Geochemical Perspectives Letters

River chemistry reveals a large decrease in dolomite abundance across the Phanerozoic

J.M. Husson, L.A. Coogan

Supplementary Information

The Supplementary Information includes:

- Data Sources
- Supplementary Tables S-1 to S-4
- Supplementary Figures S-1 to S-7
- Supplementary Information References

Data Sources

River chemistry data were combined from two structured databases. The first source was the USGS National Water Information System (NWIS) database (https://www.waterqualitydata.us/, accessed May 2022), from which measurements of dissolved Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, pH and $CO_{2(aq)}$ on filtered river water samples were downloaded. These water samples (183,195 in total) were collected from 16,350 monitoring stations. The second source was the Global River Chemistry Database (GLORICH; Hartmann *et al*., 2014), and consisted of measurements of dissolved Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, pH and alkalinity on 65,534 river water samples from 3,198 stations. Collectively, the compiled data consist of 248,729 water samples collected from 19,548 measurement stations.

Geologic maps stored in the Macrostrat database (Peters *et al*., 2018) were used to constrain the bedrock geology of these rivers. Geologic maps in this system are from heterogeneous sources, but every map consists of geologic units that minimally have chronostratigraphic age estimates and lithology descriptions. These data are linked to hierarchical vocabularies (accessible via the Macrostrat API: https://macrostrat.org/api/v2/defs/intervals and https://macrostrat.org/api/v2/defs/lithologies).

The age and lithologies of the bedrock that directly underlies each measurement station was determined via Macrostrat's API (https://macrostrat.org/api/v2/geologic_units/map), using the station's latitude and longitude. If the point intersected multiple maps in Macrostrat, data from the largest scale map (i.e., highest spatial resolution) was adopted. An upper and lower bedrock age was adopted based on the chronostratigraphic age bin assigned to the geologic unit. A 'station fraction carbonate' value (f_{stn}) was also calculated for each station as the number of listed carbonate lithologies divided by the total number of listed lithologies. In total, 8,408 map units from 172 sources were used to describe station bedrock geology.

As an alternate approach, Macrostrat was used constrain the bedrock geology of entire watersheds that contain the NWIS stations, utilizing the USGS National Watershed Boundary Dataset (WBD) that demarcates the U.S. into separate watersheds (https://www.usgs.gov/national-hydrography/watershed-boundary-dataset, USGS, 2013). Boundaries of the 5,316 different watersheds that contain NWIS stations were used to create clipped geologic maps. An area-weighted 'watershed fraction carbonate' value (f_{ws}) was calculated for each watershed – for example, if a watershed is composed of two mapped units of equal area (one igneous, and the other limestone-sandstone), its f_{ws} value would be 0.25. An upper and lower age for the watershed's bedrock was determined by an area-weighted average of the upper and lower ages of the carbonate-containing map units. In total, 198,391 map units from 16 sources were used, primarily from Horton *et al*. (2017).

Supplementary Tables

Data are provided in a supplemental Excel file, consisting of 4 tables. Each table and its fields are described below.

Table S-1 (sources) All geologic map references used to constrain the bedrock geology of measurement stations and watersheds are included in this table.

(1) source_id: unique id for a reference, assigned by the Macrostrat database. Reference metadata can be viewed via the Macrostrat API (e.g., https://macrostrat.org/api/v2/defs/sources?source_id=150).

(2) num_unit_fstn: the number of geologic map units used from a given reference to constrain the 'station fraction carbonate' value (f_{stn}, see section S-1).

(3) num_unit_fws: the number of geologic map units used from a given reference to constrain the 'watershed fraction carbonate' value (f_{ws}, see section S-1).

(4) ref_title: name of the geologic map.

(5) authors: authors of the geologic map.

(6) ref_year: year of map publication.

(7) ref_source: source of the published map (e.g., name of the survey).

(8) approximate scale: each map is classified as one of four possible (approximate) scales: 1:100,000 ('large scale'), 1:1,000,000 ('medium scale'), 1:5,000,000 ('small scale') and 1:20,000,000 ('tiny scale').

Table S-2 (samples) All compiled samples of river chemistry are included in this table.

(1) sample_id: each sample has a unique sample id. Ids for NWIS samples are adopted from the original NWIS database, in which they are called 'ActivityIdentifiers'. We generated ids for GLORICH samples during compilation for this publication.

(2) station_id: each sample was measured at a monitoring station, which has a unique station id. These ids are adopted from the original databases (called 'STAT_ID' in GLORICH and 'MonitoringLocationIdentifier' in NWIS).

(3) source: source of the sample data, either NWIS or GLORICH.

(4) latitude: latitude of measurement station in decimal degrees, relative to the WGS84 datum.

(5) longitude: longitude of measurement station in decimal degrees, relative to the WGS84 datum.

(6) Ca_umol: measured concentration of dissolved Ca^{2+} in units of μmol per litre.

(7) Mg_umol: measured concentration of dissolved Mg^{2+} in units of μmol per litre.

(8) K_umol: measured concentration of dissolved K^{+} in units of μmol per litre.

(9) Cl_umol: measured concentration of dissolved Cl^{-} in units of μmol per litre.

(10) pH: measured water sample pH.
(11) CO2aq_umol: measured concentration of aqueous CO_2 in units of μmol per litre.
(12) Alk_ueq: measured water sample alkalinity in units of μeq per litre.
(13) HCO3_umol_calc: Calculated values of dissolved HCO_3^- in units of μmol per litre. These were determined via the Python toolbox PyCO2SYS 1.8.1 (Humphreys *et al.*, 2022). To maximize the number of calculations, measured pH and measured $CO_{2(aq)}$ were used for the NWIS samples to solve the full carbonate system, and measured pH and measured alkalinity were used for the GLORICH samples.
(14) Omega_calc: Calculated values of calcite saturation (Ω) calculated via PyCO2SYS 1.8.1.
(15) f_stn: the 'station fraction carbonate' value (f_{stn}) calculated for bedrock underlying each measurement station (see section S-1).
(16) stn_t_age: upper age estimate for bedrock underlying each measurement station, in units of millions of years ago (Ma.).
(17) stn_b_age: lower age estimate for bedrock underlying each measurement station, in units of Ma.
(18) f_ws: the area-weighted 'watershed fraction carbonate' value (f_{ws}) for WBD watersheds (see section S-1).
(19) ws_t_age: the area-weighted carbonate bedrock age (upper bound) in units of Ma.
(20) ws_b_age: the area-weighted carbonate bedrock age (lower bound) in units of Ma.

Table S-3 (stations) Station-averaged data used to create Figure 2 in the main text.

(1) station_id: defined above.
(2) latitude: defined above.
(3) longitude: defined above.
(4) F_Mg_avg: mean of water sample Mg/(Ca + Mg) molar ratios (F_{Mg}) for a given measurement station.
(5) std_prcnt: one standard deviation on the F_{Mg} mean, in units of percent.
(6) num_F_Mg: number of water sample Mg/(Ca + Mg) molar ratios used to calculate 'F_Mg_avg'.
(7) Omega_calc_avg: mean water sample calcite saturation for a given measurement station, using water sample Ω values calculated via PyCO2SYS 1.8.1
(8) f_stn: defined above.
(9) stn_t_age: defined above.
(10) stn_b_age: defined above.
(11) f_ws: defined above.
(12) ws_t_age: defined above.
(13) ws_b_age: defined above.

(14) in_fig2A: boolean variable indicating whether this station-averaged data was included in Figure 2a in the main text.
(15) in_fig2B: as above, for Figure 2b.
(16) in_fig2C:as above, for Figure 2c.

Table S-4 (time series) Age-binned averages of F_{Mg} plotted in Figure 2 of the main text.

(1) t_age: upper age of the age bin, in units of Ma.
(2) b_age: lower age of the age bin, in units of Ma.
(3) age: mean age of age bin, in units of Ma.
(4) F_Mg_fig2A: age-binned F_{Mg} value plotted in Figure 2a.
(5) F_Mg_fig2B: age-binned F_{Mg} value plotted in Figure 2b.
(6) F_Mg_fig2C: age-binned F_{Mg} value plotted in Figure 2c.

Tables S-1 through S-4 (.xlsx) are available for download from the online version of this article at https://doi.org/10.7185/geochemlet.2316.

Supplementary Figures

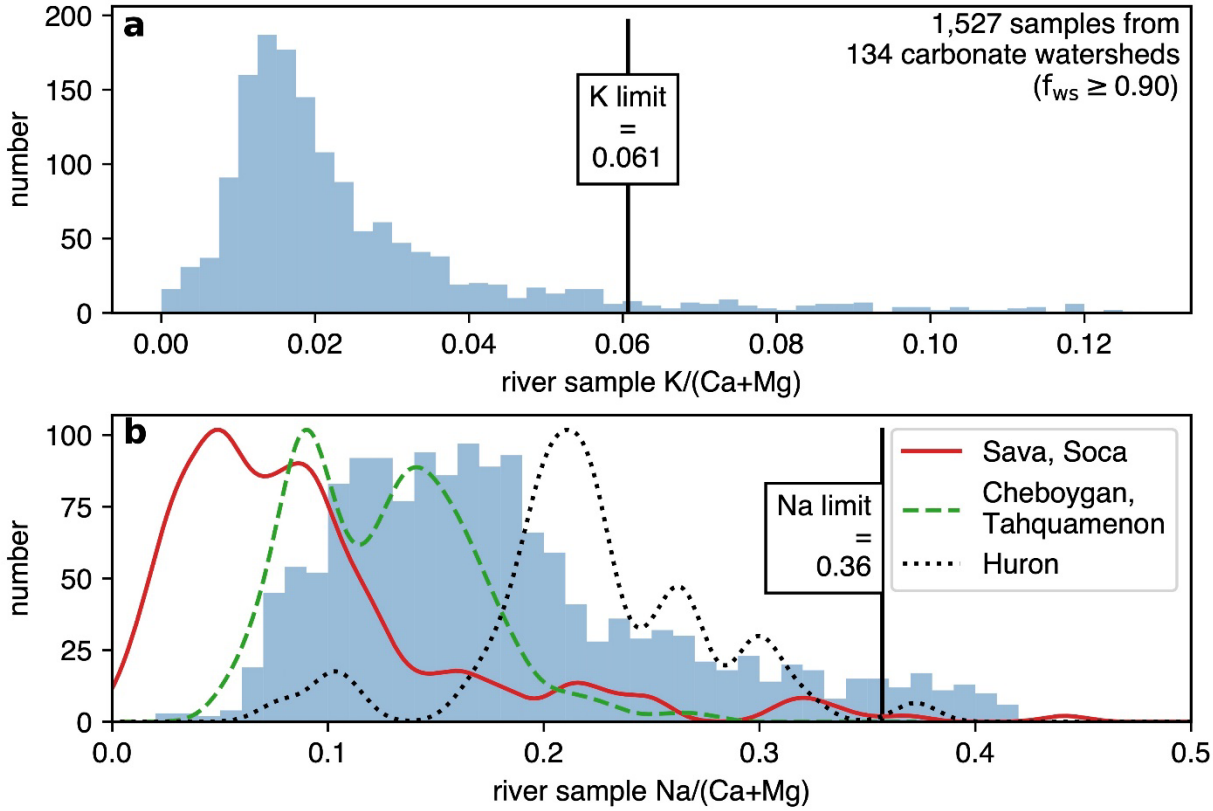

Figure S-1 The distribution of molar ratios of K/(Ca + Mg) **(a)** and Na/(Ca + Mg) **(b)** from NWIS stream samples from 134 watersheds where carbonate outcrop is ≥90 % of mapped bedrock (i.e., $f_{ws} \geq 0.9$), where carbonate weathering is expected to dominate. Each sub-plot also shows the upper limits for K/(Ca + Mg) and Na/(Ca + Mg) used to filter samples to reduce the influence of silicate weathering. Samples used to construct the time series in Figure 2b, c have K/(Ca + Mg) and Na/(Ca + Mg) ratios below these limits (0.061 and 0.36, respectively). Also shown in **(b)** are distributions of Na/(Ca + Mg) ratios from detailed studies of watersheds with ~30–70 % carbonate outcrop in Slovenia (Sava and Soča watersheds; Szramek *et al.*, 2011) and Michigan (Cheboygan, Tahquamenon and Huron watersheds; Williams *et al.*, 2007). The heights of these distributions (solid red, dashed green and dotted black lines) are arbitrarily scaled to the *y*-axis for clarity. In Michigan, the Huron watershed is more affected by anthropogenic pollution compared to Cheboygan and Tahquamenon, which can explain why its Na/(Ca + Mg) distribution is shifted to the right compared to the others (road salt (NaCl) pollution; Williams *et al.*, 2007).

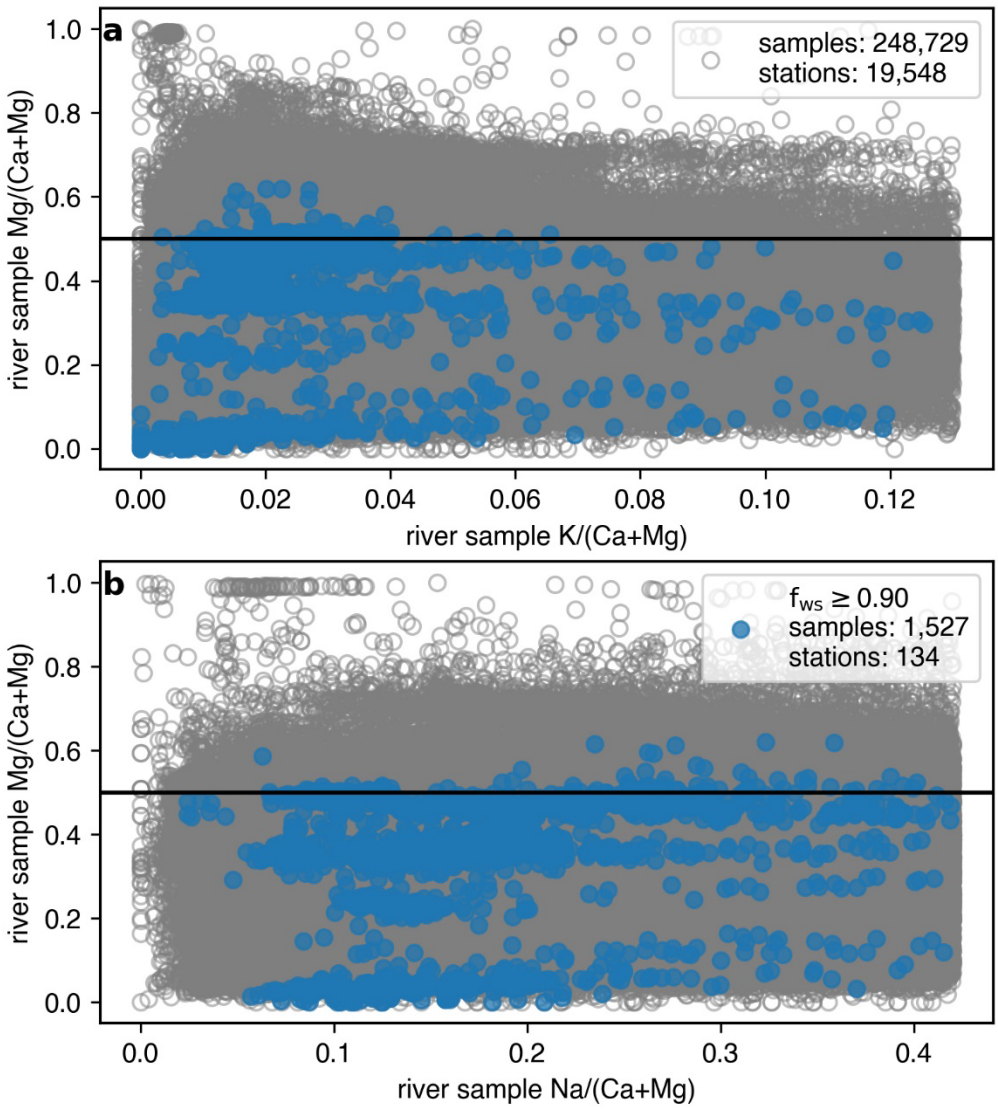

Figure S-2 Cross-plots of K/(Ca + Mg) **(a)** and Na/(Ca + Mg) **(b)** versus Mg/(Ca + Mg). Open grey circles are all the data, whereas solid blue circles are NWIS samples from the 134 watersheds with $f_{ws} \geq 0.9$, which are also plotted in Figure S-1.

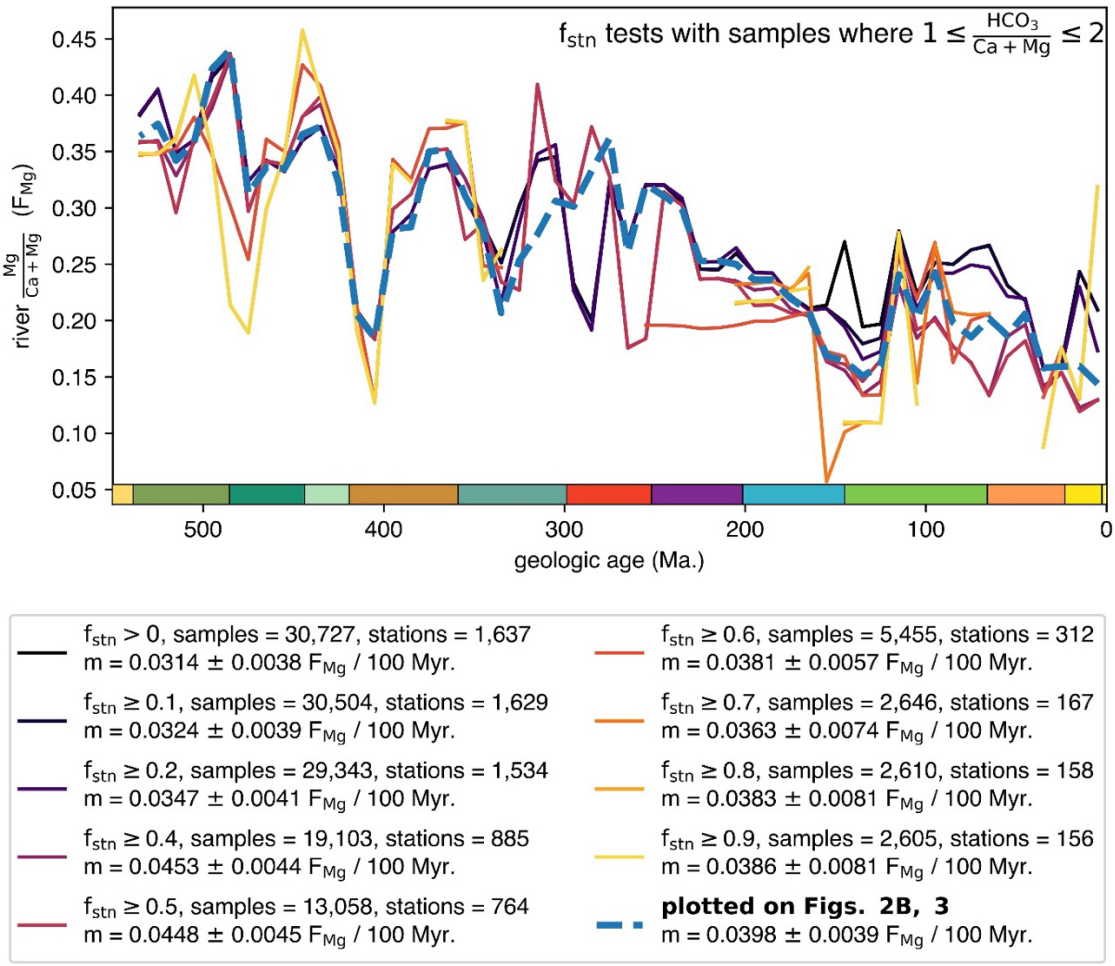

Figure S-3 Sensitivity tests of the importance of 'station fraction carbonate' (f_{stn}) as a data filter. Each solid line is an age-binned time series of Mg/(Ca + Mg) ratios (F_{Mg}) versus bedrock age, constructed using the same K/(Ca + Mg), Na/(Ca + Mg) and HCO3/(Ca + Mg) filters as were used to produce the age-binned time series in Figures 2b and 3 of the main text (replotted here as a dashed blue line). The only difference in method is that the sensitivity tests (solid lines) were constructed with different f_{stn} filter values, increasing from >0 to ≥0.9, which affects the number of samples and stations used to construct the time series. A filter value of $f_{stn} \geq 0.33$ is used in Figures 2b and 3 (dashed blue line). In the legend, each curve is labeled with the number of samples and stations in each time series and the slope of a linear fit to the data, in units of F_{Mg} per 100 Myr. (with its uncertainty reported at the 1σ level). The slope of the linear fit (~0.04 per 100 Myr.) to the dashed blue curve, which is what is discussed in the main text, overlaps with all other 9 slopes at 2σ uncertainty (and overlaps with most at 1σ uncertainty). No slopes overlap with zero within 2σ uncertainty. These results mean that significant and similar declines in F_{Mg} with decreasing age are observed in each time series.

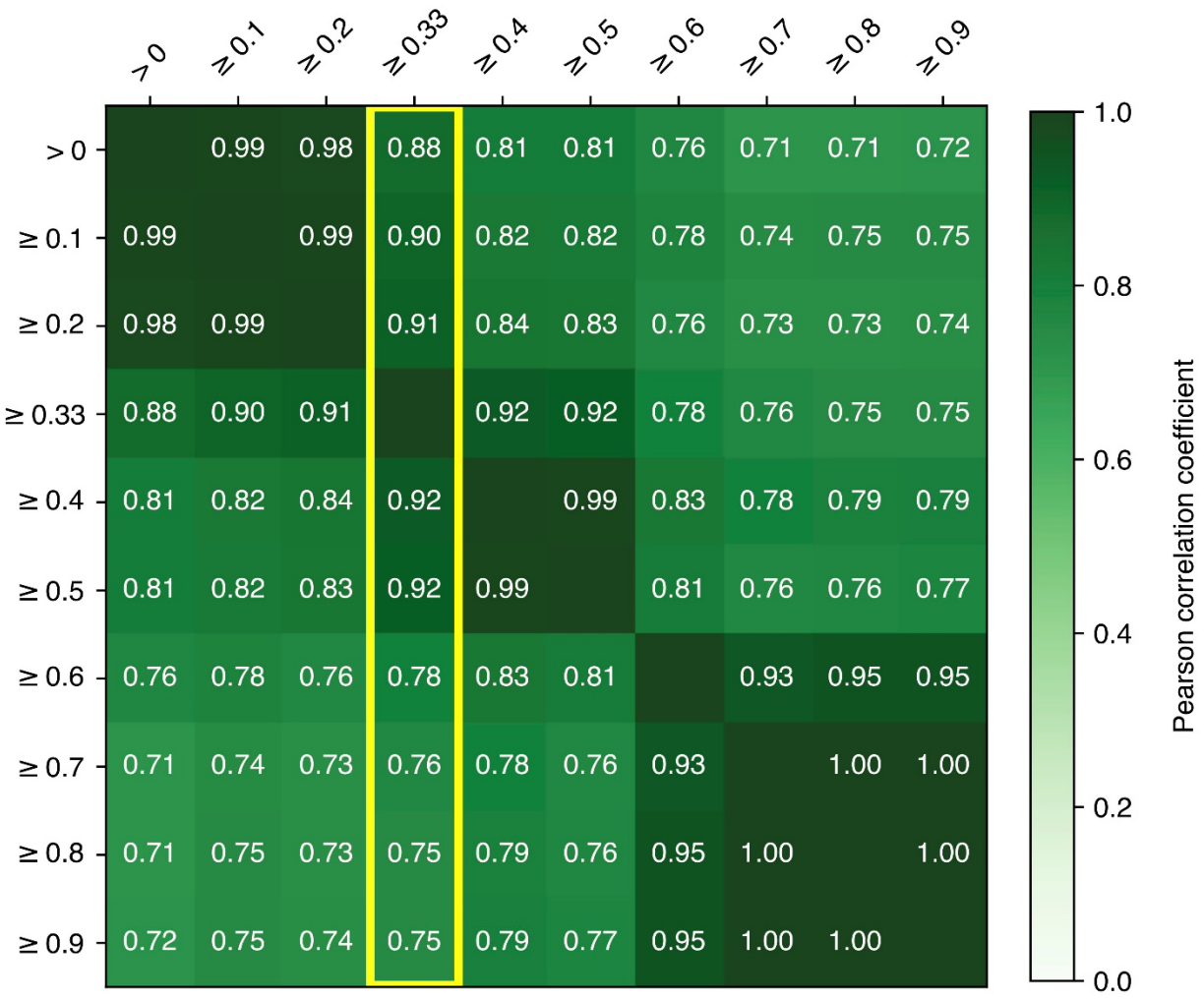

Figure S-4 Correlation coefficient matrix for the sensitivity tests shown in Figure S-3. Each colour coded cell shows the Pearson *r* value between pairs of time series (labeled with their corresponding f_{stn} filter value). Each sensitivity test is positively correlated with the time series plotted and discussed in the main text (Figs. 2b and 3, using $f_{stn} \geq 0.33$ as a filter), meaning similar results are found across a wide range of f_{stn} filter values (yellow outlined column).

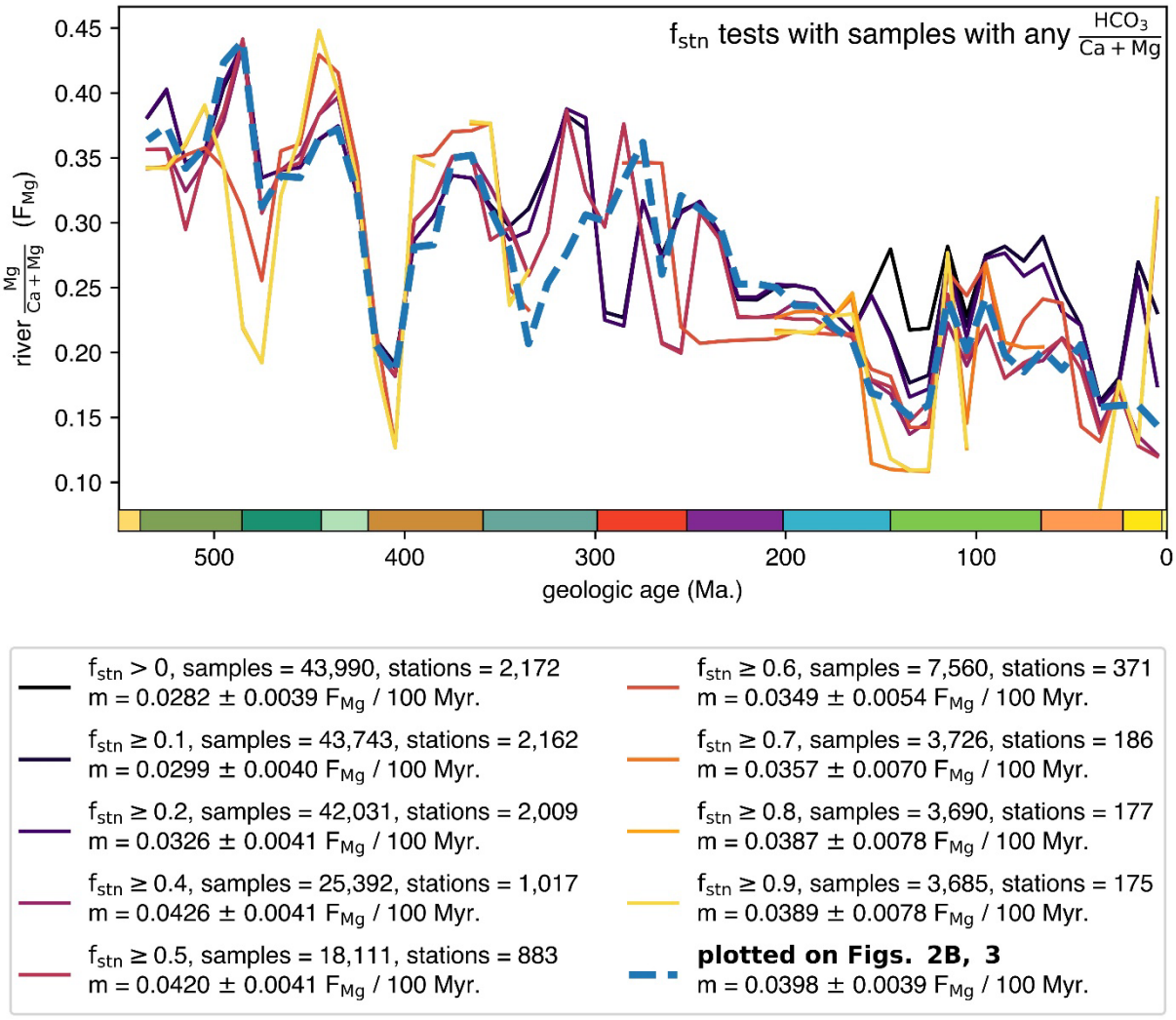

Figure S-5 Sensitivity tests for f_{stn} constructed in the exact same way as in Figure S-3, except that the $HCO_3/(Ca + Mg)$ filter was **not** used.

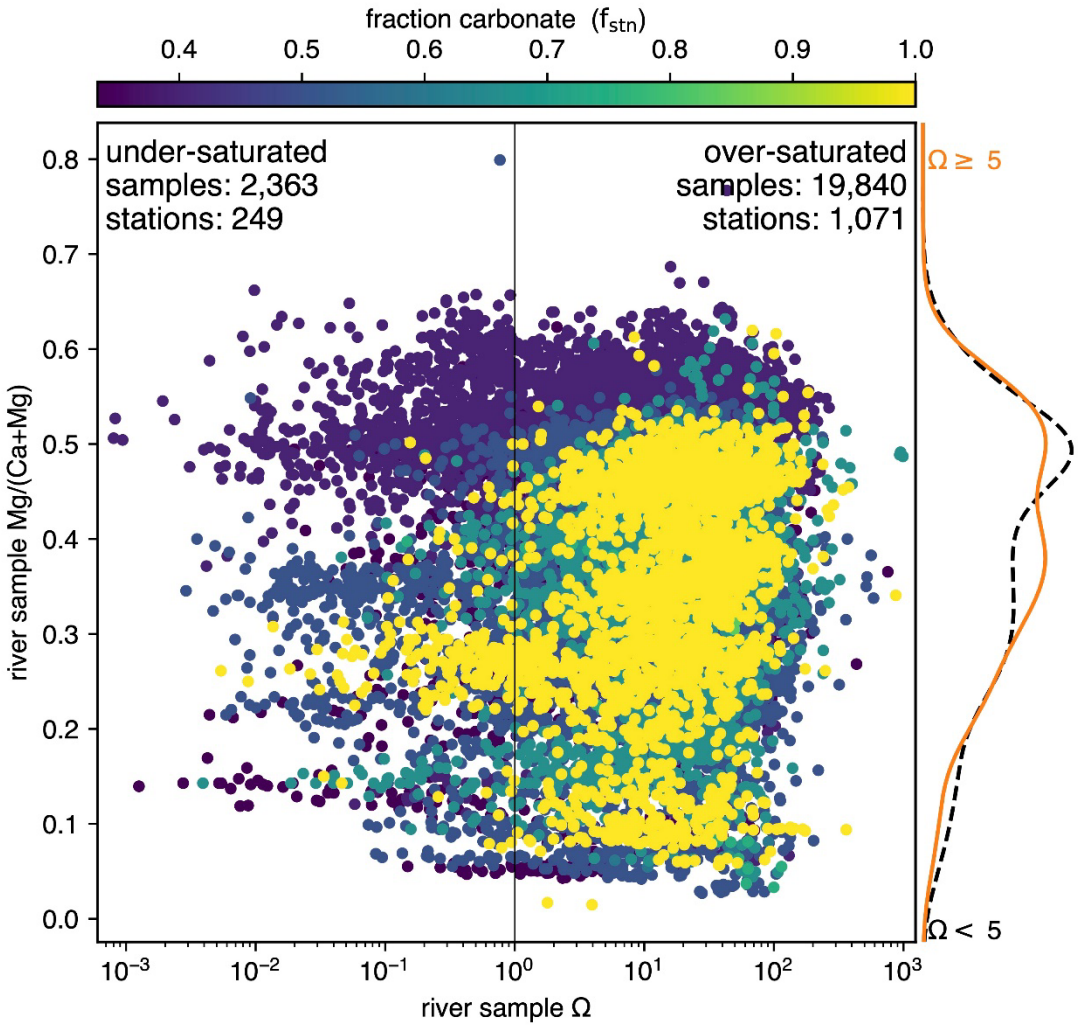

Figure S-6 Cross-plot of calcite saturation (Ω) versus Mg/(Ca + Mg), colour-coded by f_{stn} value (which range between 0.33 and 1.0). The samples included are those used in Figures 2b and 3 in the main text (22,203 in total). Most samples (90 %) are oversaturated (Ω > 1), an observation comparable to datasets from the well-studied carbonate-weathering catchments in Michigan and Slovenia (Williams *et al*., 2007; Szramek *et al*., 2011), in which 75 % of samples had Ω > 1 (and 30 % were above 5). Previous studies of calcite saturation in streams have argued that calcite precipitation does not occur until Ω is at least 5–10 (Suarez, 1983; Herman and Lorah, 1987). However, the distribution of Mg/(Ca + Mg) for NWIS and GLORICH samples with Ω ≥ 5 (orange line) looks identical to the distribution for samples with Ω < 5 (dashed black line). This observation is consistent with data from Michigan and Slovenia, which show no systematic downstream change in Mg/(Ca + Mg) compared to river headwaters that may have been driven by progressive calcite precipitation (Williams *et al*., 2007; Szramek *et al*., 2011).

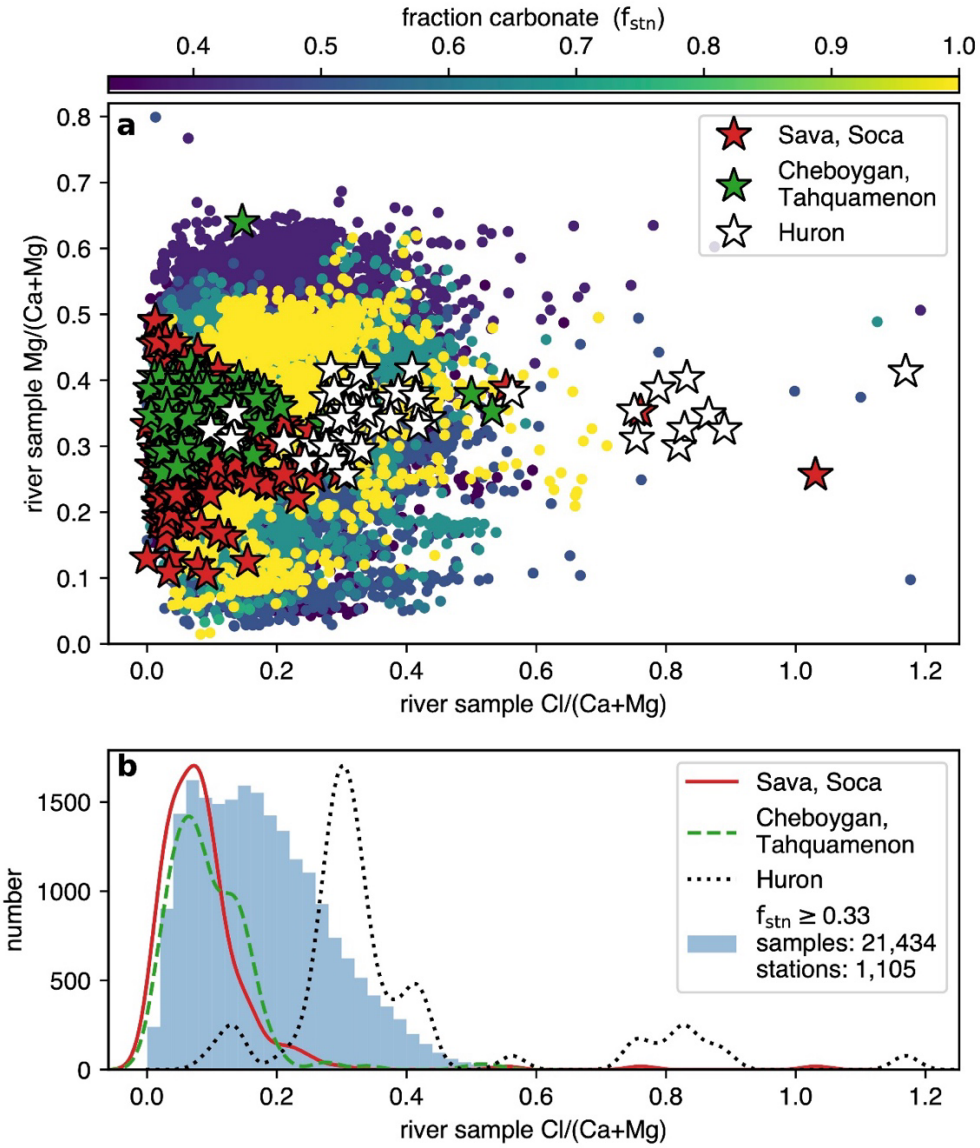

Figure S-7 Cross-plot of Cl/(Ca + Mg) versus Mg/(Ca + Mg) **(a)** and the distribution of Cl/(Ca + Mg) ratios **(b).** Data in **(a)** are colour-coded by f_{stn} value which range between 0.33 and 1.0. The samples included are those used in Figures 2b and 3 in the main text that also have measured Cl^- concentrations (21,434 in total). Also shown is data from Slovenia (Sava and Soča watersheds, Szramek *et al*., 2011) and Michigan (Cheboygan, Tahquamenon and Huron watersheds, Williams *et al*., 2007). The heights of these distributions (solid red, dashed green and dotted black lines in **(b)**) are arbitrarily scaled to the *y*-axis for clarity. As in Figure S-1b, the Cl/(Ca + Mg) distribution from the Huron watershed is shifted to the right compared to the others, which is likely a signal from increased road salt (NaCl) pollution (Williams *et al*., 2007).

Supplementary Information References

Hartmann, J., Lauerwald, R., Moosdorf, N. (2014) A brief overview of the GLObal RIver CHemistry Database, GLORICH. *Procedia Earth and Planetary Science* 10, 23–27. https://doi.org/10.1016/j.proeps.2014.08.005

Herman, J.S., Lorah, M.M. (1987) CO_2 outgassing and calcite precipitation in Falling Spring Creek, Virginia, USA. *Chemical Geology* 62, 251–262. https://doi.org/10.1016/0009-2541(87)90090-8

Horton, J.D., San Juan, C.A., Stoeser, D.B. (2017) The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States. U.S. Geological Survey Data Series 1052. https://doi.org/10.3133/ds1052

Humphreys, M.P., Lewis, E.R., Sharp, J.D., Pierrot, D. (2022) PyCO2SYS v1. 8: marine carbonate system calculations in Python. *Geoscientific Model Development* 15, 15–43. https://doi.org/10.5194/gmd-15-15-2022

Peters, S.E., Husson, J.M., Czaplewski, J. (2018) Macrostrat: a platform for geological data integration and deep-time Earth crust research. *Geochemistry, Geophysics, Geosystems* 19, 1393–1409. https://doi.org/10.1029/2018GC007467

Suarez, D. (1983) Calcite supersaturation and precipitation kinetics in the Lower Colorado River, All-American Canal and East Highline Canal. *Water Resources Research* 19, 653–661. https://doi.org/10.1029/WR019i003p00653

Szramek, K., Walter, L.M., Kanduč, T., Ogrinc, N. (2011) Dolomite Versus Calcite Weathering in Hydrogeochemically Diverse Watersheds Established on Bedded Carbonates (Sava and Soča Rivers, Slovenia). *Aquatic Geochemistry* 17, 357–396. https://doi.org/10.1007/s10498-011-9125-4

United States Geological Survey (USGS) (2013) Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD). *U.S. Geological Survey and U.S. Department of Agriculture.* https://pubs.usgs.gov/tm/11/a3/pdf/tm11-a3_4ed.pdf.

Williams, E.L., Szramek, K.J., Jin, L., Ku, T.C., Walter, L.M. (2007) The carbonate system geochemistry of shallow groundwater–surface water systems in temperate glaciated watersheds (Michigan, USA): Significance of open-system dolomite weathering. *Geological Society of America Bulletin* 119, 515–528. https://doi.org/10.1130/B25967.1