Methods (appendices will simply be data and codes)

Streamgage historical data

Historical streamflow data for all stream gage stations of the United States were downloaded from the USGS National Water Information System (NWIS) using the dataRetrieval package in the R statistical software. Daily mean discharge records do not include provisional data that have not yet been approved by the USGS, such that stations with provisional records appear inactive during these periods. Therefore, daily mean water discharge data were downloaded for the period January 1854-December 2014 while instantaneous data (15-minute interval, generally, including provisional data) were downloaded for January 2015-January 2017 (see Appendix A). Periods for which river discharge records were only estimated rather than actually measured, due to gage malfunction or lack of maintenance, were excluded from the analysis. In total, discharge data were compiled for nearly 23,000 stations.

For each year, a gage was classified as active if it recorded discharge for at least 180 days during that year. We then computed the number of active stations in each of the 352 river basins -- Hydrological Unit Code 6 (HUC6) -- of the National Water Boundary Dataset (WBD) every year from 1854 to 2016 using geographic coordinates from the NWIS (see Appendix B).

Fish biodiversity metrics

After compiling a dataset of the distribution of freshwater fish throughout the United States, we computed the endemism weighted richness/units (EWU) for each river subbasin (HUC8) nationwide following Kier et al. (2009)'s approach.

EWU = Sum of (Endemism Unit of each Species), where Endemism Unit = 1 / # HUC8s, so the most endemic species has an endemism unit of 1.

EWU was then aggregated at the basin level (HUC6) by calculating the average EWU of subbasins within each basin.

Water scarcity

To represent water scarcity for each river basin, we derived a metric from the county-level Normalized Deficit Cumulated (NDC) developed by Devineni et al. (2015). For a given county. 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 highlights the reliance of each county on outside sources (e.g. withdrawal from runoff originating outside the county) and non-renewable local sources of water (e.g. groundwater with slow recharge) -- the maximum accumulated water shortage over 61 years that needs to be provided from additional water resources. 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), it 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 (See Appendix C). We then computed for each HUC6:

$$NHD_{HUC6} = \frac{1}{Average\ rainfall_{HUC6}} \times \sum_{C=1}^{N\ county\ in\ HUC6} NDC_C \times Average\ rainfall_C \times \frac{Area_{HUC6-County\ intersection}}{Total\ area_C}$$

Average rainfall for each HUC 6 was calculated using daily gridded meteorological data for the period 1949-2010 (Maurer et al. 2002b). 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 was cumulative across counties (i.e. that maximum water deficit could be summed across county boundaries). Given that NDC is based on the maximum daily water deficit in each county, it is possible that the maximum water deficit in a given county does not occur the same day as that in other counties in the same basin. In such a case, our approach would result in our computed NDC_{HUC6} to 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 that this metric still adequately reflects relative water scarcity patterns among US regions.

Flood risk

Nationwide, 13% of the population lives within an area for which no flood hazard study exists (*this study*). Therefore, in each river basin, flood risk was computed as the percentage of the total population inhabiting areas where a flood study exists that lives in a 100-yr flood zone. It was calculated by intersecting 2010 census block level population data with the National Flood Hazard Layer, NFHL (FEMA, 2016), a digital database containing all flood hazard mapping data from the Federal Emergency Management Agency's National Flood Insurance Program (NFIP). Census-block population data were used due to their high resolution (average and median area of a census block: 0.74 km^2 and 0.02 km^2, respectively) and tendency to spatially discriminate between settled areas and waterways or uninhabited floodplains. Instead, gridded population data (Seirup and Yetman 2006) are too coarse and therefore overlap with waterways and floodplains, leading to an overestimation of the number of people living in flood zones. Despite their 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-yr flood zone (Zone A from the NFHL). When a census block did not contain any urban land cover (LC), the population living, 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-yr flood zone is equal to the total population living in a 100-yr 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 simply proportional to the area of the census block overlapping the basin.

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

$$\text{Flood riskHUC6} = \sum_{C=census\ blocks\ within\ HUC6} \frac{Population\ living\ within\ flood\ zone_C}{Population\ living\ in\ area\ with\ flood\ hazard\ study_C}$$

And,

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

Time series analysis: quasi-extinction probability and turnover rate

Table X. 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	Devineni et al. (2015)	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)	Maurer et al. (2002a)	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)	Homer et al. (2015)	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

Devineni, N., U. Lall, E. Etienne, D. Shi, and C. Xi. 2015. America's water risk: Current demand and climate variability. Geophysical Research Letters **42**:2285-2293.

Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. D. Herold, J. Wickham, and K. Megown. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing **81**:345-354.

Kier, G., H. Kreft, T. M. Lee, W. Jetz, P. L. Ibisch, C. Nowicki, J. Mutke, and W. Barthlott. 2009. A global assessment of endemism and species richness across island and mainland regions. Proceedings of the National Academy of Sciences **106**:9322-9327.

Maurer, E., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen. 2002a. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. Journal of Climate **15**:3237-3251.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen. 2002b. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. Journal of Climate **15**:3237-3251.

Seirup, L., and G. Yetman. 2006. U.S. Census Grids (Summary File 3), 2000. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY.