Supplementary material

AGU PUBLICATIONS

Geophysical Research Letters

Supporting Information for

Variability of snow and rainfall partitioning into evapotranspiration and summer runoff across nine mountainous catchments

Matthias Sprenger[1]*, Rosemary W.H. Carroll[2], James Dennedy-Frank[1], Erica R. Siirila-Woodburn[1], Michelle E. Newcomer[1], Wendy Brown[3], Alexander Newman[3], Curtis Beutler[3], Markus Bill[1], Susan S. Hubbard[4], Kenneth H. Williams[1,3]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Desert Research Institute, Reno, NV, USA

[3]Rocky Mountain Biological Laboratory, Crested Butte, CO, USA

[4]Oak Ridge National Laboratory, Oak Ridge, TN, USA

Contents of this file

Text S1: Gap filling of streamflow data

We used a random forest ensemble machine learning algorithm to regress hourly discharge (*Q*, m^3/s) between nearby discharge monitoring stations within the East River watershed network (Shortridge et al., 2016; Dwivedi et al. 2022). This method included stations monitored by the USGS from 1940-2021. All sites have some periods where major gaps exist. Hourly discharge data for long term USGS monitoring stations (Almont Station ID# 09112500, East River Below Cement Creek Station ID# 09112200, and Gunnison Station ID# 09114500) were obtained from the USGS NWIS portal using the R dataRetrieval package (Hirsch & De Cicco, 2015; USGS, 2022) and this data was merged with East River discharge monitoring locations (Carroll and Williams, 2019; Carroll et al. 2021). The machine learning algorithm within the R package randomForest implements Breiman's random forest algorithm for classification and regression (Breiman, 2001; Liaw & Wiener, 2018). We selected random forest because of the robust capabilities to handle a large number of predictor and target variables, and handle the broad temporal structure of the data (hourly over 8 years). In our approach, we allowed all available continuous discharge from any station to be used as predictor variables for the target variable (discharge station with gaps) (Jain et al., 2020). Once discharge was simulated based on a stations neighbors, we filled in the gap with the simulated values thus preserving all of the original data and the temporal structure. Additionally, this method preserves the presence of any non-stationarity within the data. Gap-filled discharge data are freely and publically available on ESS-DIVE (Newcomer et al. 2022)

Text S2: References for supplementary information

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carroll R ; Williams K (2019): Discharge data collected within the East River for the Lawrence Berkeley National Laboratory Watershed Function Science Focus Area (water years 2015-2018). Watershed Function SFA, ESS-DIVE repository. Dataset. doi:10.21952/WTR/1495380.

Carroll R ; Newman A ; Beutler C ; Williams K (2021): Stream discharge data collected within the East River, Colorado for the Lawrence Berkeley National Laboratory Watershed Function Science Focus Area (water years 2019 to 2020). Watershed Function SFA, ESS-DIVE repository. Dataset. doi:10.15485/1779721.

Dwivedi, D., Mital, U., Faybishenko, B., Dafflon, B., Varadharajan, C., Agarwal, D., et al. (2021). Imputation of Contiguous Gaps and Extremes of Subhourly Groundwater Time Series Using Random Forests. Journal of Machine Learning for Modeling and Computing. https://doi.org/10.1615/JMachLearnModelComput.2021038774

Hirsch, R. M., & De Cicco, L. (2015). User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data (version 2.0, February 2015). Retrieved from https://dx.doi.org/10.3133/tm4A10

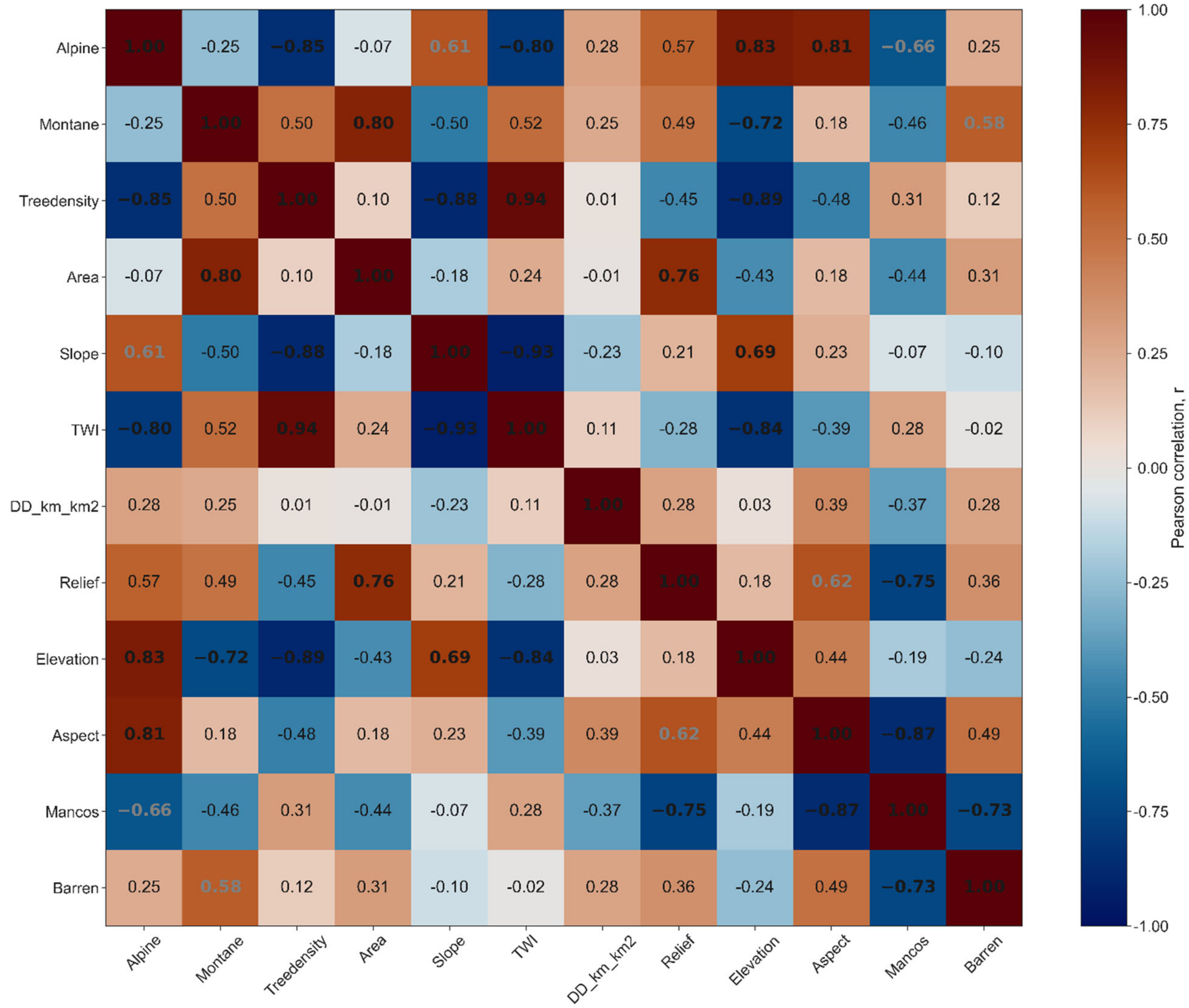
Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4), 478–505. https://doi.org/10.1139/er-2020-0019

Liaw, A., & Wiener, M. (2018). Breiman and Cutler's Random Forests for Classification and Regression. CRAN. Retrieved from https://www.stat.berkeley.edu/~breiman/RandomForests/

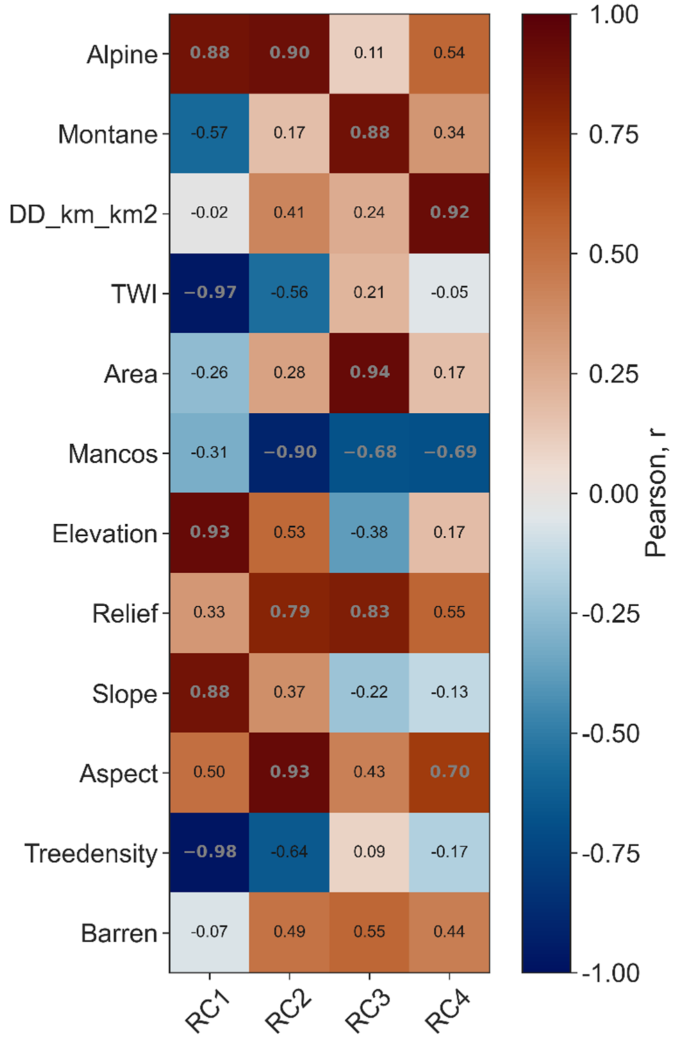
Newcomer, M.E., Williams, K.W., and Carroll R.W.H. (2022). Machine Learning Assisted Gap-Filled Discharge Data for the East River Community Watereshed for Water Years 2014-2021. [Data set]. Environmental System Science Data Infrastructure for a Virtual Ecosystem; Watershed Function SFA. https://www.dropbox.com/s/gd9jxv4f1x7blao/All_RF_Wide.csv?dl=0 (Upload to https://ess-dive.lbl.gov/ in progress and DOI fill follow)

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences, 20(7), 2611–2628. https://doi.org/10.5194/hess-20-2611-2016

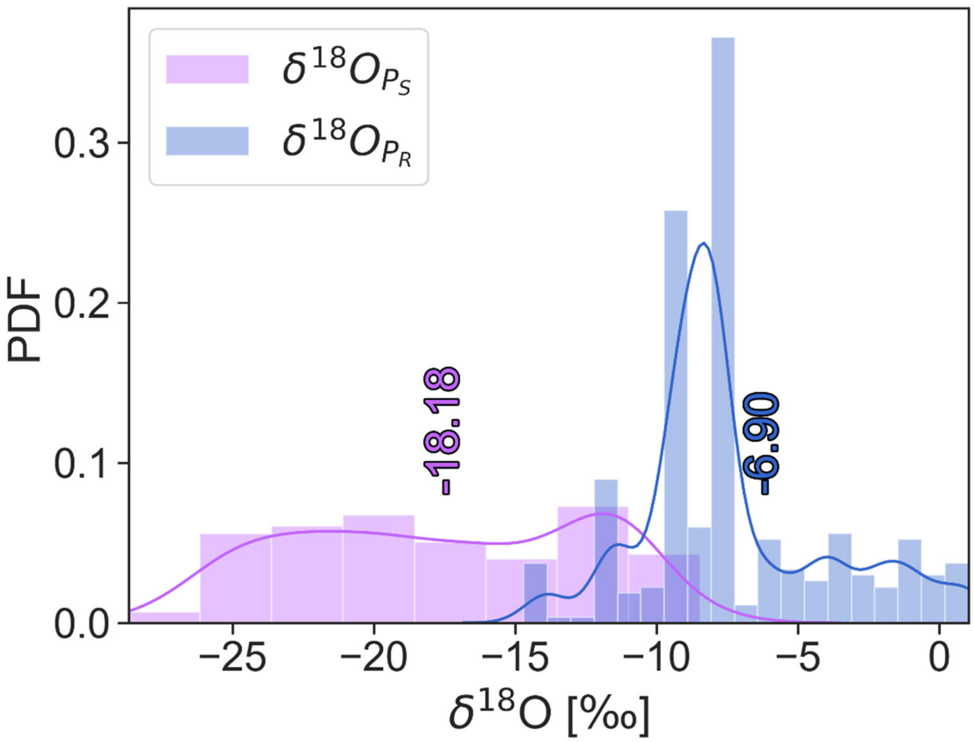
USGS. (2022). National Water Information System-Web Interface. Retrieved from https://waterdata.usgs.gov/nwis



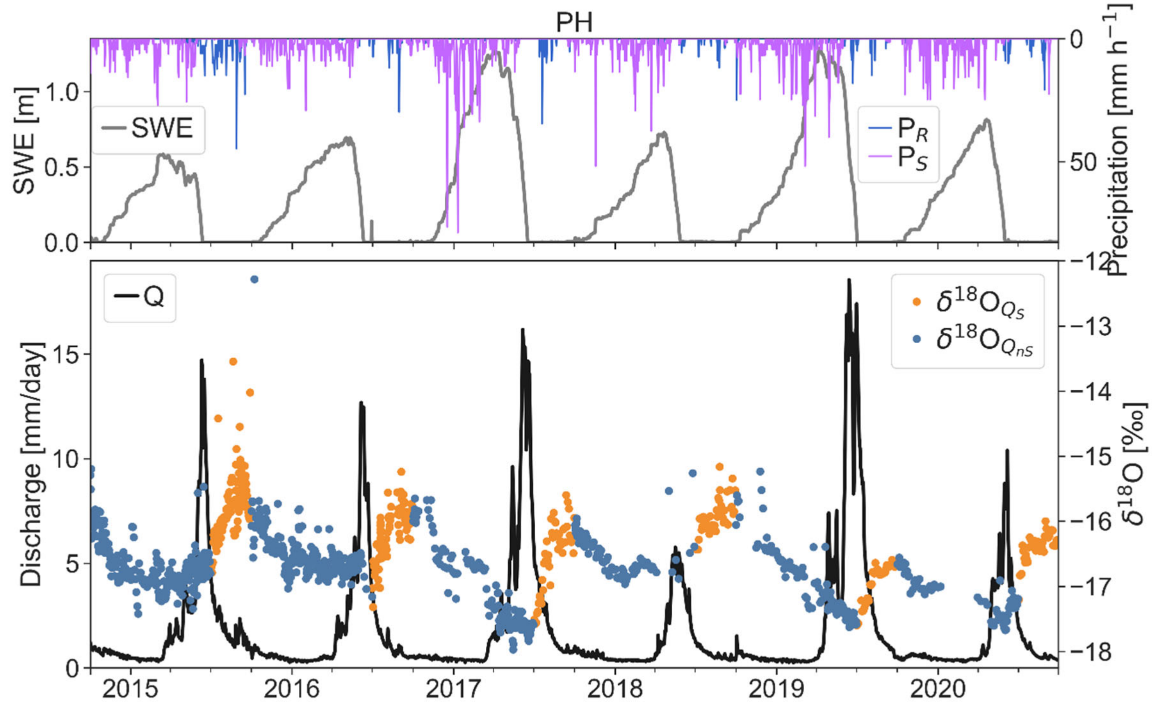
Suppl. Fig. 1 Cross-correlation of different catchment characteristics. TWI = Topographic wetness index [-]; DD_km_km2 = Drainage density [km km^{-2}]; Mancos = share of surface area with Mancos shale geology [%]; Barren = share of surface area with bare soil or rock [%]. Bold and black numbers represent significant correlations at p<0.05 and bold grey numbers for p<0.10.



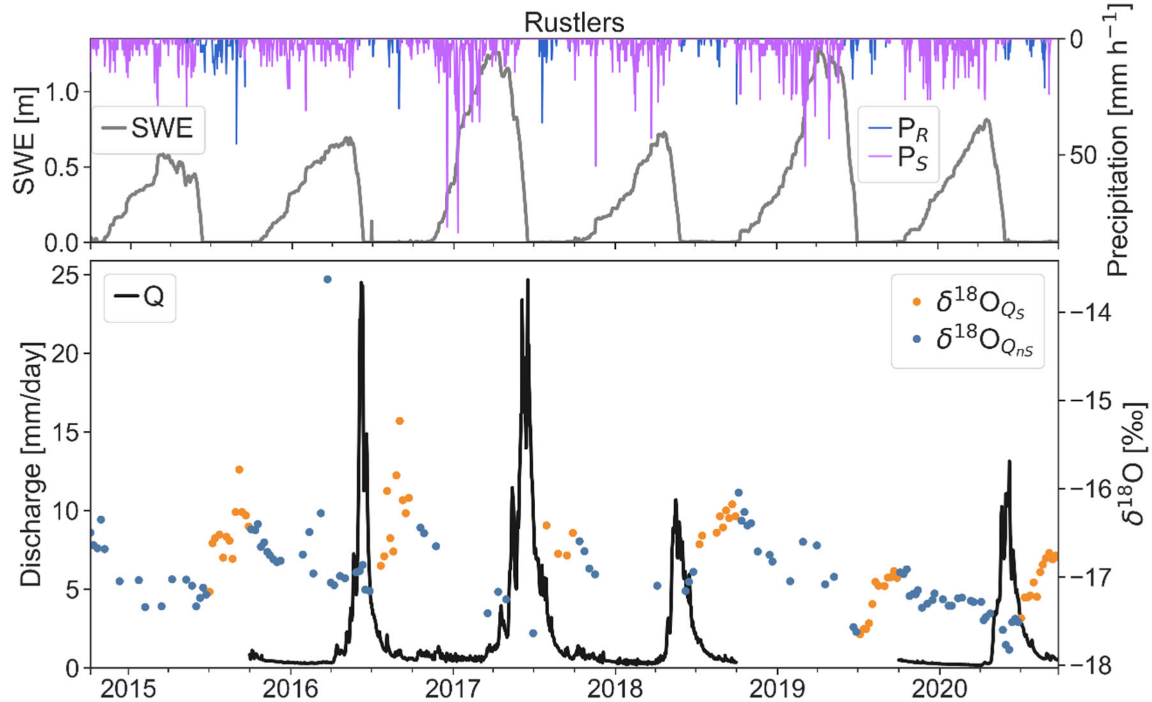
Suppl. Fig. 2 Correlation between the four components of rotated principle component analysis and their correlation with various catchment characteristics (same as in Suppl. Fig. 1). From these relationships, we chose four characteristics that represent the four components: Tree density, aspect, size area, and drainage density. Bold and black numbers represent significant correlations at p<0.05 and bold grey numbers for p<0.10.



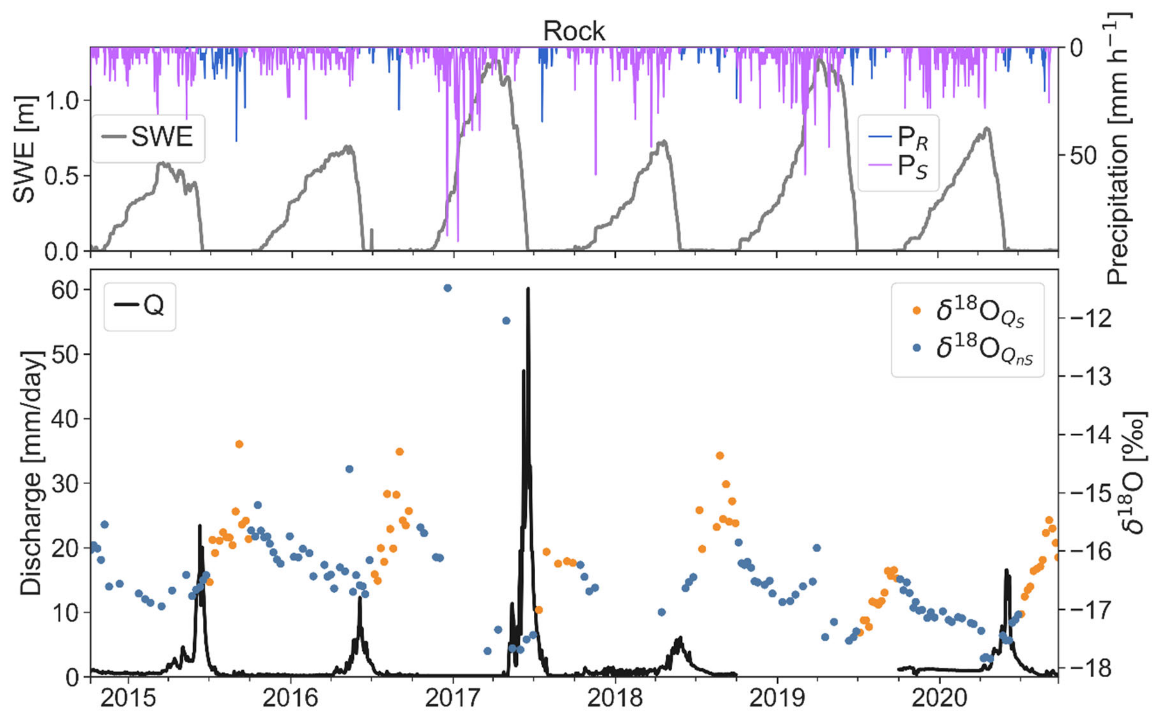
Suppl. Fig. 3 Distribution of $\delta^{18}O$ values for snow ($\delta^{18}O_{PS}$) and rain ($\delta^{18}O_{PR}$) samples taken during water year 2015 and 2016 at the location shown in Figure 1. Lines show the kernel density estimate for both histograms. Vertical values represent the weighted average values for snow and rain, respectively.



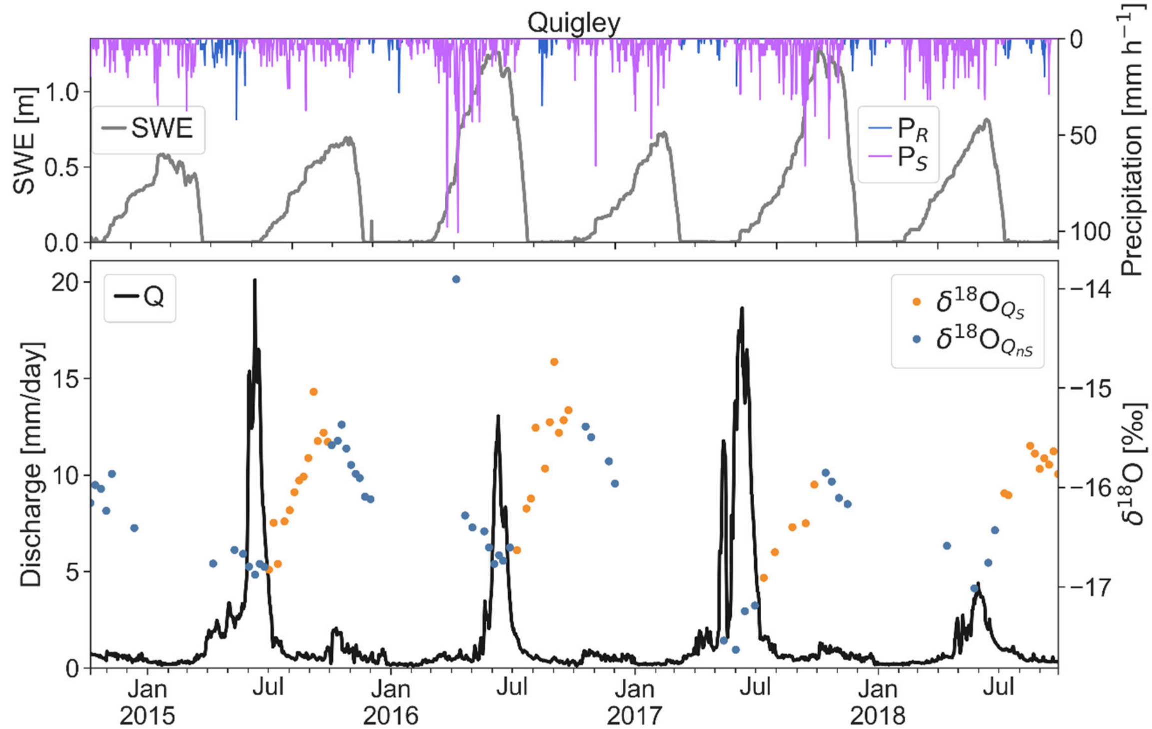
Suppl. Fig. 4 Upper panel: Snow water equivalent (SWE) measured at the SNOTEL site in Schofield (location shown in Figure 1), Rainfall (P_R) and snowfall (P_S) hourly sums for the East River catchment defined by the Pumphouse outlet (location shown in Figure 1). Lower panel: Measured discharge (Q) and $\delta^{18}O$ values for water sampled during the summer ($\delta^{18}O_S$) and non-summer ($\delta^{18}O_{nS}$) periods at the Pumphouse.



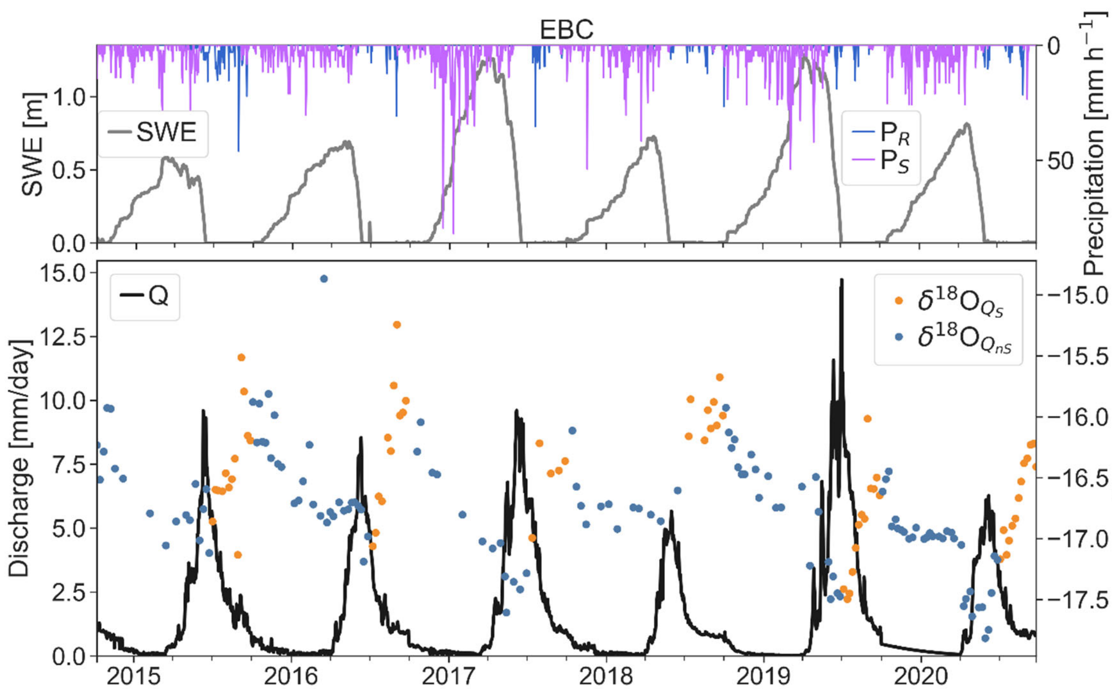
Suppl. Fig. 5 Same as in Suppl. Fig. 4, but for the Rustlers Creek catchment.



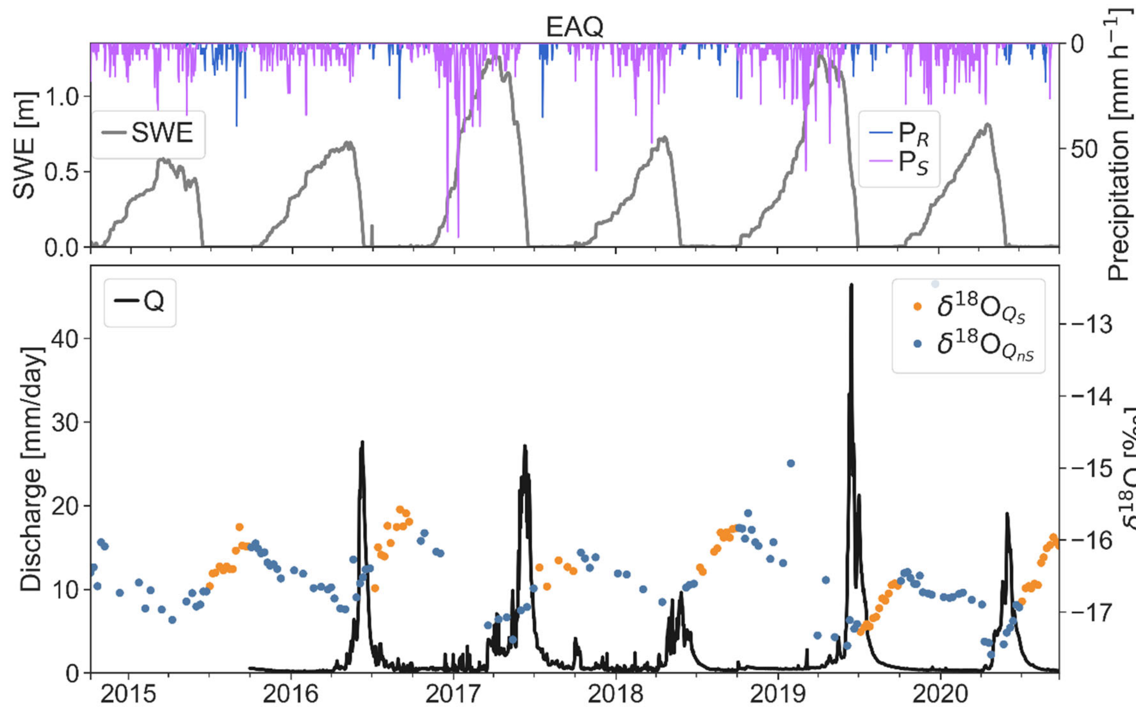
Suppl. Fig. 6 Same as in Suppl. Fig. 4, but for the Rock Creek catchment.



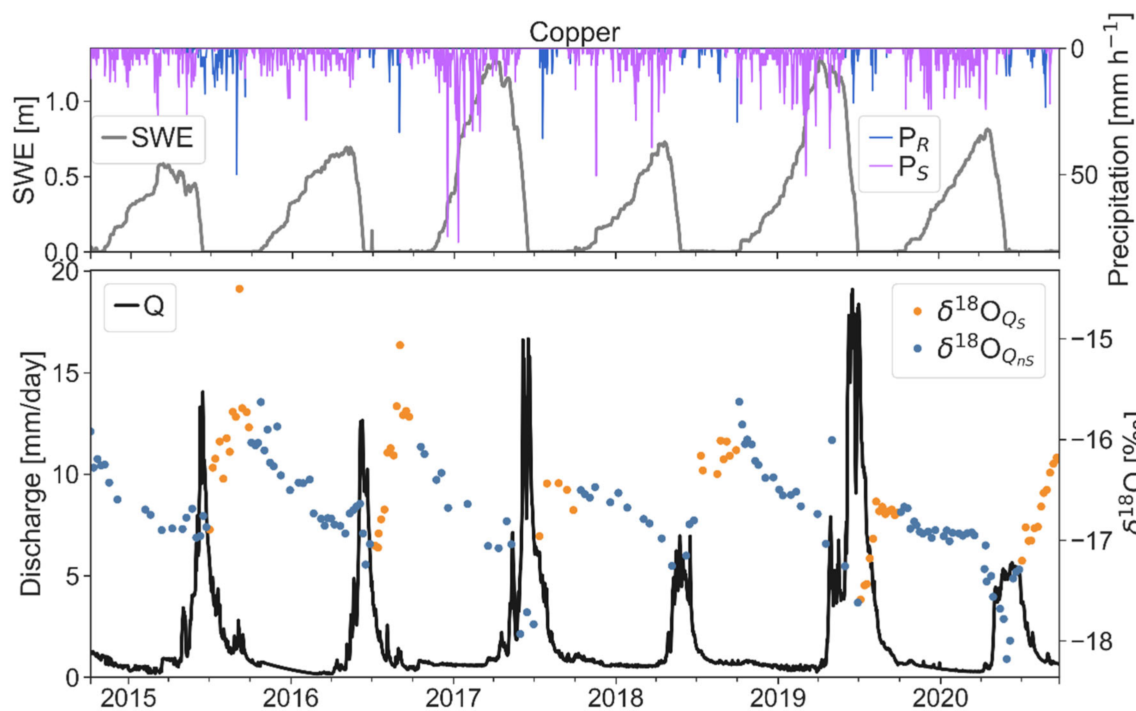
Suppl. Fig. 7 Same as in Suppl. Fig. 4, but for the Quigley Creek catchment.



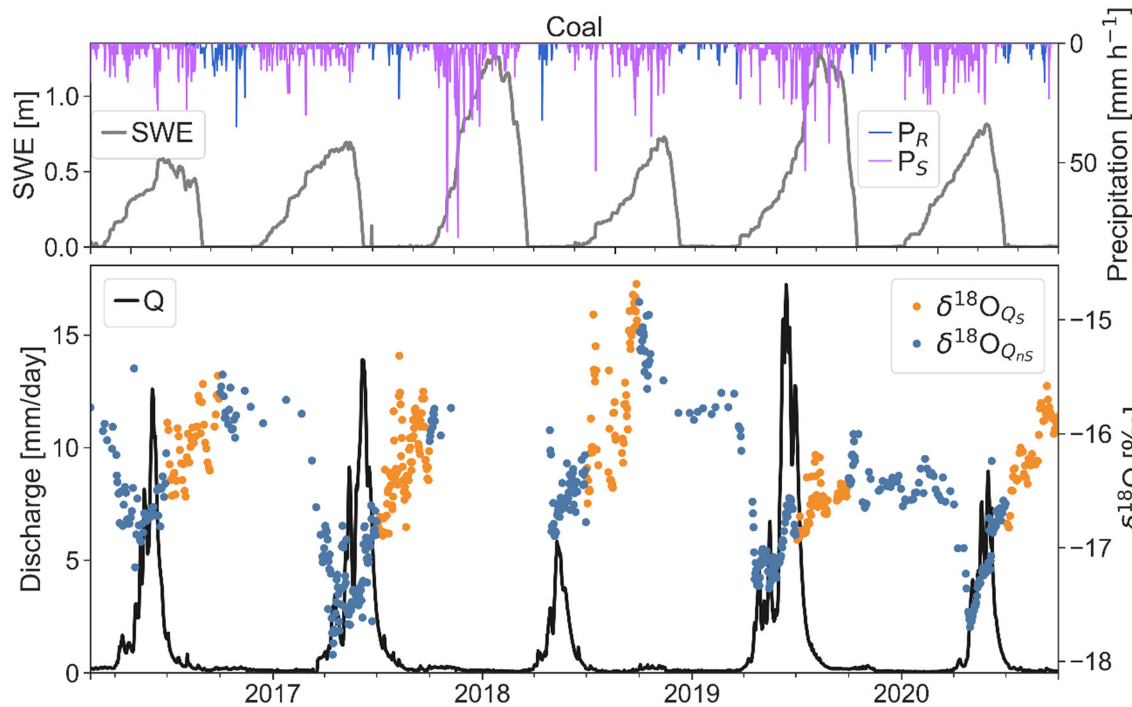
Suppl. Fig. 8 Same as in Suppl. Fig. 4, but for the East River catchment below Copper Creek tributary.



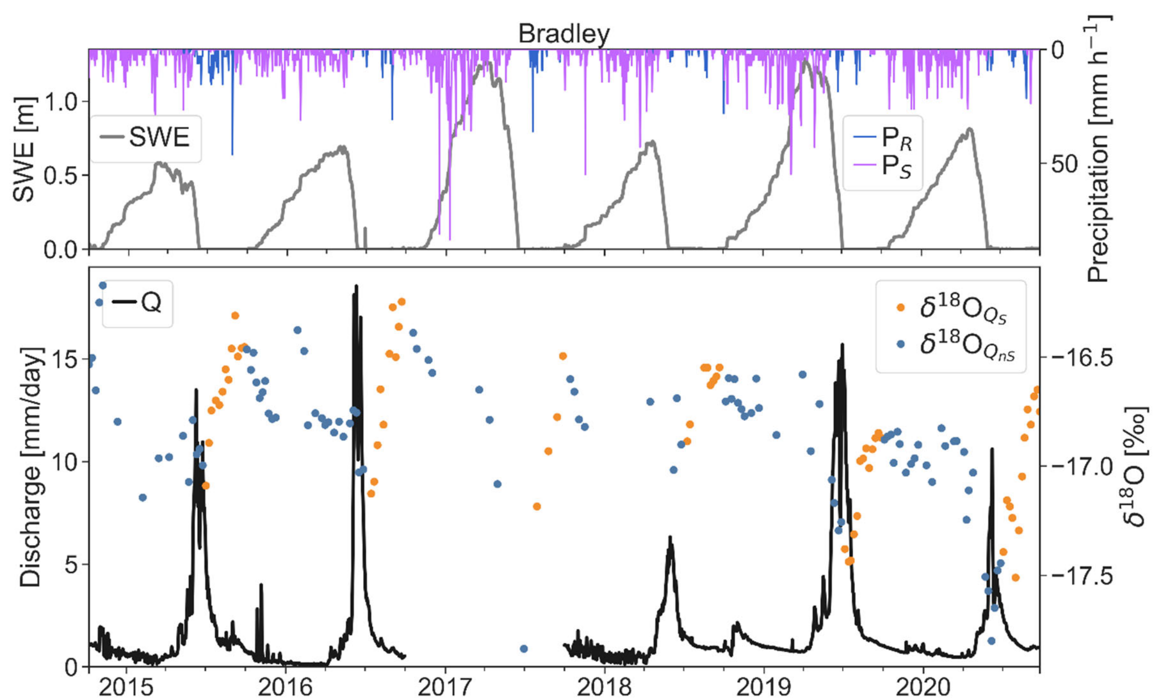
Suppl. Fig. 9 Same as in Suppl. Fig. 4, but for the East River catchment above the Quigley tributary.



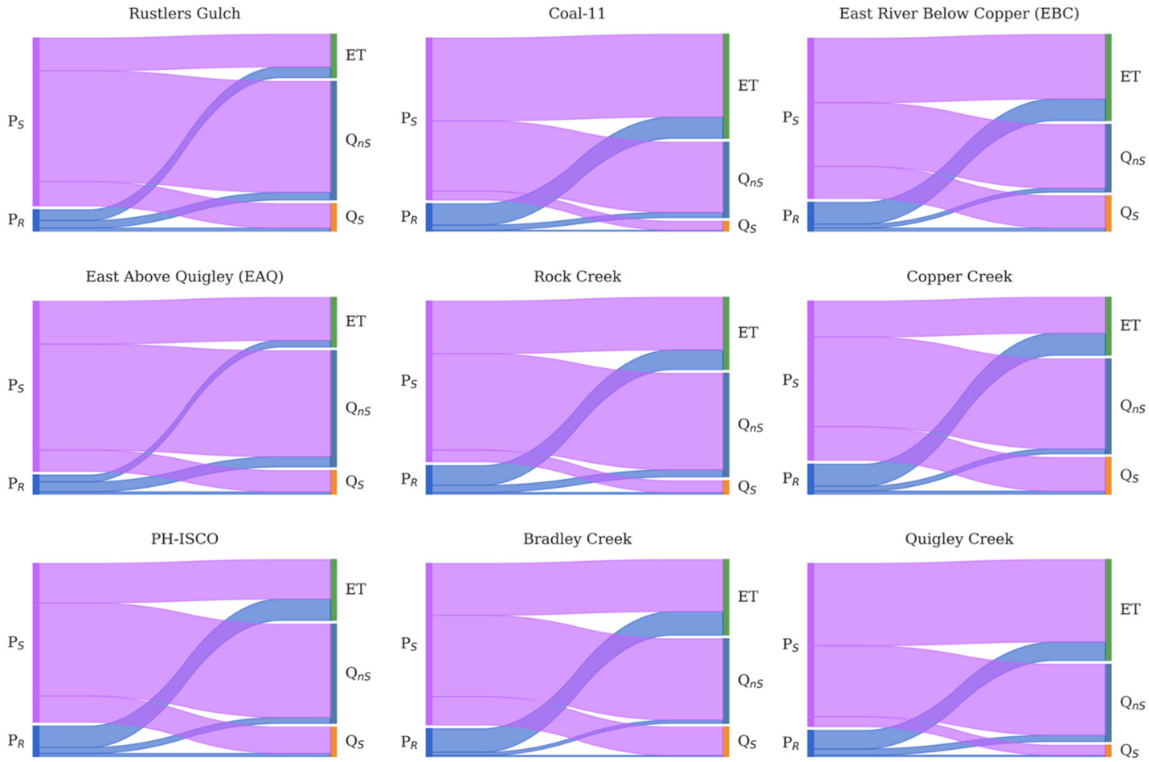
Suppl. Fig. 10 Same as in Suppl. Fig. 4, but for the Copper Creek catchment.



Suppl. Fig. 11 Same as in Suppl. Fig. 4, but for the Coal Creek catchment.



Suppl. Fig. 12 Same as in Suppl. Fig. 4, but for the Bradley Creek catchment.



Suppl. Fig. 13 Long-term (2015-2020) isotope mass balance for the nine catchments. Snow (P_S) and rain (P_R) are split into evapotranspiration (ET), non-summer (Q_{nS}) and summer season discharge (Q_S) based on endmember splitting analysis. The composition of ET, Q_{nS} and Q_S with regard to snow and rain is based on endmember mixing analysis.

Suppl. Table 1 Catchment specific characteristics: total surface area [km^2], catchment scale averages of the slope [-], aspect [°], elevation [m a.s.l.], relief [m], drainage density [km/km^2], topographic wetness index (TWI) [-], share of montane, subalpine, upper subalpine, alpine, Mancos lithology, and barren land as % of total catchment area. EAQ = East River above Quigley, EBC = East River above Copper Creek, PH = Pumphouse. We derived catchment average tree cover density using USGS Landfire at 30 m raster maps (Landfire, 2015, Existing vegetation type and cover layers. U.S. Department of the Interior, Geological Survey. http://landfire.cr.usgs.gov/viewer, accessed October 2021) weighing the different forest classes by their given tree density. Alpine > 3750 m; upper subalpine: 3700-3525 m; subalpine: 3525–3000 m, montane: <3000 m.

	Area [km^2]	Slope [-]	Aspect [°]	Elevation [m asl]	Relief [m]	DD [km/km2]	TWI [-]	Tree density [%]	Montane [%]	Subalpine [%]	upper Subalpine [%]	Alpine [%]	Mancos [%]	Barren [%]
Quigley	2.55	26	95	3,365	911	1.47	4.9	26	2	78	17	3	62	23
Rock	3.57	19	120	3,363	930	1.71	5.2	32	1	86	11	2	68	8
Bradley	3.82	26	233	3,485	1,102	1.76	4.9	20	6	45	30	19	1	40
EAQ	5.27	28	130	3,333	904	1.56	5.0	24	1	82	15	2	70	21
Rustlers	14.78	26	191	3,475	1,118	1.75	4.9	18	1	57	29	13	8	23
Copper	23.67	29	191	3,513	1,237	1.75	4.8	14	2	47	31	20	1	50
Coal	52.80	18	164	3,148	1,073	1.81	5.3	39	23	72	5	0	0	3.5
EBC	69.81	25	168	3,351	1,196	1.26	5.0	24	10	63	20	7	18	31
PH	84.73	25	169	3,331	1,362	1.85	5.0	22	18	53	20	9	18	26