GRILSS: Opening the Gateway to Global Reservoir Sedimentation Data Curation

Sanchit Minocha and Faisal Hossain

Department of Civil and Environmental Engineering, University of Washington, Seattle, WA-98105, USA

5   Corresponding Author: Faisal Hossain, fhossain@uw.edu

Abstract. Reservoir sedimentation poses a significant challenge to freshwater management, leading to declining storage capacity and inefficient reservoir operations for various purposes. However, trustworthy and independently verifiable information on declining storage capacity or sedimentation rates around the world is sparse and suffers from

10   inconsistent metadata and curation to allow global-scale archiving and analyses. The **G**lobal **R**eservoir **I**nventory of **L**ost **S**torage by **S**edimentation (GRILSS) dataset addresses this challenge by providing organized, well-curated and open-source data on sedimentation rates and capacity loss for 1,015 reservoirs in 75 major river basins across 54 countries. This publicly accessible dataset captures the complexities of reservoir sedimentation, influenced by regional factors such as climate, topography, and land use. By curating the information from numerous sources with disparate

15   formats in a homogenized data structure, GRILSS serves as an invaluable resource for water managers, policymakers, and researchers for improved sediment management strategies. The open-source nature of GRILLS promotes collaboration and contributions from the global community to grow the dataset. By providing essential reference data on sedimentation to understand the global challenge of reservoir sedimentation, this GRILLS dataset represents a gateway for the global community to share sedimentation and storage loss data for sustainable operation of world's

20   reservoirs for future generations.

1. Introduction

Artificial reservoirs created by damming a river or a lake are essential components of modern water management systems. Such reservoirs provide critical freshwater resources for irrigation, domestic and industrial use, as well as generate hydropower and moderate floods. However, these dams and their reservoirs pose significant environmental

25   challenges, including the destruction of ecosystems, loss of biodiversity, disruption of aquatic life, greenhouse gas emissions, and barriers to fish migration. Despite these drawbacks, numerous reservoirs remain in operation, as they can fulfill multiple essential functions for society. As dams age, the effectiveness of reservoirs becomes increasingly compromised by sedimentation, which gradually reduces their water storage capacity and threatens their long-term functionality (Kondolf et al., 2014 and Wisser et al., 2013).

30

Hereafter, the terms dam and reservoir will be used interchangeably as shorthand for the dam-reservoir system. Despite the ongoing planning and construction of new dams worldwide (Zarfl et al., 2015), global reservoir capacity peaked in 2006 and has apparently been declining ever since (Wisser et al., 2013). This paradox is primarily due to the accumulation of sediment at the bottom of reservoirs, which gradually reduces their storage capacity. Wisser et al.

35 (2013) estimated that global per capita reservoir capacity declined by 19% between 1987 and 2010, and when sedimentation is factored in, the decrease is even more pronounced at 21%. Moreover, most of the optimal sites for dam construction have already been utilized, and the remaining feasible locations are often less economical (Morris & Fan, 1998). As a result, the new dams being built are generally smaller (Demirbas, 2009), and the rate of construction is not keeping pace with the rate of sedimentation. Overall, the current trend highlights the pressing need

40 to address reservoir sedimentation to ensure a sustainable supply of freshwater for energy, food and flood control for future generations.

Regaining the storage capacity lost to sedimentation is essential for halting the global decline in reservoir capacity without having to build costly, uneconomical and ecosystem-damaging newer dams. One of the primary solutions is

45 to implement sediment management techniques, which requires a thorough understanding of sedimentation rates in reservoirs worldwide. Since sedimentation rates vary depending on regional factors, conducting localized studies is crucial for accurately assessing these rates. Topography, geology, vegetation and precipitation can affect the rate of soil erosion which can further affect sedimentation rates and consequently the rate of storage capacity loss. Due to the interplay of vast number of other factors such as operation procedure, reservoir shape and purpose, sedimentation rate

50 can also be significantly different from one neighboring reservoir to another.

Sedimentation rates in reservoirs can be estimated using either direct or indirect methods (Nyikadzino & Gwate, 2021). Direct methods, such as bathymetric surveys, involve measuring the reservoir's depth over time and comparing the surveys to estimate storage loss, which is attributed to sediment deposition. These methods tend to be most accurate

55 but are also more costly (Samaila–Ija et al., 2014; Patil et al., 2015). On the other hand, indirect methods use satellite remote sensing or numerical modeling to estimate sedimentation rates (Elçi et al., 2009; Hanmaiahgari et al., 2018; Pandey et al., 2016; Zhang et al., 2006). While indirect methods are less expensive, they rely on the availability of in-situ data, such as soil moisture and soil type for numerical models, or reservoir water elevation data, often obtained through in-situ or altimetry-based measurements, for remote sensing applications.

60

Due to these limitations, only a small fraction of the world's millions of reservoirs have been thoroughly assessed or gauged for sedimentation rates, and even fewer are well-documented in published literature. Countries like the United States and India have made efforts to compile sedimentation databases for a limited number of reservoirs within their national boundaries (Gray et al., 2010; Central Water Commission, 2020). However, despite these initiatives, there

65 remains no comprehensive, well-documented global dataset on reservoir sedimentation rates or storage capacity loss. While numerous regional studies have analyzed sedimentation rates for individual reservoirs, a unified, well-curated and publicly available global dataset capturing sedimentation rates and available storage capacity across the world remains absent. This lack of a cohesive data archive for the global community presents a major challenge in addressing and understanding reservoir sedimentation on a global scale (Syvitski et al., 2003; Vanmaercke et al., 2011; Wisser et

70 al., 2013; Perera et al., 2023).

In this paper, we address the need for a comprehensive, consistently curated and publicly available dataset that compiles the loss of storage capacity due to sedimentation in reservoirs worldwide. We introduce **G**lobal **R**eservoir **I**nventory of **L**ost **S**torage by **S**edimentation (GRILSS; Minocha & Hossain, 2024) dataset, the first open-source dataset of its kind, to the best of our knowledge. GRILSS includes information on over 1,000 reservoirs, detailing the amount of sedimentation and lost capacity over specific time intervals along with various meta data details. Additionally, GRILSS provides other attributes such as the height of impoundment, catchment area, and original built capacity of the reservoirs, where available.

GRILSS has the potential to be highly valuable for the global freshwater community, offering insights into sedimentation patterns in reservoirs around the world. By providing sedimentation rates and their variations across different river systems, GRILSS can potentially enhance our understanding of how sedimentation processes differ regionally. GRILSS can inform sediment management strategies by identifying how these techniques should be tailored to specific regions. As the first version of its kind, we anticipate that GRILSS will continue to grow over time, both in size and use, with future contributions on sedimentation data from the global community to expand the global coverage of reservoirs.

2. Dataset Preparation

2.1 Collection of sedimentation data

The GRILSS dataset (Minocha & Hossain, 2024) was compiled through an extensive literature review of published articles, theses, government documents, and websites, some of which were in languages other than English (Estrada et al., 2015). After carefully reviewing hundreds of papers and documents, the dataset was created by manually compiling data from 143 different sources using Microsoft Excel. Only sources that contained specific sedimentation data related to reservoirs were included. This thorough process ensured that the dataset covers sedimentation information from a wide range of regions.

Since sedimentation can be reported in various ways, we tracked metrics from different sources. Many sources directly provided sedimentation amounts in million cubic meters (MCM) along with the duration over which the sedimentation occurred. However, some sources reported sedimentation in million tons (MT) (Kim et al., 2014; Palinkas & Russ et al., 2019). Additionally, certain sources mentioned the percentage of capacity loss over a specific period but did not always include the reservoir's original capacity (Güvel, 2021; Patro et al., 2022). A few sources only provided the sedimentation rate (either in MCM or MT), with or without specifying the time frame used to calculate it (Lorsirirat, 2014; Yang et al., 2022). Due to this variability in reporting, the GRILSS dataset had to be carefully compiled and standardized to ensure consistency of data structure across the dataset (Minocha & Hossain, 2024).

For sources where the sedimentation amount was reported in million tons (MT), the bulk sediment density was obtained either from the same source or a comparable source for that region. Using this bulk density, the sedimentation volume was then estimated in million cubic meters (MCM) with the help of Eq. (1). In cases where only the

sedimentation rate was available, the rate was assumed to represent the sedimentation amount over a one-year period using Eq. (2). Unless otherwise specified, this duration was assumed representative from two years before the source's publication date to one year afterwards. For other reservoirs, the sedimentation rate was also estimated using Eq. (2).

$$Sedimentation\ volume\ (MCM) = \frac{Sediment\ weight\ (MT)}{Bulk\ sediment\ density\ \left(\frac{ton}{m^3}\right)} \tag{1}$$

$$Annual\ sedimentation\ rate\ \left(\frac{MCM}{year}\right) = \frac{Sedimentation\ volume\ (MCM)}{Time\ Duration\ (year)} \tag{2}$$

For cases where sedimentation was reported as a percentage of capacity loss, but the reservoir's original capacity was not provided, the initial capacity was extracted from other datasets or references. The sedimentation volume was then calculated using Eq. (3). Equation (3) was also used to calculate the percentage of capacity loss when sedimentation volume and original built capacity were available. The annual rate of reservoir capacity loss, expressed as a percentage of the original built capacity was estimated for all reservoir using Eq. (4). This approach ensured consistency in estimating sedimentation volumes, annual sedimentation rates, and capacity loss percentages across various reporting formats. However, sedimentation weight in million tons (MT) was not estimated for all reservoirs due to insufficient bulk sediment density data.

$$Sedimentation\ volume\ (MCM) = \frac{\%\ Capacity\ loss * Original\ built\ capacity(MCM)}{100} \tag{3}$$

$$Annual\ capacity\ loss\ rate\ \left(\frac{\%}{year}\right) = \frac{\%\ Capacity\ loss}{Time\ Duration\ (year)} \tag{4}$$

2.2 Collection of reservoir attributes

From each source, the reservoir name and the country of location were extracted and manually compiled in Microsoft Excel along with the sedimentation data. Other reservoir attributes, such as impoundment height, year of construction, original capacity, and catchment area, were also recorded when available.

While geospatial coordinates of the reservoir were rarely provided from most sources, they were included in the GRILSS dataset whenever available. In many cases, sources provided only a figure showing the approximate location of the reservoir, making it difficult to derive precise geospatial coordinates. Additionally, some sources provided only the extent of the reservoir rather than the exact coordinates of the dam impoundment. Out of the over 1,000 reservoirs for which data was collected, only around 170 had geospatial coordinates of the dam.

Since a reservoir database without georeferencing is of limited utility for analyses, addressing the challenge of mapping the reservoir attributes to their corresponding dam locations became crucial. To accomplish this, the Global Dam Tracker (GDAT), which contains more than 35,000 georeferenced dams with attributes, was utilized (Zhang &

140   Gu, 2023). The reservoir names in GRILSS were fuzzy-matched to the dam names in GDAT using the Python library "thefuzz" (Seatgeek, 2021). This method was used to account for minor variations in the spelling of reservoir names across different regions, where slight differences in spelling could prevent a direct match.

For each reservoir in the GRILSS dataset, the top 10 dam name matches, based on fuzzy-matching scores, were
145   extracted from the GDAT dataset. Since there could be multiple dams or reservoirs with the same name, a manual review of the matches was conducted. This involved comparing other attributes such as year of construction, original capacity, and country for both datasets to ensure consistency. If these attributes aligned, the match was confirmed. Through this process, we successfully identified the corresponding dams for 613 GRILSS reservoirs from the GDAT dataset.

150

In cases where attributes like impoundment height, year of construction, or original built capacity were missing in GRILSS but available in GDAT, these values were added to the dataset. However, geospatial coordinates were not always available, even for dams listed in GDAT. For all remaining reservoirs lacking dam coordinates, geospatial data was manually derived using Google Earth and Google Maps, by searching for the reservoir name in the specified
155   region as indicated by the original source. When approximate location maps were available, these figures were overlaid on actual maps with proper scaling to accurately determine the impoundment locations. Through this process, all reservoirs in the GRILSS dataset were successfully georeferenced (Minocha & Hossain, 2024).

2.3 Collection of other attributes

To increase the utility of GRILSS, additional attributes were added to the dataset. One key attribute added from each
160   source was the method used to estimate sedimentation for a given reservoir, categorized into three main types: (1) bathymetric survey, (2) remote sensing, or (3) numerical modeling.

1   Bathymetric survey (BS): This method involves repeating a survey after a specific time period and comparing it with previous surveys. The observed loss in reservoir capacity is attributed to sedimentation volume (Stauch et
165   al., 2024; Ugwu et al., 2021).

2   Remote sensing (RS): In this method, water elevation data is collected through altimetry or in-situ measurements, while the surface area of the reservoir is determined using optical satellite imagery. By combining these, an area-elevation curve of the reservoir is generated, which helps calculate the current capacity. The difference between this capacity and the original built capacity indicates the sedimentation volume (Pandey et al., 2016; Zhang et al.,
170   2006).

3   Numerical modeling (NM): Sedimentation estimates can also be derived through hydrological and sediment models. These models use in-situ data such as inflow, sediment concentration, and sediment density to simulate sediment transport and deposition in the basin, providing an estimate of sedimentation (Elçi et al., 2009; Hanmaiahgari et al., 2018).

175

Another attribute incorporated into the GRILSS dataset is the classification of reservoir storage type, indicating whether the sedimentation volume was estimated for live storage or gross storage for each reservoir. Additionally, Google Maps was utilized to verify whether the reservoir had been removed and to determine if it was associated with a creek dam or a dry dam. The website location (i.e., Uniform Resource Locator URL) for each reservoir's original data source was also documented in the dataset. For some reservoirs located in Zambia, the built year was roughly estimated based on sedimentation and capacity loss rates, with these assumptions noted in the comment section.

To ensure compatibility of the GRILSS dataset with one of the most widely used dam datasets, the Global Reservoir and Dam database (GRanD), a comparison was conducted to identify common reservoirs in both datasets (Lehner et al., 2011). This involved creating a 100-meter buffer around the dam locations in the GRILSS dataset. Dams within this buffered radius were then matched to those in the GRanD database.

As a result of the above data curation practices, the GRILSS dataset can now be effectively utilized alongside both the GDAT and GRanD datasets, with increased utility for research and analysis related to reservoir sedimentation, operations and management.

2.4 Reservoir polygons

Reservoir sedimentation varies significantly from one reservoir to another, with one key factor being the shape of the reservoir. Generally, sedimentation is higher when the reservoir's shape complexity—defined as the ratio of shoreline length (or wetted perimeter) to surface area—is greater. Conversely, reservoirs with a larger surface area but a lower wetted perimeter, indicating low shape complexity, tend to experience less sedimentation (Kantoush & Schleiss, 2014). Given the importance of reservoir geometry in sedimentation studies, we undertook efforts to extract the shape polygons for each reservoir in the GRILSS dataset. These shape polygons have the potential to provide valuable insights for future analyses by the global water community.

OpenStreetMap (OSM) was used to extract reservoir geometry shapes for each reservoir based on the dam locations (OpenStreetMap, 2017). The Overpass API was employed to fetch shapes within a 1000-meter radius of each dam. To ensure that only reservoir shapes were selected, the results were filtered based on specific tags such as "natural" = "water," "water" = "reservoir," or "water" = "lake" (Overpass, 2024). When multiple shapes were retrieved, the shape closest to the dam's point location was selected.

When applying the initial strategy, no reservoir shapes were found for around 300 reservoirs. Upon closer inspection in QGIS (QGIS, 2024), with OSM as the basemap, it was discovered that for 200 of these reservoirs, the dam coordinates—mostly sourced from GDAT—were significantly misplaced. After manually correcting these coordinates, reservoir shapes were re-extracted from OSM. However, for around 150 reservoirs, the shape remained hard to retrieve.

A quality check was then performed by reviewing each reservoir shape in QGIS using OSM and Google satellite imagery. It was found that for 150 reservoirs, no reservoir polygons existed in OSM, while for 20 reservoirs, the shapes were incomplete, and for 35 reservoirs, the wrong shapes were selected. Each of these cases was manually

215 handled for quality control. For reservoirs with incorrect shapes, the correct OSM IDs were identified using the OSM webapp, and the shapes were re-fetched. For incomplete shapes, multiple OSM shapes were joined to create a single accurate geometry.

For reservoirs without any polygons in OSM, HydroLAKES was used to match dam locations to lake geometries, providing shapes for about 50 reservoirs (Messager et al., 2016). For the remaining reservoirs, the Global Surface

220 Water Occurrence dataset (Pekel et al., 2016) was used, but the resulting shapes, derived from raster data, were pixelated. These shapes were manually refined to improve the overall quality of the GRILSS dataset. It is important to note that no reservoir geometry could be derived for dry dams, creek dams, or dams that had already been removed.

2.5 Catchment polygons of reservoirs

225 A major factor that directly impacts reservoir sedimentation is the catchment of the reservoir. Key catchment characteristics such as area, slope, land use, and land cover play a significant role in determining sedimentation rates. To enable users to assess these characteristics, we focused on extracting the catchment polygons for all reservoirs, as these shapes provide valuable insights for sedimentation analysis.

230 The Watershed (Spatial Analyst) tool in ArcGIS Pro was used to delineate catchments, utilizing a 90 m digital elevation model. To automate this process for over 1000 reservoirs, a geoprocessing workflow was created using Model Builder, a visual programming language. Snap distances were adjusted based on the original built capacity of the reservoirs, with larger reservoirs using larger snap distances (e.g., 1200 meters for those over 1 MCM and 50 meters for those under 0.1 MCM).

235

However, for a few reservoirs, particularly those in Taiwan, this method was unsuccessful. In those cases, catchment polygons were created using the Python library "pysheds" (Bartos, 2020) and HydroSHEDS flow direction data at 3 arc-seconds resolution (Lehner et al., 2008). Through this combined approach, catchment geometries were successfully obtained for all the reservoirs in the GRILSS dataset, which were then used to calculate catchment area

240 when it was not available from the original source.

3. Dataset Description

The Global Reservoir Inventory of Lost Storage by Sedimentation (GRILSS) dataset contains sedimentation data for over 1000 reservoirs worldwide. Compiling this dataset took sustained effort and methodical work over a year, involving detailed manual corrections, quality assurance, and quality control to ensure its accuracy. To the best of our

245 knowledge, GRILSS is the first dataset of its kind, aimed at supporting global studies on reservoir sedimentation and its impact on water storage. The dataset is composed of four files. The main file, an Excel spreadsheet, contains

Earth System
Science
Data
Open Access
Discussions

historical records of reservoir sedimentation, with each record assigned a unique Observation ID (OID). A single reservoir can have multiple records, reflecting sedimentation data over various time periods, and each reservoir is uniquely identified by a Reservoir ID (RID). The remaining three files consist of vector data—available in both

250   shapefile and GeoJSON formats—representing the dam, reservoir, and catchment geometries for each reservoir.

3.1 Sedimentation Dataset

The Excel file GRILSS_data.xlsx contains sedimentation data for various reservoirs. The dataset includes 35 fields describing reservoir locations, characteristics, and the sedimentation that occurred over specific time periods. Each record also provides the original source URL, allowing users to explore or learn more about the reservoir or the

255   sedimentation estimation method in detail. Missing values in the dataset are represented by blanks or empty spaces.

One of the fields includes comments that provide additional insights into each record or reservoir, such as its length and the sediment management techniques used. As detailed in Sect. 2.3, the method for estimating sediment volume is specified under the 'Survey Type' field and is categorized as 'BS' for bathymetric surveys, 'RS' for remote sensing,

260   and 'NM' for numerical modeling methods. Fields that have been manually corrected, compared to the original source, are noted under the 'Fields Corrected' field. Table 1 outlines the details of each field in the dataset.

Table 1: Overview and Description of Fields in the GRILSS Dataset (Minocha & Hossain, 2024)

S. No.	Field Name	Field Description	Units
1.	GRILSS OID	A unique identifier for each sedimentation record in the dataset	-
2.	GRILSS RID	A unique identifier for each reservoir in the dataset	-
3.	Reservoir	The name of the reservoir	-
4.	Country	The name of the country where the reservoir is located	-
5.	HYBAS_ID	The HydroSHEDS HYBAS_ID for the major basin containing the reservoir	-
6.	Major River Basin	The name of the major river basin, unique to each HYBAS_ID	-
7.	Continent	The continent where the reservoir is situated	-
8.	Capacity Loss Rate (%/year)	Annual rate of reservoir capacity loss, expressed as a percentage of the original built capacity.	%/year
9.	Sedimentation Rate (MCM/year)	Annual rate of sediment deposition in the reservoir	Million cubic meters/year
10.	Capacity Loss (%)	The percentage of the original built capacity (in MCM) that has been lost due to sediment accumulation	Percentage

| 11.	Sedimentation Amount (MT)	The total amount of sediment deposited in the reservoir	Million tons
12.	Sedimentation Amount (MCM)	The volume of sediment deposited in the reservoir	Million cubic meters
13.	Sediment Bulk Density (ton/m³)	The estimated bulk density of the sediment deposited in the reservoir	Ton/m³
14.	Observed Duration (years)	The number of years over which the sedimentation amount was estimated	Years
15.	Observation End Year	The year when the observed duration of sedimentation ended	-
16.	Observation Start Year	The year when the observed duration of sedimentation began	-
17.	Built Year	The year in which the reservoir was constructed	-
18.	Original Built Capacity (MCM)	The reservoir's capacity at the time of construction	Million cubic meters
19.	Catchment Area (km²)	The area of the reservoir's catchment	Square kilometers
20.	Height (m)	The height of the reservoir's impoundment (dam)	Meters
21.	Latitude	The latitude coordinate of the reservoir's dam	Degrees
22.	Longitude	The longitude coordinate of the reservoir's dam	Degrees
23.	Survey Type	The method used to estimate sedimentation volume in the reservoir [BS, RS, NM]	-
24.	Type of Storage	The type of reservoir storage for which sedimentation volume is estimated [Gross or Live]	-
25.	Comments	Additional information regarding the reservoir or the sedimentation estimation method	-
26.	Original Source URL	The web URL to the original source from which the record was extracted	-
27.	Fields Corrected	Names of fields that were corrected or added after manual inspection for the specific record	-
28.	Fields Computed	Names of fields that were computed for the specific record	-
29.	GDAT Dam Name	The name of the dam associated with the reservoir in the GDAT dataset	-
30.	GDAT Feature_ID	The Feature_ID corresponding to the reservoir in the GDAT dataset	-

31.	GRAND_ID	The GRAND_ID corresponding to the reservoir in the GRanD database	-
32.	GRAND Wrong Location	A binary flag indicating if the dam coordinates are incorrect in the GRanD dataset	-
33.	GDAT Wrong Location	A binary flag indicating if the dam coordinates are incorrect in the GDAT dataset	-
34.	Dam Removed or Dried	A binary flag indicating if the reservoir has been removed or is dried	-
35.	Creek Dam	A binary flag indicating if the dam is classified as a creek dam	-

3.2 Vector Dataset

265   In addition to sedimentation data, the GRILSS dataset also includes associated vector data, as described in Sect. 2. This vector data comprises three files: 1. point locations of the dams associated with the reservoirs, 2. reservoir geometry polygons, and 3. catchment geometry polygons. Each vector file contains attributes such as GRILSS RID, Reservoir name, and Country. Notably, the catchment vector file includes computed surface area for the reservoir catchments, which may differ from the catchment area reported in the sedimentation data derived from the original
270   source.

4. GRILSS Dataset Overview

The GRILSS dataset contains 1368 records for 1015 reservoirs worldwide. These reservoirs span 75 major river basins and 54 countries, representing a wide range of catchment types with varying physical characteristics such as slope, climate, land use, land cover and soil type. Figure 1 highlights the 15 countries and major river basins with the highest
275   number of reservoirs. The dataset also covers a wide range of dam construction periods, with few reservoirs dating back to the 1700s and 1800s. Most reservoirs, however, were constructed between 1950 and 2000, as shown in Fig. 2(b). The oldest reservoir in the dataset, Pareton in Spain, was built in 1713 (Erena et al., 2019) and has a capacity loss rate of 0.3% per year.
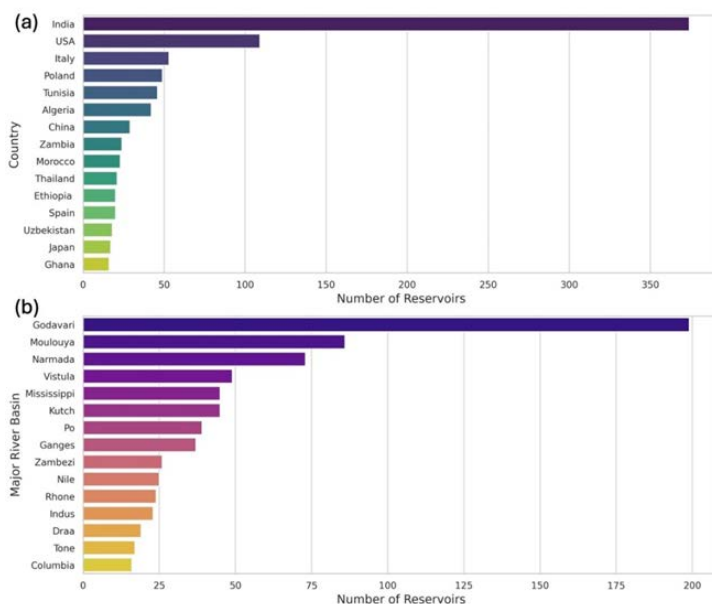
Figure 1: The top fifteen (a) countries and (b) major river basins having the highest number of reservoirs in the GRILSS dataset.
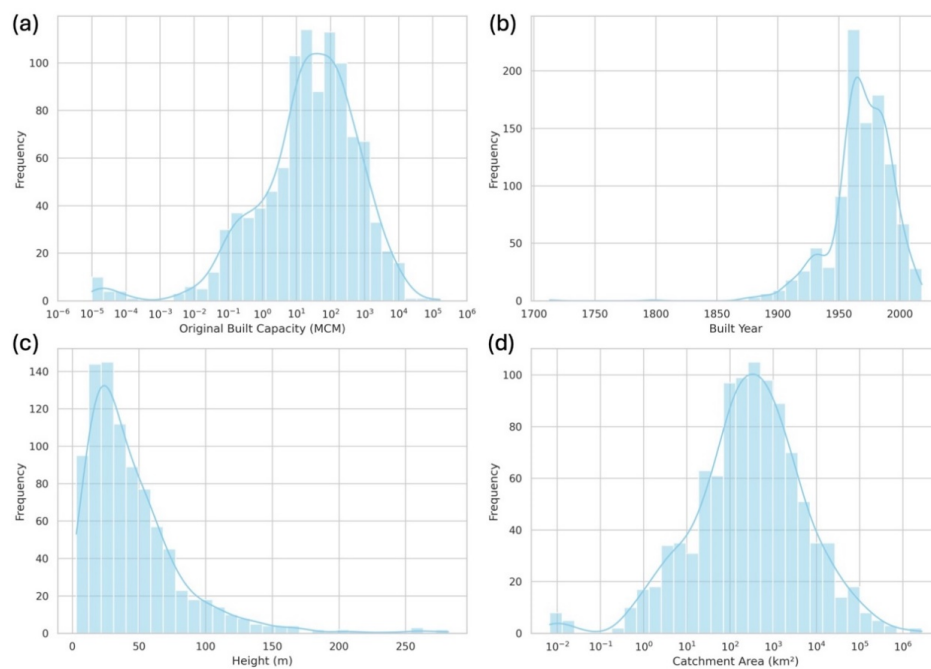

Figure 2: Frequency distribution of reservoir attributes: (a) original built capacity, (b) construction year of dams, (c) dam height, and (d) catchment area. Each panel shows the distribution of these characteristics across the reservoirs in the dataset, with the blue line representing the kernel density estimate (KDE) of the distribution.

285   In terms of survey methods, the dataset includes 1182 records based on bathymetric surveys (BS), 175 from remote sensing (RS), and 11 from numerical modeling (NM). The dataset also captures considerable variation in terms of the original built capacity of reservoirs, though reservoirs with capacities below 0.1 MCM are underrepresented (Fig. 2(a)). It includes 21 creek reservoirs, small dams constructed across creeks, all located in the United States. The smallest reservoir, a creek sediment dam, has an original built capacity of just 9.9 m³. Additionally, 13 reservoirs in

290   the dataset, all in the U.S., have either been removed or are part of dry dams. The largest reservoir in the GRILSS dataset is Lake Nasser in Egypt, with an original built capacity of 162 km³. Lake Nasser has one of the lowest capacity loss rates in the dataset, at 2.3×10^{-4} % per year.

The height of the dam is available for only 883 reservoirs in the dataset, collected from various sources as it was not

295   consistently provided in the original data source. Figure 2(c) displays the frequency distribution of dam heights, with most dams being under 75 meters tall. The average height in the dataset is 45.58 meters. Cerro Prieto in Mexico, the tallest dam in the GRILSS dataset, stands at 282 meters with a capacity loss rate of 0.31% per year. In terms of catchment area, the dataset includes reservoirs with catchment sizes ranging from less than 1 km² to over 0.1 million km². The distribution of catchment areas follows a normal pattern, with a mean of 14,759 km² (Fig. 2(d)).

300   **4.2 Capacity loss rate**

The capacity loss rate was first averaged across all records for each reservoir. From these reservoir-specific averages, the overall mean annual capacity loss rate for all reservoirs was determined to be 1.62% per year. Mayfair Lake in the USA had the highest observed loss rate at 102.24% per year. Some reservoirs even reported negative capacity loss rates due to sediment management practices like dredging (Central Water Commission, 2020). Figure 3 shows the

305   distribution of dams in GRILSS dataset and provides an overview of the rates at which reservoirs are losing capacity across different regions of the world. Figure 4(a) presents a histogram of the capacity loss rates, showing that the majority of reservoirs, 745 in total, have a loss rate of less than 1% per year.
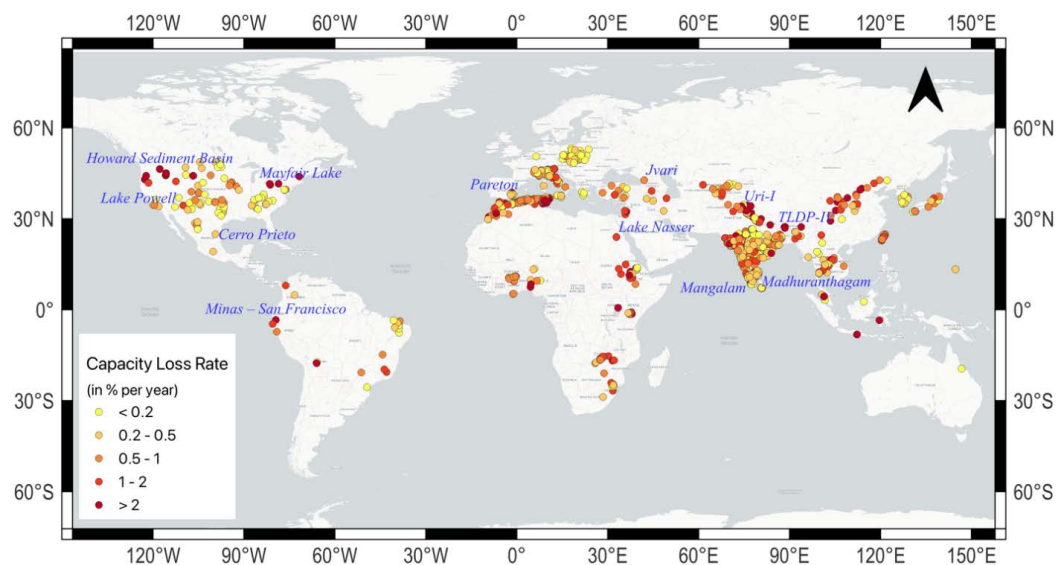
Figure 3: Map displaying the locations of all reservoirs in the GRILSS dataset. Marker colors represent the capacity loss rate (% per year), with darker shades indicating higher loss rates. Names of select reservoirs referred to in the text or other figures are also labelled for additional information.

4.2.1 Capacity loss rate by built capacity

Of the 1015 reservoirs, 174 (17.14%) have an initial built capacity of less than 1 million cubic meters (MCM), exhibiting an average capacity loss rate of 5.4% per year. This higher rate is largely attributed to creek dams, which can experience capacity loss rates as high as 132% per year due to their small size and short observed durations (around 1 year). The remaining 82.86% of reservoirs, with an initial capacity of 1 MCM or more, have an average capacity loss rate of 0.84% per year, consistent with global sedimentation rates reported by Basson (2009). Figure 4 (b) shows the distribution of capacity loss rates for small (<1 MCM) and large reservoirs (≥1 MCM), while Table 2 summarizes these statistics for the GRILSS dataset.
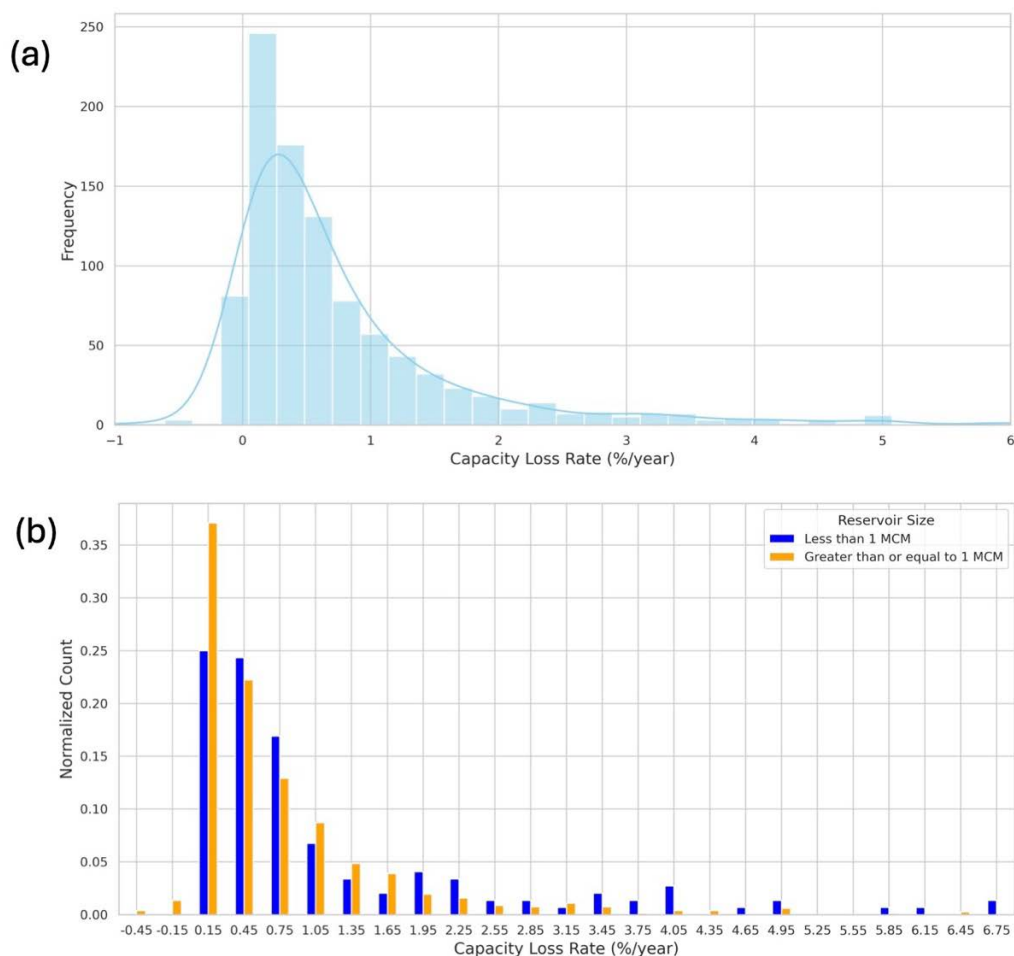
Figure 4: (a) Frequency distribution of capacity loss rates for all reservoirs, with the blue line representing the kernel density estimate (KDE). (b) Distribution of capacity loss rates for reservoirs with an original built capacity of less than 1 MCM compared to those with a capacity of 1 MCM or more, normalized by the total count in each category.

Table 2: Statistics of annual capacity loss rate expressed as percentage of initial capacity of the reservoir (% a⁻¹)

	Number of reservoirs	Min	Average	Median	Capacity-weighted mean	Max
Capacity <1 MCM	174	0.0	5.4	0.76	1.2	102.24
Capacity >1 MCM	841	-1.91	0.84	0.41	0.42	21.03

4.2.2 Capacity loss rate by observed duration

The observed duration for each record represents the time period over which sedimentation rates were estimated.

330 Similar to the capacity loss rate, this duration was averaged across all records for a reservoir. As described in Sect 2.1, if only the capacity loss rate or sedimentation rate was provided without specific duration information, it was assumed to be one year.
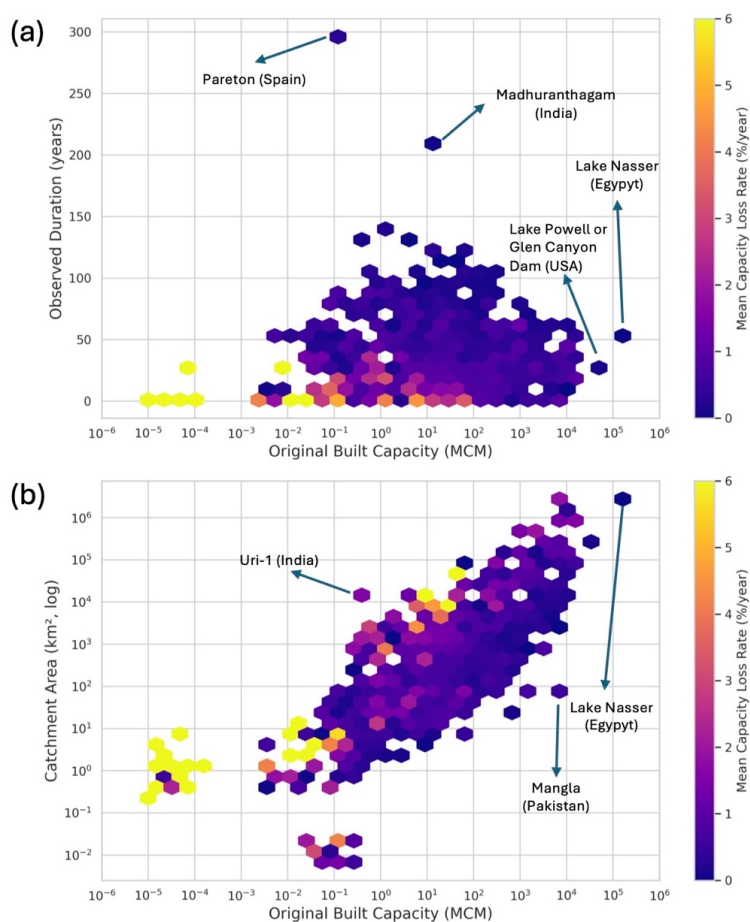

Figure 5: Density scatter plot showing hexagonal bins corresponding to specific (a) built capacity and observed duration, and (b) original built capacity and catchment area of the reservoirs. The color gradient represents the mean capacity loss rate for all reservoirs within each hexagonal bin in both panels. Name of the reservoir, along with the country is mentioned for few outliers, where there was only one reservoir in the bin.

Figure 5(a) illustrates that the capacity loss rate decreases as the observed duration increases. Additionally, for shorter observed durations, smaller reservoirs tend to exhibit higher capacity loss rates, while larger reservoirs show lower

340 rates. This relationship between observed duration and capacity loss rate may result from variable sedimentation rates

over time. Sedimentation rates are typically higher during significant storm events and floods; thus, a shorter observed
duration is more likely to encompass a substantial portion of these events, leading to increased capacity loss rates.

4.2.3 Capacity loss rate by catchment area

345   Larger reservoirs tend to have larger catchment areas, while smaller reservoirs typically have smaller catchment areas,
as illustrated in Figure 5(b). A similar relationship was observed by Walter et al. (2020) between catchment area and
the reservoirs in terms of surface area. This figure also reinforces earlier observation that smaller reservoirs experience
higher capacity loss rates. Notably, for a given built capacity, the capacity loss rate increases with the catchment area.
This trend may be attributed to a greater availability of soil for erosion in large catchment areas.

4.2.4 Capacity loss rate by height

350   Figure 6 indicates that dam height does not significantly influence the capacity loss rate. However, it is observed that
beyond a certain height—approximately 90 meters—the loss rate tends to remain low and does not reach high values.
This pattern may again be attributed to larger capacity dams, which are generally taller and, consequently, exhibit
lower loss rates.
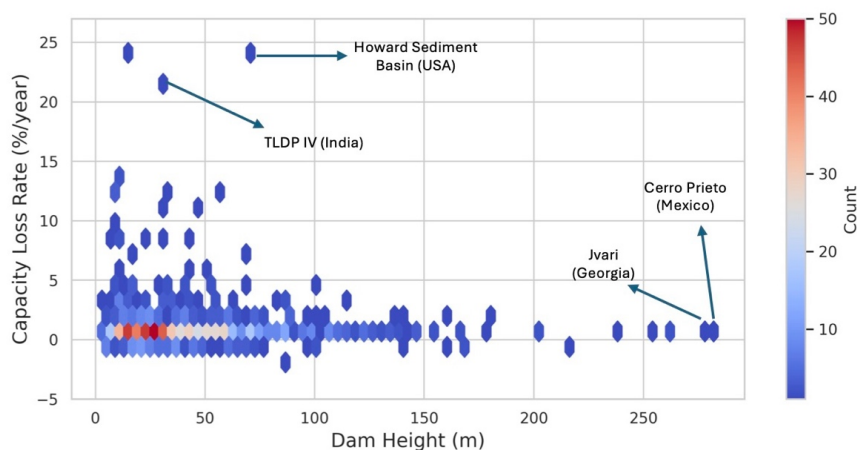
355   **Figure 6: Density scatter plot depicting the relationship between dam height (in meters) and capacity loss rate (% per year)
within each hexagonal bin. The color gradient indicates the number of reservoirs contained in each bin. Name of the
reservoir, along with the country is mentioned for few outliers, where there was only one reservoir in the bin.**

4.2.5 Capacity loss rate by built year

360   The GRILSS dataset includes two dams constructed in the 1700s, a few built in the 1800s, and many that extend into
the twentieth and twenty-first centuries. This diverse range of construction years allows for an exploration of the
relationship between the year a dam was built and its capacity loss rate. Figure 7 displays the median capacity loss
rate for each year of construction, revealing an intriguing trend: reservoirs built more recently tend to have slightly
higher capacity loss rates compared to older reservoirs. This subtle upward trend suggests that newly constructed
reservoirs may be losing capacity at an accelerated rate.

365   One possible explanation for this phenomenon is that sediment removal efforts are primarily focused on older
        reservoirs, while relatively newer ones have received less attention. Additionally, this long-term trend may be
        influenced by factors such as climate change and urbanization, and therefore needs to be further investigated.

Figure 7: Median capacity loss rate for each year of construction among the GRILSS reservoirs. The x-axis shows two
370   distinct kinks corresponding to the two reservoirs built in the 18th century. Reservoir names are indicated for those
        constructed in that century and for any that exhibit anomalously high capacity loss rates relative to the overall trend.

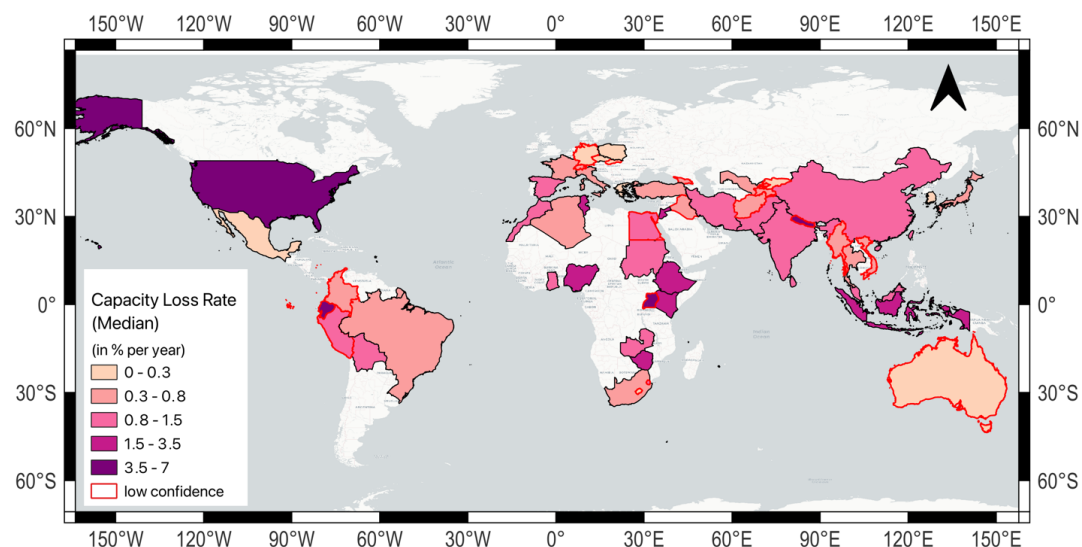### 4.2.6 Capacity loss rate by Country

        Total reservoir storage capacity is a critical concern for all countries as it determines the available freshwater for
        irrigation, domestic, and industrial use. To manage water resources effectively, it is essential for water managers to
375   understand the average capacity loss rate within their country. Figure 8 presents the median capacity loss rates for 54
        countries represented in the GRILSS dataset. Lighter shades, such as peach and light pink, indicate lower capacity
        loss rates (0–0.8% per year), while darker shades highlight countries experiencing higher losses (up to 7% per year).
        Countries where fewer than three reservoirs are available in the GRILSS dataset are outlined in red, marking the data
        for these areas as lower confidence. Notably, the United States and several countries in Southeast Asia show
380   significantly higher capacity loss rates, indicating increased sedimentation challenges in these regions. This
        observation overall aligns with the findings of Perera et al. (2023), who also reported the highest capacity losses in
        the United States.

4.3 Sedimentation rate

        While capacity loss rates are crucial for countries and water managers, understanding sedimentation rates is equally
385   important, as they indicate which regions are experiencing more intense sedimentation. As discussed in Sect. 1, factors
        such as a region's topography and climate play a key role in influencing sedimentation. For each reservoir, the
        sedimentation rate was averaged across all records. Figure 9 presents the distribution of sedimentation rates across the
        GRILSS dataset. Notably, 17 reservoirs reported negative sedimentation rates, likely due to sediment removal efforts.
        The median sedimentation rate in the dataset is 0.14 MCM per year, while the mean rate is significantly higher at 2.79
390   MCM per year. This difference is due to a few reservoirs with exceptionally high sedimentation rates, such as

Xiaolangdi and Sanmenxia in the Yellow River basin, where rates reach around 200 MCM per year. In total, 60 reservoirs report sedimentation rates exceeding 10 MCM per year.

Figure 8: World map displaying the 54 countries covered by reservoirs in the GRILSS dataset. The color fill indicates the median capacity loss rate (% per year) for each country. Countries with fewer than three reservoirs are outlined in red, signifying lower confidence in the estimated loss rate for those regions.
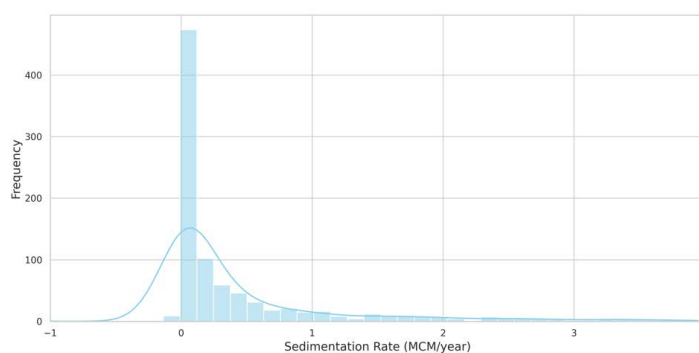

Figure 9: Frequency distribution plot of sedimentation rates (in MCM per year) for reservoirs in the GRILSS dataset, with a blue line representing the kernel density estimate.
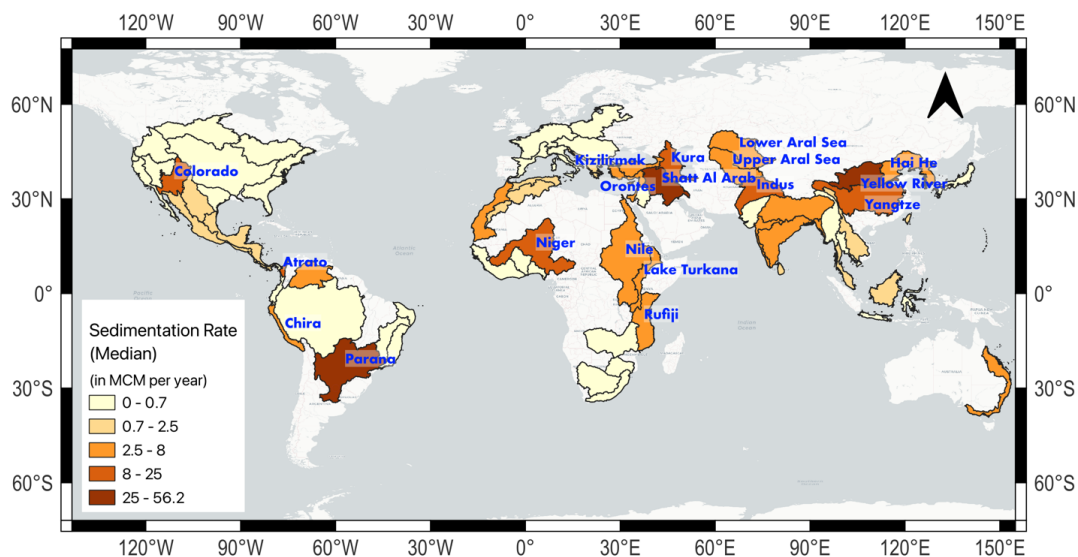
4.3.1 Sedimentation rate by river basins

The 1015 reservoirs in the GRILSS dataset are distributed across 75 major river basins worldwide, each characterized by varying geography, climate, vegetation, population density, and topography. Figure 10 illustrates the median capacity loss rates for these river basins, with lighter shades of yellow representing lower sedimentation rates (0-4 MCM per year) and darker brown shades indicating higher rates (up to 56 MCM per year). The Parana River basin in South America exhibits the highest sedimentation rate at 56.2 MCM per year, followed by the Yellow River basin at 31.4 MCM per year. All river basins with sedimentation rates exceeding 8 MCM per year are labeled on the map.

18

River basins in North America and Europe exhibit lower sedimentation rates, whereas those in Asia and Africa tend to have higher rates.

Figure 10: Map depicting the median sedimentation rates of reservoirs across 75 major river basins represented in the GRILSS dataset. River basins with sedimentation rates of 8 MCM or more are labeled on the map.

5. Discussion and Conclusion

Reservoirs play a critical role in meeting global freshwater demands for agriculture, domestic use, energy production, and industrial purposes by storing and regulating water supplies. Despite the continued construction of many new reservoirs worldwide, the total global storage capacity is in decline due to sedimentation. This loss of capacity is a growing concern, as sedimentation rates vary widely depending on factors such as the reservoir's location, catchment area, climate, topography, and soil characteristics. As a result, understanding the complexities of reservoir sedimentation requires localized studies that account for these unique environmental conditions.

While numerous studies have estimated sedimentation rates for individual reservoirs or small regions, these efforts have largely remained fragmented, having disparate curation and confined to local scales. One notable global study by Wisser et al. (2013) attempted to estimate global storage capacity loss using observed reservoir sedimentation rates. However, due to the lack of open-source data and the skewness of data toward reservoirs located in the United States (with over 1,000 records from the US and only 191 from the rest of the world), similar or follow up studies may have not been scalable. Thus, a lack of an organized, open-source dataset with a consistent curation practice that captures the reservoir sedimentation problem on a global scale, has limited the global use of sedimentation data. Until GRILLS, there has been no comprehensive resource to track and compare sedimentation rates of reservoirs worldwide, to the

430 best of our knowledge. With 1015 reservoirs spanning 75 major river basins and 54 countries, GRILSS provides critical insights into the extent of the problem worldwide. Compiled from 143 diverse and traceable sources, including published articles, government documents, and online databases, this dataset is a crucial resource for the global water community.

435 In an era where data and open access are essential, GRILSS opens new avenues for data-driven freshwater management on a global scale. Beyond sedimentation rates, the dataset also includes key reservoir attributes such as built capacity, dam height, and catchment area, along with information on the methods used to estimate sedimentation and the time period of observations. Additionally, GRILSS provides vector data on reservoir geometries and catchments, making it a versatile tool for researchers and policymakers alike.

440 Initial analysis of the GRILSS dataset suggests a concerning trend of higher capacity loss rates in more recently constructed reservoirs, indicating that newer infrastructure may be more vulnerable to sedimentation and reservoir operation efficiency. Additionally, we observed significant regional differences in sedimentation rates, with river basins such as the Parana and Yellow River experiencing some of the highest sedimentation rates globally. These insights underscore the need for targeted, region-specific sediment management strategies.

445

By organizing data on sedimentation and capacity loss that is also compatible with widely used existing reservoir datasets, GRILSS provides water managers, policymakers, and researchers a powerful dataset for planning and implementing better reservoir management practices. As the first dataset of its kind, GRILSS provides a foundation for future research into the myriad factors influencing reservoir sedimentation, including climate change, land use, 450 and sediment management practices. Its global scope opens the gateway for collaborative data curation on reservoir sedimentation, enabling the community to bridge existing gaps to sustain a growing dataset for large-scale, global studies on reservoir sedimentation (Minocha & Hossain, 2024).

455 **Dataset Availability:** All data for GRILSS is hosted on the Open Science Forum at this link: https://osf.io/w4ug8/?view_only=6a595a2d635c48339c95502a5d1d6417 Minocha, S., & Hossain, F. (2024). GRILSS: Global Reservoir Inventory of Lost Storage by Sedimentation.

Competing Interest: The contact author has declared that none of the authors has any competing interests.

References

465 Bartos, M. (2020). pysheds: simple and fast watershed delineation in python. doi:10.5281/zenodo.3822494

Basson, G. R. (2009). Management of siltation in existing and new reservoirs.

Central Water Commission. (2020). Compendium on sedimentation of reservoirs in India. Water Planning and Projects Wing, New Delhi, Government of India.

Demirbas, A. (2009). Global renewable energy projections. Energy Sources, Part B, 4(2), 212-224.

470   Elçi, Ş., Bor, A., & Çalışkan, A. (2009). Using numerical models and acoustic methods to predict reservoir sedimentation. Lake and Reservoir Management, 25(3), 297-306.

Erena, M., Atenza, J. F., García-Galiano, S., Domínguez, J. A., & Bernabé, J. M. (2019). Use of drones for the topo-bathymetric monitoring of the reservoirs of the Segura River Basin. Water, 11(3), 445.

Estrada Gutiérrez, G., Silva Hidalgo, H., Villalba, M. L., Astorga Bustillos, F., & Franco Estrada, B. (2015). Tasa de

475   acumulación de sedimentos en embalses del río Conchos, Chihuahua, México. In *1er Congreso Iberoamericano sobre sedimentos y ecología, Querétaro, Querétaro México*.

Gray, J.R., Bernard, J.M., Stewart, D.W., McFaul, E.J., Laurent, K.W., Schwarz, G.E., Stinson, J.T., Jonas, M.R., Randle, T. and Webb, J.W. (2010). Development of a national, dynamic reservoir sedimentation database. In Proceedings of the 2nd Joint Federal Interagency Conference (June 27-July 1, 2010), Las Vegas, NV (Vol. 172).

480   Güvel, Ş. P. (2021). Investigation of Sediment Deposition in Dam Reservoirs. Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 36(4), 869-878.

Hanmaiahgari, P. R., Gompa, N. R., Pal, D., & Pu, J. H. (2018). Numerical modeling of the Sakuma Dam reservoir sedimentation. Natural Hazards, 91, 1075-1096.

Kantoush, S. A., & Schleiss, A. J. (2014). Influence of geometry shape factor on trapping and flushing

485   efficiencies. Reservoir Sedimentation, 129-137.

Kim, S. M., Jang, T. I., Kang, M. S., Im, S. J., & Park, S. W. (2014). GIS-based lake sediment budget estimation taking into consideration land use change in an urbanizing catchment area. Environmental earth sciences, 71, 2155-2165.

Kondolf, G.M., Gao, Y., Annandale, G.W., Morris, G.L., Jiang, E., Zhang, J., Cao, Y., Carling, P., Fu, K., Guo, Q.

490   and Hotchkiss, R., 2014. Sustainable sediment management in reservoirs and regulated rivers: Experiences from five continents. Earth's Future, 2(5), pp.256-280.

Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J. and Nilsson, C. (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.

495   Lehner, B., Verdin, K., Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. Eos, Transactions, American Geophysical Union, 89(10): 93–94. https://doi.org/10.1029/2008eo100001

Lorsirirat, K. (2014). Models for Prediction of the Effects of Hydrological and Basin Characteristics on Reservoir Sedimentation for Water Management in Thailand. International Journal of Erosion Control Engineering, 7(3), 69-74.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O. (2016). Estimating the volume and age of water stored

500   in global lakes using a geo-statistical approach. Nature Communications, 7: 13603. https://doi.org/10.1038/ncomms13603

Minocha, S., & Hossain, F. (2024, October 12). GRILSS. Retrieved from osf.io/w4ug8 [Dataset]

Morris, G. L., & Fan, J. (1998). Reservoir sedimentation handbook: design and management of dams, reservoirs, and watersheds for sustainable use.

505 Nyikadzino, B., & Gwate, O. (2021). Estimation of Reservoir Capacity and Sedimentation Rate Using Direct and Indirect Methods. *Int J Earth Sci Geophys*, 7, 045.

OpenStreetMap contributors. (2017). Planet dump retrieved from https://planet.osm.org. Retrieved from https://www.openstreetmap.org [Last Accessed: 10-05-2024]

Overpass API. (2024, August 20). OpenStreetMap Wiki, . Retrieved 06:26, October 6, 2024
510 from https://wiki.openstreetmap.org/w/index.php?title=Overpass_API&oldid=2745759.

Pandey, A., Chaube, U. C., Mishra, S. K., & Kumar, D. (2016). Assessment of reservoir sedimentation using remote sensing and recommendations for desilting Patratu Reservoir, India. Hydrological Sciences Journal, 61(4), 711-718.

Palinkas, C. M., & Russ, E. (2019). Spatial and temporal patterns of sedimentation in an infilling reservoir. Catena, 180, 120-131.

515 Patil, R., & Shetkar, R. (2015). Sediment deposition in Koyna reservoir by integrated bathymetric survey. Int J Res Eng Technol, 4(11), 114-118.

Patro, E. R., De Michele, C., Granata, G., & Biagini, C. (2022). Assessment of current reservoir sedimentation rate and storage capacity loss: An Italian overview. Journal of environmental management, 320, 115826.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and
520 its long-term changes. Nature, 540(7633), 418-422.

Perera, D., Williams, S., & Smakhtin, V. (2023). Present and future losses of storage in large reservoirs due to sedimentation: a country-wise global assessment. *Sustainability*, *15*(1), 219.

QGIS Development Team. (2024). QGIS Geographic Information System. Retrieved from QGIS Association website: https://www.qgis.org

525 Samaila–Ija, H. A., Ajayi, O. G., Zitta, N., Kuta, A. A., Adesina, E. A., & Ibrahim, P. (2014). Bathymetric survey and volumetric analysis for sustainable management case study of Suleja Dam, Niger State, Nigeria.

Seatgeek. (2021). Seatgeek/thefuzz: Fuzzy string matching in Python. GitHub. https://github.com/seatgeek/thefuzz [Last Accessed: 10-05-2024]

Syvitski, J. P. (2003). Supply and flux of sediment along hydrological pathways: research for the 21st century. Global
530 and Planetary Change, 39(1-2), 1-11.

Ugwu, S. J., Ajoge, H. N., Abdulsalam, B., & Nwude, M. O. (2021). Bathymetry Study of the Siltation Level of Lugu Dam Reservoir in Sokoto State, Nigeria. *Nigerian Journal of Technological Development*, *18*(3), 238-243.

Vanmaercke, M., Poesen, J., Verstraeten, G., de Vente, J., & Ocakoglu, F. (2011). Sediment yield in Europe: Spatial patterns and scale dependency. Geomorphology, 130(3-4), 142-161.

535 Walter, J. A., Fleck, R., Pace, M. L., & Wilkinson, G. M. (2020). Scaling relationships between lake surface area and catchment area. *Aquatic sciences*, *82*, 1-8.

Wisser, D., Frolking, S., Hagen, S., & Bierkens, M. F. (2013). Beyond peak reservoir storage? A global estimate of declining water storage capacity in large reservoirs. Water Resources Research, 49(9), 5732-5739.

Yang, C. Y., Kang, W., Lee, J. H., & Julien, P. Y. (2022). Sediment regimes in South Korea. River Research and
540   Applications, 38(2), 209-221.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam
construction. Aquatic Sciences, 77, 161-170.

Zhang, A. T., & Gu, V. X. (2023). Global Dam Tracker: A database of more than 35,000 dams with location,
catchment, and attribute information. Scientific data, 10(1), 111.

545   Zhang, J., Xu, K., Yang, Y., Qi, L., Hayashi, S., & Watanabe, M. (2006). Measuring water storage fluctuations in
Lake Dongting, China, by Topex/Poseidon satellite altimetry. Environmental Monitoring and Assessment, 115, 23-
37.