Supplementary Materials for

Global extent of rivers and streams
George H. Allen* and Tamlin M. Pavelsky

*Corresponding author. Email: georgehenryallen@gmail.com

Published 28 June 2018 on Science First Release
DOI: 10.1126/science.aat0636

This PDF file includes:

Materials and Methods
Figs. S1 to S7
Caption for Data S1
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aat0636/DC1)

Data S1
Materials and Methods

1.0 Constructing the GRWL Database

1.1 Measuring rivers at mean annual discharge

The GRWL Database is composed of planform measurements of rivers at approximately mean discharge. While bankfull discharge usually represents the channel-forming flow (32), mean discharge occurs much more often and thus is more consistently observable from limited remotely sensed imagery. It is also more relevant for mean biogeochemical flux estimates. Studies of fluvial hydrology suggest that geomorphic relationships relating width to discharge or drainage area are similar regardless of the discharge frequency used (9, 33). Temporal fluctuations in discharge are a combined result of unpredictable events (e.g. storms and droughts) and more predictable seasonal variability in runoff. Unfortunately, no high-density, global-scale, daily discharge datasets exist to track the specific days that rivers are at their mean discharge. Instead, we approached the problem by determining the time of year that rivers in different parts of the globe are most likely to be at mean discharge.

To determine the optimal time of year to measure rivers, we used an international archive of long-term (>10 years) mean monthly discharge measurements composed of 3,693 gauges archived by the Global Runoff Data Centre (10). For each station with a complete record (i.e. no missing data values), we constructed a mean monthly hydrograph and calculated the mean and standard deviation (1σ) for all monthly measurements (fig. S1). All months with discharges that fall within one standard deviation of the mean received an optimality score,

\[ \text{score}(m) = Q_\mu - \frac{1}{2} Q_m - \frac{1}{4}(Q_{m-1} + Q_{m+1}), \]

where \( m \) is a given month being scored, \( Q_\mu \) is the mean of all monthly discharges, \( Q_m \) is the mean monthly discharge of the given month, \( Q_{m-1} \) is the mean monthly discharge one month before month \( m \), and \( Q_{m+1} \) is the mean monthly discharge of the month after month \( m \). River discharge is more likely to be at or near the overall mean value during months with lower scores. Thus, at each gauge station we produced a list of months ranked by the probability that the river was at mean discharge. To assign the monthly rankings from the gauge stations to each Landsat tile (landsat.usgs.gov), we considered both the proximity to gauge location and the monthly ranking. For example, the most highly ranked month from the nearest gauge station has the greatest weight in setting the monthly preference order at a given Landsat tile. Each Landsat tile is assigned at least one monthly preference and up to five ordered preferences (fig. S1).

Note that this interpolation method is a potential source of error in the GRWL Database. Two rivers can be close in proximity but exhibit disparate hydrographs. If both rivers are within the same Landsat tile, one river may not be sampled during conditions of mean discharge. However, this source of error is accounted for in the validation of GRWL (section 1.4) and is ultimately incorporated into the RSSA uncertainty estimate (section 2.4).

1.2 Landsat imagery acquisition

Landsat TM, ETM+ (SLC-on), and OLI scenes were acquired from two data sources. We automatically downloaded 1,071 scenes from the Global Land Cover Facility (GLCF, glcf.umd.edu) over North America. The 6,261 remaining scenes were
downloaded manually from USGS Earth Explorer website (https://earthexplorer.usgs.gov/). The highest-ranking scene was downloaded first. Upon download, each scene was visually inspected for clouds, river ice, shadows, and flooding and either kept for use or discarded. If discarded, the next highest-ranking scene was automatically downloaded. Four hundred and forty-three Landsat tiles, located primarily in the tropics, had no cloud-free scenes available. To address this problem, we developed a program in IDL (version 8.0) that identifies clouds based on their spectral signature and splices two or more complementary scenes together to produce a cloud-free composite image (34).

1.3 Image processing and GIS

To classify water, we used the modified normalized difference water index,

\[ MNDWI = \frac{\text{green} - \text{MIR}}{\text{green} + \text{MIR}}, \]

where MIR is the middle infrared band (e.g. OLI Band 6) and green is the green band (e.g. OLI Band 3) (35). We applied the MNDWI formula to all Landsat scenes, mosaicked the scenes, and clipped the resulting mosaics to 4° Latitude by 6° Longitude tiles. We then created a binary water mask by applying a dynamic threshold (36) that was visually inspected and corrected for any gaps in continuity or classification errors. These errors stem from sources including river view obstruction by topographic shadows, bridges, dams, or the erroneous inclusion of swamps, large lakes, or deltas in the river network. We used RivWidth (12) (version 0.4) to calculate a channel centerline for all river reaches longer than 10 km (Fig. 1). We visually inspected the RivWidth output for errors. Reservoirs and lakes connected to the fluvial network were labeled using GIS methods and the Global Reservoir and Dam Database (37). The portions of GRWL labeled as lakes and reservoirs were then visually inspected and labeling corrected as necessary in ArcGIS.

1.4 Database validation

We validated Landsat-derived river width measurements using stream flow and river width records at 1,250 stream gauges (Database S1) operated by the United States Geological Survey (USGS; http://nwis.waterdata.usgs.gov/nwis) and the WSC (http://www.ec.gc.ca/rhc-wsc/). At each gauge location, we estimated the river width at mean annual discharge and compared this value to the average of the five spatially closest RivWidth measurements (Fig. 2). We excluded river width measurements that: 1) were taken more than 200 m upstream or downstream from the gauge station; 2) were taken when river ice was observed; or 3) were labelled as “Poor” measurements by the USGS or WSC. We then compared the mean of all width measurements acquired when river discharge was within 10% of the mean annual discharge (red dots, Fig. 2A) to the mean width of the five nearest GRWL river width measurements. The residuals between in situ and remotely sensed width measurements are heteroscedastic, uncorrelated, and unbiased with changing in situ width.

GRWL width measurements show very little mean bias (0.97 m) relative to in situ width measurements at mean discharge, suggesting the Landsat scenes were sampled at times that, on average, matched mean discharge timing. The RMSE between GRWL and in situ widths is 25.2 m, a length similar to the minimum theoretical uncertainty of Landsat-derived river widths calculated from a binary water mask (12). The RMSE value
also incorporates several other sources of error, including differences in discharge between the remotely sensed and in situ measurements and error in the in situ width measurements.

To avoid bias from outliers, we used the Theil-Sen median estimator \( (38) \) to derive a robust linear regression between GRWL and in situ width measurements (Fig. 2C). Regression of in situ widths \( \geq 90 \) m yields a slope of very close to unity, but inclusion of all river width data \( (\geq 30 \) m) produces a slope that deviates by 17%. This deviation is expected because GRWL is more likely to include overestimates of river width compared to underestimates where river width approaches the resolution of the Landsat imagery. In other words, GRWL never includes underestimates of 30 m wide rivers because they are narrower than one Landsat pixel, but it will include overestimates of these rivers. Goodness of fit \( (r_s=0.81) \) was characterized using Spearman’s nonparametric correlation coefficient \( (39) \). Overall, comparison with in situ measurements suggests that GRWL provides, on average, an accurate representation of river widths at mean annual discharge to the extent that this is possible from Landsat imagery.

We note that the accuracy of GRWL in the United States and Canada may differ from the average accuracy of GRWL worldwide. For example, the validation analysis does not assess the performance of GRWL on tropical rivers, which comprise a substantial proportion of the global river network length. Unfortunately, we are unaware of other publicly available databases of paired width and discharge measurements similar to those available from the USGS and Water Survey Canada. To visually characterize GRWL performance in a variety of locations, figs. S2-S5 show the GRWL water mask compared with the Global Surface Water (GSW) median occurrence map (>50th percentile water occurrence within a 12-year span of Landsat imagery) from Pekel et al. (21). The GRWL water mask tends to capture more rivers than the GSW median occurrence map and does not include lentic water bodies like lakes and marshes. Figs. S2-S5 also compare the GRWL water mask with the GRWL vector product. The GRWL vector product performs best on curvilinear river morphologies (fig. S2) and on braided rivers (fig. S3). The GRWL vector product can contain gaps in complex, non-dendritic delta networks (fig. S4) or in large, complexly-shaped lakes/reservoirs (fig. S5). These gaps arise because the vectorization algorithm used here (RivWidth) was specifically optimized for rivers and not deltas or reservoirs \( (12) \). We note, however, that lakes and reservoirs are excluded from the RSSA analysis presented in this study.

2.0 Global RSSA calculation

To calculate the surface area of rivers and streams globally, we used direct GRWL measurements in combination with statistical relationships between river width, abundance, and basin-averaged climate and area. The Pareto statistical extrapolations of river surface area are based on principles of fractal river network scaling theory within basins \( (24, 25) \) as well as empirical evidence \( (2, 25, 26, 40, 41) \). The Pareto distribution can be expressed as a probability density function,

\[
PDF(x) = \frac{ax^\alpha}{x_{min}^{\alpha+1}},
\]

where \( x_{min} \) is the Pareto scale parameter and \( \alpha \) is the Pareto shape parameter. We isolated individual river networks by clipping GRWL to each highest-order basin in the HydroBASINs global hydrography database \( (22) \).
2.1 Class A basins

Class A basins contain >250,000 river measurements and exhibit well-developed fractal Pareto frequency distributions of RSSA with consistent Pareto shape parameters, (mean Kolmogorov-Smirnov goodness of fit, $D=0.10\pm0.02$, $p<0.001$, fig. S6). The departure of the empirical data from the Pareto distribution for wide rivers, apparent in Fig. 3C, is an expected outcome of the finite-size effect, or, in other words, the limiting consequence of basin size on maximum river width (42). For narrow streams, fractal stream network theory (40) and field evidence (2, 26) indicate that river length and width scale exponentially with river order down to first-order streams, so we expect RSSA, the product of these two attributes, to be Pareto distributed down to first-order streams.

Thus, in Class A basins we fitted a Pareto distribution of RSSA measurements using maximum likelihood estimation (MLE) in R (see Github code repository). To fit the distribution, we set the Pareto scale parameter, $x_m$, equal to the lowermost observed RSSA value and solved for the Pareto shape parameter. This lowermost observed RSSA, $O_{\text{min}}=3259$ m$^2$, is the product of the minimum observed river width threshold (90 m) and the mean distance between river centerline pixels in 30-m resolution Landsat imagery (36.21 m). Thus, the Pareto distribution was fit between $O_{\text{min}}$ and the maximum observed RSSA in each basin, $O_{\text{max}}$. We then extended this statistical fit to narrower rivers and streams by generating a new probability distribution with the same MLE-derived Pareto shape parameter using a new Pareto scale parameter, equal to the lowermost expected RSSA, $E_{\text{min}}=11.6\pm2.9$ m$^2$. This lowermost RSSA corresponds with the mean of the medians of first order stream widths ($0.32\pm0.08$ m) found by Allen et al. (23), multiplied by the mean distance between centerline pixels in GRWL.

The definite integral of this fitted RSSA probability density distribution from $O_{\text{min}}$ to $O_{\text{max}}$ equals the total probability density of observed RSSA. Similarly, the definite integral of the RSSA probability distribution between $E_{\text{min}}$ and $O_{\text{min}}$ equals the total probability of the density of the RSSA estimated through statistical extrapolation (represented by the light gray polygon in Fig. 3C). The ratio between the observed RSSA, $R_{A_{\text{obs}}}$, and the estimated RSSA, $R_{A_{\text{est}}}$, equals the ratio between these two definite integrals,

$$\frac{R_{A_{\text{obs}}}}{R_{A_{\text{est}}}} = \int_{O_{\text{min}}}^{O_{\text{max}}} \frac{x^\alpha}{x^{\alpha+1}} dx / \int_{E_{\text{min}}}^{O_{\text{min}}} \frac{x^\alpha}{x^{\alpha+1}} dx. \quad (4)$$

We solved for $R_{A_{\text{est}}}$ in Eq. 4 and added $R_{A_{\text{obs}}}$ to get the total RSSA contained within each Class A basin,

$$RA = R_{A_{\text{est}}} + R_{A_{\text{obs}}}. \quad (5)$$

2.2 Class B basins

Class B basins contain an intermediate amount of GRWL data, specifically between 10,000 and 250,000 river measurements. We applied the same approach in Class B basins as in Class A basins, except that rather than fitting the Pareto shape parameter to each individual basin, we used the mean and standard deviation of the Pareto shape parameter that were fit in Class A basins ($\alpha=0.90\pm0.06$; fig. S2).

2.3 Class C basins

To estimate the RSSA in Class C basins, which contain very little GRWL data, we developed an empirical power-law relationship between climate aridity, basin area, and percent basin occupied by RSSA (Fig. 3E). Class C basins, containing <10,000 GRWL
width measurements ≥90 m, tend to be small and/or dry basins. For these basins, we developed a relationship between basin percent RSSA (%RSSA), basin area (BA) (22), and aridity index (AI) (27). We used a least squares multiple linear regression of log-transformed data weighted by basin area to interpolate RSSA in Class C basins (Fig. 3E). Larger basins have a larger percent %RSSA because they contain higher-order rivers.

3.0 RSSA uncertainty estimate

The RSSA uncertainty calculation is based on errors associated with GRWL measurements, the uncertainty of the basin Pareto shape parameter (α), the lower bounds of the Pareto extrapolation, and the uncertainty of the aridity-basin area-%RSSA multiple regression. We used Monte Carlo error propagation (N ensembles = 500) to compound each source of error so that Class A basins tend to have the least uncertainty, while Class C basins, particularly humid Class C basins, tend to have the greatest uncertainty (Fig. 4B).

3.1 Class A basins

We estimated total RSSA uncertainty in Class A basins by taking into account the uncertainty of GRWL measurements and the lower boundary of the RSSA extrapolation (Fig. 3C). We represented GRWL errors by characterizing the distribution of the GRWL-in situ measurement residuals with a normal distribution (fig. S7). The fitted normal distribution overestimates the tails of the residuals, so we are conservatively overestimating the errors associated with GRWL. For each ensemble run, we perturbed all GRWL width measurements by values randomly sampled from this normal distribution (see red bars in fig. S6). Adding Gaussian noise to the GRWL data did not substantially impact the variability of α (see blue text in fig. S6) because the perturbations effectively cancel themselves out. However, adding Gaussian noise to the GRWL measurements artificially increased α due to the skewed distribution of the RSSA data and the 90 m lower threshold imposed on the GRWL width measurements. Therefore, we used the unperturbed RSSA data to establish a mean α, and used the Monte Carlo simulation to calculate the uncertainty around this mean α value. We extrapolated the Pareto distribution down to a value sampled from a normal distribution corresponding to the range of median first-order stream widths field surveyed by Allen et al. (23) (μ=0.32 m, σ=0.08). We calculated a finite definite integral for each ensemble run and determined the RSSA standard deviation in each Class A basin.

3.2 Class B basins

In Class B basins, we combined the uncertainty of GRWL measurements, the Pareto fit, and the Pareto extrapolation minimum boundary to estimate total RSSA uncertainty (Fig. 3D). The RSSA uncertainty in Class B basins was calculated in the same manner as in Class A basins except that the value of α used in the extrapolation was based on fits from Class A basins (see section 2.2).

3.3 Class C basins

Uncertainty in Class C basins is based on the RSSA uncertainty of Class A and B basins, and the 1σ confidence intervals of the climate-RSSA regression (Fig. 3E). Using the Monte Carlo ensemble outputs from Class A and Class B basins, we fit a weighted multiple linear regression on %RSSA, aridity, and basin area. For each ensemble run, we used the ensemble-specific regression to predict %RSSA in all Class C basins. We also
calculated the 1σ confidence intervals of the predicted %RSSA values in Class C basins for each ensemble. For each Class C basin, we added together the standard deviation of %RSSA for the ensemble runs and the mean 1σ confidence interval of the ensemble runs. This approach captures the uncertainty sourced from the RSSA in Class A and B basins and the uncertainty of the climate-RSSA regression itself.

A source of substantial uncertainty is the characterization of the lower end of the Pareto distribution shown in Fig. 3C. There may exist unknown confounding factors that cause the distribution of RSSA to depart from a Pareto distribution in smaller rivers and streams. For example, it is possible that a Pareto distribution may overestimate RSSA in mountainous regions where high gradients can cause fast-moving rivers to narrow. This extrapolation should be tested in future work. Allen et al. (23) conducted field surveys to quantify the variability of median first order stream widths over a range of geomorphic environments and hydrologic conditions. However, they did not survey streams in tropical or arid regions, which may not conform to the same range observed in temperate and arctic regions. Downing et al. (2) compiled a global empirical collection of perennial first-order stream widths and reported a mean of the median stream width of 1.6±1.1 m. Using this range of width as the lower limit of the Pareto RSSA extrapolation, reduces the global RSSA estimate to 620,000±60,000 km², although we emphasize that this RSSA estimate does not include the surface area of ephemeral streams. Allen et al. (23) also found that the fractal distribution of widths characterizing larger streams broke down in the smallest headwaters catchments and was better characterized by a lognormal distribution. Incorporating this more realistic shape in the lower end of the stream width distribution may yield a more accurate representation of the frequency distribution of RSSA.
Fig. S1. 
Method for determining the time of year to measure rivers. (A) Global Runoff Data Centre (GRDC) gauge locations used to estimate optimal months to sample rivers. (B) Example mean monthly hydrograph (George River, Canada). Months with discharges within one standard deviation (gray box) of the mean discharge (dashed horizontal line) are ranked (blue numbers) based on their discharge and that of their two neighboring months (red dots, Eq. 1). In this case, the best month to measure river width is September. (C) Month that rivers are most likely to be at mean discharge for each Landsat tile used to create GRWL.
Fig. S2.
Example of GRWL in a temperate river network (Danube River, 44.9°N, 19.7°E).
(A) GRWL water mask compared to the GSW occurrence mask (21) where blue is dataset overlap, red is GRWL only, and yellow is GSW only. (B) Red GRWL centerline vector product above GRWL river mask. (C) GRWL width data along river centerlines.
FIG. S3.
Example of GRWL along a braided river reach (Lena River, 66.4°N, 123.9°E). (A) GRWL water mask compared to the GSW occurrence mask (27) where blue is dataset overlap, red is GRWL only, and yellow is GSW only. (B) Red GRWL centerline vector product above GRWL river mask. (C) GRWL width data along river centerlines.
Fig. S4.
Example of GRWL in a complex deltaic environment (Mouth of the Amazon, 1.7°S, 51.8°W). (A) GRWL water mask compared to the GSW occurrence mask (21) where blue is dataset overlap, red is GRWL only, and yellow is GSW only. (B) Red GRWL centerlines above GRWL water mask. (C) GRWL width data along river centerlines. Note that the GRWL vector product has difficulty representing complex, non-dendritic channel network topologies like shown here.
Fig. S5. Example of GRWL above a large, complexly-shaped reservoir (Bratsk Reservoir, 55.3°N, 101.6°E). (A) GRWL water mask compared to the GSW occurrence mask (27) where blue is dataset overlap, red is GRWL only, and yellow is GSW only. (B) Red GRWL centerline vector product above GRWL water mask (rivers in light blue, reservoirs in dark blue). (C) GRWL width data along waterbody centerlines. The vector product has difficulty representing reservoirs with complex, non-linear geometries but we note that this study’s RSSA estimates do not include lake/reservoir data from GRWL.
Fig. S6

Pareto fits in each Class A basin. Empirical river and stream surface area (RSSA) histograms with Pareto distributions (blue lines) fitted by maximum likelihood estimation in all Class A basins. Red vertical lines represent range of uncertainty of each RSSA bin derived from errors in GRWL widths (fig. S6A). All histograms are plotted over the same limits, allowing for the Pareto shape parameters, \( \alpha \), represented by the slopes of the blue lines, to be comparable between panels. Uncertainties of \( \alpha \) are too small to be seen in these graphs (mean value = 0.0007).
Fig. S7
Histogram of residuals between GRWL and in situ river width measurements at mean discharge (Fig. 2). A normal distribution fit to residuals (red) is used to incorporate uncertainty of GRWL measurements into global estimate of RSSA (see section 3.1).
Additional Data table S1 (separate file)
In situ river width data used to validate the GRWL Database.
References and Notes


10. Global Runoff Data Centre, Long-term mean monthly discharges and annual characteristics of GRDC stations (2017); available at www.bafg.de/.


13. Materials and methods are available as supplementary materials.


