nature portfolio

Corresponding author(s): Chunqiao Song

Last updated by author(s): Apr 12, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The spatial inventory of global reservoirs and sedimentation samples was compiled from global and national databases (e.g., GRanD, GOODD, GeoDAR, GREI-p2k, RESIS-II), literature sources, and OpenStreetMap (OSM), with manual validation and refinement using high-resolution imagery in Google Earth Engine and ArcGIS Pro 3.0.
Data analysis	Reservoir data quality control and attribute editing were performed using ArcGIS Pro version 3.0. Sedimentation rate modeling was conducted using Python 3.8, with libraries such as scikit-learn, XGBoost, and shapviz. Statistical analysis and visualization were carried out using ArcGIS Pro 3.0 and RStudio. Key R packages used include ggplot2, dplyr, and sf.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Provide your data availability statement here.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study presents a high-resolution global assessment of sedimentation in over 500,000 small reservoirs. A physics-guided machine learning model (XGBoost) was used to estimate sedimentation rates and capacity loss trends globally.
Research sample	The research sample includes 6,133 field-surveyed sedimentation records from reservoirs in 56 countries, ranging in size from 0.001 km² to over 5,000 km².
Sampling strategy	A quality-controlled dataset was compiled from national surveys (e.g., China, US, India) and literature. Data were filtered to exclude pre-1930 surveys, surveys shorter than 5 years, and duplicate records (retaining the longest interval). Machine learning models were trained with 70% of data and validated using 5-fold cross-validation.
Data collection	Data were collected from government agencies (e.g., China Institute of Water Resources, US RESIS-II, India's CWC) and from a systematic literature review. Reservoir attributes were cross-validated and extracted from global databases (e.g., GRanD, GOODD, GeoDAR, OpenStreetMap), followed by manual validation using satellite imagery.
Timing and spatial scale	The GREI dataset compiled in this study provides a global-scale spatial distribution of reservoirs up to the year 2020. The estimated sedimentation rates represent long-term average rates for each reservoir.
Data exclusions	Data exclusions included pre-1930 survey data, short survey intervals (<5 years), and redundant records. Inconsistent or unmapped records were manually screened and excluded.
Reproducibility	All analysis codes are available at GitHub (https://github.com/Kailiu1989/Codes/tree/master/GREI). The GREI dataset is publicly available at https://doi.org/10.6084/m9.figshare.31067800. The full list of data sources and calculation methods is provided in the Methods section. Results are reproducible using the provided data and described procedures.
Randomization	Not applicable. The study uses observational and geospatial datasets rather than experimental treatments. No random assignment was involved. Reservoirs were included based on availability of validated data.

Blinding	Not applicable. There was no blinding during data acquisition or analysis. All data were objective, derived from remote sensing, databases, and physical measurements.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable