Check for updates

ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

TECHNICAL REPORTS: DATA

10.1029/2021WR030386

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], and Aaron Whittemore[10]

[1]Department of Earth, Marine and Environmental Sciences, University of North Carolina at Chapel Hill, 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, UK, [4]School of Earth Sciences, Ohio State University, Columbus, OH, USA, [5]Byrd Polar and Climate Research Center, Ohio State University, Columbus, OH, USA, [6]Independent Research Aid, Eindhoven, The Netherlands, [7]Independent Research Aid, Pontypool, ON, Canada, [8]Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA, [9]Geography and Environmental Science, University of Dundee, Dundee, UK, [10]Center for Sustaining Agriculture and Natural Resources, Washington State University, Pullman, WA, USA

Key Points:

- We manually identified 30,549 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

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Yang,
yangxiao@live.unc.edu

Citation:

Yang, X., Pavelsky, T. M., Ross, M. R. V., Januchowski-Hartley, S. R., Dolan, W., Altenau, E. H., et al. (2022). Mapping flow-obstructing structures on global rivers. *Water Resources Research*, *58*, e2021WR030386. https://doi.org/10.1029/2021WR030386

Received 12 MAY 2021
Accepted 23 DEC 2021

Author Contributions:

Conceptualization: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Wayana Dolan, Theodore Langhorst
Data curation: Michael Belanger, Danesha Byron, Hailey Galit, Michiel Jorissen, Eric Lawton, Riley Lynch, Katie Ann Mcquillan, Sayali Pawar, Aaron Whittemore
Formal analysis: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley, Wayana Dolan
Funding acquisition: Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley
Investigation: Elizabeth H. Altenau, Michael Belanger, Michael Durand, Ian Van Dusen, Hailey Galit, Riley Lynch
Methodology: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley, Wayana Dolan,

© 2022. American Geophysical Union. All Rights Reserved.

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 ≥30 m) globally. We based our mapping on Google Earth Engine's high resolution images, which for many places have meter-scale resolution. The resulting Global River Obstruction Database (GROD) consists of 30,549 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 (e.g., dams and locks) 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 m) to develop the Global River Obstruction Database, or GROD. GROD consists of 30,549 unique obstructions assigned to one of the six types: dam, low head dam, lock, channel dam (a 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 five other obstruction datasets, we estimate that GROD contains >90% 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

Theodore Langhorst, Sayali Pawar, Aaron Whittemore
Project Administration: Tamlin M. Pavelsky, Matthew R. V. Ross
Software: Xiao Yang, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley, Wayana Dolan, Theodore Langhorst
Supervision: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley
Validation: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley, Wayana Dolan, Elizabeth H. Altenau, Michael Belanger, Michael Durand, Ian Van Dusen, Hailey Galit, Riley Lynch, Sayali Pawar, Aaron Whittemore
Visualization: Xiao Yang, Tamlin M. Pavelsky, Stephanie R. Januchowski-Hartley
Writing – original draft: Xiao Yang
Writing – review & editing: Xiao Yang, Tamlin M. Pavelsky, Matthew R. V. Ross, Stephanie R. Januchowski-Hartley, Wayana Dolan, Elizabeth H. Altenau, Michael Belanger, Danesha Byron, Michael Durand, Ian Van Dusen, Hailey Galit, Michiel Jorissen, Theodore Langhorst, Eric Lawton, Riley Lynch, Katie Ann Mcquillan, Sayali Pawar, Aaron Whittemore

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 for 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 southeastern 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 data sets 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 (>30 m wide; Allen & Pavelsky, 2018). By manually identifying and classifying obstructions from high-resolution images in Google Earth Engine (Gorelick et al., 2017), we mapped 30,549 obstructions along 2.1 million km of river length. Aside from the data set, we describe, in chronicle order, five phases of the data set development (Figure S1 in Supporting Information 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. GRWL was chosen as it is the only observation-based (as opposed to DEM-based) global scale river centerline database and is the foundation for the satellite Surface Water and Ocean Topography (SWOT) measurements (Altenau et al., 2021), which is one of the key applications for GROD. 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. (2020), we divided the GRWL-covered land surface area into 1,039 tiles (each 12° by 6°). A typical mapping session, which starts from a new tile, contains 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, for example, 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.

In total, nine of the authors have contributed to the mapping process, conducted from 2018 to 2021. The degree and duration of participation vary among these authors and are largely depending on availability and funding.

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; see Figure S2 in Supporting Information S1 for example images of each obstruction types). In practice, classifying obstruction types follows a decision tree (see Figure S3 in Supporting Information S1). We defined the obstruction types so that they span abilities to affect water surface elevation, flow velocity, and ecological connectivity. For example, dams should have a more substantial influence on flow and connectivity than locks, which should have a heavier impact than low head dams. Relative impact on flow and connectivity for each type of obstruction was ranked based on the average obstruction size (height) and its likelihood of causing ecological fragmentation, however, such influence also depends on river flow regimes and the operational rules an obstruction follows. 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 an uncertain type later on went through secondary screening to either be removed from the data set or be incorporated as one of the six types in GROD. Note that we defined these obstruction types so that they represent common types of human-made obstructions. Other natural (e.g., riffles and pools and waterfalls) or human-made structures (e.g., bridges) were not included 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 judgment 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 data set by manually examining 1,216 potential locations. These locations were identified automatically where any two (or more) obstructions were within 200 m distance and were added to the data set by different participants. Then, we manually reexamined all obstructions that were labeled as uncertain (n > 3,000) 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 with the most experience in mapping for GROD 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 Supporting Information S1). 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 "Intermediate evaluation results" in Supporting Information S1), 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 data set because knowing their locations will be valuable to certain applications (see in Supporting Information S1).

To improve classification accuracy, following the intermediate evaluation, three of the authors have reclassified all Group I obstructions by reassigning types for them. 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 data set, 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 five independent data sources, AMBER Atlas (Belletti et al., 2020), OpenStreetMap (OSM), GRanD v1.3 (Global Reservoir and Dam Database version 1.3; Lehner et al., 2011), GOODD v1 (GlObal geOreferenced Database of Dams; Mulligan et al., 2020) and dam locations for the Amazon basin (Latrubesse et al., 2017), which contain one or more 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, 629,955 point locations from AMBER Atlas, 7,319 dam point locations from GRanD v1.3, 38,667 dam point locations from GOODD database, and 140 dam locations from Latrubesse et al. (2017). Before processing further, we constraint obstructions to only include those that were were labeled as "WEIR", "DAM", or "SLUICE" from AMBER Atlas and "dam", "weir", or "lock_gate" from OSM. Then, we narrowed down all data sets to a combined data set of 6,002 point locations that intersects the rivers used to derive GRWL.

To estimate the omission rate of GROD, we compared GROD with all 6,002 (equivalent to ~ 20% of GROD) point locations from external datasets that have similar obstruction types and in close proximity to GRWL rivers. At each location, we checked: (a) whether the point from external databases was valid for comparison with GROD and (b) if valid, whether there was a GROD point corresponding to that location. Points were considered invalid for comparison if, at that location: (a) the obstructions existed but did not correspond to one of the six types included in GROD (e.g., bridges or roads), or (b) 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), (c) no obstruction could be observed at the given location, or (d) the obstruction existed and intersected with GRWL river mask but was located beyond the coverage

Table 1
Intermediate and Final Evaluation

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

Note. Numbers in the parenthesis were from intermediate evaluation based on Group I obstructions.

of GRWL centerlines (which we labeled separately as "not_on_GRWL" in Table S2 in Supporting Information S1). For cases when external databases identified a valid obstruction but no GROD points can 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 each external database, by the total count of locations examined minus the count of invalid and locations labeled as "not_on_GRWL". At the same time, we also estimated the invalid rate for all datasets by dividing the total number of invalid cases by the total count minus the count of "not_on_GRWL" points. After estimating omission rate using external databases, we then included all missing obstructions in the final version of the GROD database.

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/1000 km") by normalizing obstruction count with 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 data set 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% points and F1 score improving on average by 23% 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.

3.1.2. External Evaluation: GROD Omission Rate

We estimated an overall GROD omission rate of 10% from comparisons of similar obstruction types in all five external databases along the same GRWL rivers, with an omission rate ranging from 3% for GOODD to 17% for the dams in the Amazon basin (Table S2 in Supporting Information S1). At the same time, we obtained substantial invalid rates for the comparison data sets, with overall invalid rate 13% (Table S2 in Supporting Information S1), meaning that approximately one in seven external database obstructions we examined was (a) not present on the image we examined at the location given, (b) not obstructing flow of the GRWL rivers or located along the GRWL rivers, or (c) not belonging to the types or the rivers studied in GROD. The geographic distribution of GROD omissions showed no obvious spatial pattern (Figure S4 in Supporting Information S1), suggesting the locations of omission are likely random. However, the omission rate we estimated is likely 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 30,549 unique obstructions globally. Of the six types, low head dams were the most abundant, accounting for 38.1% (n = 11,648) of the obstructions, followed by dams, accounting for 29.0% (n = 8,864) of the obstruction count (Table 2). Spatially, river obstructions are clustered in industrialized regions (Figure 1). Both

10.1029/2021WR030386

Table 2
Global River Obstruction Database (GROD) Obstruction Types and Counts

Obstruction type	Total number (n)	Percent (%)	Group
Dam	8,864	29.0%	I
Low head dam	11,648	38.1%	I
Lock	1,766	5.8%	I
Channel dam	1,684	5.5%	II
Partial dam 1 (<50% width)	3,687	12.1%	II
Partial dam 2 (≥50% width)	2,900	9.5%	II

Note. Total number of obstructions = 30,549.

dams and low head dams were spread widely. In contrast, the 1,766 locks identified in GROD are heavily clustered in a few regions, including western Europe, eastern China, and the eastern United States.

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, ~42% of the river length is minimally obstructed, defined as less than one Group I obstruction per 1,000 km of river length (Figure 2).

4. Discussion

In this study, we presented GROD, a data set consisting of 30,549 river obstructions along the world's large rivers. Using publicly accessible satellite images available from Google, we were able to manually identify human-made structures obstructing river flow between 2018 and 2021. 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. 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/global datasets and can serve as a baseline for regional-scale datasets, especially in places where such data are not publicly available. Indeed, GROD contains

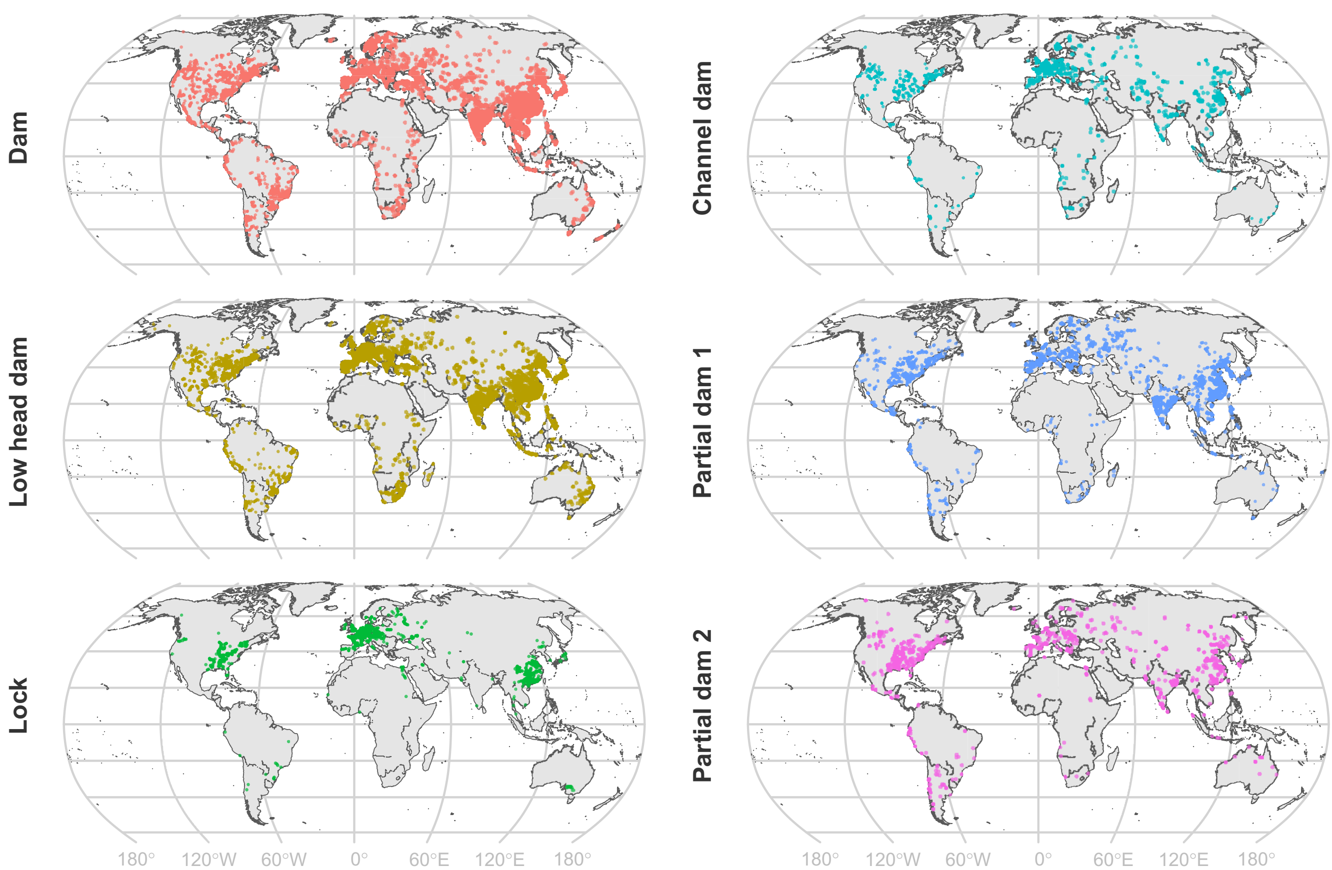

Figure 1. Spatial distribution of each obstruction type in GROD: dam, low head dam, lock, channel dam, and two types of partial dams.

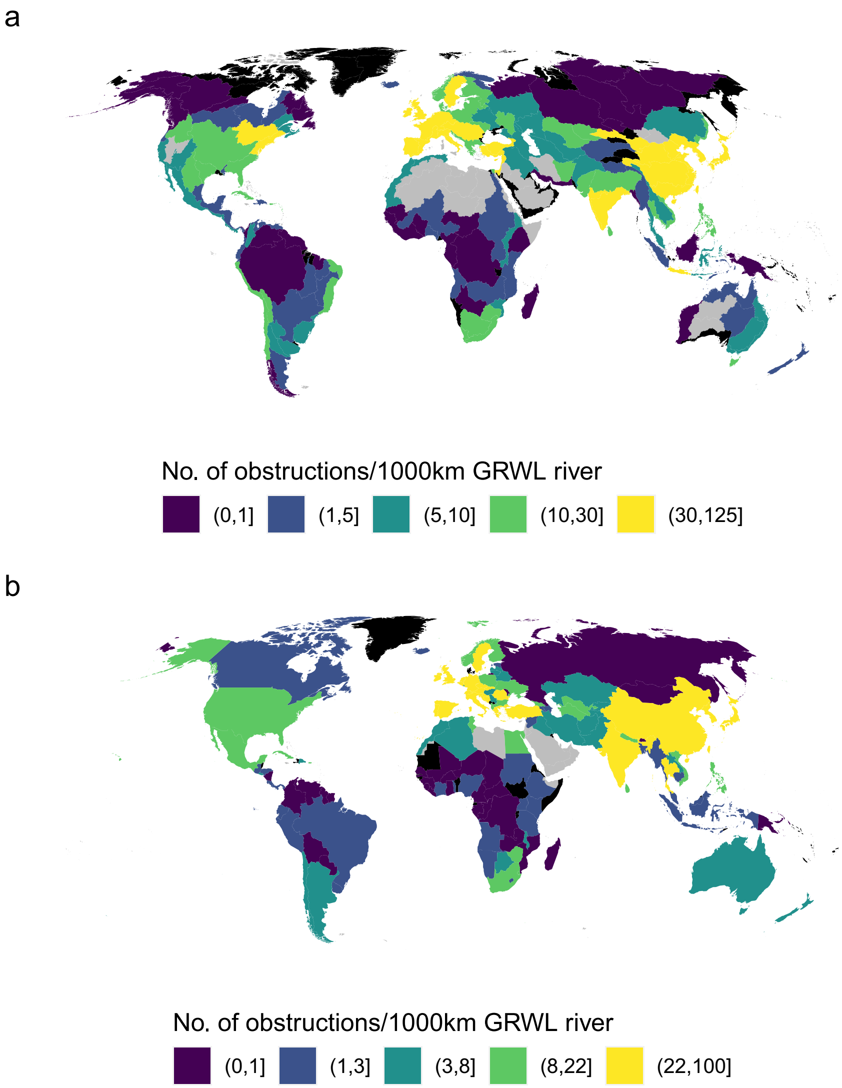

Figure 2. Obstruction density for level-3 HydroBASINS (a) and for country level (b) based on Group I obstruction locations and GRWL river lengths. Only Group I obstructions were used for calculating density on this map. For results based on all six types of obstructions, see Figures S5 and S6 in Supporting Information S1. Gray color indicates areas with no GRWL rivers and black color indicates areas with no obstructions. When looking at river obstruction density at the country level (b), the spatial patterns are similar to those based on drainage basins (a). The majority of the top 10 countries are located in Europe, with the exception of Japan (rank 2) and South Korea (rank 7) (Figure S7 in Supporting Information S1). 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 three countries being China, India, and the United States (Figure S7 in Supporting Information S1). For rankings based on all six types of obstructions, see Figure S8 in Supporting Information S1.

8,864 dams, which is substantially greater than the dam counts along GRWL rivers from the other global datasets we have compared to (from 579 in OSM to 1,535 in GRanD, see Table S2 in Supporting Information S1). The successful coordination and development of GROD demonstrate the effectiveness and efficiency of conducting mapping using publicly available cloud-based resources. This appraoch 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 data set. 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 (exact time frame of image capturing is unknown and likely to be close to the mapping period), instead of as the accurate instantaneous capture of the year of the published data set. 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) and because classification of an obstruction to a type is subjective and particularly challenging when structures are small relative to the resolution of the imagery. Additionally, the choice of GRWL as the reference river database can introduce bias toward including rivers of larger sizes. GRWL captures wider rivers (width $\geq$90 m) well, but narrower rivers (<90 m) are more likely to be missed in GRWL due to the limited Landsat resolution, difficulty in resolving narrow open water channels, and the strong seasonality in river widths in some regions.

On GRWL rivers, we estimated that GROD captured 83%–97% of obstructions by comparing it with five external databases of river obstructions. Although we did not systematically investigate causes of the missing obstructions, obstructions are likely to be missed when tracing along river centerline, which involves frequent zooming in and out. Although rare, omission can also occur along an entire reach or in an entire mapping tile. To mitigate these situations, we visually checked abnormal patterns in obstruction distribution on GRWL rivers and among mapping tiles, which can be challenging to do for places where obstructions are not abundant (e.g., in the tropical rainforest, northern high latitudes etc.). However, our estimation shows that we are less likely to miss large obstructions, as demonstrated by the much smaller omission rate (~7% and ~3%) when comparing GROD to dams in GRanD v1.3 and GOODD respectively. The missing rates provided likely represent upper bounds for how likely GROD misses obstructions, especially for places that external databases have cover well, because missed obstructions have been added to GROD after the evaluation.

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 S3 in Supporting Information S1), 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. This relatively low classification consistency can be generalized to all Group II types (Table S1 in Supporting Information S1), based on which we would caution the data users when using the Group II data. Inconsistency between repeated classifications could also arise from different participants involved in the initial global data set development (9 participants) and those who conducted the evaluations (3 participants for intermediate evaluation and 2 participants for the final evaluation). However, reclassification of Group I obstructions helped mitigate the potential biases and mistakes among 9 participants when assigning types (Table 1), which could still be present in Group II obstructions. Thus, we would recommend potential users to either merge the Group II types or to make manual checks when using the default typology. While it might be challenging to differentiate among all six types of obstructions (overall accuracy: 66%), the distinction between Group I and Group II is quite accurate (overall accuracy: 91%). Here overall accuracy was used to contrast the difference before and after grouping; balanced accuracy was not used as it is undefined as a single metric for classification of more than two types. And the advantage of having a finer typology, as we have in GROD, 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 a participatory approach, are not linked explicitly to a consistent river network. This in itself is not problematic, and should not affect completeness of the obstruction data, 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 data set 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 data set 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 archived image files for each record in GROD as png files after finalizing GROD, which are publicly available along with the GROD data set. 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. The development of GROD demonstrates that large-scale mapping projects can be planned and implemented efficiently thanks to recently available cloud-based geospatial platforms such as Google Earth Engine and to help from community scientists.

Data Availability Statement

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. GRanD v1.3 was accessed from Global Dam Watch website (http://globaldamwatch.org/grand/). 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 purposes, and will be deposited permanently for public access in Zenodo upon paper acceptance. The GROD data set can be accessed on Zenodo (https://zenodo.org/record/5793918).

Acknowledgments
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).

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. https://doi.org/10.3390/rs12091510

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. D. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water Resources Research*, *57*(7), e2021WR030054. https://doi.org/10.1029/2021wr030054

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. https://doi.org/10.1038/s41586-020-3005-2

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. https://doi.org/10.1007/s10712-015-9346-y

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*, *10*, 708–716. 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. https://doi.org/10.1016/j.rse.2017.06.031

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. https://doi.org/10.1029/1999wr900016

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. https://doi.org/10.1038/s41586-019-1111-9

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

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. https://doi.org/10.1890/120168

Jones, J., Börger, L., Tummers, J., Jones, P., Lucas, M., Kerr, J., et al. (2019). A comprehensive assessment of stream fragmentation in Great Britain. *Science of the Total Environment*, *673*, 756–762. https://doi.org/10.1016/j.scitotenv.2019.04.125

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. https://doi.org/10.1002/fee.2049

Latrubesse, E. M., Arima, E. Y., Dunne, T., Park, E., Baker, V. R., d'Horta, F. M., et al. (2017). Damming the rivers of the Amazon basin. *Nature*, *546*(7658), 363–369. https://doi.org/10.1038/nature22333

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. https://doi.org/10.1890/100125

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*, (9), 494–502.

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.

Mulligan, M., van Soesbergen, A., & Sáenz, L. (2020). GOODD, a global dataset of more than 38,000 georeferenced dams. *Scientific Data*, *7*(1), 31. https://doi.org/10.1038/s41597-020-0362-5

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. https://doi.org/10.1073/pnas.1423812112

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. https://doi.org/10.3390/su13020527

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. *Science of the Total Environment*, *299*(1–3), 21–36. https://doi.org/10.1016/s0048-9697(02)00222-x

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

References From the Supporting Information

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. https://doi.org/10.1007/978-3-642-04930-9_46

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing/IEEE Computer Society*, *7*(4), 12–18. https://doi.org/10.1109/mprv.2008.80

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