Mapping flow-obstructing structures on global rivers

Xiao Yang[1], Tamlin M Pavelsky[1], Matthew Richard Voss Ross[2], Stephanie R Januchowski-Hartley[3], Wayana Dolan[1], Elizabeth Humphries Altenau[1], Michael Belanger[1], Danesha Byron[1], Michael Durand[4], Ian Van Dusen[4], Hailey Galit[1], Michiel Jorissen[5], Theodore Langhorst[1], Eric Lawton[5], Riley Lynch[2], Katie Ann Mcquillan[6], Sayali K Pawar[7], and Aaron Whittemore[8]

[1]University of North Carolina at Chapel Hill
[2]Colorado State University
[3]Swansea University
[4]Ohio State University
[5]Independent Research Aide
[6]North Carolina State University
[7]University of Dundee
[8]Washington State University

November 24, 2022

Abstract

To help store water, facilitate navigation, generate energy, mitigate floods, and support industrial and agricultural production, people have built and continue to build obstructions to natural flow in rivers. However, due to the long and complex history of constructing and removing such obstructions, we lack a globally consistent record of their locations and types. Here, we used a consistent method to visually locate and classify obstructions on 2.1 million km of large rivers (width [?] 30m) globally. We based our mapping on Google Earth Engine's high resolution images from 2018–2020, which for many places have meter-scale resolution. The resulting dataset, the Global River Obstruction Database (GROD), consists of 29,877 unique obstructions, covering six different obstruction types: dam, lock, low head dam, channel dam, and two types of partial dams. By classifying a subset of the obstructions multiple times, we are able to show high classification consistency (87% mean balanced accuracy) for the three types of obstructions that fully intersect rivers: dams, low head dams, and locks. The classification of the three types of partial obstructions are somewhat less consistent (61% mean balanced accuracy). Overall, by comparing GROD to similar datasets, we estimate GROD likely captured 90% of the obstructions on large rivers. We anticipate that GROD will be of wide interest to the hydrological modeling, aquatic ecology, geomorphology, and water resource management communities.

Hosted file

`grod_wrr_supporting information.docx` available at https://authorea.com/users/538738/articles/602597-mapping-flow-obstructing-structures-on-global-rivers

Mapping Flow-obstructing Structures on Global Rivers

Xiao Yang[1], Tamlin M. Pavelsky[1], Matthew R. V. Ross[2], Stephanie R. Januchowski-Hartley[3], Wayana Dolan[1], Elizabeth H. Altenau[1], Michael Belanger[1], Danesha Byron[1], Michael Durand[4, 5], Ian Van Dusen[4], Hailey Galit[1], Michiel Jorissen[6], Theodore Langhorst[1], Eric Lawton[7], Riley Lynch[2], Katie Ann Mcquillan[8], Sayali Pawar[3, 9], Aaron Whittemore[10]

[1]Department of Geological Sciences, University of North Carolina at Chapel Hill, NC, USA. [2]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA. [3]Department of Biosciences, Swansea University, Swansea, Welsh, UK. [4]Schoold of Earth Sciences, Ohio State University, Columbus, OH, USA. [5]Byrd Polar and Climate Research Center, Ohio State University, Columbus, Ohio, USA. [6]Independent Research Aid, Eindhoven, The Netherlands. [7]Independent Research Aid, Pontypool, Ontario, Canada. [8]Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA. [9]Geography and Environmental Science, University of Dundee, Dundee, Scotland, UK. [10]Center for Sustaining Agriculture and Natural Resources, Washington State University, Pullman, WA, USA

Corresponding author: Xiao Yang (yangxiao@live.unc.edu)

Key Points:

- We manually identified 29,877 river obstructions on 2.1 million km of large rivers across the globe.
- The Global River Obstruction Database provides rich new context for understanding human impacts on rivers.
- GROD identifies many in-river structures missed by other global dam databases.

Abstract

To help store water, facilitate navigation, generate energy, mitigate floods, and support industrial and agricultural production, people have built and continue to build obstructions to natural flow in rivers. However, due to the long and complex history of constructing and removing such obstructions, we lack a globally consistent record of their locations and types. Here, we used a consistent method to visually locate and classify obstructions on 2.1 million km of large rivers (width ≥ 30m) globally. We based our mapping on Google Earth Engine's high resolution images from 2018–2020, which for many places have meter-scale resolution. The resulting dataset, the Global River Obstruction Database (GROD), consists of 29,877 unique obstructions, covering six different obstruction types: dam, lock, low head dam, channel dam, and two types of partial dams. By classifying a subset of the obstructions multiple times, we are able to show high classification consistency (87% mean balanced accuracy) for the three types of obstructions that fully intersect rivers: dams, low head dams, and locks. The classification of the three types of partial obstructions are somewhat less consistent (61% mean balanced accuracy). Overall, by comparing GROD to similar datasets, we estimate GROD likely captured 90% of the obstructions on large rivers. We anticipate that GROD will be of wide interest to the hydrological modeling, aquatic ecology, geomorphology, and water resource management communities.

Plain Language Summary

Many obstructions have been built on rivers across the globe to help store water, facilitate navigation, generate energy, mitigate floods, and support industrial and agricultural production. However, the lack of publicly available information on where these obstructions are reduces our ability to assess their environmental impact. In this study, we used publicly available satellite data from Google to manually identify river obstructions on all large rivers across the globe (width ≥ 30 meter) to develop the Global River Obstruction Database, or GROD. GROD consists of 29,877 unique obstructions assigned to one of the six types: dam, low head dam, lock, channel dam (dam that obstructs one channel of a multi-channel river), and two types of partial dams (dam that extends partially across a river). By repeatedly classifying subsets of GROD obstructions, we estimate high classification consistency. And by comparing GROD to two other comprehensive obstruction datasets, we estimate that GROD covers ~90% of obstructions for the rivers studied. We anticipate that the release of GROD will help people around the world better understand and manage human impacts on rivers.

1 Introduction

Globally, the study of rivers would benefit from improved data on locations and characteristics of non-reservoir-producing obstructions, such as low head or partial dams and locks (Lange et al., 2019; Mantel et al., 2017). Efforts to date have focused on mapping large dams that generate reservoirs (Lehner & Grill, 2013; Mulligan et al., 2020). Despite the existence of millions of small obstructions around the world (Belletti et al., 2020; Lehner & Grill, 2013; Smith et al., 2002) most remain undocumented or are inconsistently mapped (Mantel et al., 2017). Increasingly, non-reservoir-producing river obstructions are recognized as both individually and cumulatively affecting movement of water, sediment, and species as well as altering river habitat (Januchowski-Hartley et al., 2020; Lucas et al., 2009). Limited data on river obstructions can lead to substantial underestimation of their ecological, environmental, and socio-economic impacts.

Documenting the types, locations, and characteristics of different obstructions is critical for adaptive management and decisions regarding monitoring for safety, as well as potential removals and construction (Januchowski-Hartley et al., 2013; Lange et al., 2019; Neeson et al., 2015). Efforts are ongoing to map obstructions that do not necessarily generate reservoirs, but primarily at the regional scale (Graf, 1999; Jones et al., 2019). For example, in the last decade, dozens of projects in North America have focused on inventorying and generating decision support tools to identify priorities for obstruction removal or remediation (e.g. https://streamcontinuity.org/). In the Laurentian Great Lakes Basin, more than 275,000 dams, weirs, and road-river crossings were mapped (Januchowski-Hartley et al., 2013), and continue to be central to research, monitoring, and remediation decisions in the region (e.g. https://greatlakesconnectivity.org/). Similar projects and initiatives continue across 13 North Atlantic States and 14 Southeast States in the United States along with the Commonwealth of Puerto Rico (https://southeastaquatics.net/), including collation and maintenance of mapping and online databases of potential obstructions to hydrological and ecological connectivity determined through on-the-ground assessments. In Europe, Belletti et al. (2020) collated locations for >600,000 river obstructions (AMBER Atlas) across 36 countries. Careful compilation has been used to merge datasets of heterogeneous sources into consistent larger scale databases, though compilation becomes harder to do in regions of the world where records have not been well kept or not made public. Alternatively, participatory approaches to data collection (e.g. (Mulligan et al., 2020; Whittemore et al., 2020) can be a consistent way of mapping river obstructions that includes both large, reservoir-generating dams and smaller obstructions.

By expanding our previous work focusing on United States (Whittemore et al., 2020), we present the complete Global River Obstruction Database (GROD), which maps dams, low head dams (weirs), locks, and partial dams (e.g., wing dams)—from here termed obstructions—along the world's large rivers (>30m wide; (Allen & Pavelsky, 2018)). By manually identifying and classifying obstructions from high-resolution images in Google Earth Engine (Gorelick et al., 2017), we mapped 29,877 obstructions along 2.1 million km of river length. Aside from the dataset, we describe, in chronicle order, five phases of the dataset development (Figure S1) that we think would be valuable to similar mapping efforts in the future—Phase 1: setup and initial mapping; Phase 2: intermediate evaluation; Phase 3: revision and final internal evaluation; Phase 4: evaluation with external datasets; Phase 5: first application of GROD.

2 Methods

2.1 Phase 1: Setup and initial mapping

2.1.1 Setting up the mapping and data storage environment

We used openly available Google tools for all mapping and data management. Specifically, we used the satellite image background, a mosaic of recently-captured high-resolution images from Google Earth Engine (Gorelick et al., 2017), for project participants to map obstructions. We overlaid river centerlines from the Global River Widths from Landsat (GRWL; (Allen & Pavelsky, 2018)) on top of the satellite images to help guide the mapping process. To create an interactive mapping interface, we used Google Earth Engine's JavaScript code editor interface and geometry tools (Gorelick et al., 2017). Lastly, we used a shared Google Drive folder to organize mapping results and a shared Google Sheet to track mapping progress.

2.1.2 Mapping procedures

To allow multiple participants to work on the mapping simultaneously and reduce the chance of duplication, we divided our mapping area into tiles. Specifically, following Whittemore et al., (Whittemore et al., 2020), we divided the GRWL-covered land surface area into 1,039 tiles (each 12º by 6º). A typical mapping session, which starts from a new tile, contains the following the following action items:

1. Select a tile by inputting a unique tile ID, following which river centerlines are shown overlaying the satellite images within the selected tile from Google Earth Engine's satellite view;

2. Scroll along the river centerlines, and upon finding a structure, follow the classification criteria (see the following section) and choose an obstruction type to mark the location. Note that markers are usually placed on the structure itself and close to its center, e.g. for a dam, the marker is placed on the dam wall instead of at the center of the reservoir created by it;

3. Save the mapped data by saving the file on Google Earth Engine at the end of each mapping session;

4. Export mapped data as a csv file into a designated folder in the participant's Google Drive folder;

5. Repeat steps 1–4 until all GRWL rivers in the tile have been searched for obstructions. Then mark the tile as "completed" in the shared spreadsheet containing mapping progress.

2.1.3 Classification criteria

We followed the criteria detailed by Whittemore et al. (2020) to classify each obstruction and summarize the processes here. In general, GROD includes six types of obstructions that are the most common in rivers. They consist of three types of obstructions that cross the full width of the river (Group I obstructions): dam, lock, and low head dam, and three other types that only partially obstruct flow across the river width (Group II obstructions): channel dam, partial dams 1 (obstruction length < 50% of river width), and partial dams 2 (obstruction length ≥ 50% of river width). We defined the obstruction types so that they span abilities to obstruct flow. For example, dams should have a more substantial influence on flow than locks, which should have a heavier impact on flow than lLw head dams. Whenever a structure was composed of multiple obstruction types, we assigned it the type that affects the flow the least. For example, if an obstruction was composed of a low head dam and a lock, we labeled it as a low head dam as water can flow over the obstruction even when the lock gates were closed. An additional type named "Uncertain" was used during mapping to temporarily indicate obstructions that were hard to identify at that moment. All obstructions assigned uncertain type later on went through secondary screening to either be removed from the dataset or be incorporated as one of the six types in GROD.

Challenges of classifying three-dimensional structures based on two-dimensional satellite images from limited angles are handled using contextual information. We used features such as structure shape, extent, and material, as well as the existence of structure shadow, visible water flow on top, and relative channel width change up/downstream of the structure to determine obstruction type. Thus, the type chosen for each obstruction is based on the best judgement of how the obstruction influences the flow at the time of the image acquisition, rather than an objective determination of the design purpose of the obstruction. In some cases, additional views from other sources such as Google Street View were used to aid in the identification.

2.1.4 Data cleanup

After completing the global mapping, we conducted two data cleaning steps. First, we removed duplicates in the global dataset by manually examining 1,216 potential locations. These locations were identified automatically where any two (or more) obstructions were within 200 meters distance and were added to the dataset by different participants. Then, we manually reexamined all obstructions that were labeled as uncertain (n>3000) and either removed them or assigned them into one of the six obstruction types.

2.2 Phase 2: Intermediate evaluation

To estimate how consistently we classified GROD obstructions, three participants reclassify the same 10% random subset of obstructions (n = 3,336). The reclassified results were compared to the initial types given to these obstructions, as well as among those from the three participants. The former comparison informs us of the consistency in repeatedly classifying obstructions, and the latter comparison informs us of the consistency with which different participants classify each type of obstructions.

Throughout the evaluation, we used two metrics to report consistency between two classifications: balanced accuracy and F1 score (Van Rijsbergen, 1979), both of which can be estimated for each type of obstruction. Balanced accuracy is the mean of sensitivity and specificity, while the F1 score is calculated as the harmonic mean of sensitivity and precision (see “Evaluation metrics” in the Supporting Information). Compared to accuracy, balanced accuracy takes into account both positive and negative cases and is not affected as much by class imbalance; Values of the F1 score range from 0 to 1, with a higher F1 score indicating better agreement, with low false positive and false negative rates. In this study, we used the values of these two metrics to infer consistency of classification, rather than using them to represent the accuracy of classification, which would require us to attribute one classification as truth.

2.3 Phase 3: Revision and final internal evaluation

After Phase 2, to improve the accuracy of the GROD data, we determined that we should focus primarily on Group I obstructions since they are more likely to be classified consistently (from details on what we have found from intermediate evaluation, see Table S1 and the Supporting Information under “Intermediate evaluation results”), pose greater influence on river flow, and tend to more substantially impact the ecology (e.g., fish movement) of rivers. In comparison, Group II obstructions are more challenging to classify consistently, and their identification is more heavily affected by river flow stage. We nonetheless included them in the dataset because knowing their locations will be valuable to others (see supplementary materials).

To ensure the highest accuracy, we reclassified all Group I structures. After reclassification, we conducted our final assessment, checking classification consistency across different participants. Two authors reclassified a random subset of obstructions from Group I obstructions (N = 500 total; dams (n = 206), low head dams (n = 246), and locks (n = 48)). To evaluate the dataset, we merged the classification results from both participants and calculated the balanced accuracy and F1 against the reclassified Group I types.

2.4 Phase 4: Evaluation with external datasets

The classification accuracy from the evaluations in Phase 2 and 3 revealed how well we consistently classify obstructions of each type. However, as GROD is likely to be used to evaluate river fragmentation, it is important to also know the likelihood of obstructions being missed during the mapping process. To estimate the omission rate of GROD, we used two independent data sources, AMBER Atlas and OpenStreetMap (OSM), which contain similar obstruction types included in GROD. The details of how we extracted and cleaned the OSM data are documented in Supporting Information under "Downloading and preprocessing OpenStreetMap Global Dam Data". In total, we obtained 52,670 obstruction point locations from OSM and 629,955 point locations from AMBER Atlas. Then, we narrowed down both datasets to a combined dataset of 4,807 locations by checking whether the points overlapped the river mask used to derive GRWL. The merged data were then further constrained to only include obstruction types compatible with the types included in GROD. The final subset of data includes 2,740 points from AMBER Atlas (consisting of types "WEIR", "DAM", "SLUICE") and 550 points from OSM (consisting of types "dam", "weir", "lock_gate").

To estimate the omission rate of GROD, we compared GROD with random subsets of the cleaned OSM and AMBER Atlas database. Specifically, we randomly sampled 100 points separately from the cleaned OSM and AMBER Atlas databases and examined the 200 locations manually. At each location, we checked: (1) whether the point from OSM or AMBER Atlas was valid for comparison with GROD and (2) if valid, whether there was a GROD point corresponding to that location. Points were considered invalid for comparison if, at that location: (1) the obstructions existed but did not correspond to one of the six types included in GROD (e.g. bridges or roads), or (2) the obstruction existed but did not obstruct flow along the GRWL river flow direction (e.g. obstructions situated on the side of the main GRWL river channel and obstructing a tributary connected to the GRWL-defined river), or (3) the obstruction could not be observed at the given location. For cases when either AMBER Atlas or OSM identified a valid obstruction but no GROD points could be associated with it, we counted it as truly missing.

To estimate the omission rate, we divided the total numbers of true missing cases, separately for OSM and AMBER Atlas, by the total count of locations examined (n = 100). At the same time, we also estimated the invalid rate for these two datasets by dividing the total number of invalid cases by the total count (n = 100).

2.5 Phase 5: First application of GROD

One of the important applications of datasets like GROD is to assess and quantify how obstructions modify hydrological or ecological connectivity (Belletti et al., 2020; Grill et al., 2019; Jones et al., 2019). To demonstrate how GROD can be used to do so, we estimated obstruction density based on two different sets of spatial aggregation, using HydroBASINS (Lehner & Grill, 2013) and national boundaries. The method behind these two aggregations was the same. For each unit (a basin from HydroBASINS level-3 or a country polygon), we calculated the total river length from the GRWL database as well as the total count of obstructions. Then we estimated the density of obstructions on rivers (converted to “n/1000km”) by normalizing obstruction count by total river length.

3 Results

3.1 GROD evaluation

3.1.1 Internal evaluation: Classification consistency

Comparing the evaluation metrics for Group I obstructions between the final GROD dataset and that based on the intermediate evaluation, we found substantial increase in accuracy in all three Group I types (Table 1), with balanced accuracy improving on average by 14 percentage points and F1 score improving on average by 23 percentage points. The most improved type was the lock. The final averaged balanced accuracy for Group I obstructions was 87% (F1 score: 83%), a 20% improvement compared to the metric estimated for the Group I obstructions during the intermediate evaluation. Note that values for the intermediate evaluation were recalculated by limiting the initial data to only Group I obstructions, considering the fact that only three types of obstructions were used in the final evaluation.

Table 1. Intermediate and final evaluation. Numbers in the parenthesis were from intermediate evaluation only for Group I obstructions.

Obstruction type	**Balanced accuracy**	**F1**
Dam	(73%) 83%	(70%) 79%
Low head dam	(72%) 83%	(66%) 83%
Lock	(70%) 94%	(52%) 88%

3.1.2 External evaluation: GROD omission rate

By manually comparing GROD to a random subset of obstructions (n = 200) from the AMBER Atlas and OSM dataset, we estimated a mean GROD omission rate of 9.5% from comparisons of similar obstruction types in AMBER Atlas and OSM along the same GRWL rivers (9% for AMBER Atlas and 10% for OSM). At the same time, we obtained substantial invalid rates for the comparison datasets (20% for AMBER Atlas and 22% for OSM), meaning that one in five obstructions we examined for AMBER Atlas and OSM was either not present on the image we examined at the location given or, despite having narrowed types down to the ones included in GROD, not belonging to the types or the rivers studied in GROD. Where GROD omission occurred did not show any obvious spatial pattern, suggesting the locations of omission are likely random. However, the omission rate we estimated is likely an approximate, as the external datasets we compared GROD to have their own caveats, as indicated by their high invalid rate.

3.2 GROD and its spatial distribution

We mapped 29,877 unique obstructions globally. Of the six types, low head dams were the most abundant, accounting for 38% (n = 11,372) of the obstructions, followed by dams, accounting for 28.6% (n = 8,537) of the obstruction count (Table 2). Spatially, river obstructions are clustered in industrialized regions (Figure 1). Both dams and low head dams were spread widely in these regions. In contrast, the 1,728 locks identified in the GROD are heavily clustered in a few regions, including western Europe, eastern China, and the eastern United States (for the distribution of each obstruction type, see Figure S3).

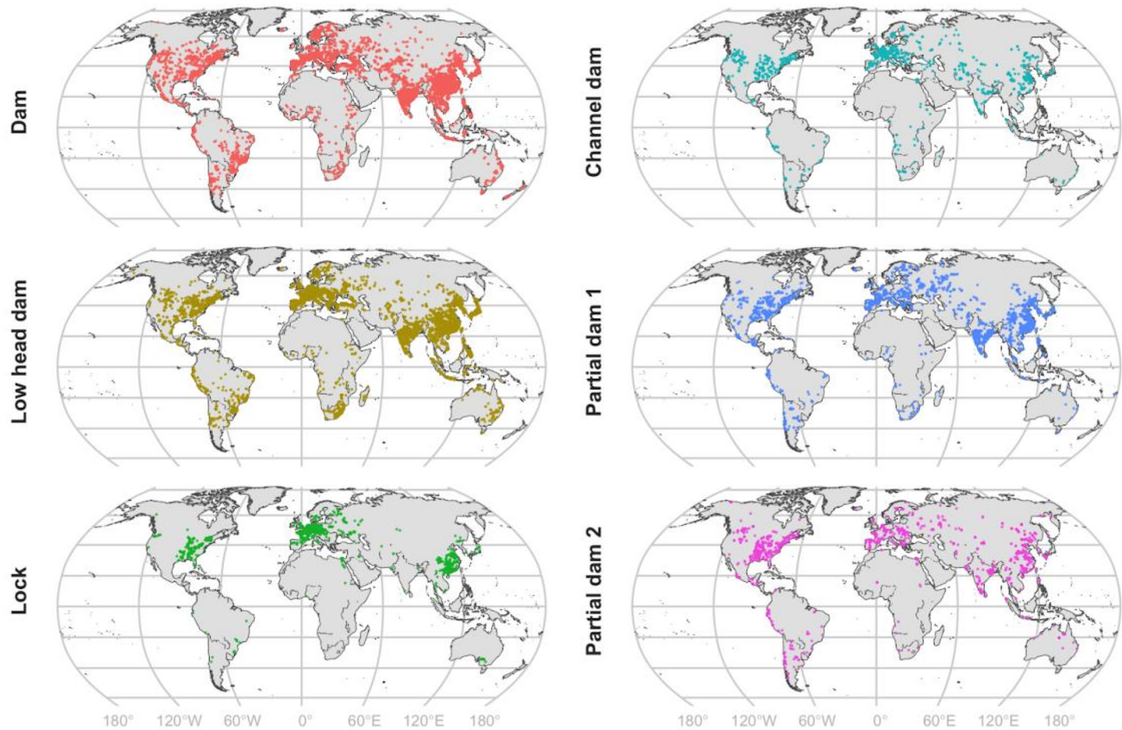

Figure 1. Spatial distribution of each obstruction type in GROD: dam, low head dam, lock, channel dam, and two types of partial dams.

Table 2. GROD obstruction types and counts. Total number of obstructions = 29,877.

Obstruction type	Total number (n)	Percent (%)	Group
Dam	8,537	28.6%	I
Low head dam	11,372	38.1%	I
Lock	1,728	5.8%	I
Channel dam	1,675	5.6%	II
Partial dam 1 (<50% width)	3,687	12.3%	II
Partial dam 2 (≥50% width)	2,878	9.6%	II

Regions with fewer obstructions are located either in high latitudes (northern and northwestern North America, northern Asia), or in other regions where industrial activity and population density are relatively low (e.g. Amazon rainforest, central Africa, and western Australia).

Overall, of the 2.1 million km of rivers studied globally, ~49% of the length is minimally obstructed, defined as less than one obstruction per 1000 km of river length.

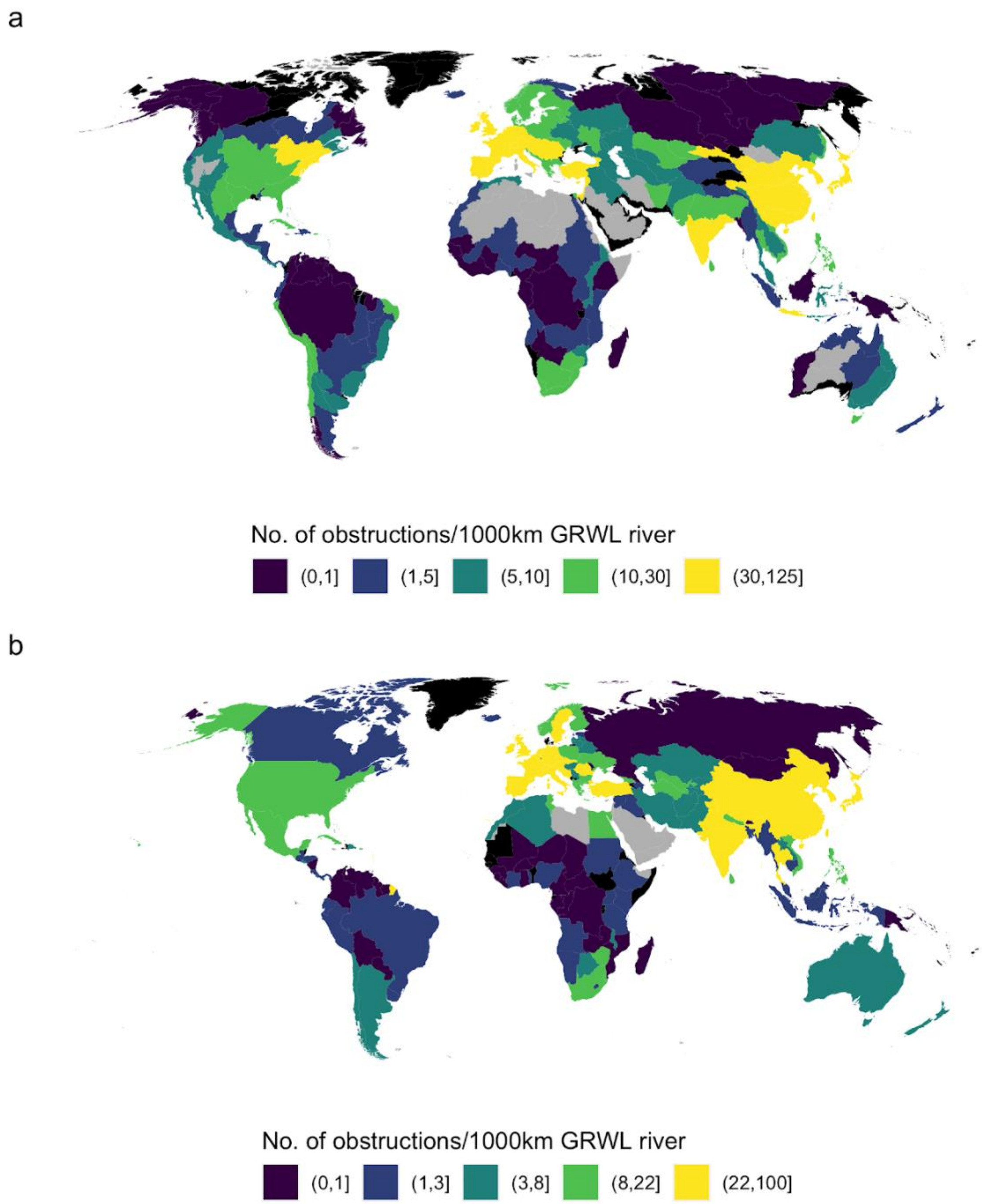

Figure 2. Obstruction density for level-3 HydroBASINS (a) and for country level (b) based on Group I obstruction locations and GRWL river lengths. Only Group I obstructions were used for calculating density on this map. For results based on all six types of obstructions, see Figures S3–4.

When looking at river obstruction density at the country level (Figure 2b), the spatial patterns are similar to those based on drainage basins (Figure 2a). The majority of the top 10 countries are located in Europe, with the exception of Japan (rank 1) and South Korea (rank 7) (Figure S5). However, if we rank the countries by their absolute number of obstructions, we see the ranking instead affected by country size and river abundance, with the top 3 countries being China, India, and the United States (Figure S5). For rankings based on all 6 types of obstructions, see Figure S6.

4 Discussions

In this study, we presented GROD, a dataset consisting of 29,877 river obstructions along the world's largest rivers. Using publicly accessible satellite images available from Google, we were able to manually identify human-made structures obstructing river flow. With five phases of dataset development, we are able to provide to the scientific community a dataset that is both comprehensive and accurate. Despite the challenges associated with classifying three-dimensional obstructions using two-dimensional image representations of the Earth's surface, we were able to achieve an averaged balanced accuracy of 87% across the Group I obstructions. We also compared GROD with other global/regional datasets to show that, by only mapping through the globe once, GROD is able to include ~90% of known obstructions on rivers ≥ 30 m wide. Additionally, as shown by Whittemore et al. (2020), GROD is likely much more comprehensive in including small river barriers compared to some of the other regional datasets and can serve as a baseline for regional-scale datasets, especially in places where such data are not publicly available. The successful coordination and development of GROD demonstrate the effectiveness and efficiency of conducting mapping using publicly available cloud-based resources. This is critical, especially for features like river obstructions that remain challenging to accurately identify using automated algorithms. We expect the release of GROD will be valuable for many research fields including those interested in changes to surface water flows and movement of aquatic migratory species.

We identified three key sources of uncertainty associated with GROD. First, the non-uniform spatial resolution of imagery across the world: structures might be harder to classify if the images have lower spatial resolution and appear blurred. However, most of the highly populated regions where river obstructions tend to cluster had high resolution images, so a disparity in resolution should have minimal effect on our dataset. Second, satellite images from Google are updated regularly, which makes it a challenge to reproduce or evaluate mapping across different time periods. This is a limitation because updated images might reflect different flow states of rivers, or different development stages of ongoing building or demolishing of obstructions: for example, obstructions presented in our initial mapping might be submerged or demolished in the follow-up mapping effort. This limitation could, in theory, be addressed by conducting all mapping in a very narrow temporal window. However, doing so is impractical, given both the amount of effort and time required to manually map obstructions globally and the limitations of the imagery itself. Obtaining cloud-free imagery in many regions is quite challenging, especially during high flow seasons or in regions with monsoons (e.g. the Indian Subcontinent or Amazon Basin) (Allen et al., 2020). For these reasons, GROD should be seen as an inventory sourced from a period of time covering several years, instead of as the accurate instantaneous capture of the year of the published dataset. Third, consistently classifying a structure, either repeatedly by the same participant or among multiple participants, proved challenging. We believe this challenge could be partly linked with the specificity of our initial obstruction typology, for example differentiating between a dam and a channel dam is not always straightforward (see Whittemore et al. 2020). Additionally, because classification of an obstruction to a type is subjective and particularly challenging when structures are small relative to the resolution of the imagery.

As identified by Whittemore et al. (2020) for regions of United States and France, we consistently found that the channel dam type had low classification consistency among participants (Table S2), likely due to the difficulties in differentiating between a dam on a single channel and a channel dam on one portion of a multichannel river. However, the effect from this limitation likely varies depending on the application: while it might be challenging to differentiate between all six types of obstructions (overall accuracy: 66%), the distinction between Group I and Group II is quite accurate (overall accuracy: 91%). And the advantage of having a finer typology is that the users can decide how to select and merge the data according to the application.

Regional records or compiled datasets like AMBER Atlas (Belletti et al., 2020), SEACAP (Martin et al., 2014) or the projects across North America (https://streamcontinuity.org/), do report obstructions on much smaller rivers and streams, but such data are only limited to regions with historical records and an ongoing effort to compile relevant data. In many parts of the world, this type of information is either not publicly available or not recorded at all (Brejão et al., 2020; Carvajal-Quintero et al., 2017; Kroon & Phillips, 2015; Shirley et al., 2021), which can lead to spatially biased representation of obstruction density if data were simply compiled together. Furthermore, existing river obstruction datasets, both regional (AMBER Atlas) and global (OSM), created by compiling existing datasets or participatory science, are not linked explicitly to a consistent river network. This in itself is not problematic, and should not affect completeness of obstructions, in the sense that these datasets aimed to include all possible obstructions resolved by the particular method used to map them. However, not linking obstruction data explicitly to a river network does limit what analyses can be carried out, and can result in additional work for end-users. For example, while comparing GROD with AMBER Atlas we frequently noticed that there are obstructions in that database that are very near to a large river but that are not actually on or obstructing flow to that channel. Given these limitations, it would be challenging to accurately calculate an equivalent to Figure 2 using AMBER Atlas or a similar dataset as it is currently available, because many of the obstructions do not pair with river channels.

Looking to the future, we anticipate that combining the type and location information from GROD with the corresponding satellite images will provide us with a labeled dataset that can be used to develop machine-learning approaches that could potentially automate the mapping process and help solve some of the limitation we face when creating GROD. In fact, we have achieved image files for each record in GROD as png files after finalizing GROD, which will be made publicly available along with the GROD dataset. We also expect the locations of GROD obstructions will help better model sediment transport in the world's large rivers. GROD also helps divide rivers into reaches within which no human-made obstructions will cause abrupt changes in flow. Knowing such locations of flow disruption is critical for remote sensing of river discharge from satellites like the upcoming Surface Water and Ocean Topology satellite

(Biancamaria et al., 2016) that aims to provide global river discharge products. Our experience with building GROD demonstrates that large-scale mapping projects can be planned efficiently and implemented thanks to recently available cloud-based geospatial platforms such as Google Earth Engine. GROD can serve as an example of how, with the help from community scientists, global mapping of important surface features can be conducted efficiently and accurately.

Acknowledgments, Samples, and Data

Work at the University of North Carolina was funded under a contract from the SWOT Project Office at the NASA/Caltech Jet Propulsion Lab. SRJ and SP acknowledge funding from the Welsh European Funding Office and European Regional Development Fund under Project 80761-SU-140 (West). External data used in this study can be found at their specific repository: GRWL centerline data used for mapping, and river mask data used for cleaning validation data from OSM and AMBER Atlas can be found at http://doi.org/10.5281/zenodo.1297434. AMBER Atlas V1 data was accessed from https://figshare.com/articles/dataset/AMBER_Atlas_of_Instream_Barriers_in_Europe/12629051. The code for the interactive Google Earth Engine interface used for mapping, and for the analysis in the paper can be found at https://github.com/GlobalHydrologyLab/GROD for review purpose, and will be deposited permanently for public access in Zenodo upon paper acceptance. The GROD dataset is provided as part of the supporting information for review purpose and will be made public permanently on Zenodo upon acceptance of the paper.

References

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*. 361(6402), 585–588. https://doi.org/10.1126/science.aat0636

Allen, G. H., Yang, X., Gardner, J., Holliman, J., David, C. H., & Ross, M. (2020). Timing of Landsat Overpasses Effectively Captures Flow Conditions of Large Rivers. *Remote Sensing*, *12*(9), 1510.

Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In *The Semantic Web - ISWC 2009* (pp. 731–746). Springer Berlin Heidelberg.

Belletti, B., Garcia de Leaniz, C., Jones, J., Bizzi, S., Börger, L., Segura, G., et al. (2020). More than one million barriers fragment Europe’s rivers. *Nature*, *588*(7838), 436–441.

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, *37*(2), 307–337.

Brejão, G. L., Teresa, F. B., & Gerhard, P. (2020). When roads cross streams: fish assemblage responses to fluvial fragmentation in lowland Amazonian streams. *Neotropical Ichthyology: Official Journal of the Sociedade Brasileira de Ictiologia*, *18*(3). https://doi.org/10.1590/1982-0224-2020-0031

Carvajal-Quintero, J. D., Januchowski-Hartley, S. R., Maldonado-Ocampo, J. A., Jézéquel, C., Delgado, J., & Tedesco, P. A. (2017). Damming Fragments Species’ Ranges and Heightens Extinction Risk. *Conservation Letters*. https://doi.org/10.1111/conl.12336

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27.

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. *Water Resources Research*, *35*(4), 1305–1311.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., et al. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, *7*(4), 12–18.

Januchowski-Hartley, S. R., McIntyre, P. B., Diebel, M., Doran, P. J., Infante, D. M., Joseph, C., & Allan, J. D. (2013). Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings. *Frontiers in Ecology and the Environment*, *11*(4), 211–217.

Januchowski-Hartley, S. R., Mantel, S., Celi, J., Hermoso, V., White, J. C., Blankenship, S., & Olden, J. D. (2020). Small instream infrastructure: Comparative methods and evidence of environmental and ecological responses. *Ecological Solutions and Evidence*, *1*(2). https://doi.org/10.1002/2688-8319.12026

Jones, J., Börger, L., Tummers, J., Jones, P., Lucas, M., Kerr, J., et al. (2019). A comprehensive assessment of stream fragmentation in Great Britain. *The Science of the Total Environment*, *673*, 756–762.

Kroon, F. J., & Phillips, S. (2015). Identification of human-made physical barriers to fish passage in the Wet Tropics region, Australia. *Marine and Freshwater Research*, *67*(5), 677–681.

Lange, K., Wehrli, B., Åberg, U., Bätz, N., Brodersen, J., Fischer, M., et al. (2019). Small hydropower goes unchecked. *Frontiers in Ecology and the Environment*, *17*(5), 256–258.

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems: GLOBAL RIVER HYDROGRAPHY AND NETWORK ROUTING. *Hydrological Processes*, *27*(15), 2171–2186.

Lucas, M. C., Bubb, D. H., Jang, M.-H., Ha, K., & Masters, J. E. G. (2009). Availability of and access to critical habitats in regulated rivers: effects of low-head barriers on threatened lampreys. *Freshwater Biology*, *54*(3), 621–634. https://doi.org//10.1111/j.1365-2427.2008.02136.x

Mantel, S. K., Rivers-Moore, N., & Ramulifho, P. (2017). Small dams need consideration in riverscape conservation assessments. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(4), 748–754. https://doi.org/10.1002/aqc.2739

Martin, E. H., Hoenke, K., Granstaff, E., Barnett, A., Kauffman, J., Robinson, S., & Apse, C. D. (2014). SEACAP: Southeast Aquatic Connectivity Assessment Project: assessing the ecological impact of dams on Southeastern rivers. *The Nature Conservancy, Eastern Division Conservation Science, Southeast Aquatic Resources Partnership*.

Mooney, P., & Corcoran, P. (2011). Annotating spatial features in OpenStreetMap. *Proceedings of the GISRUK*, 52–56.

Mulligan, M., van Soesbergen, A., & Sáenz, L. (2020). GOODD, a global dataset of more than 38,000 georeferenced dams. *Scientific Data*, *7*(1), 31.

Neeson, T. M., Ferris, M. C., Diebel, M. W., Doran, P. J., O'Hanley, J. R., & McIntyre, P. B. (2015). Enhancing ecosystem restoration efficiency through spatial and temporal coordination.

Proceedings of the National Academy of Sciences of the United States of America, *112*(19), 6236–6241.

Shirley, K., Noriega, A., Levin, D., & Barstow, C. (2021). Identifying Water Crossings in Rural Liberia and Rwanda Using Remote and Field-Based Methods. *Sustainability: Science Practice and Policy*, *13*(2), 527.

Smith, S. V., Renwick, W. H., Bartley, J. D., & Buddemeier, R. W. (2002). Distribution and significance of small, artificial water bodies across the United States landscape. *The Science of the Total Environment*, *299*(1-3), 21–36.

Van Rijsbergen, C. (1979). Information retrieval: theory and practice. In *Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems* (pp. 1–14).

Whittemore, A., Ross, M. R. V., Dolan, W., Langhorst, T., Yang, X., Pawar, S., et al. (2020). A participatory science approach to expanding instream infrastructure inventories. *Earth's Future*, *8*(11). https://doi.org/10.1029/2020ef001558