

Earth Syst. Sci. Data Discuss., referee comment RC1  
<https://doi.org/10.5194/essd-2021-278-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2021-278**

Anonymous Referee #1

---

Referee comment on "A high-resolution inland surface water body dataset for the tundra and boreal forests of North America" by Yijie Sui et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-278-RC1>, 2022

---

Review of "A high-resolution inland surface water body dataset for the tundra and boreal forests of North America" by Sui et al.

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 km<sup>2</sup> 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.

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.

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.

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?

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).

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?

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.

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

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 (Glińska-Lewczuk, 2009).

L49 – What is meant by “water density”?

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.

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?

L77 – The Canadian Cordillera generally follow the border between Alberta and B.C, and then the border between the Yukon and the Northwest 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.

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.

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.

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.

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

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

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

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

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.

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

Figure 2 – What is the scale of the images?

L122 – Please add the reference for the PeRL dataset.

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

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

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?

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.

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 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 seabottom 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.

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.

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.

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.