



OPEN

DATA DESCRIPTOR

GLASS - Global Aggregation of Stream Silica

Kathi Jo Jankowski¹✉, Keira Johnson², Nicholas J. Lyon³, Sidney A. Bush⁴, Paul Julian⁵, Lienne R. Sethna⁶, Diane M. McKnight⁷, William H. McDowell⁸, Adam S. Wymore⁸, Pirkko Kortelainen⁹, Hjalmar Laudon¹⁰, Ruth C. Heindel¹¹, Amanda E. Poste¹², Ariel Shogren¹³, Fred Worrall¹⁴, Luke Mosley¹⁵, Pamela L. Sullivan¹⁶ & Joanna C. Carey¹⁷

Riverine silicon (Si) plays a vital role in governing primary production, water quality, and carbon cycling. Climate and land cover change have altered how dissolved Si (DSi) is processed on land, transported to rivers, and cycled through aquatic ecosystems. The Global Aggregation of Stream Silica (GLASS) database was constructed to assess changes in river Si concentrations and fluxes, their relationship to other nutrients (nitrogen (N) and phosphorus (P)), and to evaluate mechanisms driving the availability of Si. GLASS includes concentrations of DSi, dissolved inorganic N (NO_3 , NO_x , and NH_4), and dissolved inorganic P (as soluble reactive P or $\text{PO}_4\text{-P}$) at daily to quarterly time steps from 1963 to 2024; daily discharge; and watershed characteristics for 421 rivers spanning eight climate zones. Original data sources are cited, data quality assurance workflows are public, and input files to a common load model are provided. GLASS offers critical data to address questions about patterns, controls, and trajectories of global river Si biogeochemistry and stoichiometry.

Background & Summary

River ecosystems fundamentally link the biogeochemical cycling of elements along the land-ocean continuum^{1–3}. This link is especially true for silicon (Si), as rivers deliver >80% of annual Si loads to global oceans^{4,5}. Dissolved Si (DSi) transported by rivers directly links to global weathering, nutrient, and carbon (C) cycles along the terrestrial-marine continuum, most notably through primary production by siliceous diatoms^{4,6} which represent ~20% of photosynthetically fixed CO_2 each year^{7–9}. Unlike other phytoplankton, freshwater, coastal and marine diatoms require Si in large quantities to grow. Marine diatoms typically require equal quantities of Si and N on a molar basis, whereas freshwater diatoms have greater Si requirements relative to nitrogen (N) and phosphorus (P)^{6,10}. In the presence of excess N and P, Si can become limiting to diatom growth, shifting phytoplankton community composition away from diatoms towards non-siliceous algae and cyanobacteria^{7,11,12}. Despite the important role of rivers in processing and supplying Si needed for diatom growth, particularly for downstream marine systems where Si is often strongly limiting, we have far less knowledge of the controls and variability in space and time of river Si exports than for other nutrients.

The total flux of river Si exported to global oceans is controlled by complex ecological, geological, and climatic processes that vary throughout the river network (Fig. 1). With the exception of river damming, which is well known to modify river Si exports^{13–15}, river Si fluxes are often assumed to be relatively stable over time,

¹U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, WI, USA. ²College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Oregon, USA. ³National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA, USA. ⁴College of Earth Ocean and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. ⁵Everglades Foundation, Palmetto Bay, FL, USA. ⁶St. Croix Watershed Research Station, Marine on St. Croix, MN, 55047, USA. ⁷Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, Colorado, 80309, USA. ⁸Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, 03824, USA. ⁹Finnish Environment Institute, Helsinki, Finland. ¹⁰Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden. ¹¹Environmental Studies Program, Kenyon College, Gambier, Ohio, 43022, USA. ¹²Department of Arctic Ecology, Norwegian Institute for Nature Research, Tromsø, Norway & Norwegian Institute for Water Research, Oslo, Norway. ¹³Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35457, USA. ¹⁴Department of Earth Sciences, University of Durham, Durham, United Kingdom. ¹⁵School of Agriculture, Food and Wine, The University of Adelaide, South Australia, Australia. ¹⁶College of Earth Ocean and Atmospheric Sciences, Oregon State University, Oregon, USA. ¹⁷Math, Analytics, Science & Technology Division. Babson College, Wellesley, MA, USA. ✉e-mail: kjankowski@usgs.gov

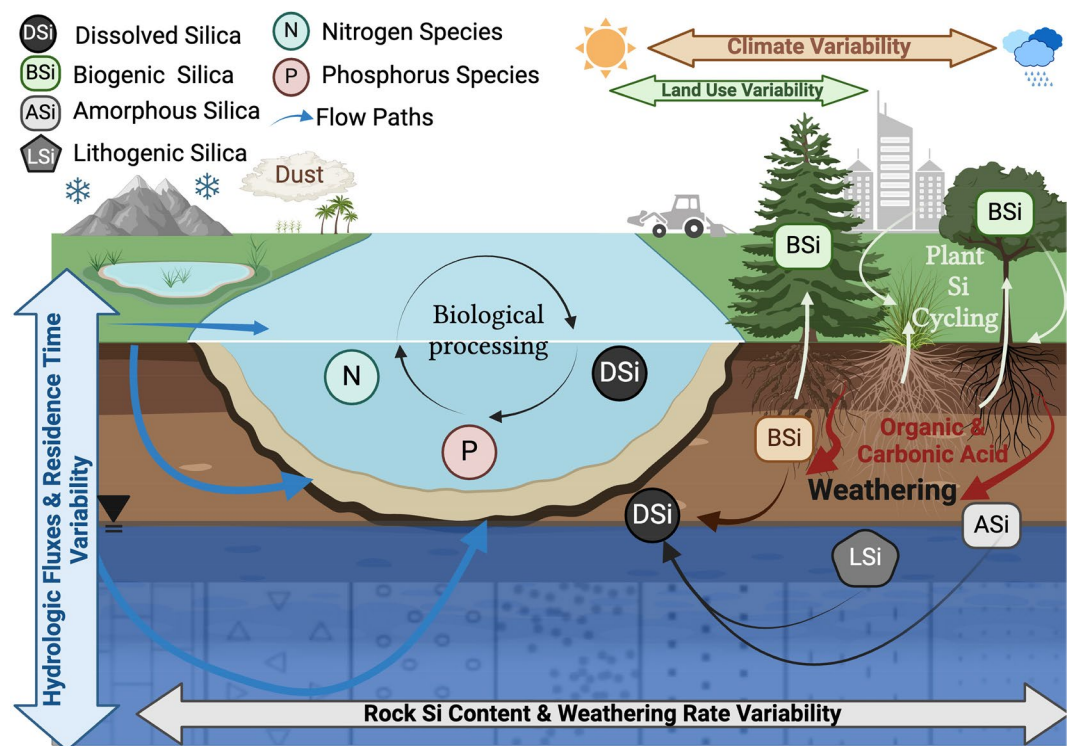


Fig. 1 Conceptual figure of the terrestrial and aquatic controls on riverine dissolved silicon (DSi) concentrations and yields. Silica derived from geogenic processes occurs either through the breakdown of rock or soil at a specific site or through the input and eventual breakdown of dust. DSi can be taken up through biotic processes (e.g., vegetation or diatoms) or precipitated as amorphous silica. DSi is transported from land to streams across various hydrologic flow paths. The interaction of the processes and their control on DSi concentrations and fluxes is dependent on the type of underlying lithology, land cover, land use, and climate. Double headed arrows in the figure indicate a wide range of variability of a given process (from Jankowski *et al.*¹⁷).

unaffected by human disturbance due to the dominant role of lithological weathering in controlling river Si exports¹⁶. Few examinations of changes in river Si over time have been completed over large spatial scales. Using a portion of the dataset presented here, dissolved Si (DSi) concentrations and yields were found to be changing over time, with the majority (62%) of the 60 rivers examined displaying significant increases in DSi yields over the past two decades¹⁷. These shifts were observed across a wide range of biome types, but most markedly in alpine and polar regions, which are particularly vulnerable to climate warming¹⁸. Watershed biogeochemical processes (e.g., changes in terrestrial vegetation, permafrost melt) were indicated as a driver of shifting fluxes, rather than simply changing streamflow.

In addition to investigating long-term (>20 year) changes in river DSi exports, GLASS has been used to characterize the seasonal cycles, or regimes, of river DSi concentrations¹⁹. Seasonal regimes are integral to understanding how river ecosystems function, as they reflect the integrated signal of hydroclimatic conditions, biological processes, and watershed characteristics, including lithology, land cover, and vegetation^{20–22}. Few studies have investigated the seasonal regimes of river Si concentration across broad spatial scales, with most prior work examining temperate rivers and identifying a fairly singular pattern of a spring drawdown and an elevated winter plateau^{23,24}. Using a subset of the dataset presented here, five distinct seasonal Si regimes across the Northern Hemisphere were identified, documenting how the seasonal timing of maximum and minimum concentrations varied widely among rivers¹⁹. Most rivers exhibited multiple regimes over time, rather than a consistent seasonal pattern. The same subset of GLASS was then used to determine the watershed-scale drivers controlling the variation in seasonal regimes²². Variation in seasonal regime was associated with a suite of climate- and ecosystem productivity-related factors, such as snow cover, temperature, green-up day, and evapotranspiration. Together, this work provides fundamental new insights about river Si cycling and highlights the diversity of processes controlling watershed Si cycling.

Assessing controls on river Si exports at large spatial scales requires relating spatially-extensive stream chemistry data to river flow and watershed climate, land cover/vegetation type, land use, and lithology characteristics. Although many agency, country, university and research monitoring efforts have collected Si data in rivers across the globe, these disparate datasets have not been combined into a single publicly-available database. Additionally, these data sources vary in terms of time span, sampling protocols, chemical species measured, documentation, data curation, and data accessibility.

To overcome this shortcoming, we developed the Global Aggregation of Stream Silica (GLASS) database to harmonize DSi datasets generated using different sampling methods, levels of documentation, and conventions

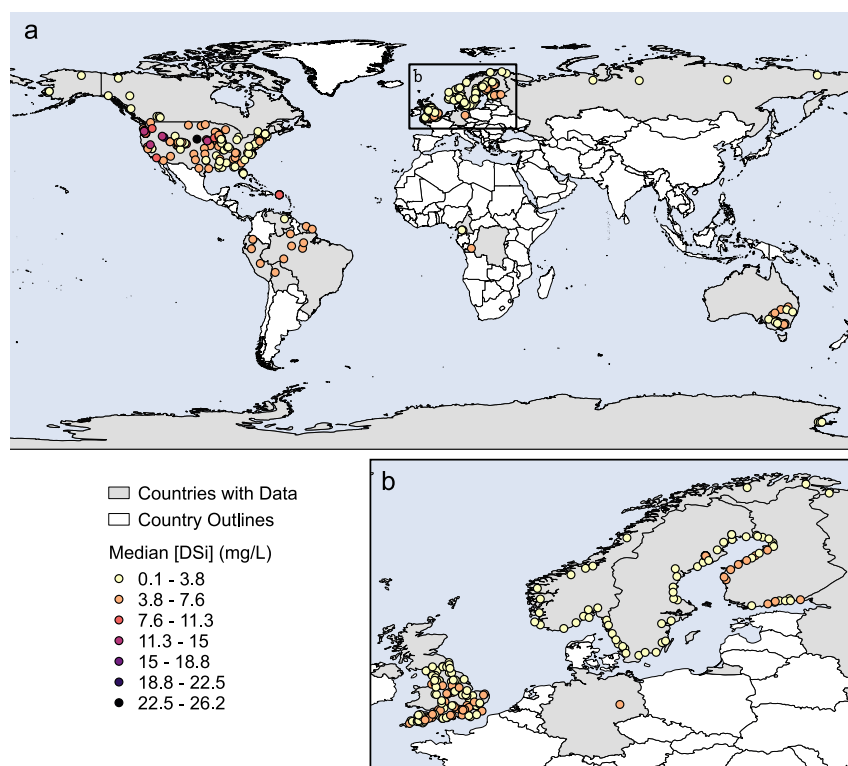


Fig. 2 Global spatial distribution (a) of rivers included in GLASS. Points indicate location of discharge and chemistry measurement, and are colored by median DSi concentration. Inset box shows Scandinavia (b).

for naming, units, and other characteristics to address critical ecological questions at large spatial scales. To understand the dominant controls on Si we integrated several additional variables such as concentrations of other solutes (e.g., N, P), river hydrology, and watershed characteristics into the GLASS database. To achieve this goal, data from national and state level monitoring programs, the U.S. LTER Network, private research institutions, and individual researchers were combined to create a georeferenced stream Si database with over 600,000 individual nutrient chemistry observations collected across 421 rivers (Fig. 2). In addition, each stream has paired daily discharge data, which is a unique feature of this dataset compared to other large river chemistry datasets that allows for estimation of loads and assessment of the role of hydrology in controlling stream Si dynamics. A shapefile delineating polygons for all watersheds is included along with the chemistry database. We also provide summarized watershed scale data including land use/land cover, lithology and soils, and climate variables for all these stream locations generated from globally consistent data sources. The database was constructed to be transparent and reproducible. We provide R code (<https://github.com/lter/lterwg-silica-data>) and additional reference files used to format and combine data sources as an appendix.

Methods

Data acquisition. *Water chemistry and discharge.* We acquired river chemistry and discharge data from published and/or publicly available datasets and through direct requests to researchers or agencies (Supplemental Table 1). Acquired river chemistry data include dissolved Si (DSi), dissolved inorganic N (DIN) and dissolved inorganic P (DIP) concentration data. DIN data include values for NH_4 , NO_3 , or NO_x . Not all sites reported the same forms of DIN, however, and thus we report data for whichever form(s) were provided in the original dataset (i.e., individual species not a total DIN value). Dissolved inorganic P is reported as “DIP” in this dataset but included data that were originally reported either as soluble reactive phosphorus (SRP) or phosphate (PO_4).

There are a total of 421 individual sites in the dataset, which all include DSi and river discharge data. 397 sites reported NO_3 or NO_x data, 196 sites report NH_4 , and 339 sites report P (Fig. 3).

Sites spanned 11 Koeppen-Geiger climate zones between -77°S and 70°N ^{25,26}, varied in drainage area from $<1\text{ km}^2$ to nearly 4 million km^2 , and in mean river discharge from $<0.01\text{ m}^3\text{ s}^{-1}$ to nearly $200,000\text{ m}^3\text{ s}^{-1}$ indicating a wide range in catchment conditions included in the dataset.

We established several criteria for including data. Each site was required to have records of daily discharge and discrete observations of dissolved silicon (DSi). We included rivers in the dataset that had a minimum of four years of data, with the period of record for rivers ranging from four to 55 years. The number of observations per year for all stream-variable combinations ranged from 1 to 178 with a median number of observations per year per stream of 14.2 and range of 2.4 to 64. Thus, some years in a dataset for a given stream did have just a single observation, but this was never the case for an entire dataset. Data were required to meet the quality assurance requirements specified by the original data source (see below for additional QA/QC information and

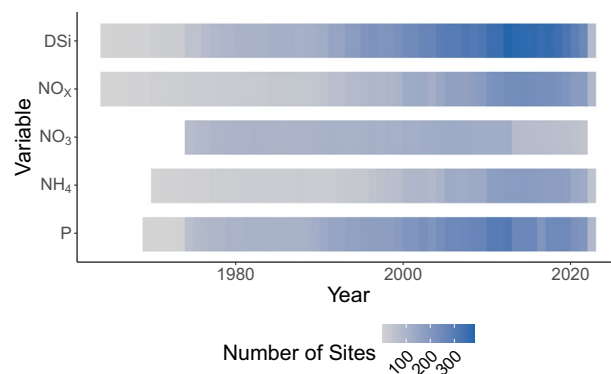


Fig. 3 Distribution of chemistry data by solute, period of record, and number of sites.

technical validation of data). All chemistry samples were derived from “regular” samples (i.e., not field or lab duplicates).

In the process of harmonization, we addressed several additional data quality issues including below detection limit and unreasonable values. DSi data are generally robust, high-quality data because of the consistency of the methods involved and the stability of dissolved Si in filtered water samples. In addition, the concentrations are usually greater than 1 mg Si/L, and thus the problems that can influence measurements of N and P at low concentrations did not occur as frequently.

Stream discharge data were required to have daily values and were all converted to cubic meters per second. Where there were gaps of less than 30 days, we linearly interpolated values and included a field in the data file to indicate whether data are measured or interpolated. Interpolated data represent less than 0.01% of the total dataset.

Watershed and climate characteristics. To characterize climate, ecosystem productivity, land cover, and lithology of contributing watersheds we acquired data from globally available modeled and remotely sensed data sources (Supplemental Table 2, Fig. 4). Specifically, we acquired precipitation, air temperature, snow-covered area, elevation, soil order, land cover, lithology, evapotranspiration, net primary productivity, green-up day, and permafrost. We used spatial data layers with global coverage to have consistent data sources across the extent of our dataset. The spatial resolution of these data sources (Supplemental Table 2) was typically coarser than it might be if we used locally available data sources but had the benefit of providing data generated using a globally-consistent methodology.

In the process of gathering spatial data from grid-based data sources, the initial step involved the procurement or creation of watershed delineations. Where feasible, we sourced existing watershed boundary shapefiles (referenced in Table S3 Johnson *et al.*²²). For six sites where the original data provider did not supply watershed shapefiles, the HydroBASINS database, as described by Lehner and Grill²⁷ was used to construct the necessary watershed delineations. HydroBASINS offers a comprehensive global network of hierarchically organized sub-basins across various scales. In its most detailed sub-basin segmentation, HydroBASINS separates a basin into two smaller sub-basins at junctures where two tributaries converge, each with a minimum upstream area of 100 km². For the compilation of watersheds pertinent to this dataset, the initial step was to pinpoint the basin that overlapped with the sampling coordinates at the most detailed HydroBASINS segmentation level, followed by the successive inclusion of all connected upstream basins. After delineating all relevant basin polygons, they were amalgamated into a unified shapefile, which then served as the definitive boundary for the watershed. Given that the smallest HydroBASINS delineations average around 100 km², which is considerably larger than some of the streams in our study, this method was not applied to basins less than 2000 km² in size.

Data were available at different spatial and temporal resolutions (Supplemental Table 2). For time-varying datasets, we primarily summarized data at an annual scale as that was most commonly available time step across datasets. Surface air temperature, precipitation, green-up day, net primary productivity, and land cover were all available as annual mean values. Evapotranspiration data were available on an 8-day time step, which we summarized to annual mean values. Snow-covered area (percent of watershed covered by snow) was also available at a daily time step, from which we extracted the maximum annual value for use in our models. The number of snow-covered days were generated by multiplying the proportion of snow-covered pixels within a watershed (binary) by the number of days those pixels were snow-covered. For example, a watershed with 10 snow-covered days could have snow cover in 100% of its pixels for 10 days within a year or have only 50% of snow cover for 20 days. This metric was highly correlated with the maximum snow-covered area. We included global land cover data at 30 m resolution²⁸. Land cover was available for 1985, 1990, 1995, and annually from 2000–2022. Years between each five-year increment (e.g., 1985–1990) were linearly interpolated. Land cover classes were lumped into forest, grassland and shrubland, wetland and marsh, tidal wetland, cropland, impervious, ice and snow, water, salt water, and bare (Table S3), and the proportion of each land cover class was reported for each watershed.

The lithology, permafrost, watershed elevation and slope data included in the database were all static values (not time varying). Lithology data were sourced from the PANGAEA dataset and lumped into volcanic,

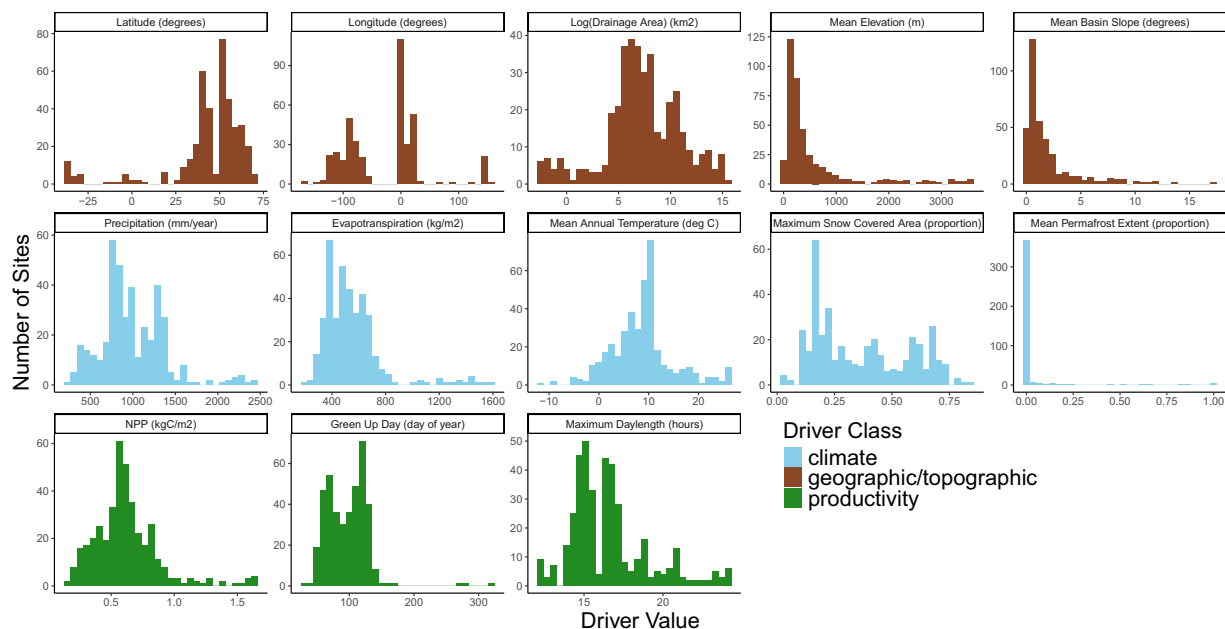


Fig. 4 Distribution of watershed characteristics across all the sites included in the dataset, colored by their class (climate, geographic/topographic, productivity). Values shown are mean values for each site over their full period of record.

sedimentary, plutonic, metamorphic, and carbonate/evaporite. Land cover and lithology categories were further refined as shown in Supplemental Tables 3 & 4. Watershed elevation was measured with the digital elevation model available on WorldClim, which is derived from the 30 second SRTM digital elevation data. Within each watershed boundary the mean, median, minimum, and maximum elevations were all calculated as well as the same summary statistics for basin slope. Permafrost probability was reported as a value between 0 and 1 and represents probability of continuous, discontinuous or sporadic permafrost²⁹.

We also include the watershed delineations for nearly all the rivers included in this dataset. This enables future users to gather additional datasets that may provide other variables at finer temporal or spatial resolutions.

Data harmonization. We built the chemistry and discharge datasets with the intention of easy integration of future additional datasets. The workflow was designed to be flexible enough to ingest datasets in many formats (i.e., wide, long, different column names, units) and to produce a single harmonized datafile with all data with the same units, date formats, variables, and column names.

All original data sources are listed in Supplemental Table 1 and the harmonization process is shown in Fig. 5. Harmonization included several QA and validation steps for both the chemistry and discharge datasets. These steps included reviewing for missing or unreasonable data values, screening and removing extreme outliers, standardizing date formats, converting units, removing duplicate values or site records and appending minimum detection limit (MDL) flags to chemistry data and gap-filled indicators to the discharge dataset. Additional reference files used to assign MDL values and select periods of the original time series (see Fig. 5) are included as Supplemental Material.

Data Records

All datasets are located in the U.S. Geological Survey ScienceBase repository³⁰.

Record 1. GLASS chemistry (Si, DIN, and DIP).

This dataset contains 421 sites from 24 different observation networks. Across all sites, periods of records for DSi, DIN, and DIP were 1964–2023, 1964–2023, and 1969–2023, respectively. For all variables across all sites, the mean number of samples per site per year was 14.2. The chemistry dataset is formatted in long format and contains all observations of all solutes for all streams across the dataset. The stream chemistry file is named `Chemistry_dat_v2.csv`.

Research_network. Name of the research network that provided data.

Stream_name. Name of the stream or stream site.

Date. Date of sample collection.

Variable. Name of constituent.

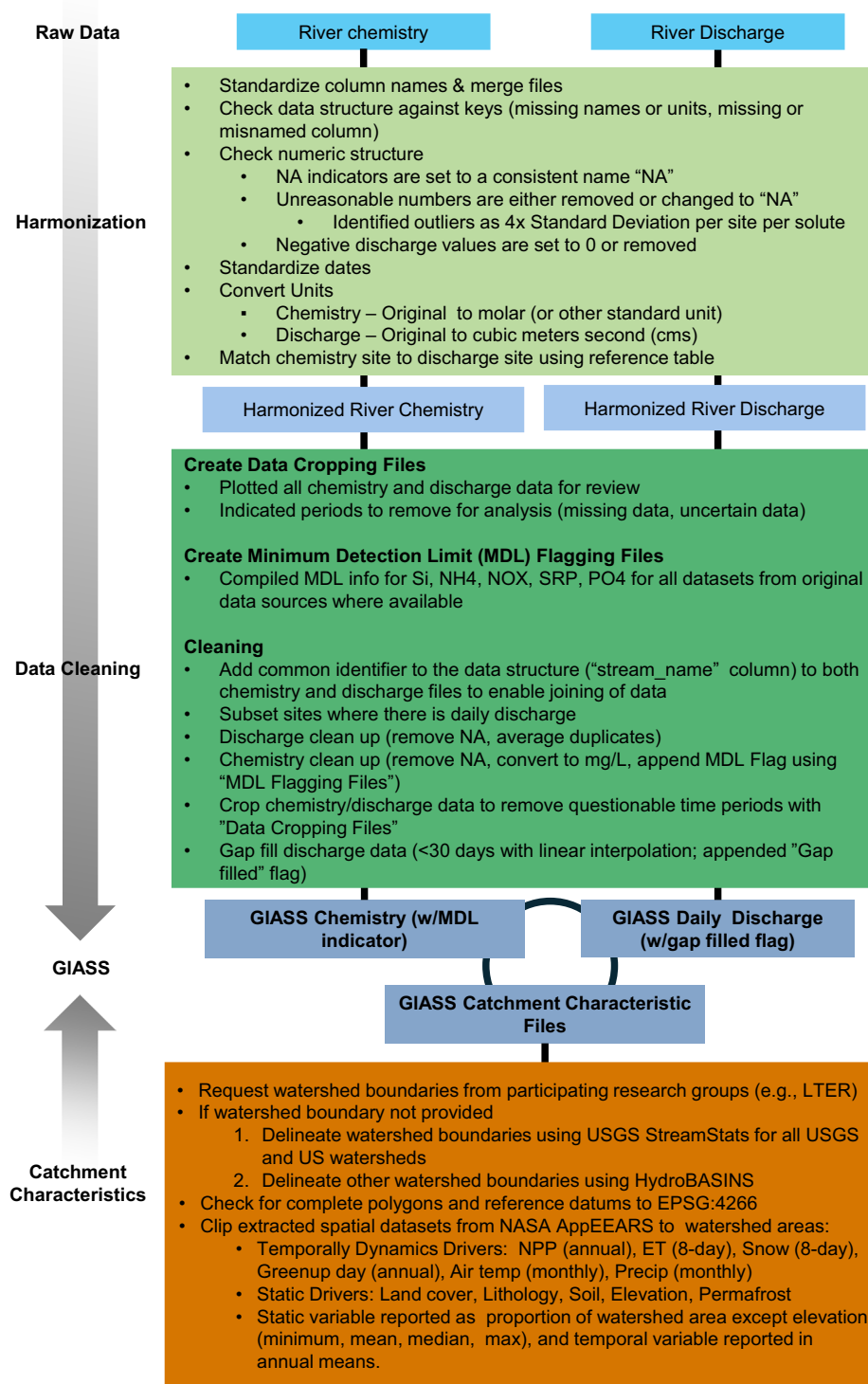


Fig. 5 Process by which chemistry, discharge, and spatial data were standardized, reviewed, and harmonized.

value_mgL. Concentration of constituent in milligrams per liter. Units are reported as concentration of Si, N or P. Specifically, DSi as Si; DIN as NO₃-N, NO_x-N, NH_x-N; and either SRP or PO₄-P as DIP).

remarks. code indicating whether the value is at or below the detection limit ("<").

Record 2. GIASS discharge datasets.

The discharge period of record ranged from 1963–2023. The discharge dataset is stored in Discharge_dat.csv and contains daily flow values for all rivers included in the dataset. Where discharge was not continuous,

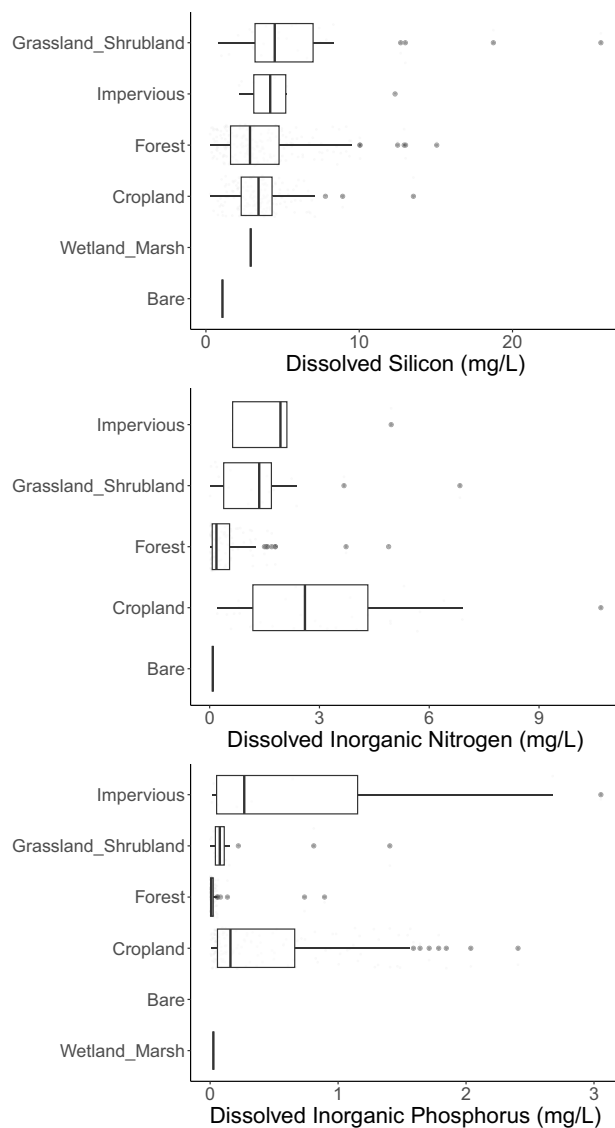


Fig. 6 Distribution of (a) DSi concentration, (b) Dissolved inorganic nitrogen (NO_x , NO_3 , and NH_4) and (c) dissolved inorganic phosphorus concentration with major land cover. Box plots depict the minimum, first quartile, median, third quartile, and maximum, with outliers depicted as single points.

discharge was gap-filled when the gap was <30 days. The column name “indicator” indicates whether the discharge was gap filled or not.

Research_network. Name of the research network that provided data.

Stream_Name. Name of the stream or stream site.

Date. Date of collection.

Discharge. Discharge in cubic meters per second.

Indicate. Indicates whether value was measured or interpolated.

Record 3. Shapefiles.

The shapefile dataset is an aggregation of all individual river shapefiles in coordinate reference EPSG:4396.

Record 4. Mean watershed and climate data parameters.

This dataset includes latitude, longitude, and long-term mean values of all parameters listed in Supplemental Table 2 for 400 sites in the dataset.

Record 5. Mean annual watershed and climate data parameters.

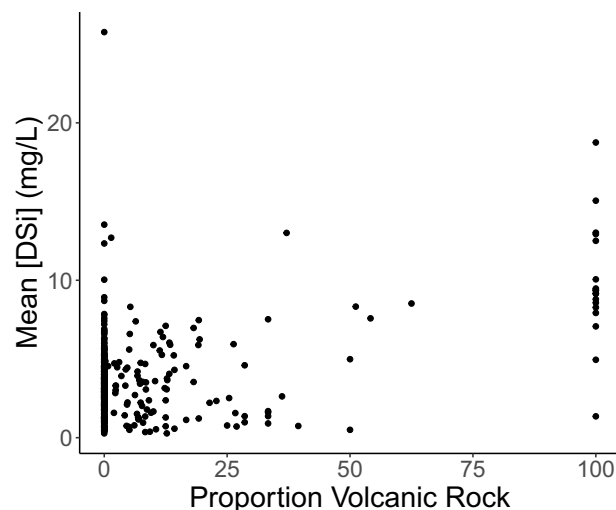


Fig. 7 Distribution of DSi concentration with the proportion of volcanic rock in the watershed.

This dataset includes latitude, longitude, and long-term mean annual values of air temperature, precipitation, net primary productivity, green-up day, snow cover, and land cover for 400 sites in the dataset (as described in Supplemental Table 2).

Technical Validation

Chemistry and discharge data. Several steps were taken to validate the data before, during, and after harmonization. Before harmonization, we acquired metadata describing data quality and limitations from the original data source where possible (e.g., USGS or EDI). We required individual data contributors to review their own data prior to submission and provide data that had been validated and QA'd according to their institution's protocols. We plotted all chemistry and discharge data to visually review and identify errors, long periods of missing data, or other data problems prior to harmonization with other datasets.

During harmonization, we removed unreasonable numbers (e.g., negative concentrations), extreme outlier values (4 times the standard deviation), and duplicate data records present across multiple data sources. We standardized site names across chemistry and discharge datasets (created unique site ID, "Stream_Name" that is common across datasets).

After data harmonization, we plotted all discharge and chemistry data and visually reviewed them for errors and reasonable ranges of values (e.g., to validate unit conversions were correct). We sent a subset of the data back to the original data contributors to review for validity and consistency. We then compared data to previously published values for sites to validate that means and ranges were similar to known values.

Finally, to ensure our data align with generally understood spatial patterns and mechanistic drivers of river chemistry and discharge, we performed a number of additional comparisons. We evaluated spatial patterns in Si, N, and P concentrations with environmental gradients known to have strong influences on their values (Figs. 6, 7). River Si concentrations typically align well with global distributions of bedrock lithology, increasing with the abundance and weathering rates of silicate rocks^{31–33}. That pattern is generally evident in this dataset, which shows the highest concentrations tend to occur in watersheds draining volcanic lithology (Fig. 7). DSi showed less variability with land cover but grassland/shrubland and urban/impervious land highest concentrations on average. River N and P concentrations tend to increase with agricultural and urban land uses^{34–36}, which is shown clearly in this dataset as well. Watersheds dominated by cropland, shrubland, and urban land use had the highest concentrations of both inorganic N and P (Fig. 6).

Spatial data. We largely relied on the original technical validation of the data done by the authors/producers of the data products (Supplemental Tables 1, 2) but did our own evaluation of the values it produced for the watersheds in this dataset.

Specifically, we reviewed values of the watershed-scale data that were generated to ensure they seemed reasonable and within expected ranges (Figs. 4, 8). For example, we verified that all proportions (e.g., land cover, lithology) added up to 100 percent for each watershed, land cover classes matched our expectations, and neighboring watersheds had similar values to one another for variables such as air temperature and precipitation. After our own internal review, we sent them to site experts to assess whether values were reasonably aligned with known or published values.

In general, we used the data as they were generated from the global products and did not modify the reported values to account for local knowledge or other available data sources. Because we did not have consistent knowledge of all parameters across all sites in our dataset, we did not modify values generated from these data products using other sources of data. We only modified or removed values if they were obviously unreasonable or wrong and largely relied on the internal QA/QC of the satellite products. To account for cases where data products reported values that appeared unlikely, we included quality flags. This was a particular issue for data from the MODIS platform likely as a result of cloud interference and pixel size. MODIS-derived data

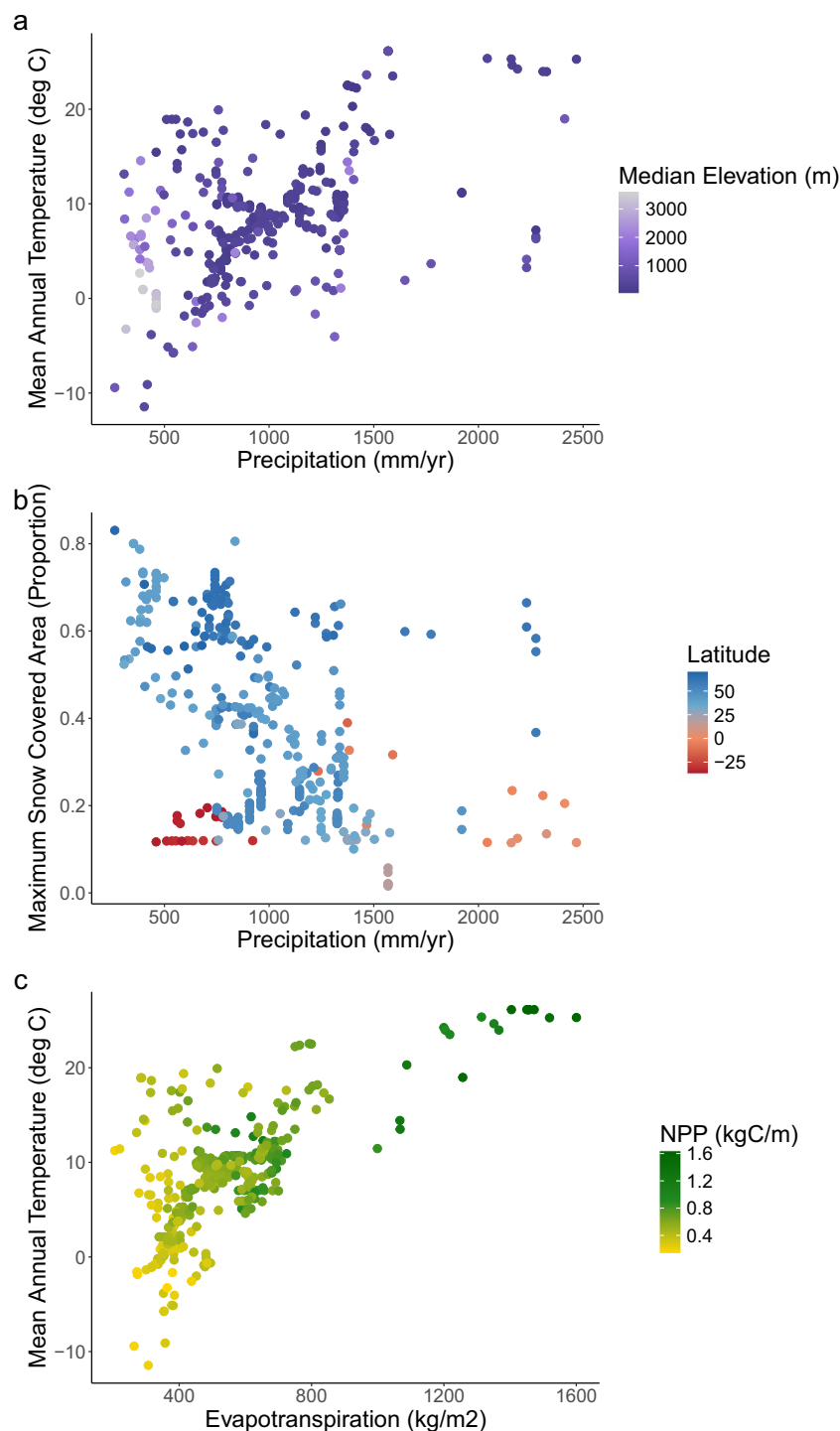


Fig. 8 Distributions and relationships among associated watershed characteristics. (a) mean annual precipitation, mean annual temperature, and median watershed elevation (Q); (b) mean annual precipitation, maximum snow-covered area, and latitude; and (c) mean annual evapotranspiration, mean annual temperature, and mean annual NPP (net primary productivity).

included snow, evapotranspiration, NPP, and green-up day. The issues were particularly clear for snow data, therefore we included a column that flagged snow data values as “unlikely” (“U”) in cases where snow cover data were reported but the site also had a mean annual temperature > 15 C, latitude < |35| degrees, and elevation < 1000 m. We also included a flag for annual land cover data to indicate greater uncertainty before 2000 because data collection occurred every five years rather than annually as it did after 2000.

Usage Notes

Limitations. *Stream chemistry.* We included the most continuous record available for stream chemistry, but some datasets have large gaps in time, change from reporting one constituent to another (e.g., early data reported as NO₃ but later data reported as NO_x) or do not extend to recent years. There are some datasets that have many values at or below the detection limit. We included a remark code to indicate that a value was below detection for a given site and constituent combination, but we did not include the actual minimum detection limit value or replace any values. The stream chemistry dataset includes dissolved forms of nutrients as those were the most frequently available across sites. There are known limitations when using concentrations of these nutrients to describe their availability, as they may not contain the entire pool available for uptake^{37–39} and low concentrations may in fact reflect high biological uptake. Integrating other datasets would be necessary to understand those types of dynamics.

There are other global databases that provide extensive stream chemistry data, including DSi^{40–42}. We reviewed these datasets and included some of the available data. Many of them did not have paired stream discharge data so were not included in this product.

Stream discharge. Discharge is provided as daily values but some of the values were interpolated. In some cases, longer records existed for discharge, but we did not include data far outside of the chemistry record as this dataset was intended for use in modeling nutrient fluxes not evaluating long-term changes in hydrology.

Watershed characteristics. There are some important limitations in the use of the watershed characteristics dataset. Given that the average size of the pixels of the original data products are large relative to the size of some individual watersheds, these data are best used to compare across watersheds at a global scale to capture large-scale environmental gradients and are not well suited to compare across small watersheds (i.e., watersheds smaller than the footprint of the pixel). In addition, the snow, ET, NPP, and green-up day values are generated using the MODIS platform, which does not perform as well in watersheds that experience a lot of cloud cover (e.g., tropical watersheds). As stated above, data flags were added to clarify where values generated by global products were uncertain or unlikely.

Data availability

All data are available at [Global Aggregation of Stream Silica \(GLASS\) \(ver. 2.0, July 2025\) - ScienceBase-Catalog](#).

Code availability

Code used to clean and harmonize dataset is located here: <https://github.com/lter/lterwg-silica-data>. This repository includes scripts that were used to process and clean the original data sources cited in Supplemental Table 1, specifically 00-harmonize_chemistry.R and 00_harmonize_discharge.R. Those harmonized data files were then further modified to produce the chemistry and discharge data described here using 01-wrtds-step02-wrangling.R. Additional scripts are included in the repository to process data through the Weighted Regression on Time, Discharge and Season model⁴³ (Hirsch *et al.* 2010).

Received: 7 October 2024; Accepted: 4 September 2025;

Published online: 20 October 2025

References

- Maranger, R., Jones, S. E. & Cotner, J. B. Stoichiometry of carbon, nitrogen, and phosphorus through the freshwater pipe. *Limnol Oceanogr Letters* **3**, 89–101 (2018).
- Turner, R. E. Water quality at the end of the Mississippi River for 120 years: the agricultural imperative. *Hydrobiologia*, <https://doi.org/10.1007/s10750-023-05383-4> (2023).
- Meybeck, M. Carbon, nitrogen and phosphorus transport by world rivers. *American Journal of Science* **282**, 401–450 (1982).
- Tréguer, P. J. & De La Rocha, C. L. The world ocean silica cycle. *Annu. Rev. Mar. Sci.* **5**, 477–501 (2013).
- Tréguer, P. J. *et al.* Reviews and syntheses: The biogeochemical cycle of silicon in the modern ocean. *Biogeosciences* **18**, 1269–1289 (2021).
- Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *Journal of Phycology* **21**, 347–357 (1985).
- Officer, C. & Ryther, J. The possible importance of silicon in marine eutrophication. *Marine Ecology Progress Series* **3**, 83–91 (1980).
- Malviya, S. *et al.* Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. USA*. **113** (2016).
- Litchman, E., Klausmeier, C. A. & Yoshiyama, K. Contrasting size evolution in marine and freshwater diatoms. *Proc. Natl. Acad. Sci. USA*. **106**, 2665–2670 (2009).
- Conley, D. J., Kilham, S. S. & Theriot, E. Differences in silica content between marine and freshwater diatoms. *Limnology & Oceanography* **34**, 205–212 (1989).
- Conley, D., Schelske, C. & Stoermer, E. Modification of the biogeochemical cycle of silica with eutrophication. *Mar. Ecol. Prog. Ser.* **101**, 179–192 (1993).
- Royer, T. V. Stoichiometry of nitrogen, phosphorus, and silica loads in the Mississippi-Atchafalaya River basin reveals spatial and temporal patterns in risk for cyanobacterial blooms. *Limnology & Oceanography* **65**, 325–335 (2020).
- Humborg, C. *et al.* Silicon retention in river basins: Far-reaching effects on biogeochemistry and aquatic food webs in coastal marine environments. *AMBIO: A Journal of the Human Environment* **29**, 45–50 (2000).
- Humborg, C. *et al.* Decreased Silica Land–sea Fluxes through Damming in the Baltic Sea Catchment – Significance of Particle Trapping and Hydrological Alterations. *Biogeochemistry* **77**, 265–281 (2006).
- Ma, N. *et al.* Effects of river damming on biogenic silica turnover: implications for biogeochemical carbon and nutrient cycles. *Acta Geochim* **36**, 626–637 (2017).
- Meybeck, M. 5.08 - Global Occurrence of Major Elements in Rivers. *Treatise on Geochemistry* **5** (2003).
- Jankowski, K. J. *et al.* Long-Term Changes in Concentration and Yield of Riverine Dissolved Silicon From the Poles to the Tropics. *Global Biogeochemical Cycles* **37**, e2022GB007678 (2023).
- IPCC. *Climate Change 2023: Synthesis Report, Contribution of Working Groups I, II and III to the Sixth Assessment Report of Hte Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]*. 184 (2023).

19. Johnson, K. *et al.* Establishing fluvial silicon regimes and their stability across the Northern Hemisphere. *Limnol Oceanogr Letters* **9**, 237–246 (2024).
20. Bernhardt, E. S. *et al.* Light and flow regimes regulate the metabolism of rivers. *Proc. Natl. Acad. Sci. USA* **119**, e2121976119 (2022).
21. Bolotin, L. A., Summers, B. M., Savoy, P. & Blaszczyk, J. R. Classifying freshwater salinity regimes in central and western U.S. streams and rivers. *Limnol Oceanogr Letters* **8**, 103–111 (2023).
22. Johnson, K. *et al.* Climate, hydrology, and nutrients control the seasonality of Si concentrations in rivers. *Journal of Geophysical Research: Biogeosciences* **129** (2024).
23. Carey, J. C. & Fulweiler, R. W. Human activities directly alter watershed dissolved silica fluxes. *Biogeochemistry* **111**, 125–138 (2012).
24. Fulweiler, R. W. & Nixon, S. W. Terrestrial vegetation and the seasonal cycle of dissolved silica in a southern New England coastal river. *Biogeochemistry* **74**, 115–130 (2005).
25. Köppen, W. The thermal zones of the Earth according to the duration of hot, moderate, and cold periods and to the impact of heat on the organic world. *Meteorologische Zeitschrift* **20**, 351–360 (2011).
26. Bryant, C., Wheeler, N., Rubel, F. & French, R. K. G. Köppen-Geiger climatic zones (2017).
27. Lehner, B. & Grill, G. Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes* **27**, 2171–2186 (2013).
28. Liu, L., Zhang, X. & Zhao, T. GLC_FCS30D: the first global 30-m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Zenodo* <https://doi.org/10.5281/zenodo.8239305> (2023).
29. Obu, J., Westerman, S., Kääb, A. & Bartsch, A. Ground Temperature Map, 2000–2016, Northern Hemisphere Permafrost [dataset]. <https://doi.org/10.1594/PANGAEA.888600> (2018).
30. Jankowski, K. J. *et al.* Global Aggregation of Stream Silica (GLASS). U.S. Geological Survey data release <https://doi.org/10.5066/P138M8AR> (2024).
31. Meybeck, M. *et al.* Nutrients (organic C, P, N, Si) in the eutrophic River Loire (France) and its estuary. *Estuarine, Coastal and Shelf Science* **27**, 595–624 (1988).
32. White, A. F. & Blum, A. E. Effects of climate on chemical weathering in watersheds. *Geochimica et Cosmochimica Acta* **59**, 1729–1747 (1995).
33. West, A., Galy, A. & Bickle, M. Tectonic and climatic controls on silicate weathering. *Earth and Planetary Science Letters* **235**, 211–228 (2005).
34. Carpenter, S. R. *et al.* Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* **8**, 559–568 (1998).
35. Stets, E. G. *et al.* Landscape Drivers of Dynamic Change in Water Quality of U.S. Rivers. *Environ. Sci. Technol.* **54**, 4336–4343 (2020).
36. Mattsson, T., Kortelainen, P. & Räike, A. Export of DOM from Boreal Catchments: Impacts of Land Use Cover and Climate. *Biogeochemistry* **76**, 373–394 (2005).
37. Dodds, W. K. Trophic state, eutrophication and nutrient criteria in streams. *Trends in Ecology & Evolution* **22**, 669–676 (2007).
38. Reinl, K. L. *et al.* The role of organic nutrients in structuring freshwater phytoplankton communities in a rapidly changing world. *Water Research* **219**, 118573 (2022).
39. Johnson, L. T., Tank, J. L. & Arango, C. P. The effect of land use on dissolved organic carbon and nitrogen uptake in streams. *Freshwater Biology* **54**, 2335–2350 (2009).
40. Hartmann, J., Lauerwald, R. & Moosdorf, N. GLORICH - Global river chemistry database [dataset]. <https://doi.org/10.1594/PANGAEA.902360> (2019).
41. Sterle, G. *et al.* CAMELS-Chem: augmenting CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) with atmospheric and stream water chemistry data. *HESS* **28**, 611–630 (2024).
42. Gaillardet, J., Dupré, B., Louvat, P. & Allègre, C. J. Global silicate weathering and CO₂ consumption rates deduced from the chemistry of large rivers. *Chemical Geology* **159**, 3–30 (1999).
43. Hirsch, R. M., Moyer, D. L. & Archfield, S. A. Weighted Regressions on Time, Discharge, and Season (WRTDS), with an application to Chesapeake Bay river inputs. *JAWRA Journal of the American Water Resources Association* **46**, 857–880 (2010).

Acknowledgements

This work was completed as part of a synthesis working group entitled: From Poles to Tropics: A multi-biome synthesis investigating the controls on river Si exports and was supported through the Long Term Ecological Research Network Office (LNO) (NSF award numbers 1545288 and 1929393) and the National Center for Ecological Analysis and Synthesis, UCSB awarded to K.J.J. and J.C.C. K.J.J. was supported by the US Army Corps of Engineers Upper Mississippi River Restoration Program. Funding was provided to J. Carey from the Babson Faculty Research Fund and the Andrew J. Butler and Debi Butler Term Chair. This material is based upon work supported by the National Science Foundation under Grant NSF 2012796 (P. L. Sullivan). A.S.W. received support from the National Science Foundation and EPSCoR project Canary in the Watershed (NSF EPS-1929148). Support was also provided by the New Hampshire Agricultural Experiment Station. This work was supported by the USDA National Institute of Food and Agriculture Hatch Multi-State Project 1022291 (A.S.W.) and McIntire-Stennis Project 1019522 (WHM). AEP was supported by the Fram Centre for High North Research Catchment to Coast (C2C) research programme. The Murray-Darling Basin Authority is thanked for the provision of the water quality and flow dataset described in Biswas and Mosley 2018. We thank Anna Lintern and Robert Sargent of Monash University for their help in gathering additional data from Australian state monitoring networks. We would like to thank Tejshree Tiwari for her work compiling and combining datasets from the Swedish Government Monitoring Program and Anya Suslova, Lindsay Scott, and Max Holmes of the Arctic Great Rivers Observatory for their help interpreting Arctic River data. Finally, we are extremely grateful to all those who collected, analyzed, and generously shared these data with the scientific community. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions

K.J.J. – conceptualization; design; data collection, harmonization, analysis, and visualization; wrote first draft, editing. K.J. – conceptualization; design; data harmonization, analysis, and visualization; writing; editing. N.L. – workflow design; data harmonization and analysis; editing. S.B. – design; data harmonization and analysis; editing. P.J. – workflow design; data harmonization; editing. L.S. – design; data harmonization and analysis; editing. D.M. – Conceptualization; design; data collection; editing. P.K. – design, data collection; editing. H.L. – design; data collection, editing. W.M. – data collection, editing. A.W. – data collection, editing. R.H. – data collection, editing. A.P. – data collection, editing. A.S. – data collection, editing. F.W. – data collection, editing.

L.M. – data collection, editing. P.L.S. – conceptualization; design; data visualization; writing; editing. J.C. – conceptualization; design; data collection; writing; editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05937-2>.

Correspondence and requests for materials should be addressed to K.J.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025