Global prevalence of non-perennial rivers and streams

Technical Documentation for data repository version 1.0

June 2021

prepared by
Mathis L. Messager (mathis.messager@mail.mcgill.ca)
Bernhard Lehner (bernhard.lehner@mcgill.ca)

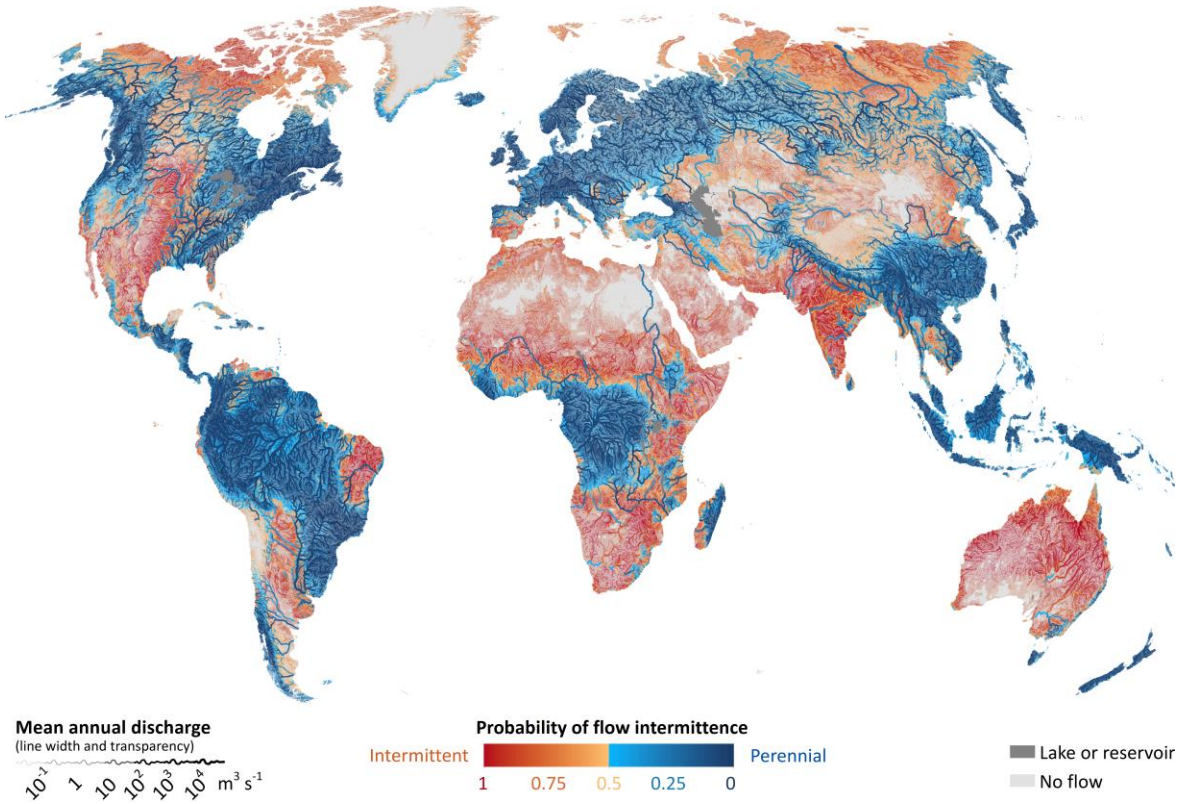

Figure 1. Global distribution of non-perennial rivers and streams. Intermittence is defined as flow cessation for at least one day per year on interannual average. From Messager et al. (2021).

1. Overview and background

This documentation describes the data produced for the research article: Messager, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C. & Datry, T. (2021). Global prevalence of non-perennial rivers and streams. *Nature*. https://doi.org/10.1038/s41586-021-03565-5

In this study, we developed a statistical Random Forest model to produce the first reach-scale estimate of the global distribution of non-perennial rivers and streams. For this purpose, we linked quality-checked observed streamflow data from 5,615 gauging stations (on 4,428 perennial and 1,187 non-perennial reaches) with 113 candidate environmental predictors available globally. Predictors included variables describing climate, physiography, land cover, soil, geology, and groundwater as well as estimates of long-term naturalised (i.e., without anthropogenic water use in the form of abstractions or impoundments) mean monthly and mean annual flow (MAF), derived from a global hydrological model (WaterGAP 2.2; Müller Schmied et al. 2014). Following model training and validation, we predicted the probability of flow intermittence for all river reaches in the RiverATLAS database (Linke et al. 2019), a digital representation of the global river network at high spatial resolution.

The data repository includes two datasets resulting from this study:

1. a geometric network of the global river system where each river segment is associated with:
   i. 113 hydro-environmental predictors used in model development and predictions, and
   ii. the probability and class of flow intermittence predicted by the model.
2. point locations of the 5,516 gauging stations used in model training/testing, where each station is associated with a line segment representing a reach in the river network, and a set of metadata.

These datasets have been generated with source code located at messamat.github.io/globalirmap/.
This documentation and the datasets described are available for download at
https://doi.org/10.6084/m9.figshare.14633022.

2. Repository content

The data repository has the following structure (for usage, see section 3. Data Format and Distribution; GIRES stands for Global Intermittent Rivers and Ephemeral Streams):

GIRES_v10_gdb.zip/ *— file geodatabase in ESRI® geodatabase format containing two feature classes (zipped)*

GIRES_v10_rivers — river network lines with attributes and model predictions, global coverage.
GIRES_v10_stations — points with streamflow summary statistics and metadata.

GIRES_v10_shp.zip/ *— directory containing ten shapefiles (zipped)*

This directory includes the same content as *GIRES_v10_gdb.zip*, but in the alternative, simplified shapefile format for users that cannot read ESRI geodatabases. Due to file size limitations for shapefiles, the global river network was tiled into nine regions. The regional extents are defined by a two-digit identifier (see **Fig. 2** for a map of regions):

GIRES_v10_rivers_af.shp — Africa **GIRES_v10_rivers_ar.shp** — North American Arctic **GIRES_v10_rivers_as.shp** — Asia **GIRES_v10_rivers_au.shp** — Australasia **GIRES_v10_rivers_eu.shp** — Europe **GIRES_v10_rivers_gr.shp** — Greenland **GIRES_v10_rivers_na.shp** — North America **GIRES_v10_rivers_sa.shp** — South America **GIRES_v10_rivers_si.shp** — Siberia	river network lines with attributes and model predictions, tiled.

GIRES_v10_stations.shp — points with streamflow summary statistics and metadata.

Other_technical_documentations.zip/ *— directory containing three documentation files (zipped)*

HydroATLAS_TechDoc_v10.pdf — original documentation for river network framework.
RiverATLAS_Catalog_v10.pdf — original documentation for river network hydro-environmental attributes.
Readme_GSIM_part1.txt — original documentation for gauging stations from the Global Streamflow Indices and Metadata (GSIM) archive.

README_Technical_documentation_GIRES_v10.pdf — this document.

Figure 2. Spatial extent of regional tiles of river network shapefiles.

3. Data format and distribution

3.1. Vector data format and projection

The geometric network (lines) and gauging stations (points) datasets are distributed both in ESRI® file geodatabase and shapefile formats. The file geodatabase contains all data and is the prime, recommended format. Shapefiles are provided as a copy for users that cannot read the geodatabase. Each shapefile consists of five main files (.dbf, .sbn, .sbx, .shp, .shx), and projection information is provided in an ASCII text file (.prj). The attribute table can be accessed as a stand-alone file in dBASE format (.dbf) which is included in the Shapefile format.
These datasets are available electronically in compressed zip file format. To use the data files, the zip files must first be decompressed.
All data layers are provided in geographic (latitude/longitude) projection, referenced to datum WGS84. In ESRI® software this projection is defined by the geographic coordinate system GCS_WGS_1984 and datum D_WGS_1984 (EPSG: 4326).

3.2. River network

File name: GIRES_v10_rivers | Dimensions: 8,099,258 rows and 120 columns.
Content: River network lines with attributes and model predictions, global coverage

The attribute table of the river network layer contains river network properties and identifiers, geometric attributes of the river reach elements (e.g., length, upstream area), hydro-environmental attributes, and model predictions from the flow intermittence model. **Table 1** provides details on all columns in the attribute table.

This geometric network and many associated attributes are originally from RiverATLAS version 1.0 (Linke et al. 2019). All river reaches from RiverATLAS with a naturalized long-term mean annual flow MAF > 0 $m^3 s^{-1}$ are included in this dataset. Refer to the associated study (Messager et al. 2021) and **Table 1** below for details on processing steps and the sources of attributes. For additional information on attributes provided by RiverATLAS, refer to https://www.hydrosheds.org/page/hydroatlas and the documentation provided in this repository (*Other_technical_documentations.zip*).

Although several attributes initially included in RiverATLAS version 1.0 have been updated for this study, the dataset provided here does not represent an established new version of RiverATLAS.

Note that Messager et al. (2021) included only rivers and streams with a MAF ≥ 0.1 $m^3 s^{-1}$ and excluded river channel sections intersecting a lake or reservoir. Therefore, to re-produce the figures found in the corresponding manuscript, a subset of this dataset must first be created.

Table 1. Attribute table of the river network.
Attribute names that differ between the geodatabase and shapefile formats (e.g., due to field name length limitations for shapefiles) are followed by (.gdb) and (.shp), respectively. "IRES study" refers to Messager et al. (2021), the study associated with this repository. The citation for RiverATLAS version 1.0 is Linke et al. (2019). The RiverATLAS v1.0 documentation is included in this repository: *Other_technical_documentations.zip/RiverATLAS_Catalog_v10.pdf*. The HydroATLAS v1.0 documentation is also included: *Other_technical_documentations.zip/HydroATLAS_Catalog_v10.pdf*. Single and Double data types refer to single- and double-precision floating-point numbers.

Name	Type	Description	Details	Source
OBJECTID (.gdb) **FID** (.shp)	OID	Default unique identifier.	Not used in analysis. Automatically created as part of the data formats to uniquely identify rows. Named OBJECTID in the geodatabase and FID in the shapefiles.	ArcGIS
Shape	Geometry	Default field describing the type of geometry of the record (here Polyline).	Not used in analysis. Automatically created as part of the data format.	ArcGIS
HYRIV_ID	Integer	Unique identifier for each river reach.	The code consists of 8 digits. The first digit represents the region: 1 = Africa; 2 = Europe; 3 = Siberia; 4 = Asia; 5 = Australia; 6 = South America; 7 = North America; 8 = Arctic; 9 = Greenland. The other 7 digits represent a unique identifier within the river network. This unique identifier can be used to link this dataset with gauging stations (GIRES_v10_stations) and with all other hydro-environmental attributes in the RiverATLAS dataset available at https://hydrosheds.org/page/hydroatlas	RiverATLAS v1.0
NEXT_DOWN	Integer	HYRIV_ID of the next downstream line segment.	This field can be used for navigation (up- and downstream) within the river network. The value '0' indicates a line with no downstream connection, i.e., the last river reach draining into the ocean or into an inland sink.	RiverATLAS v1.0
LENGTH_KM	Single	Length of the river reach segment, in kilometers.	–	RiverATLAS v1.0
CATCH_SKM	Single	Area of the catchment that contributes directly to the individual reach, in square kilometers.	The catchment only relates to the reach itself, while the contributing area of all upstream reaches is not included (see UPLAND_SKM).	RiverATLAS v1.0
UPLAND_SKM	Single	Total upstream area, in square kilometers, calculated from the headwaters to the pour point (i.e. the most downstream pixel) of the reach.	The upstream area only comprises the directly connected watershed area, i.e. it does not include endorheic regions that may be nested within the larger basin.	RiverATLAS v1.0
ORD_STRA	Integer	Indicator of river order following the Strahler ordering system.	Order 1 represents headwater streams; when two 1st order streams meet, they form a 2nd order river; when two 2nd order rivers meet, they form a 3rd order river; etc.	RiverATLAS v1.0
INLAKEPERC	Single	Proportion of the reach line that intersects with a lake.	Determined by overlapping RiverATLAS with HydroLAKES polygons from Messager et al. (2016), https://www.hydrosheds.org/page/hydrolakes .	IRES study
HYBAS_ID03	Double	HYBAS_ID of the corresponding HydroBASINS sub-basin level 3 in which the river reach resides.	This ID refers to HydroBASINS at Pfafstetter level 3 (without lakes), see HydroATLAS documentation.	HydroATLAS v1.0
PFAF_ID05	Integer	PFAF_ID of the corresponding HydroBASINS sub-basin level 5 in which the river reach resides.	This ID refers to HydroBASINS at Pfafstetter level 5 (without lakes), see HydroATLAS documentation.	HydroATLAS v1.0

MID_X	Double	Longitude of the middle of the river reach.	Referenced to datum WGS84.	IRES study
MID_Y	Double	Latitude of the middle of the river reach.	Referenced to datum WGS84.	IRES study
clz_cl_cmj	Integer	Spatial majority climate zone in reach catchment.	Source: Metzger et al. (2013).	RiverATLAS v1.0
aet_mm_cyr	Integer	Annual average actual evapotranspiration (mm) in reach catchment.	Source: Trabucco and Zomer (2010).	RiverATLAS v1.0
aet_mm_uyr	Integer	Annual average actual evapotranspiration (mm) in total watershed upstream of reach pour point.	Source: Trabucco and Zomer (2010).	RiverATLAS v1.0
pet_mm_cyr	Integer	Annual average potential evapotranspiration (mm) in reach catchment.	Source: Zomer et al. (2008).	RiverATLAS v1.0
pet_mm_uyr	Integer	Annual average potential evapotranspiration (mm) in total watershed upstream of reach pour point.	Source: Zomer et al. (2008).	RiverATLAS v1.0
snw_pc_cmx	Integer	Annual maximum snow cover extent (%) in reach catchment.	Source: Hall and Riggs (2016).	RiverATLAS v1.0
snw_pc_cyr	Integer	Annual average snow cover extent (%) in reach catchment.	Source: Hall and Riggs (2016).	RiverATLAS v1.0
snw_pc_uyr	Integer	Annual average snow cover extent (%) in total watershed upstream of reach pour point.	Source: Hall and Riggs (2016).	RiverATLAS v1.0
inu_pc_cmn	Integer	Annual minimum inundation extent (%) in reach catchment.	Source: Fluet-Chouinard et al. (2015).	RiverATLAS v1.0
inu_pc_umn	Integer	Annual minimum inundation extent (%) in total watershed upstream of reach pour point.	Source: Fluet-Chouinard et al. (2015).	RiverATLAS v1.0
inu_pc_umx	Integer	Annual maximum inundation extent (%) in total watershed upstream of reach pour point.	Source: Fluet-Chouinard et al. (2015).	RiverATLAS v1.0
lka_pc_cse	Integer	Limnicity (%) in reach catchment.	Source: Messager et al. (2016).	RiverATLAS v1.0
lka_pc_use	Integer	Limnicity (%) in total watershed upstream of reach pour point.	Source: Messager et al. (2016).	RiverATLAS v1.0
dis_m3_pmn	Single	Annual minimum natural discharge (cubic meters per second) at reach pour point.	Sources: Döll, Kaspar, and Lehner (2003); Müller Schmied et al. (2014).	RiverATLAS v1.0
dis_m3_pmx	Single	Annual maximum natural discharge (cubic meters per second) at reach pour point.	Sources: Döll, Kaspar, and Lehner (2003); Müller Schmied et al. (2014).	RiverATLAS v1.0
dis_m3_pyr	Single	Annual average natural discharge (cubic meters per second) at reach pour point.	Sources: Döll, Kaspar, and Lehner (2003); Müller Schmied et al. (2014).	RiverATLAS v1.0
run_mm_cyr	Integer	Annual average land surface runoff (mm) in reach catchment.	Sources: Döll, Kaspar, and Lehner (2003); Müller Schmied et al. (2014).	RiverATLAS v1.0
pnv_cl_cmj	Integer	Spatial majority potential natural vegetation class in reach catchment.	Source: Ramankutty and Foley (1999).	RiverATLAS v1.0
for_pc_cse	Integer	Forest cover extent (%) in reach catchment.	Source: Bartholomé and Belward (2005).	RiverATLAS v1.0
for_pc_use	Integer	Forest cover extent (%) in total watershed upstream of reach pour point.	Source: Bartholomé and Belward (2005).	RiverATLAS v1.0
glc_cl_cmj	Integer	Spatial majority land cover class in reach catchment.	Source: Bartholomé and Belward (2005).	RiverATLAS v1.0
glc_pc_c16	Integer	Spatial extent of cultivated and managed areas (%) in reach catchment.	Source: Bartholomé and Belward (2005).	RiverATLAS v1.0
glc_pc_u16	Integer	Spatial extent of cultivated and managed areas (%) in total watershed upstream of reach pour point.	Source: Bartholomé and Belward (2005).	RiverATLAS v1.0
gla_pc_cse	Integer	Glacier extent (%) in reach catchment.	Source: GLIMS and National Snow and Ice Data Center (2012).	RiverATLAS v1.0
gla_pc_use	Integer	Glacier extent (%) in total watershed upstream of reach pour point.	Source: GLIMS and National Snow and Ice Data Center (2012).	RiverATLAS v1.0
wet_pc_c07	Integer	Spatial extent of pan and brackish/saline wetlands (%) in reach catchment.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_c09	Integer	Spatial extent of intermittent wetlands/lakes (%) in reach catchment.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_cg1	Integer	Spatial extent of wetlands, lakes, reservoirs and rivers (%) in reach catchment.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_cg2	Integer	Spatial extent of wetlands excluding lakes, reservoirs and rivers (%) in reach catchment.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_u07	Integer	Spatial extent of pan and brackish/saline wetlands (%) in total watershed upstream of reach pour point.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_u09	Integer	Spatial extent of intermittent wetlands/lakes (%) in total watershed upstream of reach pour point.	Source: Lehner and Döll (2004).	RiverATLAS v1.0

wet_pc_ug1	Integer	Spatial extent of wetlands, lakes, reservoirs and rivers (%) in total watershed upstream of reach pour point.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
wet_pc_ug2	Integer	Spatial extent of wetlands excluding lakes, reservoirs and rivers (%) in total watershed upstream of reach pour point.	Source: Lehner and Döll (2004).	RiverATLAS v1.0
prm_pc_cse	Integer	Permafrost extent (%) in reach catchment.	Source: Gruber (2012).	RiverATLAS v1.0
prm_pc_use	Integer	Permafrost extent (%) in total watershed upstream of reach pour point.	Source: Gruber (2012).	RiverATLAS v1.0
slp_dg_cav	Integer	Average terrain slope (degrees x 10) in reach catchment.	Source: Robinson, Regetz, and Guralnick (2014).	RiverATLAS v1.0
slp_dg_uav	Integer	Average terrain slope (degrees x 10) in total watershed upstream of reach pour point.	Source: Robinson, Regetz, and Guralnick (2014).	RiverATLAS v1.0
lit_cl_cmj	Integer	Spatial majority lithological class in reach catchment.	Source: Hartmann and Moosdorf (2012).	RiverATLAS v1.0
swc_pc_cyr	Integer	Annual average soil water content (%) in reach catchment.	Source: Trabucco and Zomer (2010).	RiverATLAS v1.0
swc_pc_uyr	Integer	Annual average soil water content (%) in total watershed upstream of reach pour point.	Source: Trabucco and Zomer (2010).	RiverATLAS v1.0
kar_pc_cse	Integer	Karst area extent (%) in reach catchment.	Source: Williams and Ford (2006).	RiverATLAS v1.0
kar_pc_use	Integer	Karst area extent (%) in total watershed upstream of reach pour point.	Source: Williams and Ford (2006).	RiverATLAS v1.0
ari_ix_cav	Integer	Annual average global aridity index (x100) in reach catchment.	Source: Trabucco and Zomer (2018).	IRES study
ari_ix_uav	Integer	Annual average global aridity index (x100) in total watershed upstream of reach pour point.	Source: Trabucco and Zomer (2018).	IRES study
bio1_dc_cav (.gdb) **bio1_dc_c** (.shp)	Integer	Annual mean tempeature (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio1_dc_uav (.gdb) **bio1_dc_u** (.shp)	Integer	Annual mean temperature (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio2_dc_cav (.gdb) **bio2_dc_c** (.shp)	Integer	Mean diurnal range (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio2_dc_uav (.gdb) **bio2_dc_u** (.shp)	Integer	Mean diurnal range (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio3_dc_cav (.gdb) **bio3_dc_c** (.shp)	Integer	Isothermality BIO2/BIO7 x 100 (BIO3), averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio3_dc_uav (.gdb) **bio3_dc_u** (.shp)	Integer	Isothermality BIO2/BIO7 x 100 (BIO3) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio4_dc_cav (.gdb) **bio4_dc_c** (.shp)	Integer	Temperature seasonality (SD x 100 BIO4; degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio4_dc_uav (.gdb) **bio4_dc_u** (.shp)	Integer	Temperature seasonality (SD x 100 BIO4; degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio5_dc_cav (.gdb) **bio5_dc_c** (.shp)	Integer	Maximum temperature of warmest month (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio5_dc_uav (.gdb) **bio5_dc_u** (.shp)	Integer	Maximum temperature of warmest month (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio6_dc_cav (.gdb) **bio6_dc_c** (.shp)	Integer	Maximum temperature of coldest month (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio6_dc_uav (.gdb) **bio6_dc_u** (.shp)	Integer	Maximum temperature of coldest month (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study

bio7_dc_cav (.gdb) **bio7_dc_c** (.shp)	Integer	Temperature annual range (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio7_dc_uav (.gdb) **bio7_dc_u** (.shp)	Integer	Temperature annual range (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio8_dc_cav (.gdb) **bio8_dc_c** (.shp)	Integer	Mean temperature of wettest quarter (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio8_dc_uav (.gdb) **bio8_dc_u** (.shp)	Integer	Mean temperature of wettest quarter (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio9_dc_cav (.gdb) **bio9_dc_c** (.shp)	Integer	Mean temperature of driest quarter (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio9_dc_uav (.gdb) **bio9_dc_u** (.shp)	Integer	Mean temperature of driest quarter (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio10_dc_cav (.gdb) **bio10_dc_c** (.shp)	Integer	Mean temperature of warmest quarter (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio10_dc_uav (.gdb) **bio10_dc_u** (.shp)	Integer	Mean temperature of warmest quarter (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio11_dc_cav (.gdb) **bio11_dc_c** (.shp)	Integer	Mean temperature of coldest quarter (degrees celsius) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio11_dc_uav (.gdb) **bio11_dc_u** (.shp)	Integer	Mean temperature of coldest quarter (degrees celsius) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio12_mm_cav (.gdb) **bio12_mm_c** (.shp)	Integer	Mean annual precipitation (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio12_mm_uav (.gdb) **bio12_mm_u** (.shp)	Integer	Mean annual precipitation (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio13_mm_cav (.gdb) **bio13_mm_c** (.shp)	Integer	Mean precipitation of wettest month (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio13_mm_uav (.gdb) **bio13_mm_u** (.shp)	Integer	Mean precipitation of wettest month (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio14_mm_cav (.gdb) **bio14_mm_c** (.shp)	Integer	Mean precipitation of driest month (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio14_mm_uav (.gdb) **bio14_mm_u** (.shp)	Integer	Mean precipitation of driest month (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio15_mm_cav (.gdb) **bio15_mm_c** (.shp)	Integer	Precipitation seasonality averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio15_mm_uav (.gdb) **bio15_mm_u** (.shp)	Integer	Precipitation seasonality averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio16_mm_cav (.gdb) **bio16_mm_c** (.shp)	Integer	Mean precipitation of wettest quarter (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio16_mm_uav (.gdb) **bio16_mm_u** (.shp)	Integer	Mean precipitation of wettest quarter (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio17_mm_cav (.gdb) **bio17_mm_c** (.shp)	Integer	Mean precipitation of driest quarter (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio17_mm_uav (.gdb) **bio17_mm_u** (.shp)	Integer	Mean precipitation of driest quarter (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study

bio18_mm_cav (.gdb) **bio18_mm_c** (.shp)	Integer	Mean precipitation of warmest quarter (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio18_mm_uav (.gdb) **bio18_mm_u** (.shp)	Integer	Mean precipitation of warmest quarter (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
bio19_mm_cav (.gdb) **bio19_mm_c** (.shp)	Integer	Mean precipitation of coldest quarter (mm) averaged across the reach catchment.	Source: Fick and Hijmans (2017).	IRES study
bio19_mm_uav (.gdb) **bio19_mm_u** (.shp)	Integer	Mean precipitation of coldest quarter (mm) averaged across the total watershed upstream of reach pour point.	Source: Fick and Hijmans (2017).	IRES study
wdryp_pc_cav (.gdb) **wdryp_pc_c** (.shp)	Integer	Percentage of open water extent with dry period from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wdryp_pc_uav (.gdb) **wdryp_pc_u** (.shp)	Integer	Percentage of open water extent with dry period from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wfrsh_pc_cav (.gdb) **wfrsh_pc_c** (.shp)	Integer	Percentage of open water extent at anytime from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wfrsh_pc_uav (.gdb) **wfrsh_pc_u** (.shp)	Integer	Percentage of open water extent at anytime from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
whfrq_pc_cav (.gdb) **whfrq_pc_c** (.shp)	Integer	Percentage of open water extent with high frequency variations from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
whfrq_pc_uav (.gdb) **whfrq_pc_u** (.shp)	Integer	Percentage of open water extent with high frequency variations from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wloss_pc_cav (.gdb) **wloss_pc_c** (.shp)	Integer	Percentage of open water extent lost from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wloss_pc_uav (.gdb) **wloss_pc_u** (.shp)	Integer	Percentage of open water extent lost from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wperm_pc_cav (.gdb) **wperm_pc_c** (.shp)	Integer	Percentage of open water extent that was permanent from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wperm_pc_uav (.gdb) **wperm_pc_u** (.shp)	Integer	Percentage of open water extent that was permanent from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wseas_pc_cav (.gdb) **wseas_pc_c** (.shp)	Integer	Percentage of open water extent that was seasonal from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wseas_pc_uav (.gdb) **wseas_pc_u** (.shp)	Integer	Percentage of open water extent that was seasonal from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
wwetp_pc_cav (.gdb) **wwetp_pc_c** (.shp)	Integer	Percentage of open water extent with wet period from 1999 to 2019 in the reach catchment.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study

wwetp_pc_uav (.gdb) **wwetp_pc_u** (.shp)	Integer	Percentage of open water extent with wet period from 1999 to 2019 in the total watershed upstream of reach pour point.	Source: Pickens et al. (2020). The value '-9999' indicates that there was no open water cover anytime from 1999 to 2019 in the area of interest.	IRES study
cly_pc_cav	Integer	Average clay fraction in soil (0-100 cm, %) in the reach catchment.	Source: Hengl et al. (2017).	IRES study
cly_pc_uav	Integer	Average clay fraction in soil (0-100 cm, %) in the total watershed upstream of reach pour point.	Source: Hengl et al. (2017).	IRES study
slt_pc_cav	Integer	Average silt fraction in soil (0-100 cm, %) in the reach catchment.	Source: Hengl et al. (2017).	IRES study
slt_pc_uav	Integer	Average silt fraction in soil (0-100 cm, %) in the total watershed upstream of reach pour point.	Source: Hengl et al. (2017).	IRES study
snd_pc_cav	Integer	Average sand fraction in soil (0-100 cm, %) in the reach catchment.	Source: Hengl et al. (2017).	IRES study
snd_pc_uav	Integer	Average sand fraction in soil (0-100 cm, %) in the total watershed upstream of reach pour point.	Source: Hengl et al. (2017).	IRES study
predprob1	Single	Predicted probability that river reach ceases to flow for at least one day per year on average somewhere along its length.	Ranges from 0 to 1, with 1 denoting a 100% probability of flow cessation. Predicted by random forest modeling.	IRES study
predcat1	Integer	Predicted flow intermittence class (based on predprob1).	1: non-perennial, i.e., the predicted probability that the river reach ceases to flow for at least one day per year on average somewhere along its length, (predprob1) >= 50%. 0: perennial, (predprob1) < 50%.	IRES study
predprob30	Single	Predicted probability that river reach ceases to flow for at least one month (thirty days) per year on average somewhere along its length.	Ranges from 0 to 1, with 1 denoting a 100% probability of flow cessation. Predicted by random forest modeling. Note that predprob30 may be equal or higher than preprob1 in regions and climates with limited training data because these two estimates (predprob1 and predprob30) stem from independent models, each with stochastic variability.	IRES study
predcat30	Integer	Predicted flow intermittence class (based on predprob30).	1: non-perennial, i.e., the predicted probability that the river reach ceases to flow for at least one day per year on average somewhere along its length, (predprob30) >= 50%. 0: perennial, (predprob30) < 50%.	IRES study
Shape_Length (.gdb)	Double	Length of the river reach segment, in decimal degrees.	Not used in analysis. Automatically created as part of the geodatabase data format. For linear length in km, use LENGTH_KM instead.	ArcGIS

3.3. Gauging stations

File name: GIRES_v10_stations | Dimensions: 5,615 rows and 60 columns
Content: points with streamflow summary statistics and metadata

The attribute table of the streamflow gauging stations contains unique identifiers (including the identifier of the line segment representing the reach in the river network where each station is located), pre-processing metadata, streamflow summary statistics, and model predictions. **Table 2** provides details on all columns in the attribute table.

This dataset only includes gauging stations used in model training and testing. It results from the compilation and processing of subsets of (i) the Global Runoff Data Centre (GRDC 2015) database and (ii) the Global Streamflow Indices and Metadata (GSIM) archive developed by Do et al. (2018) and Gudmundsson et al. (2018). For more information on data sources, characteristics, and processing information, see the Supplementary Information of Messager et al. (2021), available at https://doi.org/10.1038/s41586-021-03565-5.

Table 2. Attribute table of the gauging stations.
"IRES study" refers to Messager et al. (2021). The citation for RiverATLAS version 1.0 is Linke et al. (2019). The RiverATLAS v1.0 documentation is included in this repository: *Other_technical_documentations.zip/RiverATLAS_Catalog_v10.pdf*. The GSIM documentation is also included in this repository *Other_technical_documentations.zip/Readme_GSIM_part1.txt*.

Name	Type	Description	Details	Source
OBJECTID(.gdb) **FID** (.shp)	OID	Default ID.	Not used in analysis. Automatically created as part of the data formats to uniquely identify rows. Named OBJECTID in the file geodatabase and FID in the shapefiles.	ArcGIS
Shape	Geom.	Default field describing the type of geometry of the record (here Point).	Not used in analysis. Automatically created as part of the data format.	ArcGIS
GAUGE_NO	String	Unique gauge ID for all gauges.	If source is GRDC: GRDC_NO. If source is GSIM: gsim_no.	IRES study
GRDC_NO	String	Unique gauge ID for GRDC gauges.	No value for GSIM gauges.	GRDC
gsim_no	String	Unique gauge ID for GSIM gauges.	No value for GRDC gauges.	GSIM
HYRIV_ID	Integer	Unique ID of the river reach from RiverATLAS associated with the gauge.	This unique identifier can be used to link this dataset to the geometric river network (GIRES_v10_rivers) and with all other hydro-environmental attributes (see RiverATLAS documentation).	RiverATLAS v1.0
area_cor	Double	Catchment size reported in reference database by gauging data provider, in square kilometers.	The suffix _cor stands for "correct" and is used because GSIM included US catchment sizes in square miles ('area' attribute in GSIM dataset). This has been corrected here.	IRES study
X	Double	Longitude.	Referenced to datum WGS84. Final coordinates after QA/QC process. See manualsnap and snap_comme fields.	IRES study
Y	Double	Latitude.	Referenced to datum WGS84. Final coordinates after QA/QC process. See manualsnap and snap_comme fields.	IRES study
refcat	String	Reference binary flow intermittence class.	1: non-perennial (if mDur >= 1, i.e., if gauging station recorded zero-flow for at least one day per year on average) 0: perennial	IRES study
predprob1	Double	Predicted probability that river reach ceases to flow for at least one day per year on average somewhere along its length.	Ranges from 0 to 1, with 1 denoting a 100% probability of flow cessation. Predicted by random forest modeling.	IRES study
predcat1	String	Predicted binary flow intermittence class (based on predprob1).	1: non-perennial, i.e., the predicted probability that the river reach ceases to flow for at least one day per year on average somewhere along its length (predprob1) >= 50%. 0: perennial, (predprob1) < 50%.	IRES study
predres1	Double	Predicted flow intermittence residuals.	Calculated as (predprob1 – refcat).	IRES study

predprob30	Single	Predicted probability that river reach ceases to flow for at least one month (thirty days) per year on average somewhere along its length.	Ranges from 0 to 1, with 1 denoting a 100% probability of flow cessation. Predicted by random forest modeling.	IRES study
predcat30	Integer	Predicted binary flow intermittence class (based on predprob30).	1: non-perennial, i.e., the predicted probability that the river reach ceases to flow for at least one day per year on average somewhere along its length (predprob1) >= 50%. 0: perennial, (predprob1) < 50%.	IRES study
predres30	Single	Predicted flow intermittence residuals .	Calculated as (predprob30 – refcat).	IRES study
manualsnap	Integer	Manual snapping code.	Null: the station was not inspected, 0: the station was inspected and found to be in the correct location, 1: the station was inspected and its position was modified manually.	IRES study
DAdiff	Double	Difference between the drainage area of the gauging station reported by the reference database and that of the RiverATLAS river reach associated with the station.	(area_cor - UPLAND_SKM from associated river reach)/UPLAND_SKM from associated river reach See Table 1 for details on UPLAND_SKM.	IRES study
firstyr	Double	Year of first daily discharge value.	–	IRES study
lastyr	Double	Year of last daily discharge value.	–	IRES study
totalyr	Integer	Total number of years with daily discharge values.	–	IRES study
firstyrin	Double	First year of data used in analysis.	Only including years with <= 20 missing daily discharge values.	IRES study
lastyrin	Double	Last year of data used in analysis.	Only including years with <= 20 missing daily discharge values.	IRES study
totalyrin	Integer	Total number of years used in analysis.	Only including years with <= 20 missing daily discharge values.	IRES study
totaldays	Double	Total number of daily discharge values used in analysis.	Only including years with <= 20 missing daily discharge values.	IRES study
integerper	Double	Proportion of daily discharge values that are integer.	Only including years with <= 20 missing daily discharge values.	IRES study
sumDur	Double	Total number of daily zero-flow values.	Only including years with <= 20 missing daily discharge values.	IRES study
mDur	Double	Mean annual number of daily zero-flow values.	Only including years with <= 20 missing daily discharge values.	IRES study
mFreq	Double	Mean annual frequency of zero-flow events.	A zero-flow event is defined as one or more consecutive days with a recorded daily discharge of zero. Only including years with <= 20 missing daily discharge values.	IRES study
Jan_mdur	Double	Mean monthly number of daily zero-flow values in January.	Only including years with <= 20 missing daily discharge values.	IRES study
Feb_mdur	Double	Mean monthly number of daily zero-flow values in February.	Only including years with <= 20 missing daily discharge values.	IRES study
Mar_mdur	Double	Mean monthly number of daily zero-flow values in March.	Only including years with <= 20 missing daily discharge values.	IRES study
Apr_mdur	Double	Mean monthly number of daily zero-flow values in April.	Only including years with <= 20 missing daily discharge values.	IRES study
May_mdur	Double	Mean monthly number of daily zero-flow values in May.	Only including years with <= 20 missing daily discharge values.	IRES study
Jun_mdur	Double	Mean monthly number of daily zero-flow values in June.	Only including years with <= 20 missing daily discharge values.	IRES study
Jul_mdur	Double	Mean monthly number of daily zero-flow values in July.	Only including years with <= 20 missing daily discharge values.	IRES study
Aug_mdur	Double	Mean monthly number of daily zero-flow values in August.	Only including years with <= 20 missing daily discharge values.	IRES study
Sep_mdur	Double	Mean monthly number of daily zero-flow values in September.	Only including years with <= 20 missing daily discharge values.	IRES study
Oct_mdur	Double	Mean monthly number of daily zero-flow values in October.	Only including years with <= 20 missing daily discharge values.	IRES study
Nov_mdur	Double	Mean monthly number of daily zero-flow values in November.	Only including years with <= 20 missing daily discharge values.	IRES study
Dec_mdur	Double	Mean monthly number of daily zero-flow values in December.	Only including years with <= 20 missing daily discharge values.	IRES study

Jan_mfreq	Double	Mean monthly frequency of zero-flow events in January.	Only including years with <= 20 missing daily discharge values.	IRES study
Feb_mfreq	Double	Mean monthly frequency of zero-flow events in February.	Only including years with <= 20 missing daily discharge values.	IRES study
Mar_mfreq	Double	Mean monthly frequency of zero-flow events in March.	Only including years with <= 20 missing daily discharge values.	IRES study
Apr_mfreq	Double	Mean monthly frequency of zero-flow events in April.	Only including years with <= 20 missing daily discharge values.	IRES study
May_mfreq	Double	Mean monthly frequency of zero-flow events in May.	Only including years with <= 20 missing daily discharge values.	IRES study
Jun_mfreq	Double	Mean monthly frequency of zero-flow events in June.	Only including years with <= 20 missing daily discharge values.	IRES study
Jul_mfreq	Double	Mean monthly frequency of zero-flow events in July.	Only including years with <= 20 missing daily discharge values.	IRES study
Aug_mfreq	Double	Mean monthly frequency of zero-flow events in August.	Only including years with <= 20 missing daily discharge values.	IRES study
Sep_mfreq	Double	Mean monthly frequency of zero-flow events in September.	Only including years with <= 20 missing daily discharge values.	IRES study
Oct_mfreq	Double	Mean monthly frequency of zero-flow events in October.	Only including years with <= 20 missing daily discharge values.	IRES study
Nov_mfreq	Double	Mean monthly frequency of zero-flow events in November.	Only including years with <= 20 missing daily discharge values.	IRES study
Dec_mfreq	Double	Mean monthly frequency of zero-flow events in December.	Only including years with <= 20 missing daily discharge values.	IRES study
winteronly	Double	Indicates whether gauge is only non-perennial during winter months.	1: if refcat == 1 AND the average annual number of zero-flow days during warm months < 1. Warm months are those with mean monthly catchment air temperature >= 10 (WorldClim v2; Fick and Hijmans 2017). 0: either perennial, or non-perennial outside of winter months. Only including years with <= 20 missing daily discharge values.	IRES study

4. License and citations

4.1. License agreement

4.2. Citations and acknowledgements.

Citations and acknowledgements of this dataset should be made as follows:

> Messager, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C. & Datry, T. (2021). Global prevalence of non-perennial rivers and streams. *Nature*. https://doi.org/10.1038/s41586-021-03565-5

We kindly ask users to cite this study in any published material produced using it. If possible, online links to this repository (https://doi.org/10.6084/m9.figshare.14633022) should also be provided.

5. References

Bartholomé, Etienne, and A. S. Belward. 2005. "GLC2000: a new approach to global land cover mapping from earth observation data." *International Journal of Remote Sensing* 26(9): 1959–77. https://www.tandfonline.com/doi/full/10.1080/01431160412331291297.

Do, Hong Xuan, Lukas Gudmundsson, Michael Leonard, and Seth Westra. 2018. "The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: the production of a daily streamflow archive and metadata." *Earth System Science Data* 10(2): 765–85. https://www.earth-syst-sci-data.net/10/765/2018/ (January 23, 2020).

Döll, Petra, Frank Kaspar, and Bernhard Lehner. 2003. "A global hydrological model for deriving water availability

indicators: model tuning and validation." *Journal of Hydrology* 270(1–2): 105–34. https://linkinghub.elsevier.com/retrieve/pii/S0022169402002834.
Fick, Stephen E, and Robert J Hijmans. 2017. "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." *International Journal of Climatology* 37(12): 4302–15. http://dx.doi.org/10.1002/joc.5086.
Fluet-Chouinard, Etienne et al. 2015. "Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data." *Remote Sensing of Environment* 158: 348–61. https://linkinghub.elsevier.com/retrieve/pii/S0034425714004258.
GLIMS and National Snow and Ice Data Center. 2012. "GLIMS Glacier Database, V1."
Global Runoff Data Centre. 2015. "In-situ river discharge data." https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html.
Gruber, S. 2012. "Derivation and analysis of a high-resolution estimate of global permafrost zonation." *The Cryosphere* 6(1): 221–33. https://tc.copernicus.org/articles/6/221/2012/.
Gudmundsson, Lukas, Hong Xuan Do, Michael Leonard, and Seth Westra. 2018. "The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: quality control, time-series indices and homogeneity assessment." *Earth System Science Data* 10(2): 787–804. https://www.earth-syst-sci-data.net/10/787/2018/ (January 23, 2020).
Hall, D. K., and G. A. Riggs. 2016. "MODIS/Aqua Snow Cover Daily L3 Global 500m SIN Grid, Version 6. [2002-2015]."
Hartmann, Jens, and Nils Moosdorf. 2012. "The new Global Lithological Map Database GLiM: a representation of rock properties at the earth surface." *Geochemistry, Geophysics, Geosystems* 13(12). http://doi.wiley.com/10.1029/2012GC004370.
Hengl, Tomislav et al. 2017. "SoilGrids250m: Global gridded soil information based on machine learning." *PLOS ONE* 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748.
Lehner, Bernhard, and Petra Döll. 2004. "Development and validation of a global database of lakes, reservoirs and wetlands." *Journal of Hydrology* 296(1–4): 1–22. https://linkinghub.elsevier.com/retrieve/pii/S0022169404001404.
Linke, Simon et al. 2019. "Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution." *Scientific data* 6(1): 283.
Messager, M.L. et al. 2016. "Estimating the volume and age of water stored in global lakes using a geo-statistical approach." *Nature Communications* 7.
Metzger, Marc J. et al. 2013. "A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring" ed. Martin Sykes. *Global Ecology and Biogeography* 22(5): 630–38. http://doi.wiley.com/10.1111/geb.12022 (February 22, 2021).
Müller Schmied, H. et al. 2014. "Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration." *Hydrology and Earth System Sciences* 18(9): 3511–38. https://www.hydrol-earth-syst-sci.net/18/3511/2014/ (October 9, 2019).
Pickens, Amy H. et al. 2020. "Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series." *Remote Sensing of Environment* 243: 111792. https://linkinghub.elsevier.com/retrieve/pii/S0034425720301620 (April 17, 2020).
Ramankutty, Navin, and Jonathan A. Foley. 1999. "Estimating historical changes in global land cover: croplands from 1700 to 1992." *Global Biogeochemical Cycles* 13(4): 997–1027. http://doi.wiley.com/10.1029/1999GB900046.
Robinson, Natalie, James Regetz, and Robert P. Guralnick. 2014. "EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 57–67. https://linkinghub.elsevier.com/retrieve/pii/S0924271613002360.
Trabucco, Antonio, and Robert Zomer. 2018. "Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database V2." https://figshare.com/articles/dataset/Global_Aridity_Index_and_Potential_Evapotranspiration_ET0_Climate_Database_v2/7504448.
Trabucco, Antonio, and Robert J. Zomer. 2010. "Global Soil Water Balance geospatial database." https://doi.org/10.6084/m9.figshare.7707605.v3.
Williams, P W, and D C Ford. 2006. "Global distribution of carbonate rocks." *Zeitschrift Fur Geomorphologie Supplementband* 147: 1.
Zomer, Robert J, Antonio Trabucco, Deborah A Bossio, and Louis V Verchot. 2008. "Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation." *Agriculture, Ecosystems & Environment* 126(1): 67–80. http://www.sciencedirect.com/science/article/pii/S0167880908000169.