https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

Open Access Earth System Science Data Discussions

A high-resolution inland surface water body dataset for the tundra and boreal forests of North America

Yijie Sui[1], Min Feng*[1,2,3] , Chunling Wang[1,3], Xin Li [1,2,3]

[1]National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China

[2]CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China

[3]University of Chinese Academy Sciences, Beijing 100049, China

Correspondence to: Feng M. (mfeng@itpcas.ac.cn)

Abstract. Inland surface waters are abundant in the tundra and boreal forests in North America, essential to environments and human societies but vulnerable to climate changes. These high-latitude water bodies differ greatly in their morphological and topological characteristics related to the formation, type, and vulnerability. In this paper we present an inland surface water body inventory (SWBI) dataset for the tundra and boreal forests of North America. Nearly 6.7 million water bodies were identified, with approximately 6 million (~90%) of them smaller than 0.1 km^2. The dataset provides geometry coverage and morphological attributes for every water body. During this study we developed an automated approach for detecting surface water extent and identifying water bodies in the 10 m resolution Sentinel-2 multispectral satellite data to enhance the capability for delineating small water bodies and their morphological attributes. The approach was applied to the Sentinel-2 data acquired in 2019 to produce the water body dataset for the entire tundra and boreal forests in North America, providing a more complete representation of the region than existing regional datasets, e.g., Permafrost Region Pond and Lake (PeRL). Total accuracy of the detected water extent by SWBI dataset was 96.36% by comparing to interpreted data for locations randomly sampled across the region. Compared to the 30 m or coarser resolution water datasets, e.g., JRC GSW yearly water history, HydroLakes, and Global Lakes and Wetlands Database (GLWD), the SWBI provided an improved ability on delineating water bodies, and reported higher accuracies in the size, number, and perimeter attributes of water body by comparing to PeRL and interpreted regional dataset. This dataset is available on the National Tibetan Plateau/Third Pole Environment Data Center (TPDC, http://data.tpdc.ac.cn): DOI: 10.11888/Hydro.tpdc.271021 (Feng et al., 2020).

1 Introduction

Inland surface waters include various types of water bodies, including rivers and streams; large and small lakes; reservoirs; and ephemeral ponds. Inland surface water occupies only 2% of the global land surface (Pekel et al., 2016), but it plays a critical role in terrestrial ecosystems. Surface water distribution varies across the landscape. More than 55% of global surface waters are located in high latitudes in the Northern Hemisphere (> 44°N), and these northern high-latitude waters are generally small and densely clustered. The high latitudes have warmed faster than other regions, with annual surface temperatures increasing > 1.4° C over the past century (IPCC 2014). The temperature of the Arctic, in particular, has risen twice as fast as the average global temperature (Graversen et al., 2008; Johannessen et al., 2004; Pachauri and Reisinger, 2007; Serreze and Francis, 2006;Li et al., 2020). This change in climate is driving changes in terrestrial ecosystems in the Arctic as well. For example, increases in vegetation productivity have been observed across the northern high latitudes (Forkel et al., 2016). Meanwhile, high-latitude water bodies have started changing since the early 1970s (Carroll et al., 2011; Carroll and Loboda, 2017; Cooley et al., 2019; Smith et al., 2005; Fayne et al., 2020; Nitze et al., 2020). Although some

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

changes are seasonal, and therefore temporary, permanent changes have been reported, and small lakes in permafrost regions are found to be more vulnerable to permanent changes in water extent (Carroll and Loboda, 2017; Karlsson et al., 2014).

As rising temperatures have been reported in permafrost (Biskaborn et al., 2019), its thawing poses a threat to the stability of inland surface waters, especially in the high latitudes, where half of the lakes are thermokarst lakes and have strong interactions with permafrost in the regions. Thawing permafrost not only leads to the formation of lakes and ponds of various sizes, but also leads to the release of organic carbon in the form of carbon dioxide (CO_2) and methane (CH_4) (Serikova et al., 2019). Changes in thermokarst formation may result in concomitant changes to the extent and connectivity of surface water bodies, which can greatly impact the sustainability of aquatic ecosystems. The shapes of the water bodies correlate to suitability of surrounding ecosystems(Grosse et al., 2013; Laird et al., 2003; Schilder et al., 2013; Sharma et al., 2019; Carpenter, 1983; Higgins et al., 2021). Shoreline complexity affects lake ice formation (Sharma et al., 2019). Lake connectivity affects fish migration (Laske et al., 2019; McCullough et al., 2019), fish habitats, and aquatic assemblages (Napiórkowski et al., 2019; Jiang et al., 2021); improves water self-purification and accelerates water cycling (Glińska-Lewczuk, 2009). Water density impacts fish density and biomass (Sandlund et al., 2016; van Zyll de Jong et al., 2017; King et al., 2021). The shape and distribution of water bodies reflect the reasons the water body formed (Laurence C. Smith et al., 2007). Furthermore, information about lake area extent can improve arctic land surface modeling (Langer et al., 2016; van Huissteden et al., 2011). For these reasons, it is critical to discern the high latitude surface water extent, as well as related morphological and topological features, including size and shape.

In the past, inland surface water was mapped at sub-hectare (i.e., 30-m) resolution using satellite data (Feng et al., 2015; Pekel et al., 2016), and these data provide unprecedented information about inland waters in the global extent, including the spatial distribution and changes of inland waters. These datasets provide data that delineates the extent of large and moderate sizes of water bodies but underrepresent or fail to include the large number of small water bodies. Coarse-resolution datasets also lead to underrepresentation in delineating complex shorelines and the shapes of surface water bodies, making it difficult to derive their morphological and topological attributes. Existing datasets containing information that describe water body shapes, such as the Global Lakes and Wetlands Database (GLWD) (Lehner and Döll, 2004) and HydroLAKES (Messager et al., 2016) are limited to water bodies larger than 0.1 km^2. In spite of these limitations, these datasets provide valuable information for improving the precision of mapping inland waters. Detecting the extent of inland surface water at finer spatial scale boosts our ability for mapping the small waters and improves the precision on delineating the shorelines of water bodies. This analysis then allows us to derive an inventory dataset of water bodies along with their morphological and topological attributes. The information allows scientists to analyze a water body as an object instead of a cluster of pixels, advancing our analysis and understanding of the water bodies' size, shoreline complexity, ecological effects, hydrological function, and vulnerability to natural and anthropogenic changes.

In this paper we present a higher resolution inland surface water body inventory (SWBI) for the tundra and boreal forests of North America. It was derived from identifying the extent of inland waters using 10-m resolution Sentinel-2 multispectral data. The dataset provides the spatial extent and morphological attributes for each identified water body. It is the first inland water inventory dataset derived at this landscape scale with the capability of delineating inland surface waters as small as 0.001 km^2.

2 Spatial extent

The SWBI dataset covers all tundra and boreal forest biomes in North America (Figure 1), with the exception of the Arctic Archipelago and Baffin Island due to their long time of snow or ice covering over water bodies. Topography of the tundra

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

and boreal forest in North America is extremely diverse, varying from mountains and rolling hills to plateaus and flat coastal plains. The eastern mountains of the Canadian Codillera are covered by numerous mountain glaciers and divide the region into east coastal plains and west plateaus. The long and narrow eastern coastal plain of this cordillera located near the Pacific Ocean is dominated by thermokarst landform and glacier lakes. The vast western plateaus belong to the stable Canadian Shield and are the result of glacial erosion. The climate of this region is characterized by long, cold winters and short, cool summers. The summer season typically lasts from June to September. The plants in the northern tundra include lichen, moss, grass, sedge, and shrub. The southern boreal forest is dominated by evergreen forests (Ritter, 2006). Lakes and ponds dominate the landscape and approximately 36% of land surface is covered with lakes. There are about 50% of the lakes and 30% of lakes by area in the total region (counted by HydroLAKES). The lakes and ponds formed by glacial erosion are abundant in the western-wide flat Canadian Shield. where the shapes of water bodies usually are thin and complex. The nearly circular water bodies distributed on the east and north coast of which are formed by freezing and thawing (Dranga et al., 2017).

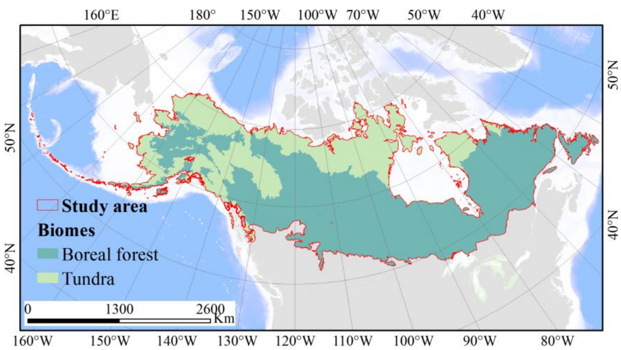

Figure 1: The extent of the study area including the tundra and boreal biomes in the North America continent excluding the Arctic Archipelago and Baffin Island.

3 Data

3.1 Sentinel-2 A/B multi-spectral images

Sentinel-2 multi-spectral images were used to delineate surface water bodies in this study. The Sentinel-2 A/B provides a short revisit cycle (2-3 days) in the high latitudes, which is critical for detecting surface water during the short, snow-free season in the region. Sentinel-2 images were obtained using the United States Geological Survey (USGS) EarthExplorer client/server interface (https://earthexplorer.usgs.gov/, last access: 7 April 2021).

Each Sentinel-2 image consists of 12 multispectral bands, including four bands at 10-m resolution, and eight others at 20-m resolution. Sentinel-2 data are distributed as collections representing different processing levels. We selected the Sentinel-2 Collection 2 data, which provides spectral bands of surface reflectance after atmospheric corrections. The 10-m Sentinel-2 bands were used for water detection to maximize spatial precision for delineating small water bodies. The "s2cloudless" (https://github.com/sentinel-hub/sentinel2-cloud-detector, last access: 7 April 2021) was applied to identify cloud-contaminated pixels, generating a probability of cloud and cirrus detection. This module includes a model generated by a Convolutional Neural Networks (CNN) trained with 6.4 million manually labeled samples. This model was validated to have 99% accuracy for identifying clouds and 84% accuracy for identifying cirrus in Sentinel-2 images (Zupanc, 2020).

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

3.2 JRC yearly water dataset

The JRC yearly water dataset (JRC GSW Yearly Water Classification History, v1.2, https://global-surface-water.appspot.com/) (Pekel et al., 2016) provides a delineation of permanent water, non-water, and seasonal water for global inland surface waters. The dataset was produced using long-term Landsat images, including Landsat TM, ETM+, and OLI images acquired from 1984 to 2019. Permanent water in the dataset was identified as water cover through the entire year, and seasonal water is identified based on occurrence during a single year.

The JRC yearly water dataset provides a reasonably accurate delineation of water distribution for 1984-2019, but its precision is limited by the 30-m spatial resolution of Landsat data. The dataset's accuracy at high latitudes is affected by the relatively poor return cycle of Landsat (16 days), cloudiness, and long periods of snow and ice in the region each year. The JRC dataset was used as a reference to overcome these limitations and improve our ability to identify and monitor inland surface water bodies, particularly small water bodies. The permanent water class in the JRC dataset was used in this analysis, while the seasonal water was excluded due to its reportedly low accuracy (Meyer et al., 2020). The maximum extent of permanent water bodies for the time period 1984-2019 were processed to fill gaps in individual years, which were then used as the reference in this study.

3.3 Permafrost Region Pond and Lake (PeRL)

The Permafrost Region Pond and Lake (PeRL) dataset was produced through a circum-Arctic effort to map ponds and lakes from modern (2002–2013) high-resolution aerial and satellite imagery with a resolution of 5-m or finer, including imagery from GeoEye, QuickBird, WorldView-1/2, the KOMPSAT-2, and TerraSAR-X. The PeRL dataset includes 69 small maps representing a wide range of environmental conditions in tundra and boreal biomes. There are 14 maps mainly distributed in five regions of North America. (Figure 2) Because of the high-resolution data, the PeRL dataset is able to delineate water bodies as small as 10^{-7} km^2, which is valuable for validating satellite-derived water datasets for regions dominated by small water bodies.

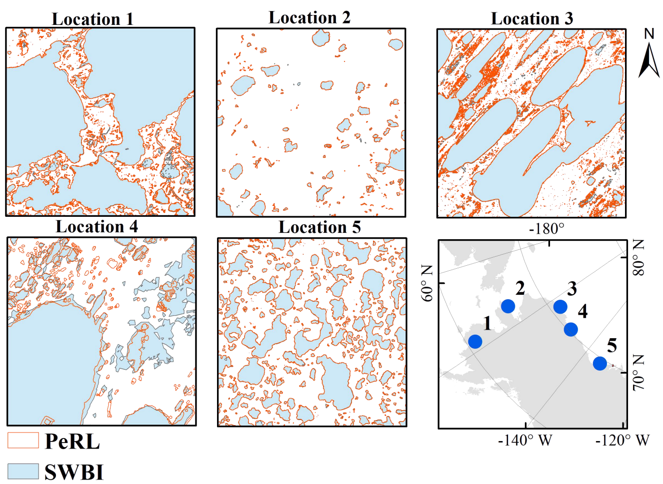

Figure 2: Water bodies identified in the SWBI and PeRL datasets, and the locations (blue dots) of the PeRL maps for the study region.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

4 Methods

The 10-m resolution Sentinel-2 A/B multispectral data are the primary source used to identify small water bodies. Machine learning models were built to detect surface water pixels in each Sentinel-2 image. The results were combined to produce a final 10-m resolution dataset of water extent for 2019 (see section 4.1). Water bodies were identified from the detected water extent using an object-based algorithm to produce the final water body inventory (see section 4.2).

4.1 Detect water extent

To reduce effects from snow cover, Sentinel-2 A/B images acquired between June and September 2019 were selected to represent the relatively snow-free season in North American tundra and boreal biomes. The pixels in each Sentinel-2 image with an estimated cloud probability higher than 65% were excluded to avoid the effects of cloud contamination. During pre-processing, water sensitive indexes were derived from each Sentinel-2 image to enhance the ability on detecting water (Figure 3). To maximize the ability to separate water from non-water, especially vegetated land, three indexes were calculated to represent water and vegetation in each image: Normalized-Difference Water Index (NDWI) (Han-Qiu, 2005), Normalized Difference Vegetation Index (NDVI) (Carlson and Ripley, 1997), and Modified Normalized-Difference Water Index (MNDWI) (McFeeters, 1996). The three indexes were calculated as follows.

$$NDWI = (B_{green} - B_{nir})/(B_{green} + B_{nir}), \quad (1)$$

$$NDVI = (B_{nir} - B_{red})/(B_{nir} + B_{red}), \quad (2)$$

$$MNDWI = (B_{vir} - B_{swir})/(B_{vir} + B_{swir}), \quad (3)$$

An HSV color space conversion was used to combine the three indexes and produce a final index for identifying water. The HSV (hue-saturation-value) color space conversion is a non-trigonometric pair of transformations from a linear red-green-blue (RGB) color space to a perceived color space (Danielson and Gesch, 2011). This method converts the three input bands into hue (color), saturation, and value components. The three indexes (NDWI, MNDWI, and NDVI) were scaled by 255, converted to a byte value type, combined into RGB color space, and then converted to HSV color space to derive a comprehensive index for identifying water.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

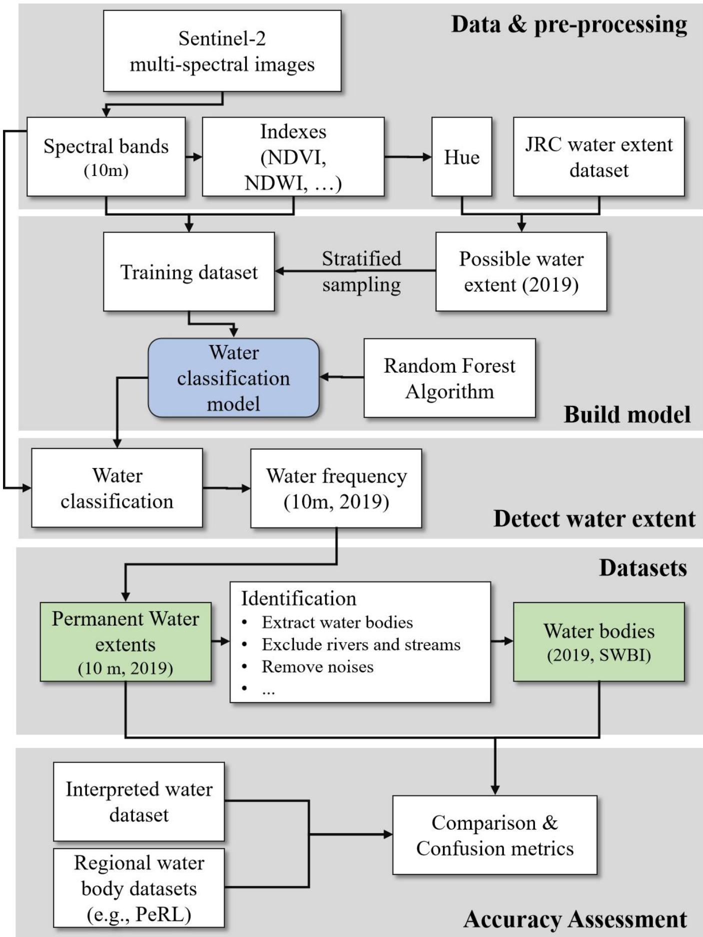

Figure 3: The flowchart of processing water extent and identifying water bodies.

Once the hue has been identified, an experimental threshold of <0.45 was applied to separate water pixels from others. The same procedure was applied to all selected Sentinel-2 images to derive temporal water extents, which were then combined to calculate the water frequency for the year. Potential water extent was then derived from the calculated water frequency data. The existing JRC water dataset provided complementary information for estimating possible water extent. The JRC permanent water records were combined with the Sentinel-2 derived water frequency dataset using a weighted linear combination.

$$A = W_s \cdot A_s + (1 - W_s) \cdot A_j, \quad (4)$$

where, A is the updated water frequency, W_s is the weight for the Sentinel-2-derived water frequency (A_s) and was 0.85 for locations with elevation < 1 km and 0.65 for higher elevations. A_j is the JRC permanent water record, which was 1.0 for permanent water and 0.0 for others. The final, combined potential water extent was identified when $A > 0.5$.

To produce training data for building a water body identification machine learning model, individual points were collected from the identified possible water extents. At this time, 250 points were randomly selected in each stratum, and a total of 1,250 points were collected. (Figure 4a) To enhance the model's ability to separate water from other land cover types in the region, the potential water extent was divided into five strata representing water, glacier, mountain, vegetation, and cloud.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

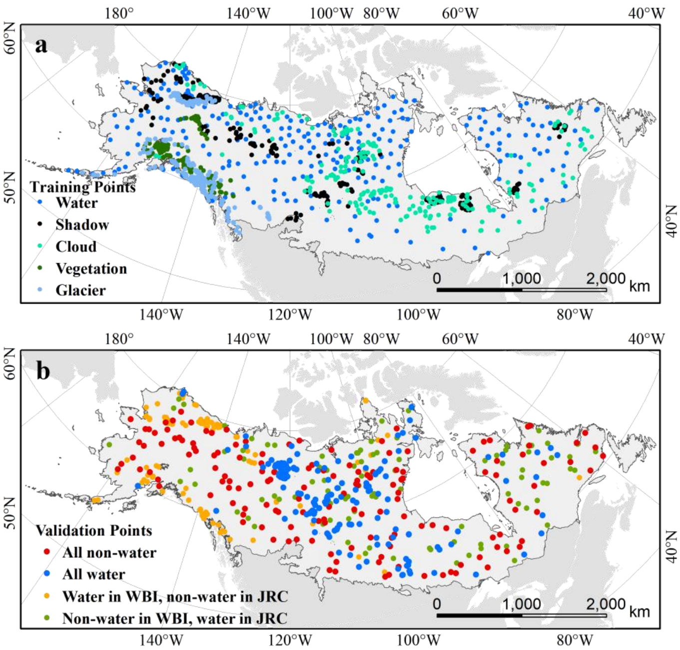

Figure 4: The training samples for random forest model building (a) and points identified for validating the accuracy of the detected water extent (b).

The five strata were established using reference datasets or customized rules. The glacier stratum was identified using the Global Land Ice Measurements from Space (GLIMS) dataset of 2017 (http://www.glims.org/, last access: 7 April 2021), which was a dataset of global glacier outlines including glacier area, geometry, surface velocity, and snow line elevation and was produced from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and the Landsat Enhanced Thematic Mapper Plus (ETM+), as well as historical information derived from maps and aerial photographs. Vegetation was identified as areas with a positive mean NDVI value calculated from the June-September Sentinel-2 images. The cloud stratum was identified as having at least 20% of mean cloud probability calculated from the selected Sentinel-2 images. The mountain stratum was identified as any elevation higher than 1-km.

The selected points were interpreted by the team to provide training data. Although we only used Sentinel-2 images during June to September 2019, points were matched with a randomly selected image at the location during the time period, providing representation for possible temporal variation. Each point was visually labeled by an interpreter after examining the image. Metrics for visible bands (red, green, and blue), NDWI, MDWI, NDVI, and hue were derived from each image to provide attributes for the point. These attributes were pooled to produce training data for building the machine learning model.

The scikit-learn Random Forest algorithm (Breiman, 2001) was adopted to build the model for surface water identification. This model was applied to the selected Sentinel-2 images to detect surface water pixels. The results were compiled temporally to produce a water frequency layer.

In this study, terrain shadows in the water frequency layer were removed with a terrain mask derived from the Global Multi-resolution Terrain Elevation Data (GMTED) (Danielson and Gesch, 2011). The mask was where the slope was greater than or equal to 7° and the elevation was over 1500 m. The elevation threshold was used to minimize the impact of the slope

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

threshold on rivers in lowlands. The method using slope to identify terrain shadows was verified to be more effective than using hill-shade (Carroll and Loboda, 2017).

Permanent water pixels were identified from the resulting water frequency layer as being those pixels with at least 50% occurrence between June and September. The resulting water pixels were then converted to vector polygons using the “Raster to Polygon” tool in ESRI ArcMap 10.2. These water polygons provided the preliminary surface water body records.

An array of geometry metrics was calculated for each water body polygon using ArcMap in the Canada_Lambert_Conformal_Conic projection (datum D_North_American_1983 and Spheroid GRS80). These metrics include area, perimeter, and a shape index (SI), which estimates the complexity of a water body polygon. The SI was calculated as:

$$SI = P_{wateri}/P_{circlei}, \quad (5)$$

where P_{wateri} is the perimeter of the water body i, $P_{circlei}$ is the perimeter of a circle that has the same area as water body i. *SI* equals 1 when a polygon is a perfect circle and greater than 1 when the polygon has a complex irregular shape.

At this point, the SI and area metrics were used to distinguish rivers and streams from lakes and ponds. Rivers and streams have long and linear feature, and we initially applied thresholds of area > 5 km^2 and SI > 10 to preliminarily separate them from lake and ponds. Then, labeled polygons were visually checked to confirm and correct misclassified water bodies.

4.2 Quality assessment

The accuracy and uncertainty of the SWBI were assessed at two levels, i.e., pixel water extent and derived water bodies, to provide a comprehensive evaluation of the dataset. We randomly selected eight square blocks with size of 10 km by 10 km in the North America tundra and boreal region (Figure 5). The selected blocks were visually interpreted by the team to identify all the water bodies within each using a high-resolution Google Earth image as reference for interpretation. Water bodies records from the PeRL were compared to the SWBI water bodies to assess the number of water bodies and spatial area of each. The interpreted dataset was also compared to the JRC-derived water body records for 2019 to assess its accuracy in terms of representing water bodies. The JRC dataset provides water/nonwatery situation for the 30-m resolution pixels, representing the distribution of water extent, but no information of spatial relationship between pixels and water bodies were provided, and we derived water bodies records from the JRC dataset using the same algorithm described in section 4.1.

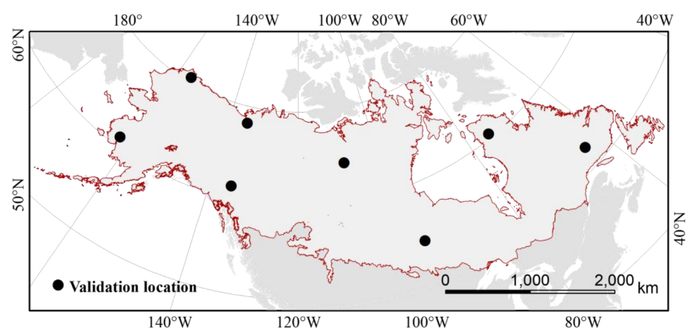

Figure 5: Locations of the 5 regions selected and interpreted for assessing the accuracy of the indicators of water bodies.

The 14 regional PeRL maps were compared to the SWBI water bodies. Although the PeRL maps were produced from high-resolution images acquired in 2002-2013, the maps show little temporal changes when comparing to the SWBI dataset in the

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

extents of the maps (Figure 2), and these maps were adopted as references for evaluating the SWBI water bodies. The PeRL maps were produced from images with 5 m resolution or finer, we excluded all water bodies in PeRL smaller than 0.0003 km^2 to ensure comparability to the scale of the SWBI.

The water extent derived from the Sentinel-2 images were assessed by manually comparing specific points between the SWBI dataset and the JRC surface water dataset. The points were collected using a stratified random sampling across the entire study region. To achieve higher sampling performance, the outcomes were divided into four strata that represent pixels that were agreed as water, disagreed as water, agreed as non-water, and disagreed as non-water. In each of the strata, 400 points were randomly selected from the dataset and manually assessed by examining the same point in the latest Google Earth image. (Figure 4b) The results from the 1600 points were compared to the derived water extent. The confusion matrix was calculated from the results.

The sampling weights were included in the calculation of the metrics as following:

$$W_s = A_s / A_{all}, \tag{6}$$

where A_s is the area of stratum s, and A_{all} is the total area of the region.

Equations of the confusion metrics with weights:

$$OA = \sum_s^4 W_s * OA_s, \tag{7}$$

$$UA = \sum_s^4 W_s * UA_s, \tag{8}$$

$$PA = \sum_s^4 W_s * PA_s, \tag{9}$$

where OA, UA and PA are the overall accuracy, user's accuracy and producer's accuracy of the entire dataset, OA_s, UA_s and PA_s are the concomitant accuracies in stratum s, and W_s is the sampling weight of stratums.

5 Results

5.1 Water bodies in tundra and boreal of North America

More than 6.65 million (6,652,015) surface water bodies were identified in the tundra and boreal forests of North America, while 90.4% of these water bodies (6,015,484) were smaller than 0.1 km^2. Those water bodies covered more than 0.8 million km^2, ~10.3% of the study area (Figure 6). The average size and perimeter of the identified water bodies were 0.12 km^2 and 1.01 km, respectively, and their average *SI* was 1.42.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

Open Access Earth System Science Data Discussions

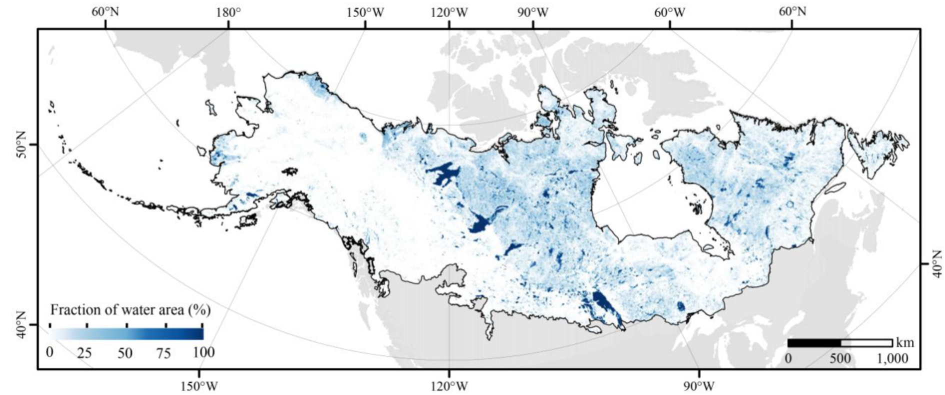

Figure 6: Percent of surface water (5 km × 5km grid) produced by aggregating the water extent for the tundra and boreal forests of North America as calculated using the SWBI.

All of the morphological indicators, including area, perimeter, and *SI*, of the identified water bodies showed great heterogeneity across the region (Figure 7). In general, the tundra biome was dominated by densely packed small water bodies with regular shapes formed by melting frozen ground (Grosse et al., 2013). In contrast, the boreal forest biome was dominated by large water bodies with complex shapes formed by glaciation (Smith et al., 2007). The number of identified water bodies in the tundra (3.32 million) and boreal forests (3.33 million) were nearly identical. However, the water extent in the boreal forest (0.57 million km^2; 70% of total water area) is more than twice that found in the tundra (0.23 million km^2; 30% of the total water area), suggesting again that the water bodies in the tundra are smaller than those in the boreal. This finding was confirmed by reviewing the water body perimeters for the two biomes. The average perimeter of water bodies in boreal forests was 1.2 km, compared to a much smaller 0.8 km average perimeter for water bodies in the tundra. The average *SI* for water bodies in the boreal was 1.46, longer than the 1.37 average *SI* for the tundra water bodies, suggesting the boreal water bodies have much more complex shorelines, while the tundra water bodies are more circular.

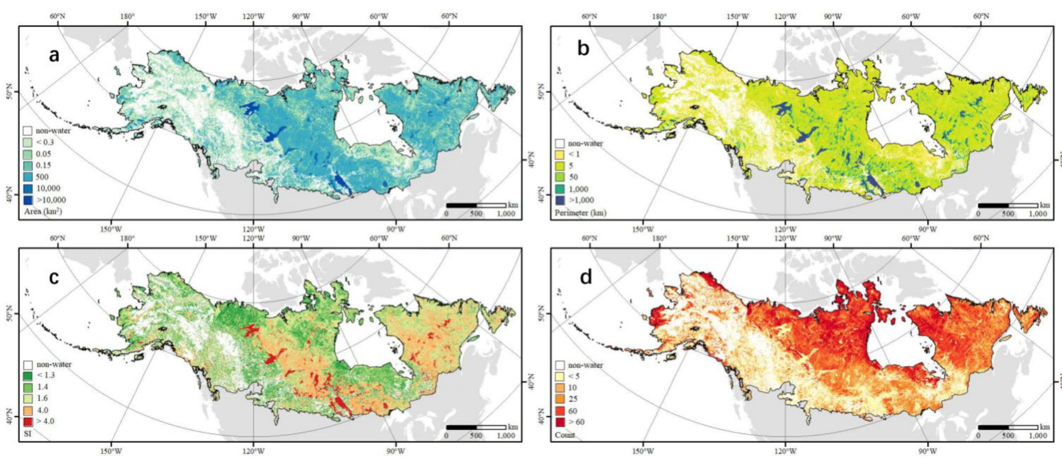

Figure 7: The average area, perimeter, *SI*, and number of identified water bodies in the study area aggregated to 5 km × 5 km grids for visualization.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

Open Access Earth System Science Data Discussions

Inland water in the region is mainly concentrated in the Canadian Shield, i.e., about 0.79 million km^2 of water (98% of water extent in the study region). In addition, most large water bodies were located in the Canadian Shield, including 75% of the identified large water bodies (sizes ≤ $1km^2$). The shorelines of the water bodies in the Canadian Shield were also more complex than those in other areas, especially south of the Laurentian Plateau near the Great Lakes.

5.2 Accuracy assessment

The overall accuracy of the SWBI's water extent was 96.36%, while the producer's accuracy was 99.9%, and the user's accuracy was 96.36%. Misclassifications were primarily found in shadows of the Mackenzie Mountains, where the east-west high-elevation mountain range cast constant shadows on the northern slopes.

Both the JRC and SWBI accurately identified the size of larger water bodies. However, the SWBI performed better than the JRC, and the advantage of the SWBI was demonstrated for smaller water bodies (Figure 8). For small water bodies (size ≤ 0.02 km ^{2}), the average area of the SWBI water bodies was 72% of those manually digitized over high-resolution Google Earth images, compared to only 45% with the water area detected by the JRC (Figure 8a). For medium water bodies (between 0.02 km^2 and 0.05 km^2), the average area of SWBI water bodies was about 85% times that of manually digitized water bodies, compared to 67% with the water area detected by the JRC (Figure 8b). For water bodies larger than 0.05 km^2, the water areas of SWBI were highly consistent (98%) with that of manually digitized. While the water area of JRC was slightly lower (about 87%) for water bodies in the category (Figure 8c).

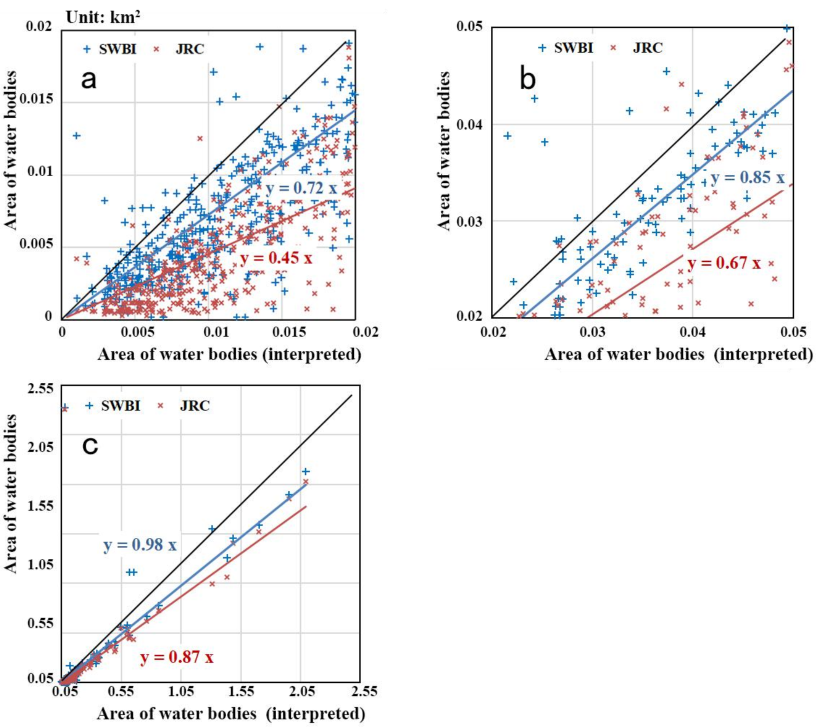

Figure 8: Comparisons of the water body area identified by the JRC, SWBI, and interpreted water maps. The 1:1 lines are in black. The red crosses represent the JRC water bodies, and the blue pluses represent the SWBI water bodies, in comparison with the manually interpreted water bodies. The water bodies are compared in groups of sizes, i.e., (a) small water bodies with sizes < 0.02 km^2; (b) medium water bodies with sizes between 0.02 km^2 and 0.05 km^2; (c) large water bodies with sizes > 0.05 km^2. The R^2 for the SWBI and JRC identified water bodies were similar, i.e., 0.6 for small water bodies, 0.5 for medium water bodies, and 0.9 for large water bodies.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

The comparison between the water bodies identified by SWBI and the PeRL were largely consistent for the derived indicators of water area, perimeter, and number (Figure 9). Linear correlations between the water bodies identified by SWBI and the PeRL water bodies reported R^2 higher than 0.99 for all the three indicators. The slopes of the linear regressions reported that the water area showed the least bias when compared to the PeRL (slope=0.98), followed by the number of water bodies (slope=0.78), and finally the perimeter of the water bodies (slope=0.62).

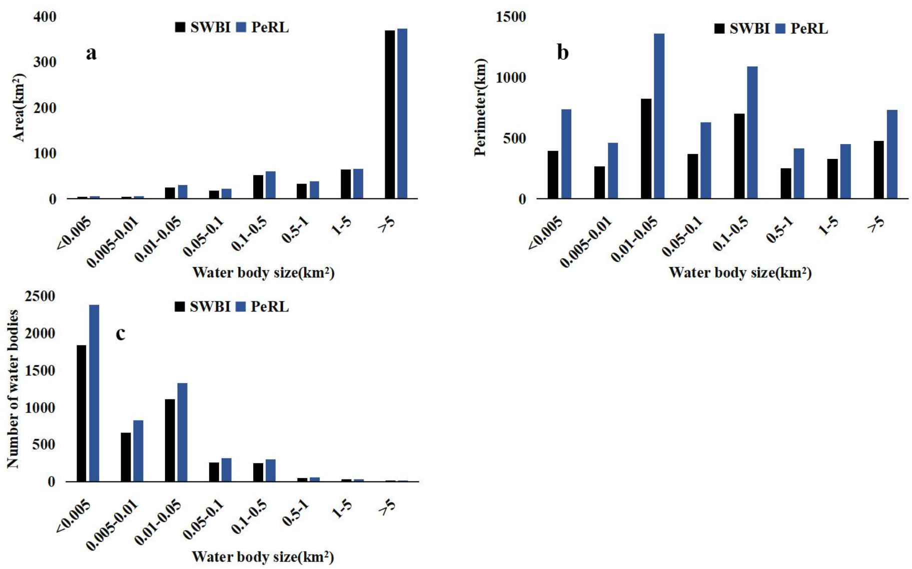

Figure 9: The area, perimeter, and number of the water bodies identified by the PeRL and SWBI dataset.

6 Discussion

6.1 A high-resolution water body dataset for the continental tundra and boreal

The SWBI dataset provides the first known delineation of water bodies at 10-m resolution for the continental tundra and boreal forest of North America, which is one of the highest concentrations of the global inland water especially the small sized water bodies. The dataset not only maps the extent of inland water during 2019 but also identifies the water bodies and their morphological metrics, which are critical for understanding and modeling freshwater lentic ecosystems (Downing, 2009; Heathcote et al., 2015; Kuhn and Butman, 2021; MacIntyre et al., 2009; Muster et al., 2013). The SWBI was produced using Sentinel-2 satellite data to take advantage of the high resolution and 2-3-day revisit time of Sentinel-2 satellites. Sentinel-2's revisit time allows the SWBI to have sufficient observations during the snow-free season, which is critical for mapping inland surface water in this high latitude region with long periods of snow coverage.

The SWBI's 10-m resolution provided the capability for detecting water bodies as small as 0.001 km^2. The validation showed that the WBI dataset had a high overall accuracy and significantly improved upon the ability of the existing global JRC water maps for detecting small water (e.g., smaller than 0.006 km^2) than the existing global JRC water maps. These small water bodies consist of nearly half the total water bodies in the tundra and boreal forest regions of North America, and generally experience faster cycling of water, material, and energy than larger water bodies (Winslow et al., 2014; Carroll et

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

al., 2011; Messager et al., 2016). The improved SWBI dataset may provide more accurate inputs for hydrological estimates, which are vital components for understanding and modeling the pan-Arctic hydrological, biochemical, and energy cycling.

The higher resolution of SWBI also provides the ability to delineate the number, area, shoreline complexity of water bodies. Our comparison confirmed that SWBI-derived water areas and shorelines were similar to those from the regional 5-m or finer resolution PeRL dataset. Meanwhile, the number of water bodies identified in the SWBI was consistent with those of other datasets, including HydroLAKES and GLWD (Figure 10). The numbers of water bodies were roughly identical for the SWBI, HydroLAKES, and GLWD for water bodies larger than 1 km^2. For the water bodies between 0.1 and 1 km^2, the SWBI and HydroLAKES reported similar numbers (Figure 10), but the number reported by GLWD was considerably lower, suggesting that the omission error of GLWD was higher for water bodies smaller than 1 km^2, as noted by Lehner and Döll (2004). Unfortunately, both the HydroLAKES and GLWD datasets only provide records of water bodies larger than 0.1 km^2 (Messager et al., 2016; Lehner and Döll, 2004), and are thus missing records for what we estimate to be 90% of the total number of water bodies in the region. The SWBI is able to extend these indicators to much smaller water bodies than HydroLAKES and GLWD, providing a much more complete record of water bodies in the region. This estimate of the number and extent for small water bodies can improve our understanding of continental freshwater sources stressing the importance of small water bodies in continental biochemical and energy cycling, potentially correcting a misconception that large lakes are most important (Downing, 2010).

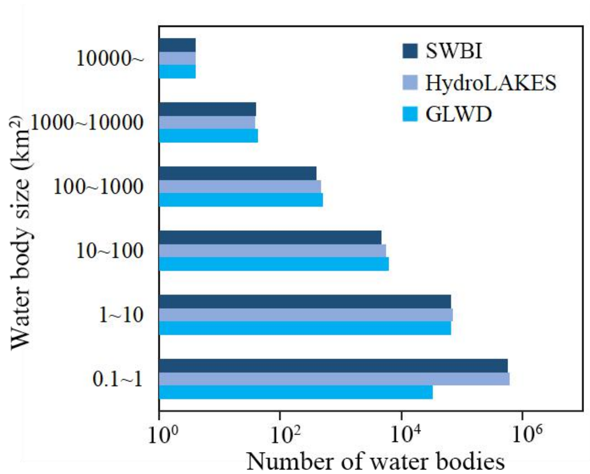

Figure 10: Comparing the number of water bodies identified by the SWBI and by other datasets based on size class.

6.2 Distribution of the water bodies

The largest and most complex water bodies are distributed primarily in the Canadian Shield. These lakes in the Canadian Shield formed through processes such as erosion and glaciation (Smith et al., 2007). Erosion and glaciation formed water bodies with complex shapes, which may contribute to the higher *SI* (1.48) reported by the SWBI for the region. During the most recent Wisconsin glaciation, the Canadian Shield was covered by the Laurentide Ice Sheet, a giant, 3-km thick expanse of ice. When the ice sheet retreated north, it carved out the five Great Lakes as well as thousands of small lakes throughout the Canadian Shield (Dyke and Prest, 1987). Currently, 98% of the water extent in the tundra and boreal forests are distributed in this particular region. For example, the largest lake in the region - Great Bear Lake - has a surface area of 30,227 km^2 with a long, complex shoreline (the perimeter is 5,705 km and the *SI* of the lake is 9.3). It was formed by ice erosion during the Pleistocene (Johnson, 1975).

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

The tundra, on the other hand, is dominated by small, regular shaped water bodies, which is related to the thawing and freezing of permafrost (Grosse et al., 2013). During the winter, water in the soil can freeze into ice. The freezing soil becomes puffy, forming a hilly structure. In the summer, this hilly structure melts and settles, forming a thermokarst lake. This hilly structure is small and regular, resulting in small, circular thermokarst lakes (Grosse et al., 2013). Numerous thermokarst lakes are experiencing dramatic changes, which is considered as an indicator for permafrost degradation (Smith et al., 2005; Karlsson et al., 2012, 2014). The small thermokarst lakes were also found experiencing stronger changes comparing to the large lakes (Karlsson et al., 2014; Carroll and Loboda, 2017). Monitoring water extent without discriminating lake sizes could not precisely reflect those strong changes in the small lakes due to the area dominance of large lakes. Additionally, the small thermokarst lakes are the primary source of permafrost carbon emissions (Kuhn et al., 2018; Walter Anthony et al., 2016; Yvon-Durocher et al., 2017), and the small water bodies were found to be a major uncertainty in estimating greenhouse gas emissions (Holgerson and Raymond, 2016). The SWBI could provide critical information for investigating thermokarst lakes, especially the small thermokarst lakes and ponds, and estimating their effects on carbon emission and permafrost sustainability in the tundra and boreal forests in North America. As reported by the analysis of the SWBI, 3.32 million small water bodies were found in the tundra in 2019 with an average size of 0.07 km2 and average SI of 1.37, much smaller than the SI of the lakes in the boreal. Teshekpuk Lake is the largest thermokarst lake in the world with a relatively smooth shoreline (SI = 5.4), considerably smaller than the SI of the Great Bear Lake in the boreal (Markon and Derksen, 1994).

6.3 Limitations

The data and methods used to derive the10-m resolution SWBI dataset are able to detect water bodies smaller than the 30-m or coarser resolution satellite derived datasets, but have difficulty identifying water bodies smaller than 0.001 km^2, and the capability can be further improved by incorporating higher resolution satellite data, such as from Planet, WorldView, QuickBird, and Gaofen (Veremeeva and Günther, 2017; Sun et al., 2020; Watson et al., 2016; Andresen and Lougheed, 2015). Errors in the satellite data provide substantial sources of uncertainty, including an inability to separate rivers and streams because the resolution is too coarse, bias in estimates of water extent resulting from temporal gaps in data, and misclassifications resulting from spectral resolution. The misclassifications impacted by terrain (e.g., mountain shadows) still exist even though they have been substantially reduced during data processing. Further processing may be possible to further reduce these errors. This dataset was produced using satellite data acquired in 2019, and it does not reflect changes of the water bodies in the region. Further efforts can be carried out to produce an inland water dataset for multiple time periods using these methods to capture the seasonal and multi-year dynamics of inland water in the region.

7 Data availability

This dataset can be accessed via the website of the National Tibetan Plateau/Third Pole Environment Data Center (TPDC, http://data.tpdc.ac.cn): DOI: 10.11888/Hydro.tpdc.271021 (Feng et al., 2020). The dataset is provided in ESRI Geodatabase format. The volume of this dataset is about 1.5 GB.

8 Conclusions

This study presents an inland surface water body dataset of tundra and boreal forest biomes of the northern latitudes of North America. The SWBI dataset was generated using Sentinel-2 data with machine learning methods and an object-based algorithm. Three morphological metrics (area, perimeter, and *SI*) were calculated for each water body. Accuracy of the dataset was carefully assessed with respect to detecting inland surface water extent (or pixel level) and identifying water

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

bodies. The dataset's overall accuracy for water extent reached 96.36%. In addition, the WBI showed a high consistency with high resolution images in terms of water area, perimeter, and quantity.

To our knowledge, the SWBI dataset provided the most complete inventory of inland surface water bodies for the tundra and boreal forest of North America. Overall, 6.65 million water bodies were identified, covering 10.3% of the region. Small water bodies were dominance in the region with ~90.4% were smaller than 0.1 km^2. Results from an analysis of the SWBI indicate that the tundra biome is dominated by densely small water bodies with regular shapes (the average *SI* was 1.37) while the boreal forest biome is dominated by large water bodies with complex shapes (the average *SI* was 1.46). The WBI is expected to be able to provide supporting data for modeling hydrologic, biochemical, and energy cycling in these areas.

Acknowledgements

This work was supported by Basic Science Center for Tibetan Plateau Earth System (BSCTPES, NSFC project No. 41988101)

Reference

Andresen, C. G. and Lougheed, V. L.: Disappearing Arctic tundra ponds: Fine-scale analysis of surface hydrology in drained thaw lake basins over a 65 year period (1948–2013), J. Geophys. Res. Biogeosciences, 120, 466–479, 2015.
Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., and Abramov, A.: Permafrost is warming at a global scale, Nat. Commun., 10, 1–11, 2019.
Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.
Carlson, T. N. and Ripley, D. A.: On the relation between NDVI, fractional vegetation cover, and leaf area index, Remote Sens. Environ., 62, 241–252, 1997.
Carpenter, S. R.: Lake geometry: implications for production and sediment accretion rates, J. Theor. Biol., 105, 273–286, 1983.
Carroll, M. and Loboda, T.: Multi-Decadal Surface Water Dynamics in North American Tundra, Remote Sens., 9, https://doi.org/10.3390/rs9050497, 2017.
Carroll, M. L., Townshend, J. R. G., DiMiceli, C. M., Loboda, T., and Sohlberg, R. A.: Shrinking lakes of the Arctic: Spatial relationships and trajectory of change, Geophys. Res. Lett., 38, 2011.
Cooley, S. W., Smith, L. C., Ryan, J. C., Pitcher, L. H., and Pavelsky, T. M.: Arctic-Boreal Lake Dynamics Revealed Using CubeSat Imagery, Geophys. Res. Lett., 46, 2111–2120, https://doi.org/10.1029/2018gl081584, 2019.
Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), 2011.
Downing, J. A.: Global limnology: Up-scaling aquatic services and processes to planet Earth, Int. Ver. Für Theor. Angew. Limnol. Verhandlungen, 30, 1149–1166, 2009.
Downing, J. A.: Emerging global role of small lakes and ponds: little things mean a lot, Limnetica, 29, 0009–0024, 2010.
Dranga, S. A., Hayles, S., and Gajewski, K.: Synthesis of limnological data from lakes and ponds across Arctic and Boreal Canada, Arct. Sci., 4, 167–185, https://doi.org/10.1139/as-2017-0039, 2017.
Dyke, A. and Prest, V.: Late Wisconsinan and Holocene history of the Laurentide ice sheet, Géographie Phys. Quat., 41, 237–263, 1987.
Fayne, J. V., Smith, L. C., Pitcher, L. H., Kyzivat, E. D., Cooley, S. W., Cooper, M. G., Denbina, M. W., Chen, A. C., Chen, C. W., and Pavelsky, T. M.: Airborne observations of arctic-boreal water surface elevations from AirSWOT Ka-Band InSAR and LVIS LiDAR, Environ. Res. Lett., 15, 105005, 2020.
Feng, M., Sexton, J. O., Channan, S., and Townshend, J. R.: A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic – spectral classification algorithm, Int. J. Digit. Earth, 9, 113 – 133, https://doi.org/10.1080/17538947.2015.1026420, 2015.
Feng, M., Sui, Y.: High resolution inland surface water dataset for the tundra and boreal in North America, https://doi.org/10.11888/Hydro.tpdc.271021, 24 October 2020.
Forkel, M., Carvalhais, N., Rödenbeck, C., Keeling, R., Heimann, M., Thonicke, K., Zaehle, S., and Reichstein, M.: Enhanced seasonal CO2 exchange caused by amplified plant productivity in northern ecosystems, Science, 351, 696–699, 2016.
Glińska-Lewczuk, K.: Water quality dynamics of oxbow lakes in young glacial landscape of NE Poland in relation to their hydrological connectivity, Ecol. Eng., 35, 25–37, https://doi.org/10.1016/j.ecoleng.2008.08.012, 2009.
Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., and Svensson, G.: Vertical structure of recent Arctic warming, Nature, 451, 53–56, 2008.

Grosse, G., Jones, B., and Arp, C.: 8.21 Thermokarst Lakes, Drainage, and Drained Basins, in: Treatise on Geomorphology, edited by: Shroder, J. F., Academic Press, San Diego, 325 – 353, https://doi.org/10.1016/B978-0-12-374739-6.00216-5, 2013.

Han-Qiu, X.: A study on information extraction of water body with the modified normalized difference water index (MNDWI), J. Remote Sens., 5, 589–595, 2005.

Heathcote, A. J., del Giorgio, P. A., and Prairie, Y. T.: Predicting bathymetric features of lakes from the topography of their surrounding landscape, Can. J. Fish. Aquat. Sci., 72, 643–650, 2015.

Higgins, S., Desjardins, C., Drouin, H., Hrenchuk, L., and Van der Sanden, J.: The role of climate and lake size in regulating the ice phenology of boreal lakes, J. Geophys. Res. Biogeosciences, 126, e2020JG005898, 2021.

Holgerson, M. A. and Raymond, P. A.: Large contribution to inland water CO2 and CH4 emissions from very small ponds, Nat. Geosci., 9, 222–226, https://doi.org/10.1038/ngeo2654, 2016.

van Huissteden, J., Berrittella, C., Parmentier, F. J. W., Mi, Y., Maximov, T. C., and Dolman, A. J.: Methane emissions from permafrost thaw lakes limited by lake drainage, Nat. Clim. Change, 1, 119 – 123, https://doi.org/10.1038/nclimate1101, 2011.

Jiang, X., Zheng, P., Cao, L., and Pan, B.: Effects of long-term floodplain disconnection on multiple facets of lake fish biodiversity: Decline of alpha diversity leads to a regional differentiation through time, Sci. Total Environ., 763, 144177, 2021.

Johannessen, O. M., Bengtsson, L., Miles, M. W., Kuzmina, S. I., Semenov, V. A., Alekseev, G. V., Nagurnyi, A. P., Zakharov, V. F., Bobylev, L. P., and Pettersson, L. H.: Arctic climate change: observed and modelled temperature and sea-ice variability, Tellus Dyn. Meteorol. Oceanogr., 56, 328–341, 2004.

Johnson, L.: The Great Bear Lake: its place in history, Arctic, 28, 231–244, 1975.

Karlsson, J., Lyon, S., and Destouni, G.: Temporal Behavior of Lake Size-Distribution in a Thawing Permafrost Landscape in Northwestern Siberia, Remote Sens., 6, 621–636, https://doi.org/10.3390/rs6010621, 2014.

Karlsson, J. M., Lyon, S. W., and Destouni, G.: Thermokarst lake, hydrological flow and water balance indicators of permafrost change in Western Siberia, J. Hydrol., 464–465, 459–466, https://doi.org/10.1016/j.jhydrol.2012.07.037, 2012.

King, K. B., Bremigan, M. T., Infante, D., and Cheruvelil, K. S.: Surface water connectivity affects lake and stream fish species richness and composition, Can. J. Fish. Aquat. Sci., 78, 433–443, 2021.

Kuhn, C. and Butman, D.: Declining greenness in Arctic-boreal lakes, Proc. Natl. Acad. Sci., 118, 2021.

Kuhn, M., Lundin, E. J., Giesler, R., Johansson, M., and Karlsson, J.: Emissions from thaw ponds largely offset the carbon sink of northern permafrost wetlands, Sci. Rep., 8, 9535, https://doi.org/10.1038/s41598-018-27770-x, 2018.

Laird, N. F., Walsh, J. E., and Kristovich, D. A.: Model simulations examining the relationship of lake-effect morphology to lake shape, wind direction, and wind speed, Mon. Weather Rev., 131, 2102–2111, 2003.

Langer, M., Westermann, S., Boike, J., Kirillin, G., Grosse, G., Peng, S., and Krinner, G.: Rapid degradation of permafrost underneath waterbodies in tundra landscapes—Toward a representation of thermokarst in land surface models, J. Geophys. Res. Earth Surf., 121, 2446–2470, https://doi.org/10.1002/2016jf003956, 2016.

Laske, S. M., Rosenberger, A. E., Wipfli, M. S., and Zimmerman, C. E.: Surface water connectivity controls fish food web structure and complexity across local- and meta-food webs in Arctic Coastal Plain lakes, Food Webs, 21, https://doi.org/10.1016/j.fooweb.2019.e00123, 2019.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, J. Hydrol., 296, 1–22, https://doi.org/10.1016/j.jhydrol.2004.03.028, 2004.

Li, X., Che, T., Li, X., Wang, L., Duan, A., Shangguan, D., Pan, X., Fang, M., and Bao, Q.: CASEarth poles: big data for the three poles, Bull. Am. Meteorol. Soc., 101, E1475–E1491, 2020.

Lindgren, P. R., Farquharson, L. M., Romanovsky, V. E., and Grosse, G.: Landsat-based lake distribution and changes in western Alaska permafrost regions between the 1970s and 2010s, Environ. Res. Lett., 16, 025006, https://doi.org/10.1088/1748-9326/abd270, 2021.

MacIntyre, S., Fram, J. P., Kushner, P. J., Bettez, N. D., O'brien, W., Hobbie, J., and Kling, G. W.: Climate-related variations in mixing dynamics in an Alaskan arctic lake, Limnol. Oceanogr., 54, 2401–2417, 2009.

Markon, C. J. and Derksen, D. V.: Identification of tundra land cover near Teshekpuk Lake, Alaska using SPOT satellite data, Arctic, 222–231, 1994.

McCullough, I. M., King, K. B. S., Stachelek, J., Diaz, J., Soranno, P. A., and Cheruvelil, K. S.: Applying the patch-matrix model to lakes: a connectivity-based conservation framework, Landsc. Ecol., 34, 2703 – 2718, https://doi.org/10.1007/s10980-019-00915-7, 2019.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, Int. J. Remote Sens., 17, 1425–1432, 1996.

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, Nat. Commun., 7, 1–11, 2016.

Meyer, M. F., Labou, S. G., Cramer, A. N., Brousil, M. R., and Luff, B. T.: The global lake area, climate, and population dataset, Sci Data, 7, 174, https://doi.org/10.1038/s41597-020-0517-4, 2020.

Muster, S., Heim, B., Abnizova, A., and Boike, J.: Water body distributions across scales: A remote sensing based comparison of three arctic tundra wetlands, Remote Sens., 5, 1498–1523, 2013.

Napiórkowski, Bąkowska, Mrozińska, Szymańska, Kolarova, and Obolewski: The Effect of Hydrological Connectivity on the Zooplankton Structure in Floodplain Lakes of a Regulated Large River (the Lower Vistula, Poland), Water, 11, https://doi.org/10.3390/w11091924, 2019.

https://doi.org/10.5194/essd-2021-278
Preprint. Discussion started: 10 December 2021

Nitze, I., Cooley, S. W., Duguay, C. R., Jones, B. M., and Grosse, G.: The catastrophic thermokarst lake drainage events of 2018 in northwestern Alaska: Fast-forward into the future, The Cryosphere, 14, 4279–4297, 2020.

Pachauri, R. K. and Reisinger, A.: IPCC fourth assessment report, IPCC Geneva, 2007, 2007.

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, 418–422, https://doi.org/10.1038/nature20584, 2016.

Ritter, M. E.: The physical environment: An introduction to physical geography, Date Visit. July, 25, 2008, 2006.

Sandlund, O. T., Eloranta, A. P., Borgstrøm, R., Hesthagen, T., Johnsen, S. I., Museth, J., and Rognerud, S.: The trophic niche of Arctic charr in large southern Scandinavian lakes is determined by fish community and lake morphometry, Hydrobiologia, 783, 117–130, https://doi.org/10.1007/s10750-016-2646-5, 2016.

Schilder, J., Bastviken, D., van Hardenbroek, M., Kankaala, P., Rinta, P., Stötter, T., and Heiri, O.: Spatial heterogeneity and lake morphology affect diffusive greenhouse gas emission estimates of lakes, Geophys. Res. Lett., 40, 5752–5756, 2013.

Serikova, S., Pokrovsky, O. S., Laudon, H., Krickov, I., Lim, A. G., Manasypov, R. M., and Karlsson, J.: High carbon emissions from thermokarst lakes of Western Siberia, Nat. Commun., 10, 1–7, 2019.

Serreze, M. C. and Francis, J. A.: The Arctic amplification debate, Clim. Change, 76, 241–264, 2006.

Sharma, S., Blagrave, 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, Nat. Clim. Change, 9, 227–231, https://doi.org/10.1038/s41558-018-0393-5, 2019.

Smith, L. C., Sheng, Y., MacDonald, G. M., and Hinzman, L. D.: Disappearing arctic lakes, Science, 308, 1429–1429, 2005.

Smith, L. C., Sheng, Y., and MacDonald, G. M.: A first pan-Arctic assessment of the influence of glaciation, permafrost, topography and peatlands on northern hemisphere lake distribution, Permafr. Periglac. Process., 18, 201 – 208, https://doi.org/10.1002/ppp.581, 2007.

Sun, J., Wang, G., He, G., Pu, D., Jiang, W., Li, T., and Niu, X.: Study on the Water Body Extraction Using GF-1 Data Based on Adaboost Integrated Learning Algorithm, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., 42, 641–648, 2020.

Van Gerven, M. and Bohte, S.: Artificial neural networks as models of neural information processing, Front. Comput. Neurosci., 11, 114, 2017.

Veremeeva, A. A. and Günther, F.: Thermokarst lake and baydzherakh area changes on Yedoma uplands, Yakutian coastal lowlands: repeat inventory using high resolution imagery, 2017.

Walter Anthony, K., Daanen, R., Anthony, P., Schneider von Deimling, T., Ping, C.-L., Chanton, J. P., and Grosse, G.: Methane emissions proportional to permafrost carbon thawed in Arctic lakes since the 1950s, Nat. Geosci., 9, 679–682, https://doi.org/10.1038/ngeo2795, 2016.

Watson, C. S., Quincey, D. J., Carrivick, J. L., and Smith, M. W.: The dynamics of supraglacial ponds in the Everest region, central Himalaya, Glob. Planet. Change, 142, 14–27, 2016.

Winslow, L. A., Read, J. S., Hanson, P. C., and Stanley, E. H.: Lake shoreline in the contiguous United States: quantity, distribution and sensitivity to observation resolution, Freshw. Biol., 59, 213–223, 2014.

Yvon-Durocher, G., Hulatt, C. J., Woodward, G., and Trimmer, M.: Long-term warming amplifies shifts in the carbon cycle of experimental ponds, Nat. Clim. Change, 7, 209–213, https://doi.org/10.1038/nclimate3229, 2017.