Supplementary Materials for

Losing the pulse of the Earth's fresh waters

Albert Ruhi, Mathis L. Messager, Julian D. Olden

correspondence to: aruhi@sesync.org

This PDF file includes:

Materials and Methods

Streamgage historical data

Historical streamflow data for all streamgage stations of the United States were compiled from the USGS National Water Information System (NWIS) using the 'dataRetrieval' R package (*1*). Daily mean discharge records do not include provisional data that have not been approved by the USGS. Consequently, stations with provisional records appear as inactive during these periods, resulting in an artificial drop in the number of active stations in recent years. To compensate for this, daily mean water discharge data were downloaded for the period January 1854 to December 2014, while instantaneous data (15-minute interval, generally), including provisional data, were downloaded for January 2015-January 2017. Periods for which river discharge records were estimated rather than measured (e.g., due to gage malfunction) were excluded from the analysis. In total, discharge data were compiled for nearly 23,000 stations.

For each year, a streamgage was classified as 'active' if it recorded discharge for at least 180 days (not necessarily consequentially) during that year, and as 'inactive' if it did not. We then computed the number of active stations in each of the 352 river basins (Hydrological Unit Code or HUC 6) of the National Water Boundary Dataset (WBD), every year from 1854 to 2016, using the geographic location of the gaging station reported in the NWIS.

Water scarcity

We represented water scarcity for each river basin according to the county-level Normalized Deficit Cumulated (NDC) (*3*). The NDC is equal to the maximum cumulative deficit between average daily water demand in 2010 and local daily renewable supply from to 1949 to 2010, divided by the average annual rainfall volume (county area x average depth of precipitation) in that county. This index does not include exogenous sources of water such as rivers and canals flowing through each county or water transfers from outside as renewable supply in its calculation. By only including endogenous sources of renewable water (i.e. runoff and groundwater recharge from precipitation in that county), this metric reflects the reliance of each county on outside sources (e.g. withdrawal from runoff originating outside the county) and on non-renewable local sources of water (e.g. groundwater with slow recharge). Moreover, by using the maximum cumulative deficit over the entire study period rather than the maximum yearly deficit (reset at the beginning of each year), this index captures the potential long-term impact of continued overconsumption of water and multiyear droughts on water resources in each county. Finally, by computing the ratio of water deficit to annual rainfall volume, NDC is normalized across climate regions to relate each county's deficit to its available renewable endogenous water budget.

We computed the HUC6-specific NDC from county-level NDC by first intersecting counties and river basins (HUC6), as administrative boundaries rarely follow watershed boundaries. We then computed for each HUC6, as:

$$NDC_{HUC6} = \frac{1}{\bar{R}_{HUC6}} \times \sum_{County=1}^{\substack{N\ county \\ in\ HUC6}} NDC_{County} \times \bar{R}_{County} \times \frac{Area_{HUC6-County\ intersection}}{Total\ area_{County}}$$

, where $\bar{R}$ is the average annual rainfall (mm) for each HUC 6 calculated using daily gridded meteorological data for the period 1949-2010 (*4*). By aggregating NDC across counties intersecting with a given river basin based on the proportion of their area located in that basin, our main assumptions were that water deficit is distributed uniformly throughout each county, and that NDC is cumulative across counties (i.e., that maximum water deficit can be summed across county boundaries). Given that NDC is based on the maximum daily water deficit in each county, the maximum water deficit in a given county may not occur the same day in other counties in the same basin. In such a case, our NDC_{HUC6} would overestimate the actual maximum water deficit for the basin. Nevertheless, because agricultural patterns and climate are relatively homogenous within river basins across multiple counties (average HUC6 area: 24,000 km^2), we believe this metric adequately reflects relative water scarcity patterns among regions in the U.S.

Flood risk

We calculated flood risk as the percentage of the total human population that lives within the 100-year flood zone where flood hazard studies existed. To compute flood risk, 2010 census block level population data were intersected with 100-year flood zone polygons from the National Flood Hazard Layer (NFHL), a digital database containing all flood hazard mapping data from the Federal Emergency Management Agency's National Flood Insurance Program (FEMA NFIP). Census-block population data was used because of its high resolution (average and median area of a census block: 0.74 km^2 and 0.02 km^2, respectively) and its ability to spatially discriminate between settled areas and waterways or uninhabited floodplains. Despite its high resolution, the section of census blocks overlapping flood zones often did not include any housing. Therefore, we calculated the number of people in each census block living in a flood zone as the product of the census-block population size and the percentage of the census block urbanized area (calculated from 2011 National Land Cover Data) overlapping the 100-year flood zone (Zone A from the NFHL). When a census block did not contain any urban land cover (LC), we computed the population living in a flood zone as the product of the census block population and the percentage of the census block intersecting a flood zone. The population in each census block living in an area for which a flood hazard study exists was computed the same way. Finally, the percentage of the population in each river basin living in a 100-year flood zone is equal to the total population living in a 100-year flood zone across all census blocks of a river basin divided by the total population living in an area for which a flood hazard study exists. When a census block straddled the boundary between two river basins, the population size living in a flood zone within each basin was proportional to the area of the census block overlapping the basin.

For each HUC6, flood risk was therefore computed as follows:

$$\text{Flood risk}_{\text{HUC6}} = \sum_{C=census\ blocks\ within\ HUC6} \frac{Population\ living\ within\ flood\ zone_C}{Population\ living\ in\ area\ with\ flood\ hazard\ study_C}$$

, where

$$\text{Flood risk}_{\text{HUC6}} = \frac{\sum_{C=1}^{N\ blocks\ with\ urban\ LC} \frac{Total\ area\ within\ HUC6_C}{Total\ area_C} \times \frac{Population_C \times Urban\ area\ within\ flood\ zone_C}{Total\ urban\ area_C}}{\sum_{C=1}^{N\ blocks\ with\ urban\ LC} \frac{Total\ area\ within\ HUC6_C}{Total\ area_C} \times \frac{Population_C \times Urban\ area\ with\ FEMA\ status_C}{Total\ urban\ area_C}} + \frac{\sum_{C=1}^{N\ blocks\ with\ no\ urban\ LC} \frac{Total\ area\ within\ HUC6_C}{Total\ area_C} \times \frac{Population_C \times Area\ within\ flood\ zone_C}{Total\ area_C}}{\sum_{C=1}^{N\ blocks\ with\ no\ urban\ LC} \frac{Total\ area\ within\ HUC6_C}{Total\ area_C} \times \frac{Population_C \times Area\ with\ FEMA\ status_C}{Total\ area_C}}$$

Fish biodiversity metrics

We compiled a dataset of the distribution of freshwater fish at the subbasin level (HUC8) throughout the United States based on the NatureServ and ancillary local and regional sources, and then computed the endemism weighted richness/unit (EWU) for each river subbasin nationwide, following (*2*):

EWU_{HUC8} = Sum of each species’ Endemism Unit

, where

Endemism Unit = 1 / Number of HUC8s where the species is present

Accordingly, the most endemic species in a HUC8 has an Endemism Unit = 1. EWU_{HUC8} were then aggregated at the river basin level (HUC6) by calculating the average EWU of subbasins within each basin.

Time-series analysis of gage 'population' dynamics

We analyzed streamgaging dynamics within each HUC6 using statistical time-series models, namely univariate, first-order, autoregressive or AR(1) models, using the 'MARSS' R package (*5*). Autoregressive models are frequently used in population viability analysis to estimate probabilities of hitting critical decline thresholds or 'quasiextinction' (*5*). Here we similarly applied this concept to streamgage 'populations' within each HUC6. In its simplest form (i.e., without considering exogenous covariates or density-dependent effects), an AR(1) model can be expressed as:

$$x_{i,t} = x_{i,t-1} + a_i + w_{i,t}$$

, where $x_{i,t}$ is the number of gages active in HUC6 *i* at time step *t*, *a* represents their long-term trend (akin to a species intrinsic growth rate), and $w_{i,t}$ represents the process error, or the stochastic component that affects the gage population from one time step to the next one (assumed to be drawn from a normal distribution with a mean of 0 and variance *q*). We estimated *a* and *q* based on observed streamgaging data over the last 70 years (1947-2016), and we then used these coefficients to compute the probability that a gage population would hit a 50% decline threshold within the following 6 years ('decline risk') [after (*5*)]. Model residuals were examined via the autocorrelation function. Only the 232 river basins (HUC6) with a median number >10 gages over the period 1947-2012 were considered for this analysis. Because the relative influence of losing a gage in low-density basins is higher than in high-density basins, this approach provides a conservative estimate of gaging decline risks. We chose the forecast to be 6 years long because that time horizon coincides with the end of the current USGS Water Science strategic plan 2012-2022 (6).

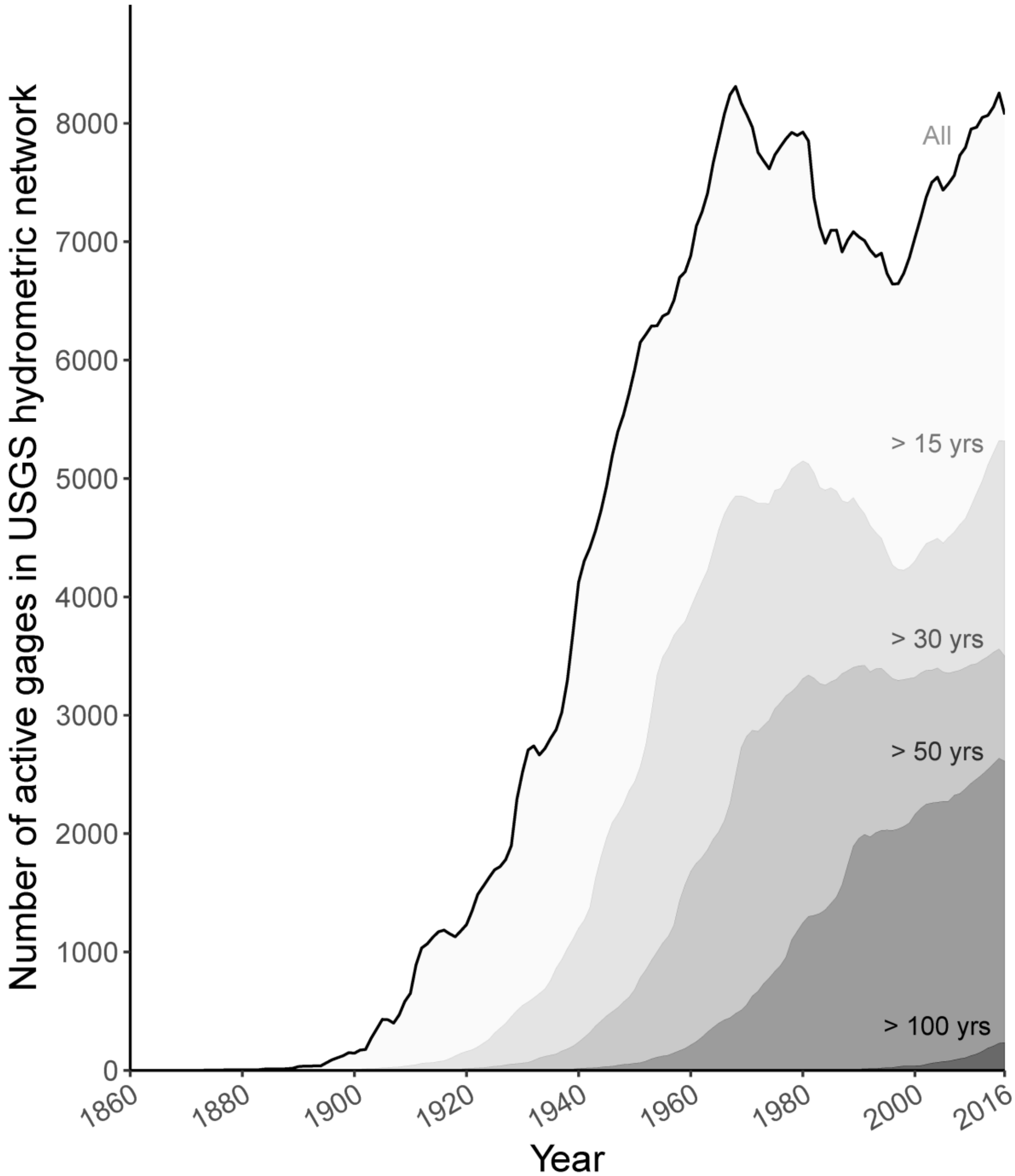

Fig. S1. Trends in the number of active streamgaging stations in the USGS hydrometric network, classified by age class (length of the daily discharge record, in years).

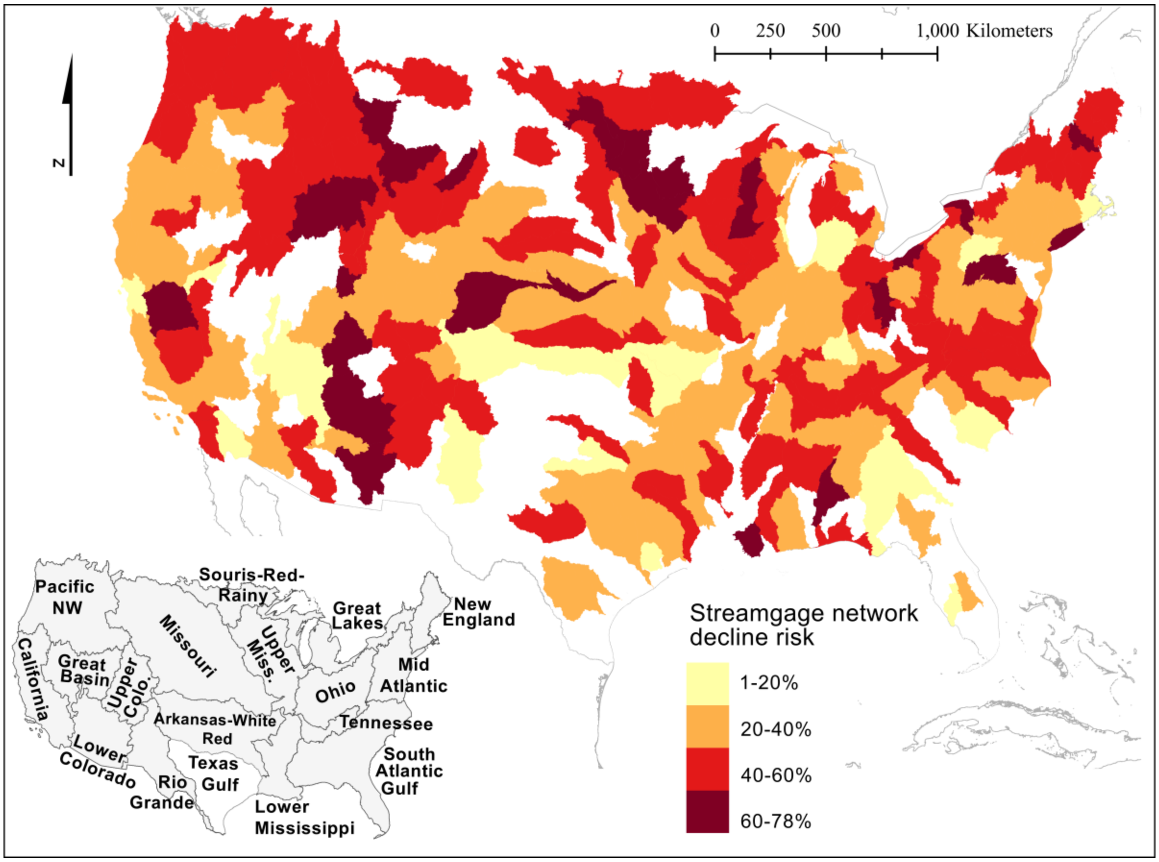

Figure S2. Spatial distribution of risk in streamgaging decline, inferred via time-series analysis of USGS streamgaging history (1947-2016, see Supplementary Materials and Methods for details). Within each river basin (Hydrological Unit Code 6), decline risk was measured as the probability of halving gage density over the next 6 years, that is, by the end the current USGS Water Science strategic plan 2012-2022 (*6*).

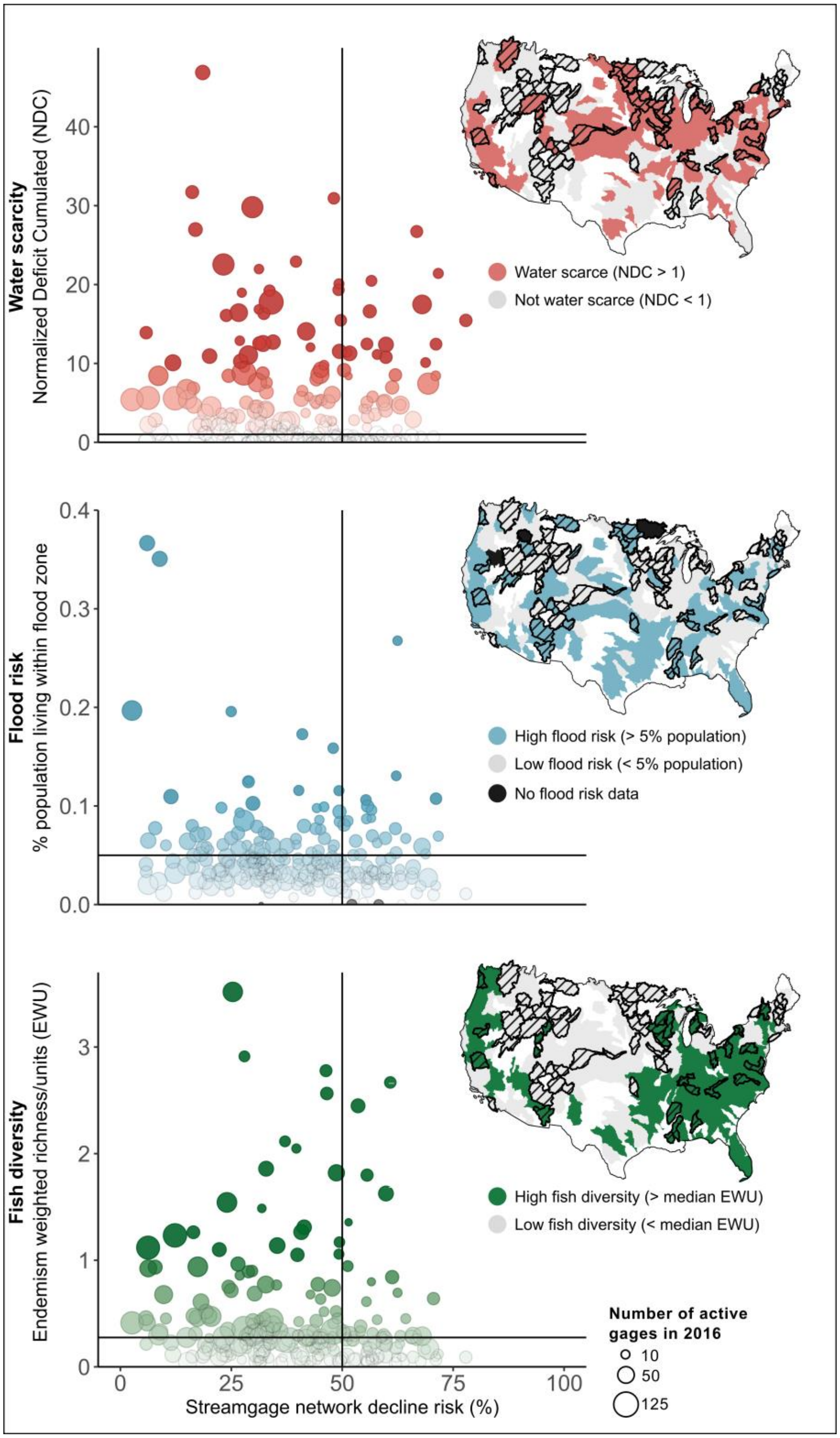

Figure S3. Streamgaging decline risk (see Fig. S2) *vs.* socioenvironmental monitoring needs: water scarcity (measured as Normalized Cumulative Deficit, NDC), flood risk to human populations (measured as % of population living within a 100-year flood zone), and fish biodiversity (measured as endemism weighted richness). See Supplementary Materials and Methods for details.

Table S1. Data sources.

Theme	Dataset	Source	Access
Gages	National Water Information System (NWIS) - Discharge data	US Geological Survey	https://waterdata.usgs.gov/nwis/sw
Fish diversity	Digital Distribution of Native U.S. Fishes	NatureServ	http://www.natureserve.org/conservation-tools/data-maps-tools/digital-distribution-native-us-fishes-watershed
	ESA species listing	US Fish and Wildlife Service	https://ecos.fws.gov/ecp0/reports/ad-hoc-species-report?kingdom=V&kingdom=I&status=E&status=T&status=EmE&status=EmT&status=EXPE&status=EXPN&status=SAE&status=SAT&fcrithab=on&fstatus=on&fspecrule=on&finvpop=on&fgroup=on&header=Listed+Animals
Water Scarcity	US long term cumulative water stress 1949-2010	(*3*)	Per author request
	US counties (2010 Census)	US Census Bureau	https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html
	Daily 1/8-degree gridded meteorological data (1 Jan 1949 - 31 Dec 2010)	(*4*)	http://www.engr.scu.edu/~emaurer/gridded_obs/index_gridded_obs.html
Flood exposure	National Flood Hazard Layer (NFHL, updated May 2016)	Federal Emergency Management Agency	https://catalog.data.gov/dataset/national-flood-hazard-layer-nfhl/resource/89b88927-fc8e-4557-a97f-3f3729aad36d
	2010 Census - Population & housing unit counts - Blocks	US Census Bureau	https://www.census.gov/geo/maps-data/data/tiger-data.html
	National Land Cover Data 2011 (NLCD 2011)	(*7*)	https://www.mrlc.gov/nlcd2011.php
All	National Hydrography Dataset plus version 2 (NHDplusV2) - WBD snapshot	US Environmental Protection Agency (EPA) and US Geological Survey (USGS)	http://www.horizon-systems.com/NHDPlus/V2NationalData.php

References

1. R. M. Hirsch, L. A. De Cicco, "User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data" (US Geological Survey, 2015).

2. G. Kier *et al.*, A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl. Acad. Sci.* **106**, 9322–9327 (2009).

3. N. Devineni, U. Lall, E. Etienne, D. Shi, C. Xi, America's water risk: Current demand and climate variability. *Geophys. Res. Lett.* **42**, 2285–2293 (2015).

4. E. P. Maurer, A. W. Wood, J. C. Adam, D. P. Lettenmaier, B. Nijssen, A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Clim.* **15**, 3237–3251 (2002).

5. E. R. Holmes, E. Ward, K. Wills, Package "MARSS" (2014), (available at http://cran.r-project.org/web/package/MARSS).

6. E. J. Evenson *et al.*, "Strategic directions for US Geological Survey water science, 2012-2022-Observing, understanding, predicting, and delivering water science to the Nation" (US Geological Survey, 2012).

7. C. Homer *et al.*, Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* **81**, 345–354 (2015).