical imagery for providing a refined characterization of Arctic lake ice phenology as well as its multi-decadal dynamics.

**Author Contributions:** Conceptualization, C.L., F.H. and X.C.; methodology, C.L. and H.H.; software, C.L. and Z.Z.; validation, C.L. and Z.Z.; formal analysis, C.L.; investigation, C.L. and Z.Z.; writing—original draft preparation, C.L.; writing—review and editing, C.L., H.H., F.H. and X.C.; visualization, C.L.; supervision, X.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (No. 2019YFC1509104, No. 2018YFC1407103) and Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021008).

**Data Availability Statement:** The Pan-Arctic lake ice-off phenology map for the year 2019 at 20 m resolution is publicly available from the corresponding author on reasonable request.

**Acknowledgments:** We are grateful to the European Space Agency for providing the Sentinel-2 data, European Centre for Medium-Range Weather Forecasts for providing the ERA5 data, World Wildlife Fund for providing the HydroLAKES dataset, and the Finnish Environmental Institute for providing in situ lake ice phenology measurements. This work would not have been possible without the
