

General Comments:

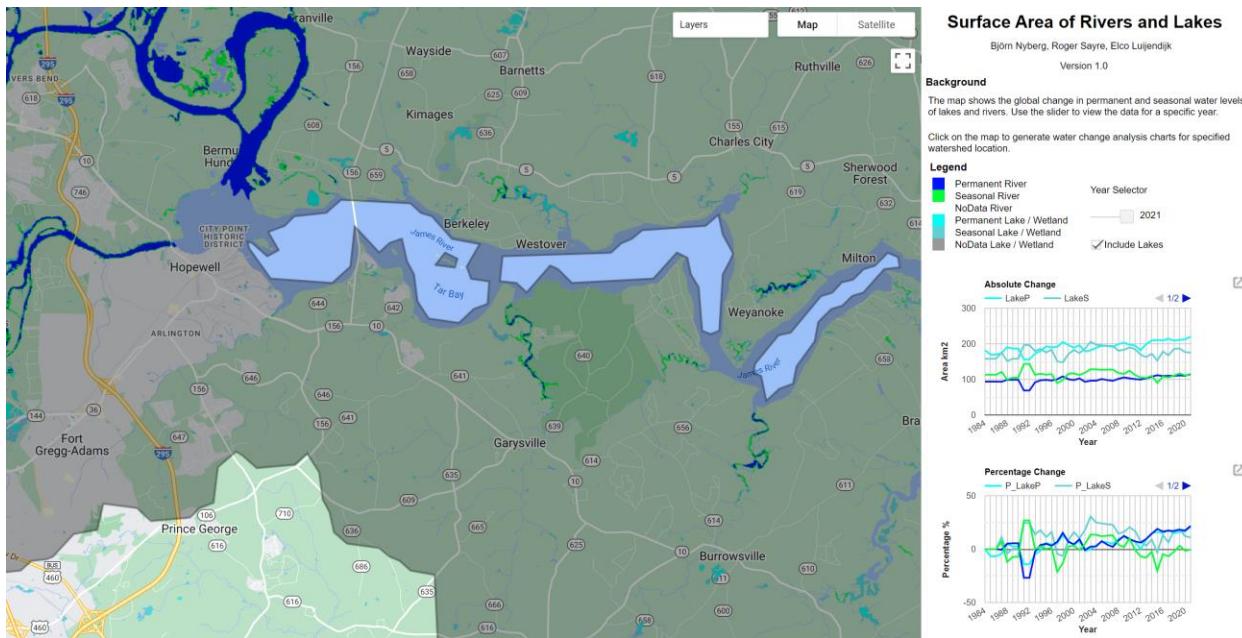
Surface water area plays an important role in ecosystems, the carbon cycle, flood and drought risks, and water resource management. Global surface water datasets built on moderate resolution satellite imagery (e.g., Landsat with 30-m, 16-day resolutions) have been developed, but they do not differentiate types of surface water bodies. The authors created a new dataset of the surface area of rivers and lakes (SARL) with seasonal and permanent surface water data for each year from 1984-2022. They developed this new dataset by combining data from the Global Surface Water dataset (Pekel et al., 2016) and from the global extent of river channel belts dataset (Nyberg et al., 2023). They were also able to assess changes from 1984 to 2022 in permanent and seasonal surface water areas in watersheds across the globe. They found seasonal surface water area had increased for both rivers and lakes, while the global total permanent surface water area was relatively unchanged. This dataset, showing where and when there are changes in lake and river permanent and seasonal surface water areas, can be helpful for water resource management. Such management practices are used to help limit the adverse impacts of extreme events such as floods and droughts, which are becoming more common due to climate change. The SARL dataset can help identify changes in the seasonality or permanence of surface water areas of key stormwater or drinking water reservoirs and help inform decision making about these water bodies in the future.

Specific Comments:

Figure 1 B: i) The caption suggest image B is showing the change in permanent water extent from 1984 to 2020; however, the legend in the image suggests it is the permanent and seasonal extents for one year, i.e., there is no legend entry to indicate a pixel has changed classification. Please change either the figure or the caption to clarify what you are showing. ii) it is hard to distinguish the Permanent Lake from the Seasonal Lake. Consider using a different color for one of them. iii) There is already green in the Landsat image, so it is hard to distinguish the green for Seasonal River. Consider showing the Landsat image using false color or changing the Seasonal River color to something easier to identify.

Line 111-113: How did you use the HydroLakes and OpenStreetMap datasets? I'm guessing you removed GCB pixels outside of the HydroLakes and OpenStreetMap bounds. Also, I'm assuming the data product of GCB is pixel-based rather than shapefiles, but I'm not clear on that.

Also, have you looked at what areas are being removed from these bounds? By looking at your GEE app, I found an example of portions of the James River outside of Richmond, Virginia in the United States being excluded from the watershed. The exclusion of such large areas of what I would assume is permanent river area may be impacting your results. I think a discussion of this limitation would be valuable to add.



Line 124: The GSW database does not identify waterbody type and the missing/no data values can be non-water, as well as permanent or seasonal surface water. How does your method account for this possibility?

Line 125-126: Because you are suggesting that seasonal water surface area has increased between 1984 and 2022, why do you think using the average ratio of seasonal to permanent waterbody extents between 2015 and 2017 is a good ratio to use across all of the years? (aside from those years lack of missing values.) I would suggest moving the sentence from 145-147 up to this paragraph.

Line 135-136/Equation 1: Please clarify the units of the values in the equations. It reads like pArea, pO, sArea, and sO are all measures of area and therefore the units are something like ha or m². However, the nD appears to be a count of no data values (or pixels) in the catchment. I expect you are not adding area and pixel counts together, to find pArea and sArea, but please clarify.

Line 171-172: because you are using the same surface water dataset to generate the validation dataset as you used to generate the SARL dataset (GSW), this validation assessment seems to only be validating the GCB dataset. If you want to validate the SARL dataset, you might want to consider classifying the Landsat 8 images you used to manually delineate channel belts into permanent, seasonal and non-water classes; use the manually delineate channel belts to classify the permanent and seasonal water into riverine and lacustrine; and then compare these validation data to the SARL data.

Line 169-176: This is the only paragraph where you refer to the SARL dataset as being generated through an “automated method”. What I assume this means, is that the GCB data used to create the SARL dataset is produced using an automated method, as compared to the manual delineation of the channel belt used to generate the validation dataset. The final sentence of the paragraph supports my assumption that this validation analysis is focused on validating how well the GCB dataset performs, rather than validating the SARL dataset. This is fine, but if that is the goal, consider restructuring the aim of the paper from “to compile existing information to map the historical change in water surface area...” to “to assess the utility of a new, global river extent dataset (Nyberg et al., 2023) in mapping the historical change in water surface area...”

Line 222: What are the “independent estimates of permanent and seasonal lake coverage” to which you are comparing the SARN dataset?

Technical Corrections:

Line 43: typo -- “seasonall” to “seasonal”

Line 46: “their biodiversity” whose biodiversity? The biodiversity of human or ecosystems or something else?

Line 46: use of an Oxford comma where in the previous sentence one is not used. Recommend picking one (use it or don’t use it) and make it consistent throughout the manuscript

Line 63: typo/subject-verb agreement: “provides” to “provide”

Line 78: If you are getting the GSW data from GEE, 1) which GSW data product are you using? I assume you are using the JRC Yearly Water Classification History, v1.4 (“JRC/GSW1_4/YearlyHistory”). 2) if that is the case, that product only runs through 2022, not 2023. Please include which version of the GSW product you are using and how you accessed it.

Line 91: did you evaluate the model in this current study, or is the evaluation from Nyberg et al., 2023? If the latter, consider rephrasing to “The model reported … (Nyberg et al., 2023).” or “The authors reported the model had a 94% accuracy …”

Line 93-94: consider rephrasing “surface area of lakes and rivers (SARN)” to “surface area of and rivers lakes (SARN)”

Line 94: is the SARN dataset from 1984-2022 (as in the title) or 2023?

Line 96: consider rephrasing the last sentence of this paragraph. It sounds like the beginning and end of the sentence are saying the same thing. You could probably end the sentence after “wetland regions”, unless I’m not understanding what you’re trying to say.

Figure 1/Line 100: is the GCB a model or a dataset produced by a model?

Line 106: rephrase for methods clarity. From “classification of Nyberg et al., (2023) by utilizing a 10% confidence on the GCB prediction… and a 50% confidence…” to “classification of Nyberg et al. (2023) by using GCB pixels with a reported confidence of 10% or higher… and a 50% or higher confidence…” or “classification of Nyberg et al. (2023) by utilizing GCB pixels with a $\geq 10\%$ confidence or higher… and a $\geq 50\%$ confidence…”.

Line 111-113: use of passive voice (“datasets were included” and “these steps were processed”) after using active voice in the preceding paragraphs. I recommend you pick one and stick with it throughout the manuscript.

Section 2.1: I suggest a slight restructuring of this section to clarify the steps taken to generate the SARN dataset. Currently, it reads as:

- a) Describe GSW (Pekel et al., 2016)
- b) Describe GCB (Nyberg et al., 2023)
 - a. Discuss combining GSW and GCB
- c) Describe how GCB was subset using confidence thresholds and other lake/reservoir datasets

I suggest reordering it so it flows more like:

- a) Describe GSW (Pekel et al., 2016)
- b) Describe GCB (Nyberg et al., 2023)
- c) Describe how GCB was subset using confidence thresholds and other lake/reservoir datasets
- d) Describe how GSW and the subset GCB were combined to create SARL

Line 153: How many HydroSHEDS catchments are there globally versus how many did you all use in your analyses? Phrased another way, how many catchments are you removing with the qualifications stated in this sentence?

Line 174: You suggest that you're creating this validation dataset using GSW data from 1984-2023, but you are only using the data from 2022 for validation (see lines 160-161). Consider rephrasing this sentence to clarify how much of the GSW data you are using for validation.

Line 224-225/Figure 4: Figure 4 shows Permanent River accuracy is between 94-98% and the second highest accuracy range, but your sentence says it has the lowest accuracy range (84-91%). Please check which is correct and update.

Line 262: suggest removing “e.g.”

Line 273: I'm unclear on how area and mean annual water discharge are linked here. Consider rephrasing for clarity.

Line 282: suggest changing “entails” to “indicates”

Section 4.3: This section is titled “Water Resource Management” but I am not seeing much discussion about how the SARL dataset will be used for water resource management. The paragraphs have lots of good information, but it would be nice to have the connection to the SARL dataset spelled out a little more explicitly.

Line 310: change “river” to “rivers”

Line 318: Change “show” to “shows”; delete “the” before “seasonal extent”

Line 319: Change “is attributed” to “correlates”

Line 321: Add comma between “significant” and “increase”

Line 325: Add comma between “well” and “which”

Line 338: Are you talking about catchments with any kind of statistically significant change (i.e., permanent lake or seasonal river) or are you talking about catchments with a specific kind of statistically significant change? Please clarify.

Minor rephrasing suggestions

These comments are just suggestions of rephrasing for clarity and do not need to be addressed.

Line 80: consider rephrasing “...with at least one month during which water was detected” to “where water was detected for at least one month” (optional, it’s okay as-is but rephrasing might make it a little clearer)

Line 85-87: consider rephrasing “Nyberg et al. (2023) quantified...” to “We used a dataset that quantified the global extent of river channel belts (GCBs) ... produced by Nyberg et al. (2023). In this dataset, the river channel belt...”

Line 89: consider rephrasing “The model reports” to “The model used to produce this dataset reports...” or “The model used by Nyberg et al. (2023) reports...”

Line 159: suggest rephrasing “available water surface change data” to “available surface water data”

Line 169: suggest removing the term “proposed”. You have utilized the method, so I do not think you are proposing it. I would rephrase the sentence to “We used the same method to classify the validation dataset as we used to classify the SARL dataset (Figure 1), where waterbodies defined...”

Line 291: I suggest rephrasing “The decreasing permanent extent of many lakes and rivers (Figure 2) suggests that...” to “Because we found many (XX and XX%) lakes and rivers have decreasing permanent extents (Figure 2), our results suggest that...” Right now, the phrasing sounds like the majority of the lakes and rivers are decreasing in permanent area. Whereas in Figure 3, it looks like there is a somewhat even split between statistically significant increasing and decreasing permanent water area for lakes and rivers, with somewhat more showing increases.

Line 308-309: Same idea as above; the use of the term “many” seems to imply that at least close to the majority if not the majority of permanent water in lakes and rivers is decreasing.

Line 314-316: consider rephrasing to “There are also rivers that have experienced relatively stable permanent river water levels while also showing a decrease in annual seasonal flooding, e.g., the Ob River in Siberia (Zemtsov, 2019).”