



Supplement of

Predicting streamflow drought in the conterminous United States using machine learning and a donor-gage approach, 1982–2020

Aaron Heldmyer et al.

Correspondence to: Aaron Heldmyer (aheldmyer@usgs.gov)

The copyright of individual parts of the supplement might differ from the article licence.

Supplements

Section S1.

For each predictor variable, a single netCDF file for one year between 1980 and 2020 was downloaded (Table 1) and converted from raster to vector (polygon) format. This only needed to be done once for each variable because all source netCDF files for each variable were concurrent and orthogonal (i.e., cover the same extent with the same raster resolution). Next, a weighting file was created for each predictor variable by using polygons of the drainage basins associated with the streamgages of interest (described in section 2.1) and overlaying the polygon layer representing the predictor variable values. This union was performed to attribute the area of each raster cell polygon that lies within an associated streamgage basin. The weighting file was then used to calculate daily averages of the predictor variables for all streamgage basins. The value of each predictor variable polygon was multiplied by the proportion of the polygon that overlapped with the streamgage basin to produce an area-weighted value. These values were used to compute an area-weighted mean for all predictor variables in all streamgage basins.

Table S1. GAGES-II climatic and physiographic variables used for the donor gage regressions.

GAGES-II Variable	Variable Type	Description
FST32F_BASIN	Climate	Watershed average of mean day of the year of first freeze
LST32F_BASIN	Climate	Watershed average of mean day of the year of last freeze
PET	Climate	Mean-annual potential evapotranspiration (PET)
PPTAVG_BASIN	Climate	Mean annual precipitation (cm) for the watershed
PRECIP_SEAS_IND	Climate	Index of how much annual precipitation falls seasonally (high values) or over the year (low values). Based on monthly precipitation values for 30-year POR (1971-2000), from PRISM data. Ranges from 0 (precip spread out exactly evenly in each month) to 1 (all precip falls in a single month).
RH_BASIN	Climate	Watershed average relative humidity (percent)
SNOW_PCT_PRECIP	Climate	Snow percent of total precipitation estimate
T_AVG_BASIN	Climate	Average annual air temperature for the watershed, in degrees Celsius (C)
WDMAX_BASIN	Climate	Watershed average of monthly maximum number of days (days) of measurable precipitation
WDMIN_BASIN	Climate	Watershed average of monthly minimum number of days (days) of measurable precipitation
BAS_COMPACTNESS	Hydrologic	Watershed compactness ratio=area/perimeter^2 * 100; higher number=more compact shape.
BFI_AVE	Hydrologic	Base Flow Index (BFI). The BFI is the ratio of base flow to total streamflow, expressed as a percentage and ranging from 0 to 100
CONTACT	Hydrologic	Subsurface flow contact time index

MAINSTEM_SINUOSITY	Hydrologic	Sinuosity of mainstem stream line
PCT_1ST_ORDER	Hydrologic	Percent of stream lengths in the watershed which are first-order streams (Strahler order)
PCT_2ND_ORDER	Hydrologic	Percent of stream lengths in the watershed which are second-order streams (Strahler order)
PCT_3RD_ORDER	Hydrologic	Percent of stream lengths in the watershed which are third-order streams (Strahler order)
PCT_4TH_ORDER	Hydrologic	Percent of stream lengths in the watershed which are fourth-order streams (Strahler order)
PCT_5TH_ORDER	Hydrologic	Percent of stream lengths in the watershed which are fifth-order streams (Strahler order)
PCT_6TH_ORDER_OR_MORE	Hydrologic	Percent of stream lengths in the watershed which are sixth or greater-order streams (Strahler order)
PERDUN	Hydrologic	Dunne overland flow, also known as saturation overland flow
PERHOR	Hydrologic	Horton overland flow, also known as infiltration-excess overland flow
STRAHLER_MAX	Hydrologic	Maximum Strahler stream order in watershed
STREAMS_KM_SQ_KM	Hydrologic	Stream density, km of streams per watershed sq km
TOPWET	Hydrologic	Topographic wetness index, $\ln(a/S)$; where "ln" is the natural logarithm, "a" is the upslope area per unit contour length and "S" is the slope at that point
DEVNLCD06	Land Cover	Watershed percent "developed" (urban)
FORESTNLCD06	Land Cover	Watershed percent "forest,"
PLANTNLCD06	Land Cover	Watershed percent "planted/cultivated" (agriculture)
AWCAVE	Soil	Average value for the range of available water capacity for the soil layer or horizon (inches of water per inches of soil depth)
BDAVE	Soil	Average value of bulk density (grams per cubic centimeter)
CLAYAVE	Soil	Average value of clay content (percentage)
HGA	Soil	Percentage of soils in hydrologic group A. Hydrologic group A soils have high infiltration rates. Soils are deep and well drained and, typically, have high sand and gravel content.
HGAC	Soil	Percentage of soils in hydrologic group A/C. Hydrologic group A/C soils have group A characteristics (high infiltration rates) when artificially drained and have group C characteristics (slow infiltration rates) when not drained.
HGAD	Soil	Percentage of soils in hydrologic group A/D. Hydrologic group A/D soils have group A characteristics (high infiltration rates) when artificially drained and have group D characteristics (very slow infiltration rates) when not drained.
HGB	Soil	Percentage of soils in hydrologic group B. Hydrologic group B soils have moderate infiltration rates. Soils are moderately deep, moderately well drained, and moderately coarse in texture.
HGBC	Soil	Percentage of soils in hydrologic group B/C. Hydrologic group B/C soils have group B characteristics (moderate infiltration rates) when artificially drained and have group C characteristics (slow infiltration rates) when not drained.
HGBD	Soil	Percentage of soils in hydrologic group B/D. Hydrologic group B/D soils have group B characteristics (moderate infiltration rates) when artificially drained and have group D characteristics (very slow infiltration rates) when not drained.
HGC	Soil	Percentage of soils in hydrologic group C. Hydrologic group C soils have slow soil infiltration rates. The soil profiles include layers impeding downward movement of water and, typically, have moderately fine or fine texture.
HGD	Soil	Percentage of soils in hydrologic group D. Hydrologic group D soils have very slow infiltration rates. Soils are clayey, have a high water table, or have a shallow impervious layer.
HGVAR	Soil	Percentage of soils in hydrologic group VAR. Hydrologic group VAR soils have variable drainage characteristics.
KFACT_UP	Soil	Average K-factor value for the uppermost soil horizon in each soil component. K-factor is an erodibility factor which quantifies the susceptibility of soil particles to detachment and movement by water. The K-factor is used in the Universal Soil Loss Equation (USLE) to estimate soil loss by water. Higher values of K-factor indicate greater potential for erosion

NO10AVE	Soil	Average value of percent by weight of soil material less than 3 inches in size and passing a No. 10 sieve (2 mm)
NO200AVE	Soil	Average value of percent by weight of soil material less than 3 inches in size and passing a No. 200 sieve (0.074 mm)
NO4AVE	Soil	Average value of percent by weight of soil material less than 3 inches in size and passing a No. 4 sieve (5 mm)
OMAVE	Soil	Average value of organic matter content (percent by weight).
PERMAVE	Soil	Average permeability (inches/hour)
RFACT	Soil	Rainfall and Runoff factor ("R factor" of Universal Soil Loss Equation); average annual value for period 1971-2000
ROCKDEPAVE	Soil	Average value of total soil thickness examined (inches)
SANDAVE	Soil	Average value of sand content (percentage)
SILTAVE	Soil	Average value of silt content (percentage)
WTDEPAVE	Soil	Average value of depth to seasonally high water table (feet)
ASPECT_DEGREES	Topographic	Mean watershed aspect, degrees (degrees of the compass, 0-360)
ASPECT_EASTNESS	Topographic	Aspect "eastness." Ranges from -1 to 1. Value of 1 indicates watershed is facing/draining due east; value of -1 indicates watershed is facing/draining due west.
ASPECT_NORTHNESS	Topographic	Aspect "northness." Ranges from -1 to 1. Value of 1 indicates watershed is facing/draining due north; value of -1 indicates watershed is facing/draining due south.
ELEV_MEAN_M_BASIN	Topographic	Mean watershed elevation (m)
RRMEAN	Topographic	Dimensionless elevation-relief ratio, calculated as (ELEV_MEAN - ELEV_MIN)/(ELEV_MAX - ELEV_MIN).
SLOPE_PCT	Topographic	Mean watershed slope, in percent
WATERNLCD06	Topographic	Watershed percent Open Water (class 11)

Figure S1. Boxplots of importance values for all at-site random forest (RF) models used in this analysis (n=1,900).

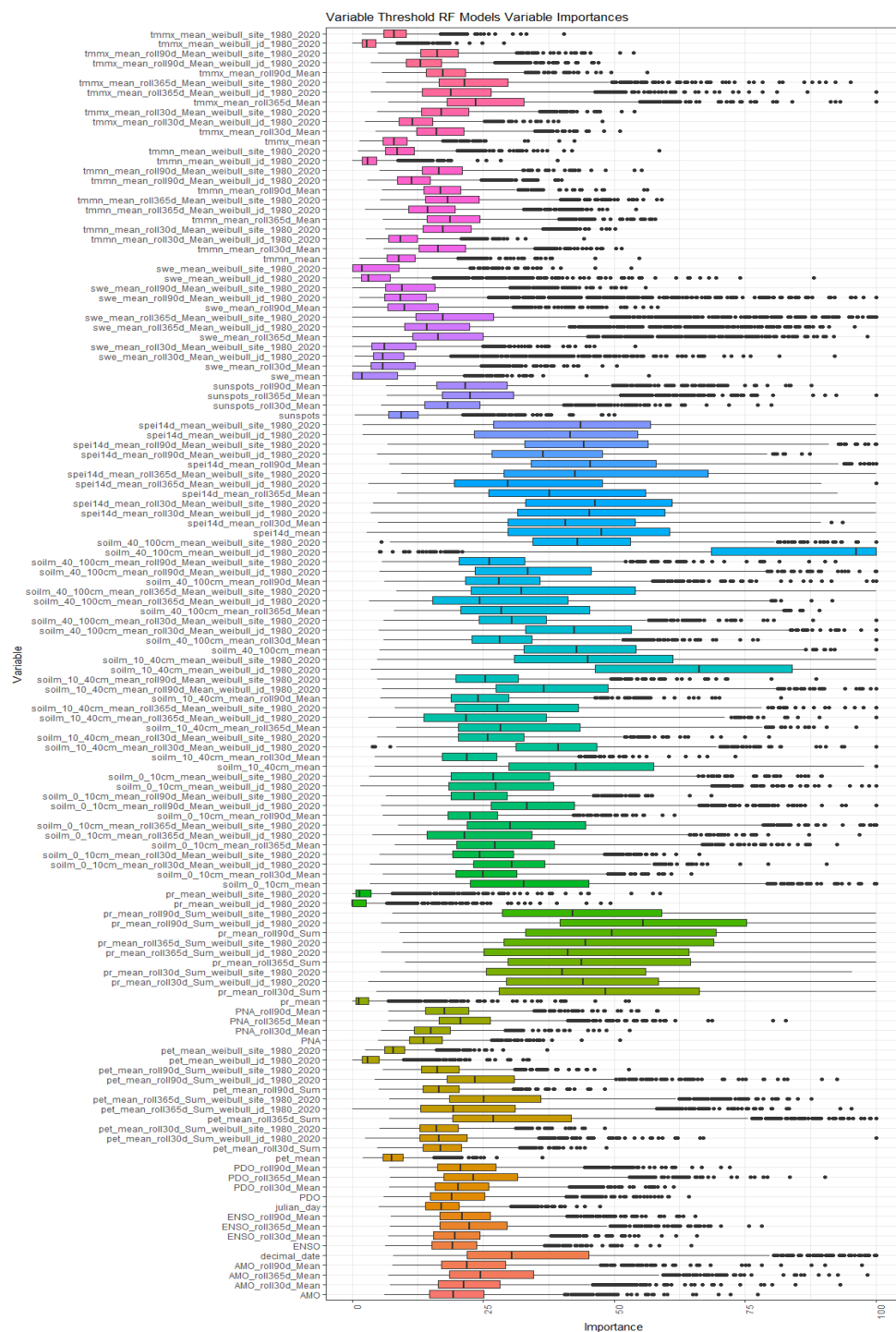


Table S2. Regression statistics for all 27 principal components.

PC	R2	Adjusted R2	Residual Standard Error	F-statistic	F-critical	Degrees of Freedom	p-value	AIC*
1	0.42	0.41	5.26	36.84	37	1862	0	6348.68
2	0.67	0.67	3.05	101.25	38	1861	0	4270.88
3	0.74	0.74	1.73	143.46	37	1862	0	2112.95
4	0.46	0.45	2.1	40.82	39	1860	0	2867.25
5	0.4	0.39	1.54	35.03	35	1864	0	1665.32
6	0.49	0.48	1.35	41.18	43	1856	0	1178.34
7	0.48	0.47	1.3	47.67	36	1863	0	1045.44
8	0.35	0.34	1.24	27.13	37	1862	0	851.3
9	0.33	0.32	1.2	31.35	29	1870	0	728.47
10	0.45	0.44	0.98	39.02	39	1860	0	-23.75
11	0.35	0.34	1	30.51	33	1866	0	52.01
12	0.3	0.28	0.95	18.17	43	1856	0	-132.8
13	0.26	0.24	0.93	16.49	39	1860	0	-227.83
14	0.24	0.23	0.94	16.76	35	1864	0	-213.63
15	0.2	0.19	0.85	13.14	36	1863	0	-570.53
16	0.2	0.19	0.82	15.07	31	1868	0	-704.76
17	0.23	0.21	0.78	16.46	33	1866	0	-914.29
18	0.26	0.25	0.73	18.08	37	1862	0	-1161.81
19	0.16	0.14	0.75	11.68	30	1869	0	-1059.55
20	0.18	0.17	0.72	11.75	35	1864	0	-1228.9
21	0.13	0.12	0.69	8.13	35	1864	0	-1389.93
22	0.14	0.13	0.66	9.92	31	1868	0	-1534.08
23	0.1	0.09	0.65	9.05	24	1875	0	-1591.39
24	0.15	0.13	0.63	8.97	37	1862	0	-1738.13
25	0.16	0.15	0.61	13.19	27	1872	0	-1863.06
26	0.15	0.14	0.56	10.3	32	1867	0	-2166.34
27	0.2	0.19	0.51	13.72	34	1865	0	-2491.11

* Akaike Information Criterion