Dear Editor,

We appreciate the encouraging comments and valuable suggestions from the reviewers to our manuscript titled "A high-resolution inland surface water body dataset for the tundra and boreal forests of North America." We have made extensive revisions in the manuscript in response to the comments and suggestions. The major revisions are:

1) as a reviewer suggested, an additional analysis was added in revision to examine the difference between the lake numbers observed in our data and those estimated from the power-law relationship. Although the relationship has been adopted widely for providing a critical parameter to research such as greenhouse gas emission, the relationship has significantly overestimated the number of small lakes, and our presenting dataset could provide the key dataset for correcting the issue.
2) The water body mapping method was revised extensively. To clarify the logic of mapping procedures, we have reorganized the structure of the Methods section, and the original section of water detection was divided into three sections for describing the key steps. Moreover, the flowchart was updated to provide a clearer description of the method.
3) considering the extreme difficulty of identifying water bodies connected to rivers or streams, we applied a method to estimate the water bodies that could possibly be these, providing a reference for users to consider when they need to separate the water bodies.
4) Figure 1 was updated to include names of the key regions and lakes to help users follow the discussion when these geographic places are involved.
5) The name of the dataset was changed from SWBI to WBD-NAHL to be more clearly associated with the North American part of the biomes.

We believe revision has been much improved. In addition to the revision in the manuscript, we have responded to the comments individually. You may find the details of the comments and our responses in the attached document.

Would you please let me know if you have questions or further suggestions. We are looking forward to hearing from you.

Yours sincerely,

Min Feng on behalf of all co-authors

The following are our responses to the review comments. We keep the comments from the reviewers in regular font and our responses in bold and blue colour.

Referee #1

The presented dataset includes polygons of 6.7 million water bodies in northern North America, where the novelty is the inclusion of small water-bodies <0.1 km2 in size (90% of the included water bodies). I can see a value in this dataset, especially for studies of impacts of climate change on high latitude greenhouse gas exchange – where small aquatic ecosystems are likely to have a very disproportional influence.

Response: We appreciate the reviewer's comment on the value of the dataset. We completely agree that the dataset could provide information on the enormous amount of small water bodies, which are critical for understanding these water bodies and their impact on the climate, particularly the greenhouse gas exchange.

The dataset is based on analysis of 10m resolution Sentinel-2 data from 2019. The approach to delineate and validate the polygons for water bodies seems robust, but I suggest that the authors discuss and perhaps analyze what the implications are of using satellite data from a single year. The spatial extent of many high-latitude water bodies, especially smaller water bodies in relatively flat regions, can vary significantly both inter- and intra-annually. Many regions of northern North America have relatively dry climates, and thus can experience significant multi-year drought or flood conditions. For example, it would be good to know what the deviations from normal conditions were in terms of cumulative precipitation for the 2 years prior to the collection of the remote sensing data. If some regions had relatively higher or lower long-term precipitation prior to 2019, this could indicate that the resulting dataset either over- or under-estimates the number and areas of small water bodies.

Response:

We agree with the reviewer. Mapping with a single year of data could only represent the status of the corresponding year, and changes could occur in other years. As the reviewer

suggested, we expanded the discussion to discuss the implication of a single year mapping. We also investigated the precipitation of 2019 in comparison with that of the past decade (2010-2019) and found that 2019 was very close to the normal of the decade. We also mentioned that in the discussion. (Line 406-411: " The WBD-NAHL dataset was produced based on Sentinel-2 data acquired in the summer of 2019 and represents the distribution of surface water in the corresponding year. The mean total precipitation in 2019 in the region was 438.5 mm, which was close to the historical average from 2010 to 2019 (mean: 435.9 mm, standard deviation: 11.5 mm) (Huffman et al. 2019). Although 2019 can be considered a normal year of the past decade in terms of precipitation, the spatial extent of high-latitude water bodies, especially smaller water bodies, can still vary significantly both inter- and intra-annually locally. Nevertheless, it would be interesting to explore water bodies' changes using observations from multiple years. ")

It is not explained why the dataset doesn't also include high latitude regions in Eurasia. I'm assuming it's about computing power? A full Pan-Arctic lake dataset would have additional uses compared to one that only focuses on North America, as it e.g. would allow for global estimates of high latitude aquatic greenhouse gas emissions.

Response:

We thank the reviewer for the suggestion. As the reviewer pointed out, the current dataset was only focused on the biomes in North America mainly due to limitations on computational and funding. We agree that covering the biomes in Eurasia would provide further value for research, such as high latitude aquatic greenhouse gas emissions. We are planning to move forward to Eurasia following this dataset. We also discussed the point in the manuscript (Line 413: The WBD-NAHL dataset focused on the tundra and boreal forest regions of North America. The methodology can be extended to Eurasia to provide a complete representation of the biomes.)

A minor suggestion could be to change the name of the dataset so it is more precise, so that the name of the dataset indicates the region which it covers. Perhaps "The North America High Latitude Water Body Dataset", NAHL-WBD?

Response: We agree with the reviewer and changed the name of the dataset to "WBD-NAHL" in this revision to be more precise.

The manuscript has significant deficiencies in its discussion/explanation of the physiography of North America, and the dominant lake types in different regions. Detailed comments are below in the "specific comments" section. The description of different regions within North America is confusing and hard to follow what regions actually are considered. The maps used in the manuscript do not indicate any of the key regions that are discussed. Other key points are that the reduction of lakes types in the discussion into just glacial and thermokarst lakes is insufficient (other types of lake are important, including those in peatland regions), and that focusing on differences in lake characteristics between just the boreal and tundra regions likely misses more important controls on lake morphology and size distribution (e.g. that the influence of surficial and bedrock geology and is more important for lakes than the terrestrial biome they are located in).

Response:

Extensive revisions were made to the manuscript following the reviewer's comments. For example, Figure 1 was updated to indicate key regions, and the influence of geology on the formation of water bodies was also discussed. Detailed edits and responses were explained in the following comments.

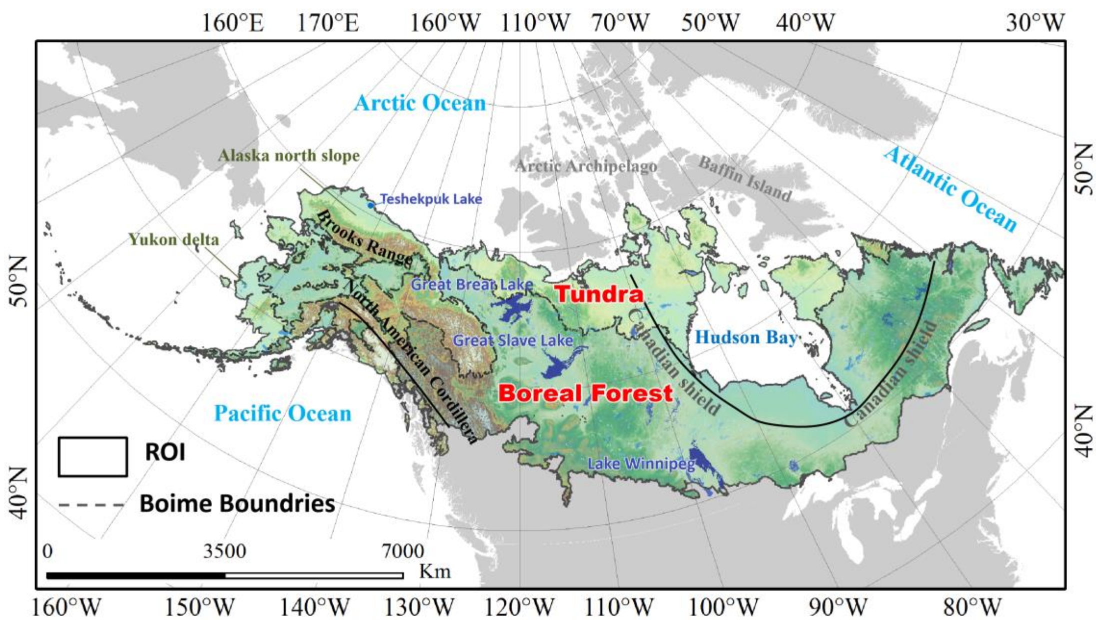


Figure 1: The extent of the study area including the tundra and boreal biomes in the North America continent excluding the Arctic Archipelago and Baffin Island.

Several previous attempts at assessing the importance of aquatic ecosystems for greenhouse gas emissions have relied on databases that only included larger water bodies, along with assumptions of the number and area of smaller lakes (e.g. see Holgersson et al., 2016 and Cael and Seekell 2016). It would be interesting to see how the number and areas of small waterbodies in the presented dataset compares to the assumptions made in these previous studies. While not necessary for the study, you could include an analysis of the relationships between number and areas of water bodies for different regions of study domain, (contrast the lake size/area distribution for the Canadian Shield with that of coastal lowlands and peatland regions). Are these observed distributions similar to, or different from those previous attempts which were based on assumptions rather than observations?

Response:

We appreciate the valuable suggestion! We conducted an extra analysis in this revision to examine the difference between the lake numbers observed in our data and those estimated from the power-law distribution of the lake area. The analysis revealed that the relationship significantly overestimated the number of small lakes. As the reviewer pointed out, we believe that our presenting dataset could contribute to addressing such issues and improving greenhouse gas emissions. The following sentences were added (Line 349-359: An empirical power-law distribution was found between lake areas and lake numbers (Messager et al., 2016; Downing et al., 2006), and the distribution was applied to estimate the number of small lakes, which were used for estimating greenhouse gas emissions (Holgersson et al., 2016). According to the power-law distribution and HydroLAKES, the number of water bodies larger than 0.1 km^2 was estimated to be about 798,895, which was close to the 629,130 water bodies reported by WBD-NAHL (Figure 11). However, the number of water bodies sized between 0.1 and 0.01 km2 was estimated to be about 10.2 million, 4.8 times higher than estimated by WBD-NAHL. Furthermore, the water bodies sized between 0.01-0.001 km^2 were estimated to be about 126.1 million, 33.6 times higher than what was estimated by WBD-NAHL, suggesting that the power-

law distribution significantly overestimates the number of small lakes. A similar finding was reported by Seekell et al. (2016). Estimating the number small water bodies using a power-law distribution could introduce considerable uncertainties in the estimation of the contribution of small water bodies to greenhouse gas emissions. Accurately identifying small water bodies could correct this overestimation and improve greenhouse gas emission estimates (Holgersson et al., 2016).

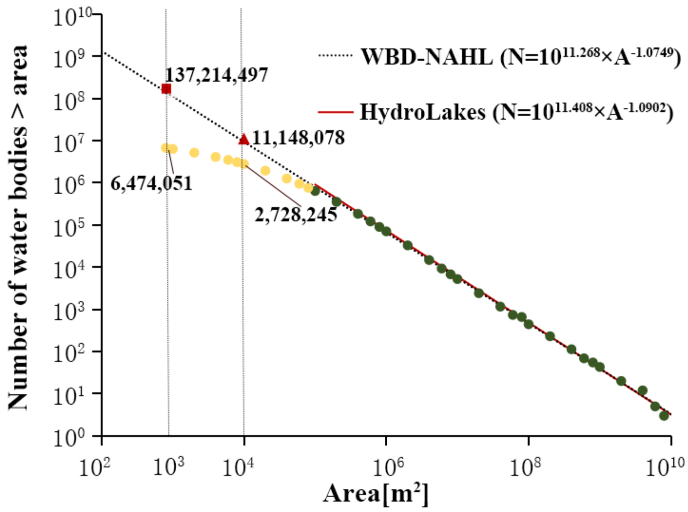


Figure 11: Distribution of the total numbers of water bodies in relation to the area of North American tundra and boreal forests water bodies. The circles represent the number of water bodies provided by WBD-NAHL. The black line is the power-law distribution modeled using water bodies > 0.1 km^2 from WBD-NAHL. The red line is the power-law distribution modeled using HydroLAKES in the study region. The red triangle and square represent, respectively, the extrapolated number of water bodies > 0.01 km^2 and > 0.001 km^2 based on the power-law distribution modeled from HydroLAKES.

Specific Comments:

L40 – What is the reference for the statement "especially in the high latitudes, where half of the lakes are thermokarst lakes"? Depending on what you define as "high latitudes", this statement may not be accurate. The most lake-rich region in the boreal and arctic region is the Canadian Shield, which predominately is not affected by thermokarst the same way as the coastal tundra plains in Russia and Alaska. L40

Response:

The sentence has been changed to avoid confusion. Line 40: " especially in arctic lowland surface areas, where most of the water bodies could be thermokarst lakes (Jones et al., 2011; Olefeldt et al., 2016)."

L44 – I don't understand what is meant by "The shapes of the water bodies correlate to suitability of surrounding ecosystems." Please clarify.

Response:

The statement was to point out that the shape of water bodies could correspond to the cause of their formation. For example, Schilder et al. (2013) suggested that the wind speed was related to lake size and shape. Sharma et al. (2019) found that the shoreline complexity of lakes played an important role in governing ice cover. This sentence has been restated in this revision. Line 45: The morphology of the water bodies could be shaped by the surrounding environment.

L46 – The second part of this sentence after the semicolon is not clear in what it refers to: "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 (GliÅ" ska- Lewczuk, 2009).

Response: The sentence was edited to avoid confusion. Line 47: " 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), water self-purification and accelerates water cycling (Glińska - Lewczuk, 2009; Vaideliene & Michailov, 2008; Xiong et al., 2017)."

Increased lake connectivity improves water self-purification and accelerates water cycling (GliÅ" ska- Lewczuk, 2009).

L49 – What is meant by "water density"?

Response: It was changed to "density of water bodies" in this revision. Line 49: "density of water bodies"

L39-53. This paragraph is meandering – it starts out with a focus on lake thermokarst, and then it pivots to focus on the ecological influences from lake size and shape. The connection between these two topics is not clear.

Response: The reviewer was right. We split the paragraph into two paragraphs to be more inner consistent.

L69 – Why was the focus on North America, why not also include the Eurasian boreal and tundra regions to have a consistent high latitude freshwater dataset? Was the limitation computational?

Response:

As explained in the previous comment, the current dataset was only focused on the biomes in North America mainly due to limitations on computational and funding, and we also agree that covering the biomes in Eurasia would provide further value for research such as high latitude aquatic greenhouse gas emissions. The manuscript was edited to make the point (Line 413: The WBD-NAHL dataset focused on the tundra and boreal forest regions of North America. The methodology can be extended to Eurasia to provide a complete representation of the biomes.)

L77 – The Canadian Cordillera generally follow the border between Alberta and B.C, and then the border between the Yukon and the Northwester Territories. I can't see how the Canadian Cordillera can be described as "eastern mountains" when it is located in western Canada. I also do not understand what is referred to when it is said that the Canadian Cordillera separates the continent into east coastal plains and west plateaus.

Response:

We would like to thank the reviewer for pointing it out. We realized that it was confusing and mistakenly stated. We revised the sentence and also replaced "Canadian Cordillera" with "North American Cordillera" following the mountain list of Cordillera (https://en.wikipedia.org/wiki/American_Cordillera). Line 77: " The mountains of the North American Cordillera are covered by numerous mountain glaciers and also a large number of glacial lakes. A large number of thermokarst lakes were found in lowland

tundra areas, e.g., the Yukon Delta and the Alaska North Slope (Olefeldt et al., 2016). The vast Canadian Shield also has a high density of lakes."

L78 – The geographical description of the "eastern coastal plain … located near the Pacific Ocean" doesn't make sense to me – I don't understand what region is referred to.

Response:

We realized that it was mistakenly stated. The sentence was changed to "A large number of thermokarst lakes were found in lowland tundra areas, e.g., the Yukon Delta and the Alaska North Slope (Olefeldt et al., 2016)." at Line 78.

L80 – The climate description of the Canadian Shield doesn't make sense. The Canadian Shield stretches from the great lakes to the Arctic Ocean (north-south), so there is a very large difference in climate between different parts of the Canadian Shield.

Response:

The climate description was stated for the whole study region. As the reviewer pointed out, there is a very large difference in climate between different parts of the Canadian Shield, as well as in the study region. Thus, we removed the description about summer months and changed the sentences to " The climate of this study region is characterized by long, cold winters and short, cool summers." at Line 79-80.

L86 – I don't understand what regions are referred to when described as "east and north coast".

Response: As the reviewer's suggestion below, the lakes in the study region were not only formed by thaw or glaciation. Thus, the confusing sentence was removed in this version.

L82 – I think it is inaccurate to say that lakes dominate the landscape when you then state that they cover 36% of the land surface. The word "dominate" indicates >50% in my mind.

Response: Good point! We corrected this expression at Line 81-82: "Lakes are widely distributed in the study region and approximately 36% of the land surface is covered by water (Messager et al., 2016)."

L83 – It is not clear what landscape is referred to in this sentence, the full study domain or the Canadian Shield?

Response: It meant the whole study region. We clarified it in this version: "Lakes widely distributed in the study region."

L83 – The statement on 36% lake cover is not referenced. Where does this data come from?

Response: A citation was substantiated in this revision: " Lakes are widely distributed in the study region and approximately 36% of the land surface is covered by water (Messager et al., 2016)" at Line 81-82.

L84 – Now the lake area is 30%, which contradicts the stated 36% in the previous sentence?

Response: We corrected the misstatement at Line 82-84: " The number of lakes in this region accounts for 50% of the global lakes, and the area of lakes accounts for 30% of the global lakes in the region, indicating the region to be one of the richest areas of surface water bodies (Messager et al., 2016)."

L74-87. The geographical description of lakes in boreal and tundra biomes of North America is very poor in this paragraph. Please use better geographical names and descriptions to make a reader be able to follow which regions you are referring to. It is also very simplistic to only consider two types of lakes, glacial and thermokarst lakes, and it is not correct. Very important in this study area also organic lakes (not affected by permafrost), and there are also fluvial lakes, meteorite lakes, volcanogenic lakes, and anthropogenic lakes (dams) in the region.

Response: We restated the geographical description in the "Spatial extent" section and recreated Figure 1 to indicate key geographical names:

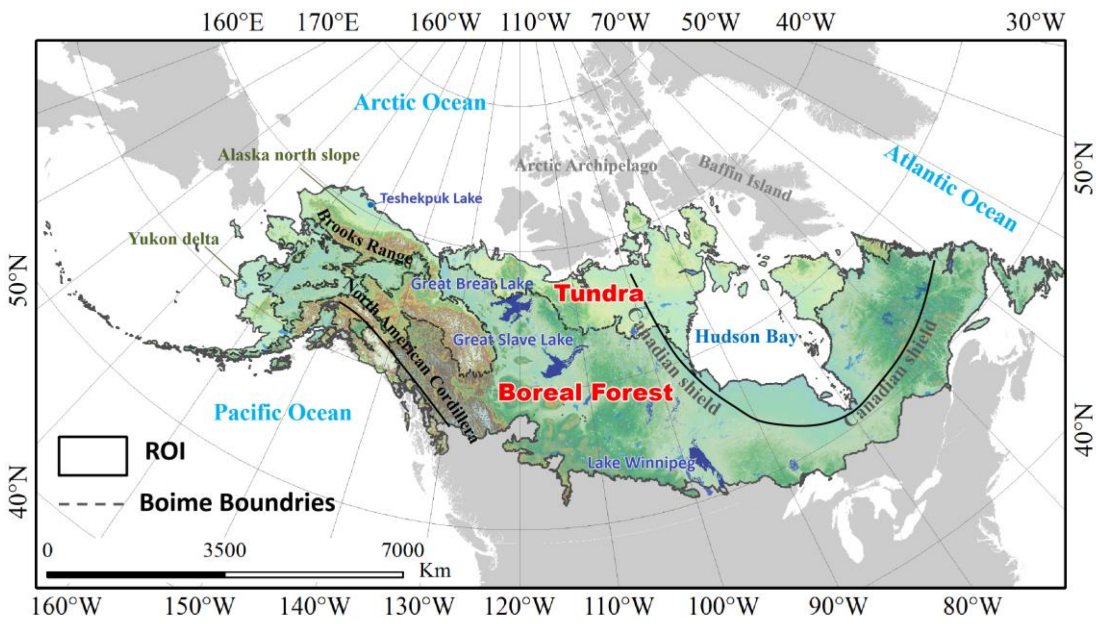


Figure 1: The extent of the study area including the tundra and boreal biomes in the North America continent excluding the Arctic Archipelago and Baffin Island.

We agree with the reviewer's comment that various types of lake were located in the region besides glacial and thermokarst lakes. We changed the sentences in this version. Line 84-86: Various types of lakes, including organic, fluvial, meteorite, volcanogenic, and anthropogenic lakes, are distribute in the study region and feature very different sizes and shapes (Dranga et al., 2017).

L105 – What does the abbreviation JRC stand for? Not explained. JRC

Response: We added the full name in this version: "3.2 Joint Research Centre (JRC) yearly water dataset"

Figure 2 – What is the scale of the images?

Response: We added the scales in the images.

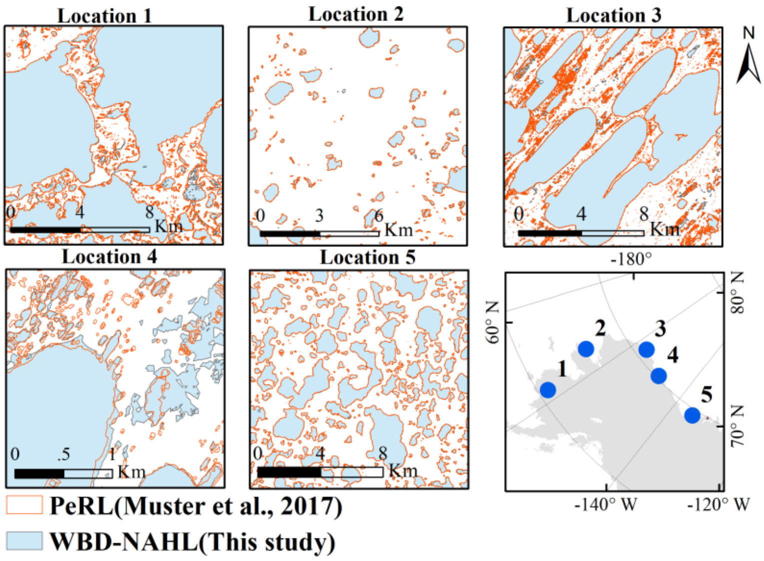


Figure 2: Water bodies identified in the WBD-NAHL (this study) and PeRL datasets (Muster et al., 2017), and the locations (blue dots) of the PeRL maps for the study region.

L122 – Please add the reference for the PeRL dataset.

Response: A reference for the PeRL dataset was added as suggested.

L128 – Perhaps include references in the figure legend – (This study) for SWBI and (Muster et al., 2017)

Response: A reference as included in the figure legend as suggested:

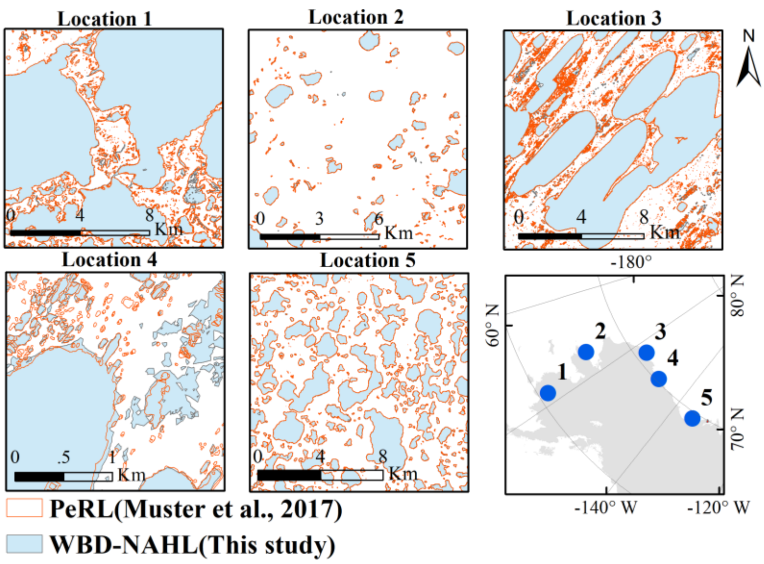


Figure 4 – You are using a different abbreviation here for your data product – WBI, should it not be SWBI?

Response: The name was corrected and further changed to WBD-NAHL along with the adoption of the new name for the dataset.

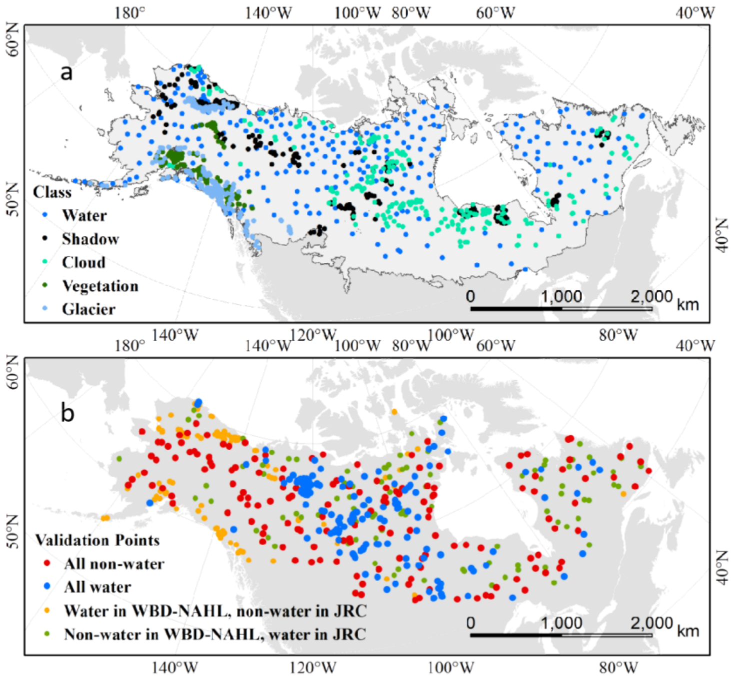


Figure 4: The training samples for random forest model building (a) and points identified for validating the accuracy of the detected water extent (b).

L179 – There are substantial areas of boreal forest at elevations > 1 km, especially in the Rocky Mountains and in interior B.C. and the Yukon. How was the 1 km cut-off chosen?

Response:

The empirical 1 km threshold was selected in this study to provide a stratum to ensure sufficient samples were collected to represent mountain shadows. The stratum was not used as a mask in the production of the dataset but used for collecting samples in favour of representing mountain shadows. Therefore, we adopted the empirical threshold as a general guideline for the stratum. To avoid confusion, the sentence was changed to: " The mountain shadow stratum was identified as any elevation higher than 1-km and slope greater than or equal to 3°. " at Line 192.

L204-206 – The exclusion of water bodies that were large and elongated; did you check how this exclusion worked on the Canadian Shield? Many lakes on the shield are extremely elongated and may here fall in the excluded category? Also to be noted, the distinction between lakes and rivers can be hard to define on the Canadian Shield as many of the very elongated "lakes" really are part of the watershed drainage network and convey water downstream.

Response:

We completely agree with the comment! Yes, it is extremely difficult to distinguish between lakes and rivers, as the reviewer pointed out. We adopted a relatively conservative rule for identifying and removing large rivers to avoid mistakenly excluding lakes. In this revision, we adopted an approach for further identifying rivers to respond to comments from other reviewers. However, these identified rivers were not removed from the dataset, but a field named "river" was added instead, and these identified water bodies were labeled in this new field. This column could be used as a reference for identifying rivers when it is needed.

We also updated the manuscript accordingly: Line 219-224: "The derived water body morphological metrics (i.e., the SI and area) and the HydroRIVERS were used to identify rivers and streams in the WBD-NAHL water bodies. Rivers and streams tend to have long, narrow, and linear shapes. We applied area thresholds > 5 km2 and SI > 10 in combination with visual examination to exclude large rivers and streams in WBD-NAHL. Considering the extreme difficulties in distinguishing small rivers and streams, water bodies that could possibly be rivers and streams were further identified by selecting long and linear water bodies (SI > 3) located close to the rivers and streams (< 100 m), as indicated by HydroRIVERS."

L250-253 – I think the interpretation that Tundra region has circular lakes formed by thaw, while the Boreal region has irregular-shaped lakes formed by glaciation is lacking. It is not Tundra vs Boreal biome that determines the shape of lakes, but is it much better described as being determined by the surficial geology, which is not tied to borders of biomes. The irregular-shaped lakes are found on the Canadian Shield, where the there is only thin surficial geology and most of the lakes are incised into the bedrock – and this can be found both in Tundra and Boreal biomes. Similarly, the more circular-shaped lakes are found in regions with thick overburden – either a result from being unglaciated (including aeolian deposits), or from being former sea bottom that has been rising through isostatic rebound, or by being located in regions with thick moraines or widespread peatlands. Especially I would point out the extensive peatland regions of the Hudson Bay lowlands and the Mackenzie river basin – where large boreal regions are found in SWBI to have circular-shaped lakes.

Response:

Thanks for the reviewer's constructive suggestion. Indeed, it is not biomes that determine the shape of lakes, and it would be more suitable to use surficial geology to explain the shape and size of lakes. However, for the extent of the entire study region, we still think that the comparison analysis of biomes could provide an informative insight into the water body distribution. Therefore, we keep the comparison but also expand the discussion to provide detailed elaboration of the relationship between the shape and the formation of water bodies.

We removed the interpretation that the tundra region has circular lakes formed by thaw, while the Boreal region has irregular-shaped lakes formed by glaciation at Line 268-269. In the 6.2 section, we added the limitation of the biome-based analysis. Line 390-395:

"The biome-based analysis provided insights into the distribution of water body shapes across the study area; additionally, it revealed complex relationships between the shapes and the surface geology of the water bodies. For example, circular-shaped lakes can be found in regions with thick overburden – possibly as a result from remaining unglaciated, from aeolian deposits or from rising from the sea bottom through isostatic rebound; These circular-shaped lakes can be found in regions with thick moraines or widespread peatlands in the boreal Hudson Bay lowlands and the Mackenzie River Basin. The high-resolution WBD-NAHL could help further explore the distribution of water bodies by size and shape."

L256 – This is a generalization which I don't think needs to be adjusted similar to the comment above. That is, lake size is likely more linked to surficial geology rather than terrestrial biome.

Also, the average lake size of the boreal biome is likely strongly influenced by the series of very large lakes that are found along the transition from the interior plains onto the Canadian shield – Great Bear Lake, Great Slave Lake, Lake Athabasca, Lake Winnipeg, Manitoba, Winnipegosis. Again – i.e. the location of these lakes is not linked to which terrestrial biome they are in but rather due to specific geological transitions. I would again also further emphasize that the vast boreal peatland regions seem to be very distinct in terms of lake size and shape when compared to the Canadian Shield.

Response: We agree with the comment. The confusing sentence was changed in this version. Line 271-273: "However, the water extent in the boreal forest (0.57 million km^2; 71% of total water area) is more than twice that found in the tundra (0.23 million km^2; 29% of the total water area), indicating that the average size of water bodies in the boreal area are larger than those in the tundra"

L339-340 – As noted above, round lakes is not only due to permafrost processes, but common for lakes in regions with thick surficial geology, which is not only restricted to the tundra biome.

Response: As suggested, we updated the description of lake size and shape in section 6.2. Line 376-377: The tundra, on the other hand, has a large number of small, regularly-shaped water bodies, which could be related to the thick peatland and thermokarst landscape....

L365-366 – I'd like to see a bit expanded discussion on the use of a single year to determine lake area. Especially for small lakes this is likely to be a source of uncertainty, especially for high latitudes where there often is both important seasonal trends in landscape inundation, as well as very pronounced interannual variabilities due to multi-year dry or wet conditions. Do you have data that can shed light on whether 2019 (and the fall and winter of 2018 that led into 2019) was a year with normal, dry or wet conditions for different parts of North America? That could give an indication on whether these estimates are likely to be higher or lower than the long term normal.

Response: As we responded to the previous comment, we expanded the discussion on the use of a single year to determine the lake area in this version. We analyzed the

precipitation in the past ten years and found that 2019 was not too dry or too wet in the study area. The discussion was presented in the section "Limitations". Line 406-410: "The WBD-NAHL dataset was produced based on Sentinel-2 data acquired in the summer of 2019 and represents the distribution of surface water in the corresponding year. The mean total precipitation in 2019 in the region was 438.5 mm, which was close to the historical average from 2010 to 2019 (mean: 435.9 mm, standard deviation: 11.5 mm) (Huffman et al. 2019). Although 2019 can be considered a normal year of the past decade in terms of precipitation, the spatial extent of high-latitude water bodies, especially smaller water bodies, can still vary significantly both inter- and intra-annually locally. Nevertheless, it would be interesting to explore water bodies' changes using observations from multiple years. "

Referee #2

This manuscript presents a data set of inland surface water bodies (SWBI) for the tundra and boreal forests of North America. The data inventory is generated by an automated approach of mapping the 10 m resolution Sentinel-2 multispectral satellite imagery. The resulting SWBI includes approximately 6.7 million water bodies > 0.001 km^2, of which there are 6 million (~90%) smaller than 0.1 km^2. The data set is also compared with other earlier regional or global water body products (e.g., JRC GSW, PeRL, GLWD, and HydroLAKES) and manually interpreted data. The data set, if with good quality, can provide finer-scale water body distribution information over the tundra and boreal forest regions of NA Arctic, which is critical for studying Arctic surface water bodies in response to changing climate and thawing permafrost. However, several fundamental problems related to the remote sensing mapping method and insufficient data quality check prevent the paper from being considered for publication.

Response: We appreciate the reviewer for the encouraging comment on the value of the dataset. Indeed, having a fine-scale water body dataset would be critical for studying the surface water in the pan-Arctic region in response to the changing climate and thawing permafrost. We also thank the reviewer for pointing out the issues in the method

description and data quality, particularly regarding rivers/streams in the identified water bodies. We have extensively revised the Methods section to clarify the process of water probability layer, water extent detection, and the final water body identification. A new procedure was introduced to further examine whether the water bodies are small rivers and streams. Details of the changes can be found in the response to the specific relative comments.

1. Mapping method

(1) The water body mapping method in section 4.1 is not described clearly. The logic of mapping procedures, including mapping updated water frequency, applying the machine learning model for water body identification, and deriving the final water body map is not described sufficiently, confusing the audience about the rationality of the methodology. **Response: The section has been extensively revised in response to the comment. We expanded the beginning paragraph to provide a general logic description of the method**

Line 131-136: The 10-m resolution Sentinel-2 A/B multispectral data are the primary source used to identify small water bodies. An approach was developed to produce a water probability layer for 2019 by combining the water-sensitive indexes derived from the Sentinel-2 bands and the 30-m resolution JRC water dataset (section 4.1). A machine learning model was trained to retrieve water extent from the Sentinel-2 images from possible water extent restricted by the water probability layer (section 4.2) (Figure 3). Water bodies were finally identified from the water extent using an object-based algorithm to produce the final water body inventory (see section 4.3).

The original section of water detection was divided into three sections to provide a clear structure for describing the key steps. In addition to the general logic description of the whole method, we also provided sentences at the beginning of each of these three new sections to provide a further logical description of each key step.

Line 140-143: A water probability layer was derived to represent the likelihood of a pixel to correspond to permanent water during the summer of 2019. The 10-m resolution water-sensitive indexes calculated from the Sentinel-2 multispectral bands were used as the main

input. The other reference water dataset (e.g., the JRC water dataset) was adopted as a supplemental input and fused with the main input to produce the water probability estimate at each 10-m resolution pixel.

Line 177-180: Although the possible water extent estimated the likelihood of a pixel to correspond to water, confusion with shadow, ice, or cloud contamination in area with complex environments is still possible due to the limitations of water indexes with similar spectra (Isikdogan et al., 2017). A random forest model was trained with points collected through visual interpretations to further detect water within the areas indicated as possible water.

Line 209-211: Permanent water pixels were identified from the resulting water frequency layer (*f*) 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.

Furthermore, Figure 3 was updated to provide a clearer description of the method in a flowchart.

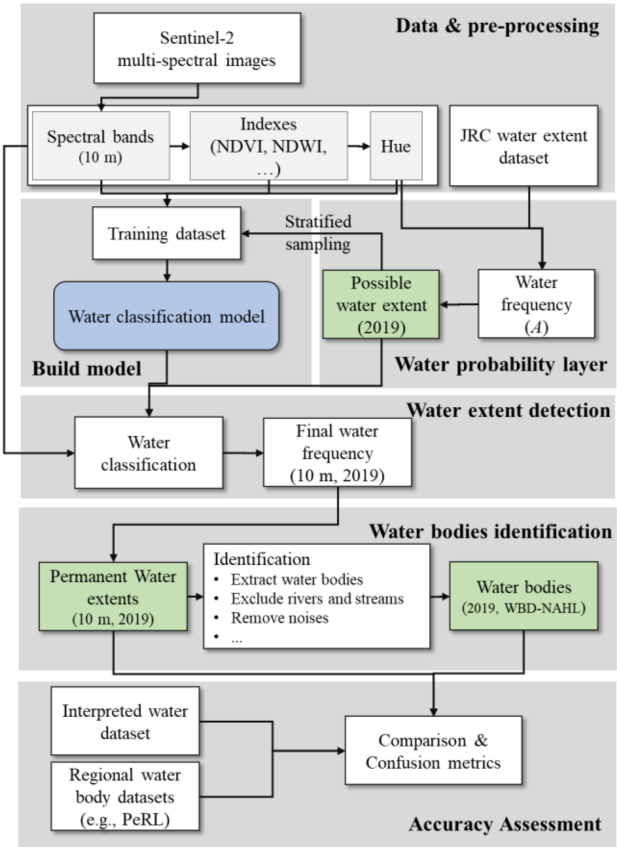


Figure 3: Flowchart for processing water extent and identifying water bodies.

(2) The mapping method integrates the Sentinel-2 derived water frequency and JRC water dataset. However, the JRC dataset generated from Landsat imagery has a spatial resolution of 30m, which is inconsistent with 10-m Sentinel-2 data. How to create a 10-m SWBI data set with the water body size as small as 0.001 km^2, about one pixel for Landsat imagery? Another water frequency data product based on Landsat imagery, the GLAD group's Global surface water dynamics 1999-2020 (https://geog.umd.edu/news/new-global-surface-water-dynamics-maps-published-remote-sensing-environment), may be also considered for the integrated mapping.

Response: The JRC dataset was resampled from 30 m to 10 m to match the resolution of the Sentinel-2 derived dataset. We believe that the reviewer's concern was valid; however, the impact of the coarser resolution of the JRC to the final dataset was minor because of the following: (1) the JRC dataset was used as a complementary information to the Sentinel-2 derived dataset as it was assigned a lower weight in the combination. If a small water body was missed by the JRC, it could still be captured by the Sentinel-2 dataset; (2)

The possible water extent that provided from the combination of the Sentinel-2 dataset and the JRC was only used as a reference, and the final water detection was carried out from only the 10-m resolution Sentinel-2 dataset using a machine learning model.

In response to the comment, we have revised the manuscript to clarify the role of the JRC in producing the possible water extent as well as the role of the possible water extent in the final water extent detection.

We thank the reviewer for the suggestion of the GLAD dataset, which is another excellent water dataset. However, we did not adopt the dataset for this research because of the following: (1) the JRC dataset provided acceptable quality for the region at similar spatial resolution as the GLAD; (2) our research started at the end of 2019 before the publication of the GLAD article in RSE. We are confident that the GLAD dataset would be an excellent reference for our later studies and **we added the citation of GLAD in the Introduce section.**

(3) The mapping algorithm (Eq. 4) adopts a weighted linear combination. The weights and water frequency thresholds seem to be determined arbitrarily. In addition, whether the training set size (1250 points) is sufficient to establish the machine learning models for mapping approximately 6.7 million water bodies requires more evidence.

Response: The weight and thresholds of water frequency were based on our multiple experiments on water detection in different conditions. These threshold values were intended to be loose to ensure that all water bodies were included in the possible water extent layer, which then provide the area restriction for the machine learning based water detection. As responded to the previous comments, we have revised the Methods section to clarify the general logic of the method, particularly regarding the role of the possible water extent, which is result that these weights and thresholds applied to.

It's an excellent point! We realized that we failed to make it clear on the logic of the water detection method. The final water extent was detected from the combined effects from the possible water extent and the machine learning model, which was trained from the 1250 points. We could admit that 1250 points cannot be considered as a large population for machine learning. However, the model was built to detect water only in the area that

already been identified to be possible water, and the variation of conditions have been significantly reduced in comparison to a regular land cover classification; Moreover, the points were collected from five-strata to increase the performance of the sampling; Last, these points were visually interpreted to ensure the high quality of the reference data for training the water detection model. Therefore, we believe that the 1250 points could still provide a sufficient training data to the machine learning model, and the quality assessment confirmed the performance of the model in section 4.4.

2. Data quality control

The reviewer roughly went over the SWBI dataset. Many mapping errors exist in the data product, e.g., mixing with ocean water areas near the coastline, remaining river segments, the rough and tumble polylines, etc. The flowchart (Figure 3) suggests that an identification procedure was conducted, excluding rivers/streams and removing noises. However, many river segments and coastal waters still exist in the data set. How to separate the multiple lakes linked by river/stream? This will influence the number and morphologic calculation of water bodies in the data product. I suggest severe quality assurance before publishing the data set.

Response:

The suggestion is valuable for improving the quality of the presenting dataset. As the other reviewer also pointed out, the distinction between lakes and rivers can be hard to define, particularly in the Canadian Shield where exists large number of very elongated "lakes" really are part of the watershed drainage network and convey water downstream. In this revision, we have applied further procedures to examine and address the issues regarding the issues as the reviewer pointed out. The large rivers and the ocean water areas that remained in the last version were removed manually. For the small segments of rivers, we carefully examined and adopted an approach for further identifying rivers to respond to comments from other reviewers, and 177,964 water bodies were identified. However, due to the previously mentioned difficulty, there is a risk to falsely exclude very elongated lakes, and we were cautious on deleting them. Although these water bodies were excluded from the analysis in the revised manuscript, they were kept in the dataset, but

labelled in a new field named "river", which could be used as a reference by users on applying their own water body type identification.

The manuscript was updated to introduce the method at Line 219-224: " The derived water body morphological metrics (i.e., the *SI* and area) and the HydroRIVERS were used to identify rivers and streams in the WBD-NAHL water bodies. Rivers and streams tend to have long, narrow, and linear shapes. We applied area thresholds > 5 km^2 and SI > 10 in combination with visual examination to exclude large rivers and streams in WBD-NAHL. Considering the extreme difficulties in distinguishing small rivers and streams, water bodies that could possibly be rivers and streams were further identified by selecting long and linear water bodies (*SI* > 3) located close to the rivers and streams (< 100 m), as indicated by HydroRIVERS." Those water bodies (177,964) were not removed but marked in the attribute table (field "river") to provide an extra reference to the users."

3. Data validation

In this study, the data quality assessment includes two aspects, the comparison with other data products, e.g., JRC GSW data, HydroLAKES, and GLWD, and the validation by manual interpreted data. However, the earlier data products have inconsistent water body definitions with the SWBI. For example, the raster-format JRC GSW contains all water bodies, e.g., lake, river/stream, fish ponds, etc., while the HydroLAKES and GLWD only include lakes (and reservoirs). It is therefore not valid to attribute the differences to mapping quality.

Response:

Indeed, the water body definitions were inconsistent among these datasets. However, these comparisons with different products offered a comprehensive evaluation of our dataset. The coarser products, e.g., JRC GSW data, HydroLAKES, and GLWD provided references for evaluating WBD-NAHL in the extent of the entire region. Although the HydroLAKES only include lakes and reservoirs, they could be comparable with the large water bodies identified in WBD-NAHL. To compare with HydroLAKES and GLWD, we only selected the water bodies larger than 0.1 km^2 for analysis of accuracy to avoid the interference of small river segments that remained in this dataset. Moreover, the

interpreted data and PeRL derived from high-resolution images provided a reference for assessing the morphological parameters of water bodies derived by WBD-NAHL.

Other comments:

Line18: This data set does not include water bodies in Eurasia Arctic. How to say it is a more complete representation than the PeRL data?

Response: In North America, the data set in this study is a more complete representation than the PeRL data. The confusing sentence was changed in this version. Line 18: The dataset provided a more complete representation of the region than existing regional datasets for North America, e.g., Permafrost Region Pond and Lake (PeRL)

Line19: I would like to recommend the public share and open access of the manually interpreted data, which can be helpful for the quality assessment of the future published water body data sets.

Response: In this version, we uploaded both the validation data to help for the quality assessment of the future published water body data sets.

Line82: "Lakes and ponds" used here have different means?

Response: We corrected "Lakes and ponds" as "lakes" at Line 82: " Lakes are widely distributed in the study region and approximately 36% of the land surface is covered by water (Messager et al., 2016)."

Line83: "…about 50% of the lakes and 30% of lakes by area…" should be "…about 50% of the lakes by count and 30% of lakes by area…"?

Response: We corrected this misrepresentation as " The number of lakes in this region accounts for 50% of the global lakes, and the area of lakes accounts for 30% of the global lakes in the region, indicating the region to be one of the richest areas of surface water bodies (Messager et al., 2016)" at Line 83.

Line97-98: For the Sentinel-2 sensors, there are three bands at the 60-m resolution among the total of 12 bands.

Response: We changed the sentence at Line 96: Each Sentinel-2 image consists of 13 multispectral bands, including four bands at 10-m resolution, six bands at 20-m resolution, and three others at 60-m resolution.

Line131: The MNDWI calculation requires the input of SWIR band, with a resolution of 20m for Sentinel-2 imagery. How to process the different resolutions for different spectral bands?

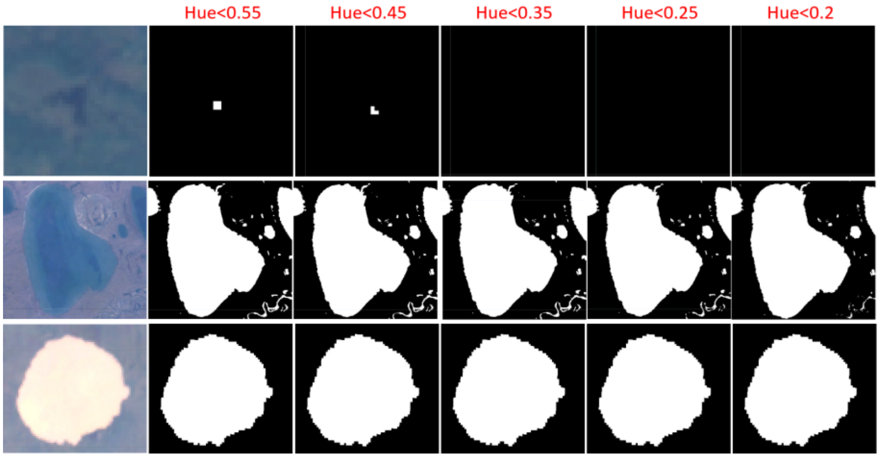
Response: The SWIR band was pan-sharpened to 10 m using the À Trous Wavelet Transform (ATWT) algorithm as recommended by Du et al., (2016). We clarified the resolution and the pan-sharpening method in the revision. Line 156-157: These bands have 10-m resolution except B_{swir}, which has 20-m resolution and was pan-sharpened using the À Trous Wavelet Transform (ATWT) algorithm as recommended by Du et al., (2016).

Line141-143: The references for the NDWI (McFeeters, 1996) and MNDWI (Xu, 2006) were mistakenly cited.

Response: We corrected the citation at Line 149-151: "Normalized-Difference Water Index (NDWI) (McFeeters, 1996), Normalized Difference Vegetation Index (NDVI) (Carlson and Ripley, 1997), and Modified Normalized-Difference Water Index (MNDWI) (Xu, 2006). The three indexes were calculated as follows."

Line155: how to determine the threshold of "hue <0.45" for extracting water pixels? Please test the threshold sensitivity for water bodies under different conditions and images.

Response: The threshold of "hue <0.45" was determined on our multiple trials. The threshold sensitivity for water bodies was presented as follows:



The hue was set loosely to guarantee that the water extent was included in the possible water map. Because the hue threshold was not used to generate the final water map, the threshold sensitivity analysis was not mentioned in this manuscript but explained as an experiment threshold: " Once the hue has been identified, an experimental threshold of < 0.45 was applied to identify the water pixels." at Line 164.

Line161: A is the updated water frequency. What is the final mapping result of water body data set?

Response: *A* is the water frequency or water score, ranging from 0~1, presenting the possibility of being water in the target year. The possible water extent was generated by extracting the pixel where *A* is greater than 0.5. We further clarify this at Line 175: The final, combined possible water extent was identified when *A* > 0.5.

Line162: how to derive the Sentinel-2-derived water frequency (As) here? By the machine learning model as introduced in the following part?

Response: The Sentinel-2-derived water frequency (A_s) was the rate of being identified as water by the satellite images acquired for the same location. In this version, we clarified the generation method of A_s in this version. The Sentinel-2-derived water frequency (A_s) was derived by the threshold method. Line 164: Once the hue has been identified, an experimental threshold of < 0.45 was applied to identify the water pixels. The same procedure was applied to derive temporal water extents from all selected Sentinel-2

images. All the water extents were then combined to calculate the water frequency (A_s) for the year.

Line163: the threshold setting by combining water frequency and elevation looks a little weird. What is the rationality for doing this?

Response: The hue derived from the Sentinel-2 images was found to be more heavily affected by the snow and ice on mountains, particularly on the north slope. JRC did provide a better result on excluding snow and ice in high elevations. A higher weight was assigned to the JRC to reduce the effect of snow and ice in the mountains. An explanation was added in this revision for clarification at Line 171-174: *A* is the updated water frequency, W_s is the weight for the Sentinel-2-derived water frequency (A_s) and set to 0.85 to ensure that the 10-m measurements were the main input for the final water probability estimate. However, W_s was decreased to 0.65 in high elevations pixels (elevation > 1 km) to reduce the effect of snow and ice on the Sentinel-2-derived hue over mountains.

Line167: "(Figure 4a)" ---the wrong inserting place.

Response: We corrected this mistake: " Then, 250 points were randomly selected in each stratum, for a total of 1,250 points (Figure 4a)."

Line190-191: the terrain shadows have few influences on the water body mapping for the areas with an elevation below 1500 m?

Response: The low altitudes are usually flat and away from the mountains. Applying the slope-based shadow mask could falsely exclude true waters due to errors and outdated presentation of the terrain in the DEM data. Moreover, the minimum 0.5 water frequency filter could remove locations affected by temporary snow, ice, and shadow. Therefore, the terrain shadow filter was only applied to high elevations (>1 km) where the mountains are distributed.

Line197-200: As shown in the data set, the polygons of mapped water bodies were generalized by GIS tool. Does the simplification tolerance affect the calculation of geometry metrics of polygons?

Response: The threshold for the simplification was chose according to the minimum size of water bodies recognizable by the results to avoid over-simplifying the results. The results were also evaluated by comparing to morphological metrics of the polygons to these derived by the high-resolution images, as presented in Figure 2.

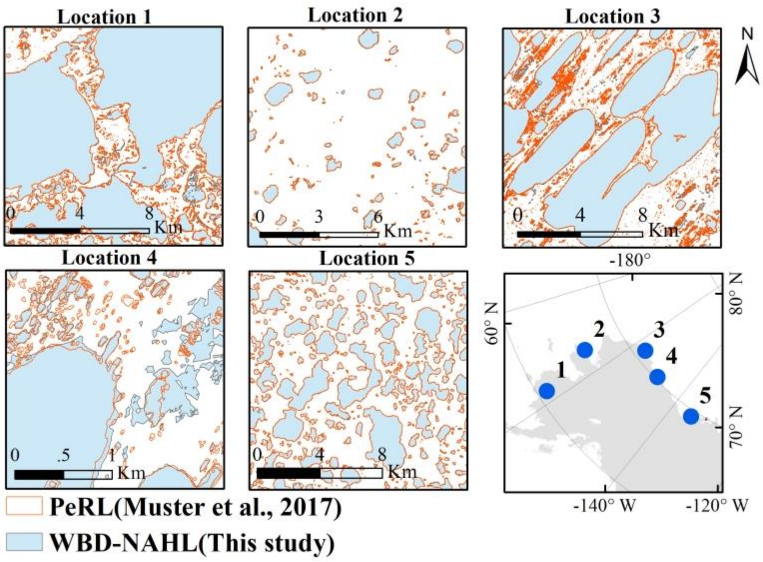


Figure 2

Line243-246: The analyses of water body abundance (the power-law statistics), the counts of different lake size levels for the SWBI, should be added.

Response: We appreciate the valuable suggestion! An extra analysis in this revision was added to examine the difference between the lake numbers observed in our data and these estimated from the power-law distribution of lake area. The analysis revealed that the relationship significantly overestimated the number of small lakes. The following sentences were added (Line 349-359: An empirical power-law distribution was found between lake areas and lake numbers (Messager et al., 2016; Downing et al., 2006), and the distribution was applied to estimate the number of small lakes, which were used for estimating greenhouse gas emissions (Holgersson et al., 2016). According to the power-law distribution and HydroLAKES, the number of water bodies larger than 0.1 km^2 was estimated to be about 798,895, which was close to the 629,130 water bodies reported by WBD-NAHL (Figure 11). However, the number of water bodies sized between 0.1 and 0.01 km2 was estimated to be about 10.2 million, 4.8 times higher than estimated by WBD-NAHL. Furthermore, the water bodies sized between 0.01-0.001 km^2 were estimated to be about 126.1 million, 33.6 times higher than what was estimated by WBD-NAHL, suggesting that the power-law distribution significantly overestimates the number of small lakes. A similar finding was reported by Seekell et al. (2016). Estimating the number small water bodies using a power-law distribution could introduce considerable uncertainties in the estimation of the contribution of small water bodies to greenhouse gas emissions. Accurately identifying small water bodies could correct this overestimation and improve greenhouse gas emission estimates (Holgersson et al., 2016).

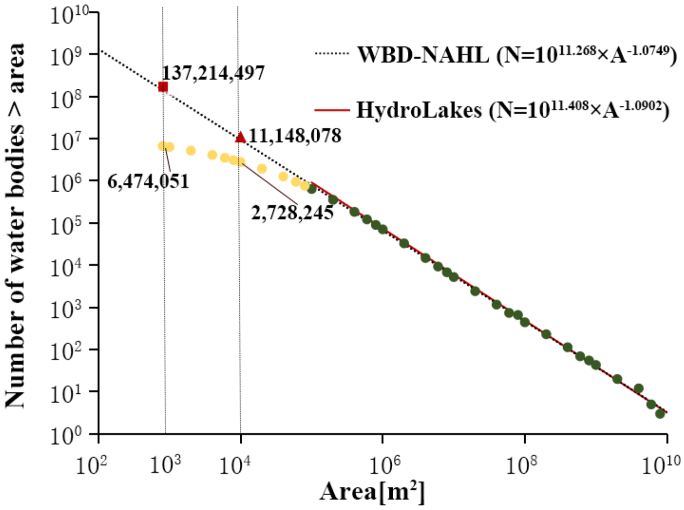


Figure 11: Distribution of the total numbers of water bodies in relation to the area of North American tundra and boreal forests water bodies. The circles represent the number of water bodies provided by WBD-NAHL. The black line is the power-law distribution modeled using water bodies > 0.1 km^2 from WBD-NAHL. The red line is the power-law distribution modeled using HydroLAKES in the study region. The red triangle and square represent, respectively, the extrapolated number of water bodies > 0.01 km^2 and > 0.001 km^2 based on the power-law distribution modeled from HydroLAKES."

Besides, the counts of different lake size levels for our dataset were added in the Figure 6:

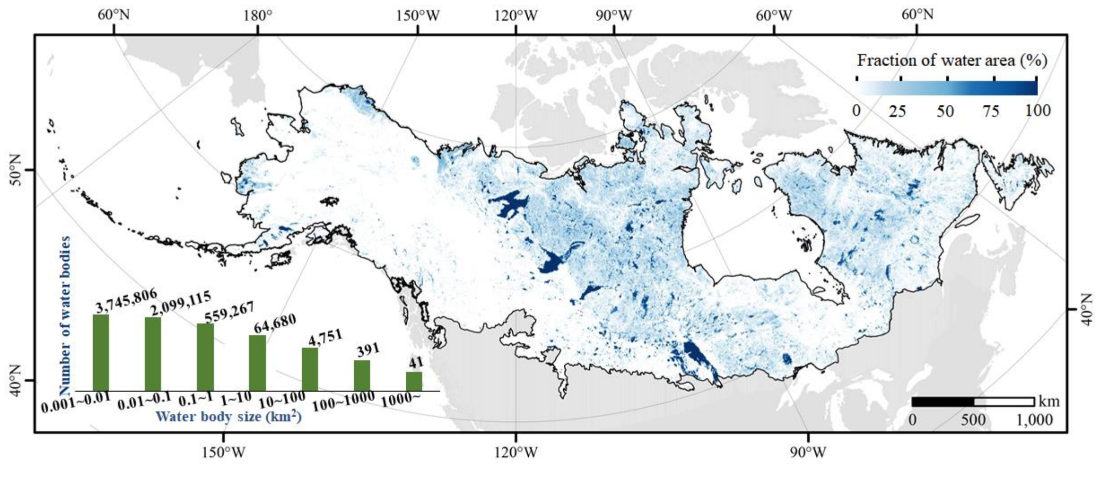


Figure 6: Percent of surface water (5 km × 5 km grid) produced by aggregating the water extent for the tundra and boreal forests of North America as calculated using the WBD-NAHL.

Figure 7a: the 5 km × 5 km grid can contain the water area >500 km2???

Response: We thank the reviewer for pointing it out. The values at each pixel in the grid were calculated by selecting the water bodies' interest with the extent of the pixel and then counted or calculated the mean of the targeted parameter (e.g., area, *SI*, and perimeter) from these selected water bodies. We updated the caption of Figure 7 to clarify the calculation: The aggregated distribution of area (a), perimeter (b), and *SI* (c), and the number (d) of the identified water bodies in the study area. The values at each 5 km x 5 km pixel in the grid were calculated by selecting the water bodies' interest with the extent of the pixel and then counted or calculated the mean of the targeted parameter (e.g., area, *SI*, and perimeter) from these selected water bodies.

Line262-263: please indicate the sub-title for maps (a, b, c, and d).

Response: Revised as suggested.

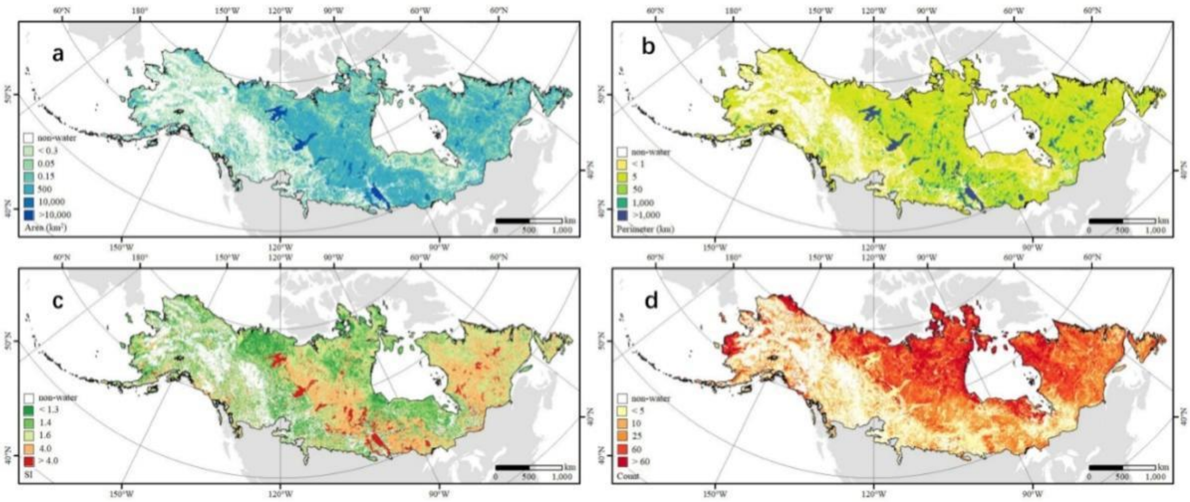


Figure 7: The aggregated distribution of area (a), perimeter (b), and *SI* (c), and the number (d) of the identified water bodies in the study area. The values at each 5 km x 5 km pixel in the grid were calculated by selecting the water bodies' interest with the extent of the pixel and then counted or calculated the mean of the targeted parameter (e.g., area, *SI*, and perimeter) from these selected water bodies.

Line273-280: why are the mapped water body areas mostly smaller than the manually interpreted data?

Response: Our investigation found out that the smaller water body area was mainly caused by the conservation of identifying water over mixed pixels located at the edge of the water bodies. As indicated in the comparison in Figure 8, our 10 m resolution WBD-NAHL dataset provided closer area estimates to these from the 3 m interpretation than the 30 m resolution JRC, suggesting that the higher resolution of this dataset provided higher accuracy area estimates than coarser resolution water datasets. In this version, we stressed this limitation at Line 291: Both the JRC and WBD-NAHL datasets accurately identified the size of larger water bodies. For mixed water pixels, the area estimates of both datasets were more conservative than the reference data.

Line315-316: the number of SWBI water bodies for the size levels (100~1000, 10~100, 1~10, 0.1~1) are all slightly smaller than that of HydroLAKES, why? for lake changes (disappearances) during different mapping periods, or mapping uncertainty?

Response: Despite the very small difference between SWBI and HydroLAKES, our investigation found that the different water was likely due to the cause that HydroLAKES identified a few clustered small water bodies (<0.1 km^2) as one big water body (> 0.1 km^2); hence, these water bodies were occurred in HydroLAKES but were excluded from the SWBI because smaller than the 0.1 km^2 threshold. A few examples are presented in the figure below: