https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences Discussions

Integrating remotely sensed surface water dynamics into hydrologic signature modelling

Melanie K. Vanderhoof[1*], Peter Nieuwlandt[2], Heather E. Golden[3], Charles R. Lane[4], Jay R. Christensen[3], William Keenan[1], Wayana Dolan[1]

[1]U.S. Geological Survey, Geosciences and Environmental Change Science Center, PO Box 25046, MS 980, Denver Federal Center, Denver Colorado 80225, USA
[2]Delaware Water Gap National Recreation Area, 1978 River Rd, Bushkill, PA 18324, USA
[3]Office of Research and Development, U.S. Environmental Protection Agency, 26 W. Martin Luther King Dr., Cincinnati, Ohio, 45268, USA
[4]Office of Research and Development, U.S. Environmental Protection Agency, 980 College Station Road, Athens, Georgia 30605, USA

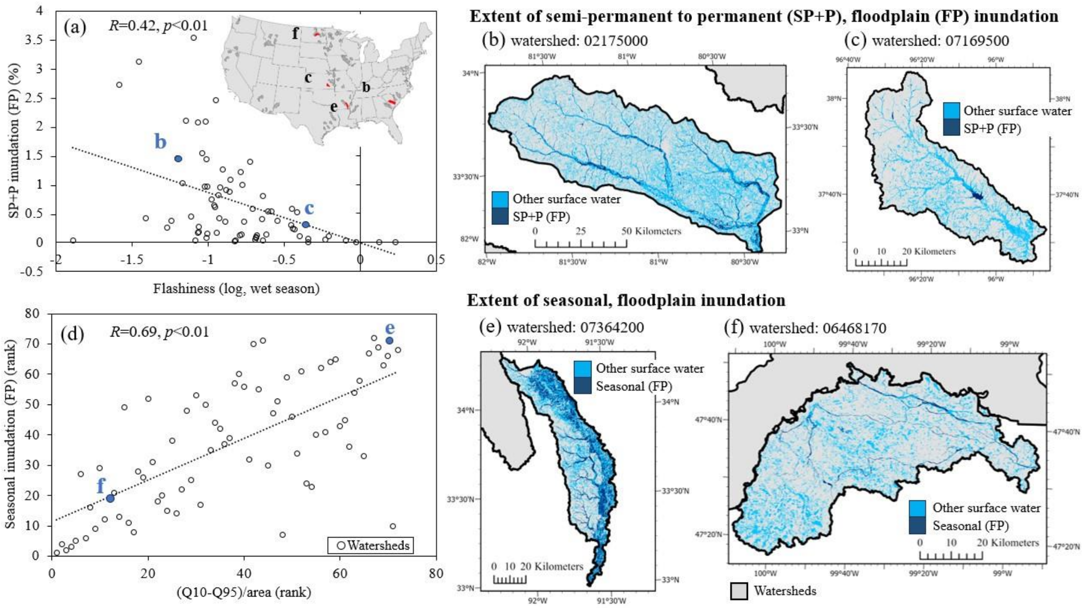
Correspondence to: Melanie Vanderhoof (mvanderhoof@usgs.gov)

Abstract. Extreme flow conditions in rivers have far-reaching environmental and economic consequences. The retention of surface water in lakes, wetlands, and floodplains can potentially modify the timing, duration, and magnitude of flow. However, efforts to explore the impact of surface water storage on discharge regimes have been limited in geographic extent. In this analysis, we calculated six hydrologic signatures, reflecting flashiness and high and low flow conditions, at 72 gaged watersheds across the conterminous United States. In addition to traditionally considered variables representing climate, land cover, topography, and soil, we incorporated a novel remote sensing (Sentinel-1 & 2) approach to study the contribution of surface water storage dynamics when modelling spatial variability in hydrologic signatures using random forest models. While climate variables explained much of the variability in the hydrologic signatures, models for five of the six signatures showed some degree of improvement in model performance when landscape characteristics were added with adjusted R^2 improving 1.75 to 11.69% and Akaike information criterions improving 0.24% to 6.69%. Automated variable selection can be indicative of the relative importance of certain variables over others. Using a forward selection process, five of the six signature models selected remotely sensed inundation variables with all five variables showing a significant (p<0.01) contribution to the respective model. More semi-permanent and permanent inundation within the floodplain (i.e., lakes along rivers), for example, was associated with lower wet season and annual flashiness. Further, greater seasonal floodplain inundation extent was associated with increases in peak flows, so that floodplain water storage was relevant to both flashiness and high flow signatures. Additionally, spatial variability in the amount of semi-permanent and permanent non-floodplain water significantly contributed to explaining spatial variability in the baseflow index. These findings suggest that surface water storage dynamics may help explain variability in streamflow signatures. Watershed management will benefit from an improved understanding of how surface water storage influences stream behaviour.

Keywords: drought, floodplain, floods, hydrologic signatures, inundation, lakes, non-floodplain wetlands, stream discharge metrics, wetlands

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Graphical Abstract



Short Summary

Streamflow signatures can help characterize a watershed's response to meteorological conditions. We explored if surface water storage-related variables, which are typically excluded from streamflow signature analyses, may help explain spatial variability in streamflow signatures. We found that remotely sensed surface water storage extent and duration were correlated with and explained a portion of the variability in many of the hydrologic signatures across the 72 streamflow gages.

1. Introduction

The response of streamflow to climate extremes has important environmental and economic implications. Drought events limit streamflow available for agriculture, drinking water, and wildlife (Stewart et al., 2020; Apurv and Cai, 2021), and have cost the United States $53 billion in just the past five years (2019-2023) (NOAA, 2024). Flood events, meanwhile, can endanger property, infrastructure, and human lives, and have caused global economic damages exceeding $1 trillion between 1980 and 2013 (Winsemius et al., 2016). Climate change is altering the frequency of these hydroclimatic extremes (Heidari et al., 2020) and may also alter how climate extremes propagate to impact runoff (Wu et al., 2022). In recent years, several studies have shown that surface water storage (e.g., wetlands, lakes, ponds), at least in some watersheds, can potentially increase baseflow and decrease peak flows (Rajib et al., 2020; Wu et al., 2020; Zeng et al., 2020), implying that consideration of surface water storage and storage dynamics in models could improve predictions of flood and drought impacts (Golden et al., 2021). However, surface water storage is typically excluded from both hydrological models (Golden et al., 2014; Jones et al., 2019) and analyses of river and stream hydrologic signatures (Addor et al., 2018; McMillan, 2019). Therefore, our understanding of when and where surface water storage influences river discharge is still very limited.

Hydrologic signatures are quantitative metrics, typically calculated from daily timeseries of discharge, that can describe the magnitude, timing, rate of change, duration, and frequency of flow conditions (Richter et al., 1996; Daigle et al., 2011; McMillan et al., 2019). Hydrologic signatures are often selected for a specific hydrological or ecological application or objective. For example, some studies have developed signatures that reflect wet conditions such as flashiness or seasonal flooding (Hannaford and March, 2008; Hendry et al., 2019), while others have focused on applying hydrologic signatures to characterize late-season, low flow regimes (Daigle et al., 2011; Kelly and White, 2016), or alternatively, the impact of hydrologic alterations, such as groundwater pumping, flow diversions, or land use conversion (Richter et al., 1996). The relationship between hydrologic signatures and watershed characteristics, such as climate and topography, has been characterized using statistical techniques such as correlation analyses (Berghuijs et al., 2016; Kuentz et al., 2017), random forest models (Trancoso et al., 2016; Addor et al., 2018; Oppel and Schumann, 2020) and regression functions (van Dijk, 2010; Beck et al., 2015; Kuentz et al., 2017), with studies finding variability in the model strength between different signatures (Beck et al., 2015; Addor et al., 2018).

Previous research has shown that drivers of hydrologic signatures can reflect specific aspects of flow. For example, signatures that reflect high flow events are often best predicted by climate, including precipitation (van Dijk, 2010; Kuentz et al., 2017), while signatures reflecting baseflow are often linked to geology (Kuentz et al., 2017), as well as potential evapotranspiration (van Dijk, 2010; Beck et al., 2013). Generally, hydrologic signatures are best explained by climate variables, such as aridity, precipitation, and snowfall (Beck et al., 2015; Addor et al., 2018). Land cover, such as proportion forest, often acts as a secondary controlling process (Kuentz et al., 2017; Trancoso et al., 2016; Addor et al., 2018). While Beck et al. (2013) found baseflow to be positively correlated with the average proportion of each basin classified as open water, and Beck et al. (2015) found slope, which can be indicative of potential water storage capacity, to be helpful in explaining multiple signatures, efforts to model drivers of hydrologic signatures have rarely included or considered surface water storage capacity, and have not, to our knowledge, considered surface water extent dynamics or hydroperiod.

Despite surface water storage being infrequently considered in the analysis of hydrologic signatures, it is widely accepted that wetlands and lakes have a significant influence on the hydrologic cycle (Bullock and Acreman, 2003). In watersheds lacking surface water storage (e.g., lakes, ponds, reservoirs, and wetlands) when precipitation falls, it is captured by vegetation, infiltrates the soils, or is transported downgradient as infiltration-excess or saturation-excess runoff (Eamus et al., 2006). Conversely, in watersheds where surface storage availability exists, precipitation, snow water equivalent and runoff can be stored and gradually released through time from both floodplain and non-floodplain storage - via groundwater baseflow, fill-spill surface runoff, or merging with streams via fill-and-spill mechanisms (Rains et al., 2016; Fritz et al., 2018; Lane et al., 2018; Stepchinski et al., 2023), creating a less “flashy” system (Shaw et al., 2012; Kuppel et al., 2015). Surface storage areas, both within and outside of the floodplain, can also contribute to streamflow when stream-connected surface-water stages rise, subsuming nearby, previously disconnected storage systems, e.g., non-floodplain wetlands (Vanderhoof et al., 2016). The influence of these disconnected systems, e.g., upland wetlands, can depend on the position of the wetlands relative to the stream network as well as watershed

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

characteristics (Fritz et al., 2018; Lane et al., 2018; Wu et al., 2020). Although we know that lakes and wetlands can withhold and contribute water to river networks, it is less clear if surface water storage across multiple watersheds and regions has a measurable impact on river discharge dynamics.

Our limited understanding of how surface water storage dynamics impact river discharge is in part attributable to surface water storage being traditionally ignored by hydrologic models (Golden et al., 2014; Jones et al., 2019). In recent years, studies have shown that integrating wetlands, particularly non-floodplain wetlands, into hydrologic models can improve streamflow simulation accuracy (Rajib et al., 2020; Golden et al., 2021). While recent modelling studies have been limited in spatial extents, have simplified wetland volume estimates, and have relied, most commonly, on topographic estimates of potential water storage, each have demonstrated that surface water storage, at the scale of an individual basin, can potentially increase baseflow (McLaughlin et al., 2014; Zeng et al., 2020) as well as potentially reduce peak flow and flood duration (Evenson et al., 2018; Ameli and Creed, 2019; Wu et al., 2020).

Further research is needed to improve our understanding of when and where dynamic surface water storage influences river discharge across multiple diverse watersheds and regions. Here, we calculated a suite of hydrologic signatures to characterize variability in flow flashiness and high and low flow conditions across 72 diverse watersheds in the contiguous United States (CONUS). We developed two random forest models for each flow signature: one representing climate variables only and one representing climate, land cover, geology, topographic, and surface water storage input variables. This approach helped us to assess the relative ability of climate alone, compared to catchment characteristics that uniquely included novel remotely sensed surface water extent and hydroperiod, to explain the variability in hydrologic signatures. Specifically, our research questions were: (1) What are the dominant explanatory variables explaining the variability in flow flashiness and high and low flow condition-related hydrologic signatures across watersheds representing different climates, topography, and land covers? and (2) To what extent do surface water storage-related variables correlate with or help explain variability in these selected hydrologic signatures?

2. Materials and Methods

2.1 Watersheds

A total of 72 U.S. Geological Survey (USGS) stream gages and associated watersheds (Fig. 1) were selected across the conterminous U.S. (CONUS) from the GAGES-II dataset (Falcone, 2011). Catchment size influences runoff (Pilgrim et al. 1982), therefore we prioritized selecting non-nested watersheds within a bounded size class. Most watersheds, 80%, were between 1500 km^2 and 5000 km^2, while the full-size range was 292 km^2 to 9918 km^2. In comparison, 74 of the CAMEL watersheds are between 1500 and 5000 km^2 (Newman et al., 2014). Secondly, gaged watersheds, to the extent possible, were selected to be approximately co-located with regions used to train the Sentinel-1 and Sentinel-2 satellite-based surface water algorithms to maximize the accuracy of the algorithms (Vanderhoof et al., 2023). The algorithms were used to map surface water extent over time at each of the watersheds. The intensity of computing resources needed to process Sentinel-1 and Sentinel-2 imagery into surface water extent also limited the number of watersheds that was feasible to include. Watersheds with tidal wetlands were excluded to focus on freshwater aquatic systems. Further, potential watersheds were reviewed to minimize the inclusion of major dams,

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
Discussions

defined as dams 15.2 meters or more in height (storage capacity of 6.17 million cubic meters) near watershed outlets (National Atlas of the United States, 2006).

Across the selected watersheds, stream density, as calculated from the National Hydrography Dataset (NHDplus) high resolution dataset (USGS, 2022), ranged from 259 m km^{-2} to 4182 m km^{-2} across the selected watersheds, with a median density of 1461 m km^{-2} (Table A1). The proportion of each watershed classified as wetland by the National Wetland Inventory (NWI) dataset (USFWS, 2019) ranged from 1.1% to 48.7% with a median wetland proportion of 5.6% (Table A1). Mean annual precipitation (2016-2023) ranged from 325 mm to 1659 mm, with a median annual average of 967 mm (GRIDMET; Abatzoglou, 2013). In addition, the dominant landcover class was cultivated crops or hay/pasture for 36 of the watersheds, with other dominant classes including forest (18 watersheds) and grassland-shrub/scrub (13 watersheds) (Homer et al., 2020; Table A1). The watersheds were grouped by U.S. region, including West, Southwest, North Central, Gulf Coast, Midwest, and East, to facilitate data interpretation (Fig. 1).

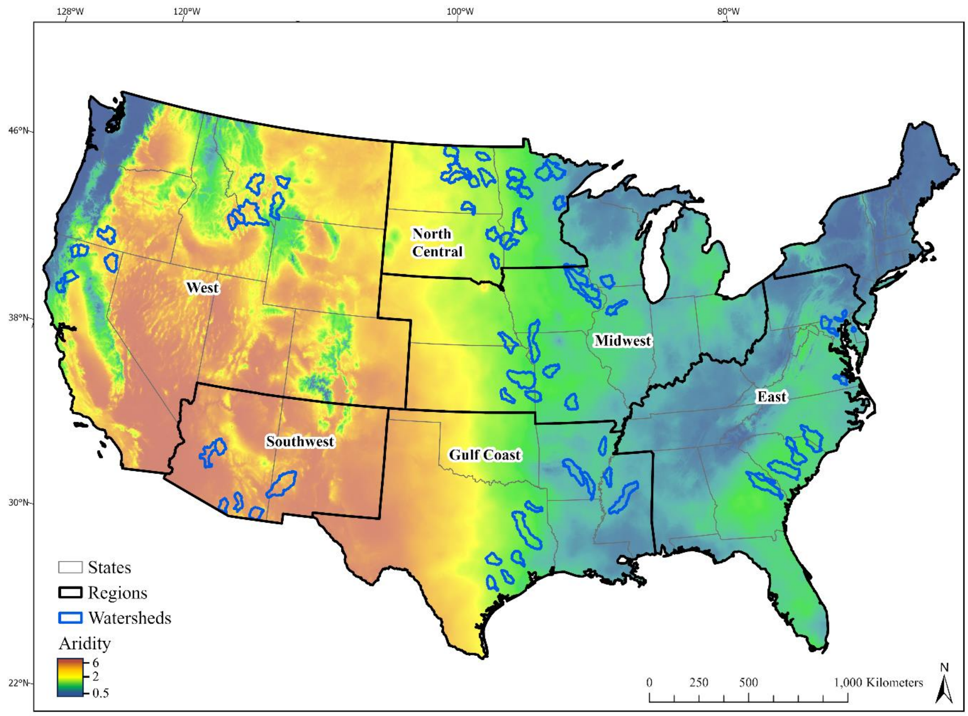


Figure 1. Selected U.S. Geological Survey (USGS) gaged watersheds in relation to aridity (2016-2023), where maroon/orange indicates arid conditions and blue indicates less arid conditions. Legend values indicate median values for the corresponding colours and a histogram equalize stretch was applied.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

2.2 Hydrologic signatures: response variables

Hydrologic signatures were calculated from daily discharge at each gage and were used as the response variables in our statistical analyses (Table 1). Daily rate of stream discharge was acquired from the USGS National Water Information System for 2016-2023 (USGS, 2024). The period was limited by the temporal availability of Sentinel-2 imagery (Sentinel-2a and -2b launched in June 2015 and March 2017, respectively), required for the surface water algorithm. Signatures were selected from the literature to represent discharge extremes (high flow and low flow) as well as variability in discharge and were calculated using the calendar year. Signatures related to characterizing high flow conditions included a (1) wet season flashiness index, where flashiness reflected daily variability in discharge within the wet season, defined as the three months in each year with the highest average discharge (Baker et al., 2004). (2) The maximum annual 30-day flow per drainage area (km^2) (MAX30/area) reflected seasonal peaks in discharge (Hannaford and Marsh, 2008); and (3) discharge exceeded 10% of the time, within a given year (Q10) minus discharge exceeded 95% of the time (Q95), within a given year ((Q10-Q95)/area) and averaged over multiple years, or the difference between high flows and the baseflow regime (National River Flow Archive, 2024). The (4) flashiness index, which reflected daily variability in discharge across seasons, was included as a metric on how rapidly a watershed responds to precipitation or snowmelt events (Baker et al., 2004). Low flow conditions were characterized using (5) a baseflow index, calculated as the ratio of the average annual baseflow volumes to the average annual flow volumes (USFS, 2022), and (6) the average driest month discharge per area (DryMonth/area, Daigle et al., 2011) (Table 1). We explored shorter time scales (i.e., 7-day instead of 30-day) for MAX30/area and DryMonth/area but as similar patterns were documented between the two time periods, only the 30-day version was included. We also considered including signatures based on the coefficient of variation, but decided they were more challenging to interpret hydrologically, since variability could reflect episodic or seasonal variability. Signatures were either calculated to be unitless or divided by the drainage area (km^2) so that they could be compared across watersheds (Daigle et al., 2011). The distribution of hydrologic signature values was evaluated using the Shapiro-Wilk test for normality. Variables with extreme outliers were normalized using log10 transform (Beck et al., 2015) and included the flashiness index and wet season flashiness index.

Table 1. Hydrological signatures included in the analysis. MAX: maximum

Signature	Targeted flow regime	Calculation	Units	Median	Min	Max	Source
Flashiness index	All flows	The sum of the absolute value of the changes in discharge from the day prior to the current day (discharge t_2 – discharge t_1) divided by the sum of the daily discharge values (log normalized).	Unitless	-0.81	-1.63	0.23	(Baker et al., 2004)
Flashiness index (wet season)	High flows	The sum of the absolute value of the changes in discharge from the day prior in the three wettest months (highest discharge) divided by the sum of daily discharge values in those months (log normalized).	Unitless	-0.84	-1.89	0.23	(Baker et al., 2004)
MAX30/ area	High flows	The flow rate for the 30 days per year with the highest flow rate, summed over the 30 days, and averaged per year, divided by the watershed area.	m^3/sec/km^2	0.94	0.01	3.48	(Hannaford and Marsh, 2008)
(Q10-Q95)/area	High flows	Discharge exceeded 10% of the time (Q10) minus discharge exceeded 95% of the time (Q95), divided by watershed area.	m^3/sec/km^2	0.016	0.000	0.056	(National River Flow Archive, 2024)

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
Discussions

DryMonth/ area	Low flows	Average annual discharge in the driest month (excluding snow cover months) divided by watershed area.	$m^3/sec/km^2$	0.0019	0.0000	0.0112	(Daigle et al., 2011)
Baseflow index	Low flows	The ratio of the average daily flow during the lowest annual 7-day flow (excluding snow cover conditions) to the annual average daily flow.	Unitless	0.19	0.00	0.70	(USFS, 2022)

2.3 Dependence of hydrologic signatures on selected period

While the selected period was limited by the available Sentinel-1 and Sentinel-2 image record, signature uncertainty can increase when using shorter flow records (Kennard et al., 2010). To evaluate potential uncertainty in the hydrologic signature values based on the selected period of analysis, the signatures from our 8-year period (2016-2023) were contrasted with hydrologic signature values derived from a longer 24-year period (2000-2023), using Pearson correlation, generated using the Hmsic package in R, and relative bias. Between-site variability in the hydrologic signatures derived from the 8-year period, was highly correlated with the between-site variability from a longer, 24- year period (2000-2023) (Table 3). The median value of hydrologic signatures showed some differences between the 8-year period (2016-2023) and the longer 24-year period (2000-2023). While both flashiness indices had a bias of <1%, the MAX30/area and (Q10-Q95)/area had a relative bias of 13.5% and 8.7%, respectively, indicating that average peak wetness conditions were wetter within the 8-year period, relative to the longer period. Additionally, while the DryMonth/area bias was minimal, the baseflow index showed a relative bias of -11.8%, potentially reflecting a higher volume of water coming from high flows within the 8-year period, relative to the longer period (Table 3). While the hydrologic signatures of the high and low flow conditions were amplified during the selected period, the signature values between the two periods were highly correlated, with correlation values ranging from 0.94 to 0.99 (Table 2). This suggests that the relative variations in hydrologic signature values between the long-term flow records (24 years) compared to the study period (8 years) are tightly associated.

Table 2. Correlation values comparing the 2016-2023 hydrologic signatures with the same signatures derived from the 2000-2023 period. The relative bias compares the paired signature values from each watershed. All R values were significant at $p<0.01$. MAX: maximum

Metric	***R* (2016-2023 vs 2000-2023)**	**Median relative bias (%)**
Flashiness index	0.99	0.9
Flashiness index (wet season)	0.99	0.2
MAX30/area	0.97	13.5
(Q10-Q95)/area	0.98	8.7
DryMonth/area	0.94	-2.2
Baseflow index	0.95	-11.8

We also contextualized the study period's meteorological conditions using the GRIDMET 5-day Palmer Drought Severity Index values (PDSI; Abatzoglou, 2013). Specifically, we converted PDSI for 1980-2023 to a rank percentile, where 50% represents the median PDSI for the 1980-2023 period. We examined the minimum (i.e., driest), maximum (i.e., wettest) and median per watershed PDSI rank percentile that occurred between 2016-2023 to understand the range of PDSI conditions that this 8-year period represents. The 2016-2023 period averaged 5%, 100%, and 62%, for the minimum, maximum, and median PDSI conditions, respectively (Table A1). This indicated that the period was slightly wetter, on average, relative to the longer 44-year period, but that most watersheds exhibited a large range of PDSI conditions (maximum – minimum) over the 2016-2023 period.

2.4 Independent variables

2.4.1 Climate variables

Climate variables were averaged over the 2016-2023 period. Total annual, average precipitation and actual evapotranspiration (ET), using grass as the reference vegetation ("eto"), were derived from the daily University of Idaho Gridded Surface Meteorological Dataset (GRIDMET, 4 km resolution; Abatzoglou, 2013; Table 3). Water demand was correspondingly evaluated as annual precipitation – annual ET (Abatzoglou, 2013). An aridity index was calculated as the annual total potential evapotranspiration (PET) divided by the annual total precipitation (TerraClimate, 4.6 km; Abatzoglou et al., 2018; Fig. 1), where higher values represent arid watersheds and lower values represent less arid watersheds (Budyko, 1958). Since only approximately half of the watersheds experience snow, a snowmelt only variable, like snow water equivalent, was not included, and instead snowmelt was represented by a precipitation coefficient of variation (CV), precipitation seasonality, and maximum monthly (January-December) precipitation variables, calculated using DAYMET daily precipitation, which includes daily estimates of snow-water equivalent (Table 3). Seasonality was defined as the difference between average summer (June, July, August) and average winter (December, January, February). A Rainfall and Runoff Factor (RFACT), referred to as rainfall intensity, was included to reflect the long-term average of rainfall amount and peak intensity for storm events, and was derived from PRISM climate data (1971-2000) (Falcone, 2011). Maximum daily temperature was derived from DAYMET, which has been found to outperform GRIDMET for temperature (Mehdipoor et al., 2018), and variables included temperature seasonality as well as the maximum temperature CV. Both CV variables were calculated from a monthly time step. DAYMET variables relied on 2016-2022 data, as 2023 was not yet available at the time of the analysis.

2.4.2 Land cover, soils, topography, and wetland variables

Vegetation was represented by the 2019 National Land Cover Database (NLCD), as the proportion of each watershed classified as (1) forest (evergreen, deciduous, or mixed), (2) developed (low, medium, and high intensity) and (3) cultivated crops (Homer et al., 2020). Annual minimum depth to water table, depth to bedrock, geologic permeability, fraction clay, fraction silt, and fraction sand were derived from the Soil Survey Geographic Database (SSURGO; Falcone, 2011). To represent topography, the mean percent slope and elevation range divided by mean elevation were derived using the 10 m USGS Digital Elevation Model (DEM) (Table 3). The mean watershed topographic diversity was also considered, calculated from the multi-scale Topographic Position Index (mTPI) and the Continuous Heat-Insolation Load Index (CHILI, 30 m; Theobald et al., 2015). Stream density was calculated using the total stream length, defined by the NHDplus high resolution dataset (USGS, 2022). The National Wetland Inventory dataset (USFWS, 2019) was used to calculate the proportion of each watershed mapped as wetlands. The floodplain variable was defined as the proportion of each watershed classified as within the 100-year floodplain (Woznicki et al., 2019). Lastly, the connectivity of wetlands to streams can influence the timing of water moving into the stream network, so the proportion of each watershed mapped as geographically isolated wetlands (GIWs;

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

Leibowitz, 2003), or non-floodplain wetlands (NFW), that are surrounded by upland, as well as the proportion of total wetland area mapped as GIWs was considered (Lane and D'Amico, 2016).

2.4.3 Inundation variables

In addition to including static water variables, such as wetland area, remote sensing platforms allow us to include variables that characterize the hydroperiod of surface water stored within watersheds, including lakes, ponds, wetlands, and temporary inundation in flood prone areas. Although Landsat can provide a longer temporal record of surface water dynamics, observations are limited to periods free of clouds, snow, and ice, which can limit the accuracy of temporary and seasonal patterns of inundation. Alternatively, the more frequent Sentinel-2 revisit, and incorporation of a synthetic aperture radar (SAR) satellite, like Sentinel-1, can help bypass these limitations. Sentinel-1 and Sentinel-2 based algorithms that map non-water, open water and vegetated water were previously developed using gradient boosted classifier algorithms for 12 sites across the conterminous U.S. (20 m resolution; Vanderhoof et al., 2023). Details on the surface water algorithms can be found in Vanderhoof et al., (2023). In this effort individual Sentinel-1 and Sentinel-2 images, collected between January 1, 2016, and December 31, 2023, overlapping each of the gaged watersheds (n=72) were classified into open water, vegetated water, and non-water. The classified Sentinel-1 (S1) and classified Sentinel-2 (S2) time series were consolidated at a 14-day time step where pixel values were assigned as the majority classification, water (defined as open water plus vegetated water), or non-water (Fig. 2). If observations of water and non-water were equal, then open water was prioritized followed by non-water, and lastly vegetated water (Fig. 2), consistent with the higher accuracy of the open water class relative to the vegetated water class (Vanderhoof et al., 2023). Where no valid observations were present in the 14-day period, pixels were gap-filled using observations from the t-1 and t+1 timestep, as shown in Fig. 2.

To limit commission error in the surface water time series, a water mask, defined as the maximum allowable surface water extent, was derived for each watershed, and applied across the time series. To generate each water mask, the Sentinel-1 open water and vegetated water, and Sentinel-2 open water, and vegetated water percentile rasters were manually reviewed for each watershed (Fig. 2). Percentile thresholds were selected, below which the frequency of erroneously classified water pixels visually exceeded the frequency of correctly classified water pixels (Table A2). To help inform the threshold selection, ancillary data were used including the NWI dataset (USFWS, 2019), the 2019 NLCD (Homer et al., 2020), and base map imagery, delivered through ArcMap. The spatial extent where water pixels were retained was defined as pixels located within the 100-year floodplain (Woznicki et al., 2019), to account for short-term flood events, or pixels where the water percentile was greater than the selected threshold in any of the four 5-year percentile rasters (Table A2). Pixels classified as water outside of the water mask were re-classified as non-water. The Sentinel-1 algorithm has a documented omission and commission error of 3.1% and 0.9% for open water, and a 28.4% and 16.0% commission error for vegetated water, respectively, while the Sentinel-2 algorithm has an omission and commission error of 3.1% and 0.5% for open water, and a 10.7% and 7.9% commission error for vegetated water, respectively, when validated against 36 high-resolution images (i.e., WorldView-2, WorldView-3, PlanetScope) (Vanderhoof et al., 2023). When consolidated at a monthly time-step to a S1-S2 water, non-water classification, errors of omission and commission for monthly surface water extent averaged 1.6% and 10.4%, respectively, when validated against 64 PlanetScope images (Vanderhoof et al., 2024). The use of a water mask was

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences Discussions

previously shown to reduce commission error, resulting in errors of omission and commission of 1.9% and 6.5%, respectively for the monthly surface water extent (Vanderhoof et al., 2024).

After gap-filling and applying the water masks, the time series for each watershed was then consolidated into an 8-year percentile. Categories of surface water, using the percent of watershed area, were defined in reference to the 100-year floodplain (Woznicki et al., 2019), and included, (1) temporarily flooded, defined as an average of ≥3 days but <1 month per year (Cowardin et al., 1979; Scott et al., 2019), (2) seasonally flooded, defined as inundated >1 month but <6 months per year, on average, and (3) semi-permanently and permanently inundated, defined as >6 months per year, on average (Cowardin et al., 1979; Donnelly et al., 2019) (Table 3). The total amount of inundation of any hydroperiod within the 100-year floodplain, and outside of the 100-year floodplain was also included, as was the proportion of inundation that was seasonal (Table 3). Examples of variability in inundation patterns between watersheds are shown in Fig. 3. The terms surface water extent and inundation are used interchangeably in this analysis.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Table 3. Independent variables considered modelling hydrological signatures. DEM: Digital elevation model, SRTM: Shuttle Radar Topography Mission, NLCD: National Land Cover Database, SSURGO: Soil Survey Geographic Database, NHD: National Hydrography Dataset, CV: coefficient of variation, USFWS: U.S. Fish and Wildlife Service

Variable Type	Variable	Units	Min	Max	Median	Source
Climate	Precipitation (P, annual)	mm	325.3	1659.1	967.4	GRIDMET (Abatzoglou, 2013)
	Evapotranspiration (ET, annual)	mm	714	1934.1	1181.1	GRIDMET (Abatzoglou, 2013)
	Aridity index (PET/P, annual)	unitless	0.8	6.63	1.21	TerraClimate (Abatzoglou et al., 2018)
	Water demand (P - ET, annual)	mm	-1586	265.6	-247.4	GRIDMET (Abatzoglou, 2013)
	Precipitation seasonality	mm	-396	276.6	105	DAYMET (Thornton et al., 2020)
	Precipitation CV	unitless	0.41	1.31	0.64	DAYMET (Thornton et al., 2020)
	Rainfall intensity	~	12.1	412.5	139.2	SSURGO (Falcone, 2011)
	Maximum monthly precipitation	mm	53.9	230.8	131.6	DAYMET (Thornton et al., 2020)
	Temperature seasonality	ºC	15.6	34.2	23	DAYMET (Thornton et al., 2020)
	Temperature CV	unitless	0.23	1.3	0.48	DAYMET (Thornton et al., 2020)
Land Cover	Forest (evergreen, deciduous, mixed)	% of area	0.059	56.1	17.5	NLCD (2019; Homer et al., 2020)
	Developed (low, medium, high intensity, open space)	% of area	0.323	35.7	4.69	NLCD (2019; Homer et al., 2020)
	Cultivated crops	% of area	0.0	84.7	17.9	NLCD (2019; Homer et al., 2020)
	Stream density	$m\ km^{2}$	259.2	4181.6	1460.9	NHDPlus High Res. (USGS, 2022)
Soil and Geology	Clay fraction	fraction	0.08	0.47	0.23	SSURGO (Falcone, 2011)
	Sand fraction	fraction	0.07	0.74	0.33	SSURGO (Falcone, 2011)
	Silt fraction	fraction	0.17	0.72	0.44	SSURGO (Falcone, 2011)
	Depth to bedrock	cm	81.3	152.4	145.8	SSURGO (Falcone, 2011)
	Annual min depth to water table	meters	0.49	1.83	1.40	SSURGO (Falcone, 2011)
	Geological permeability	$cm\ day^{-1}$	0.5	8.7	2.2	SSURGO (Falcone, 2011)
Topography	Slope	%	0.5	32.5	3.7	DEM (Gesch et al., 2002)
	$(Elevation_{max} - Elevation_{min}) / Elevation_{average}$	unitless	0.2	4.9	1.0	DEM (Gesch et al., 2002)
	Global SRTM topographic diversity	unitless	0.03	0.7	0.1	(Theobald et al., 2015)
Inundation Dynamics	Temporarily flooded, floodplain (3 days - 1 month)	% of area	0.07	4.16	0.65	(Vanderhoof et al., 2023)
	Temporarily inundated, non-floodplain (3 days - 1 month)	% of area	0.03	5..85	1.29	(Vanderhoof et al., 2023)
	Seasonally inundated, floodplain (1 - 6 month)	% of area	0.04	8.58	1.77	(Vanderhoof et al., 2023)
	Seasonally inundated, non-floodplain (1 - 6 month)	% of area	0.01	45.81	4.07	(Vanderhoof et al., 2023)
	Semi-permanently and permanently inundated, floodplain (>6 month)	% of area	0	3.54	0.39	(Vanderhoof et al., 2023)
	Semi-permanently and permanently inundated, non-floodplain (>6 month)	% of area	0	5.55	0.44	(Vanderhoof et al., 2023)
	Total floodplain inundation	% of area	0.42	15.46	3.08	(Vanderhoof et al., 2023)
	Total non-floodplain inundation	% of area	0.04	52.59	6.06	(Vanderhoof et al., 2023)
	Proportion of inundation that is seasonally inundated, floodplain (1 - 6 months)	% of inundation	3.15	53.56	18.34	(Vanderhoof et al., 2023)
	Proportion of inundation that is seasonally inundated, non-floodplain (1 - 6 months)	% of inundation	2.16	77.44	39.64	(Vanderhoof et al., 2023)
Wetland	Geographically Isolated Wetlands (GIW)	% of area	0.0	9.4	0.6	(Lane and D'Amico 2016)
	Proportion of wetland area identified as GIW	% of area	0.6	80.9	11.4	(Lane and D'Amico 2016; USFWS 2019)
	Floodplain	% of area	1.2	36.8	7.7	(Woznicki, et al., 2019)
	National Wetland Inventory (NWI) wetlands	% of area	1.1	48.7	5.6	NWI (USFWS 2019)

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

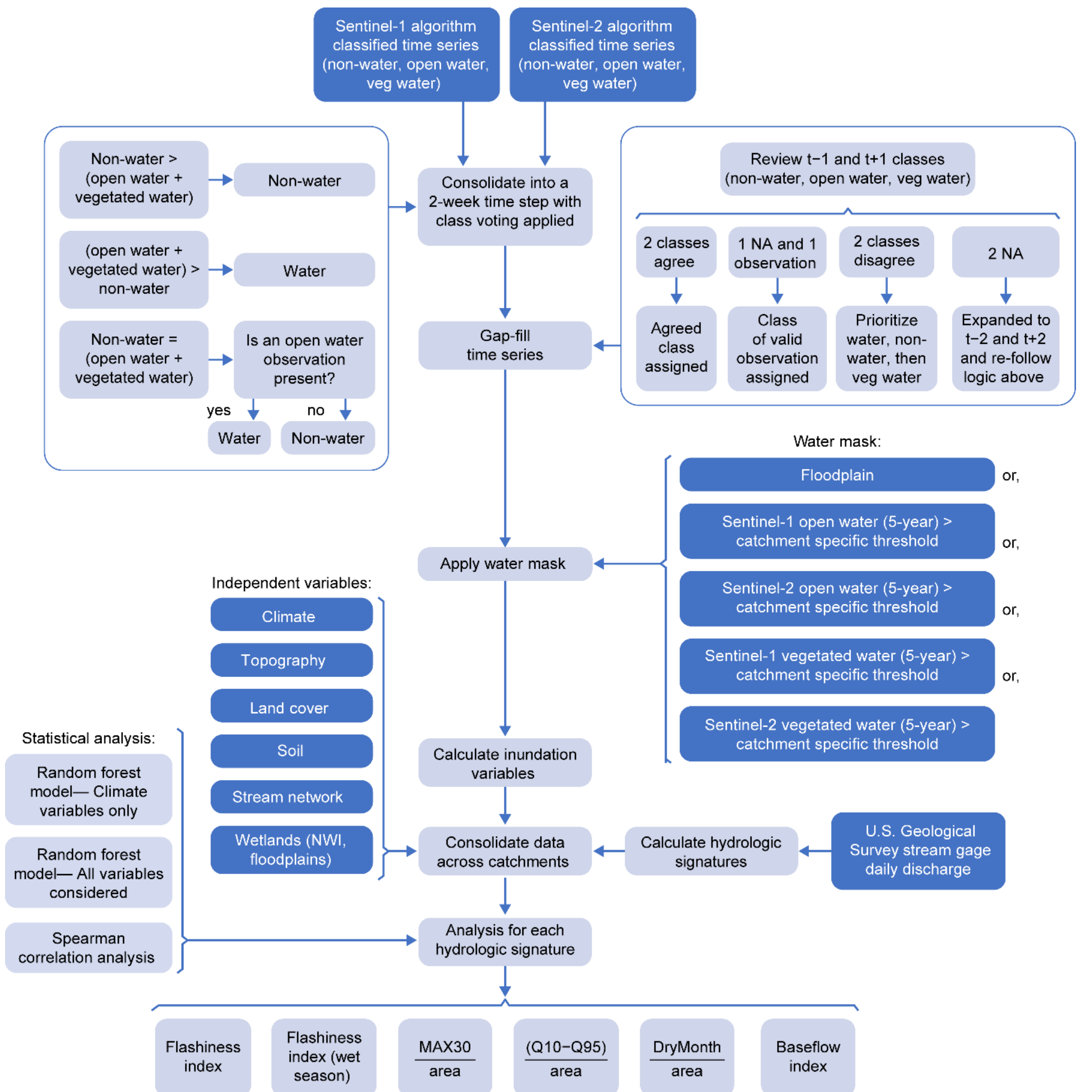


Figure 2. Flowchart of steps to generate the surface water variables and data analysis.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

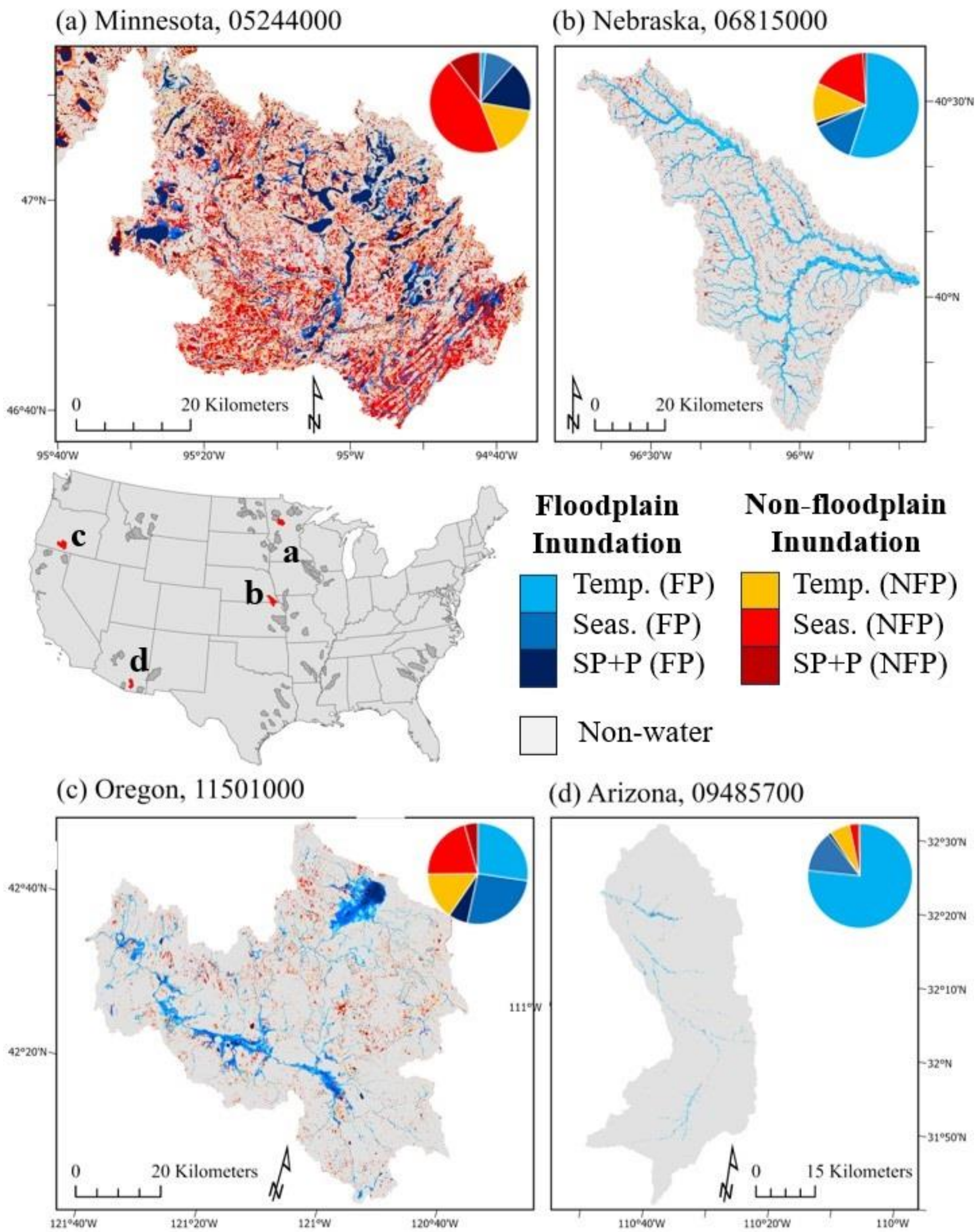


Figure 3. Examples of between watershed variability in the abundance of inundation variables, with the relative distribution of inundation variables shown with pie charts, including (a) MN5, (b) NE1, (c) OR1, (d) AZ2, where the numbers indicate the gage number. Temp: temporary, Seas: seasonal, SP+P: semi-permanent to permanent, FP: floodplain, NFP: non-floodplain

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences Discussions

2.5 Modelling analysis

The relationships between multiple predictor variables and hydrologic signatures were modelled with random forest regressions developing using the scikit-learn python package (Pedregosa et al., 2011). For each hydrologic signature, random forest models were generated that (1) considered the inclusion of climate-related variables only ($M_{Climate}$), and (2) considered inclusion of all variables, including climate, topographic, land cover, and wetland and inundation related variables (M_{All}) (Table 2). The multi-model approach furthered our ability to quantify the relative contribution of different variable types to explain variability in the hydrologic signatures.

Random forest models use a bootstrapping approach to generate hundreds of regression trees and make no prior assumptions about cause-and-effect relationships or correlations among variables (Hastie et al., 2009). They have also been previously used in the analysis of hydrologic signatures (e.g., Trancoso et al., 2016; Addor et al., 2018; Oppel and Schumann, 2020). While random forest techniques are generally insensitive to multicollinearity, the inclusion of highly correlated variables can make it more challenging to identify the most predictive variables, deflate or bias variable importance values, and complicate model interpretation (Murphy et al., 2010; Gregorutti et al., 2016). Conversely, an automated variable selection can be indicative of the relative importance of certain variables over others (Murphy et al., 2010). A stepwise forward selection routine was implemented where the set of potential predictors were sequentially tested. The predictor that contributed most to reducing the RMSE was selected. During each step, the remaining predictors were removed if they had a correlation value of 0.85 or greater with any of the selected predictors. This process was iterated until the improvement in the model's RMSE was <0.001 with any additional variables (Sherrouse and Hawbaker, 2023).

For each model the variable and hyperparameter selection process were concurrently run, where the potential models were compared using a nested cross-validation, KFold with 6 splits (Cawley and Talbot, 2010). The hyperparameters tested were n_estimators (the number of trees in the forest with tested values of 300, 500, 700, and 1000), max_depth (the maximum depth of a tree with tested values of 2, 3, and 4). For all models, max_features (the number of features to consider when looking for the best split) was set at the square root of the number of features, and max_samples (the proportion of samples selected to train each estimator) was set at 0.8. The model with the highest cross-validated adjusted R^2 was selected.

Random forest models do not consider the spatial pattern between samples, therefore any clustering of the watersheds included in the analysis could potentially bias model predictions (Hengl et al., 2018). The residuals of each selected model were tested for spatial autocorrelation using Moran's I (Klute et al., 2002). Of the random forest model residuals, only 1 of 12 showed significant (p<0.01) spatial autocorrelation, the (Q10-Q95)/area (M_{All}). An autocovariate, or additional model term, representing the mean neighborhood (defined as within 500 km of the catchment center, reflecting catchment clusters) model residual value, was included in this model to account for spatial dependency (Betts et al., 2006). Performance of final random forest models was evaluated using the leave-one-out cross validation to account for the limited sample size (n=72) (Vabalas et al., 2019), and the cross-validated model RMSE, R^2, and adjusted R^2, to account for differences in the number of variables selected. The Mean Square Error (MSE) and Akaike information criterion (AIC) were also calculated from the observed and model predicted values,

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences Discussions

where decreases in both values indicate model improvement (Portet, 2020). Variable importance was calculated with Python Scikit-learn as the permutation importance. Significance of model selected variables and the corresponding decrease in MSE with the exclusion of each variable was calculated using the rfPermute package in R using 100 repetitions. Single variable correlations between the hydrologic signatures and the predictor variables were calculated using the non-parametric Spearman Rank Correlation Coefficient, generated in R using the Hmisc package. Because of the number of comparisons, a Bonferroni correction was applied before significance was determined (Emerson, 2020).

3. Results

3.1 Flashiness signatures

The flashiness and wet season flashiness signatures reflect how quickly discharge changes in response to episodic rainfall and snowmelt events, over the course of the year and within the wet season, respectively. Despite representing different portions of the year, the two signatures were highly correlated ($R = 0.97$, $p<0.01$). Flashiness and wet season flashiness were highest, on average, in the Southwest watersheds, and lowest in the West and North Central watersheds (Table A3, Fig. 4). Watershed flashiness and wet season flashiness were significantly correlated with very few of the independent variables considered. Most prominently, both significantly ($p<0.01$) decreased with an increased area mapped as semi-permanently and permanently inundated within the floodplain, and with increased total area classified as wetland by the NWI dataset (Table 4). Correlations with climate variables were weaker relative to the other hydrologic signatures explored. The flashiness index and wet season flashiness index M_{All} models saw improvement in explanatory power and associated decreases in the RMSE, MSE and AIC relative to $M_{Climate}$, or when landscape and water variables were added for consideration (Table 5). Adjusted R^2, for example improved by 11.57% and 8.72% for the flashiness and wet season flashiness, respectively, while MSE decreased by 10.4% and 5.13%, respectively (Table 6). Variability in the flashiness signature was best explained by the evapotranspiration and the amount of semi-permanent-permanent (SP+P) inundation within the floodplain. In the wet season flashiness $M_{All,}$ model the amount of SP+P floodplain inundation had the greatest variable importance (Table 7; Fig. 5). Lower flashiness was associated with greater SP+P inundation within the floodplain, both across the year as well as in the wet season (Fig. 6a) and showed strong variable importance (Table 7) over both time periods. Table 6 consolidates the information on the role of inundation variables. For the flashiness signatures, both signatures had an inundation variable selected, which was significant ($p<0.01$), and their potential exclusion had a projected increase in MSE of 18% (Table 6). Both flashiness signatures saw consistent improvement in model performance, across metrics, for M_{All} relative to $M_{Climate}$, although these improvements were minor to moderate (3.99% to 11.69%), and lastly, the inundation variable selected for inclusion in M_{All} was consistent with the inundation variables that were significantly ($p<0.01$) correlated with the flashiness signatures (Table 6).

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

3.2 Peak flow signatures

The peak flow signatures, MAX30/area and (Q10-Q95)/area, were highest, on average, within the Gulf Coast watersheds, and lower, on average, within the Southwest, North Central, and West watersheds, although both signatures saw a higher degree of variability across the West region (Table A3, Fig. 4). The two signatures were positively correlated (R = 0.93, p<0.01). In relation to the independent variables considered, both signatures, MAX30/area and (Q10-Q95)/area, were most highly positively correlated with precipitation and water demand (P-ET), and negatively correlated with aridity (Table 4). The MAX30/area and (Q10-Q95)/area were also both significantly correlated with four of the remotely sensed inundation variables. An example of the correlation of (Q10-Q95)/area in relation to seasonally inundated area in the floodplain (R=0.69, p<0.01) is shown in Fig. 5. The high flow signatures had a positive, significant (p<0.01) correlation with the total amount of inundation within the floodplain, the amount of seasonal inundation in the floodplain, and the amount of temporary inundation outside of the floodplain (Table 4). These correlation values were equivalent to or exceeded correlation with existing water variables, specifically the 100-year floodplain (Table 4). The MAX30/area M_{All} model was best explained by the maximum monthly precipitation and aridity index, while the (Q10-Q95)/area M_{All} model was best explained by annual precipitation and water demand (Table 7). Despite the high explanatory power of climate variables for both high flow signatures, the M_{All} models still showed some minor improvement with the adjusted R^2 improving by 5.44% and 1.75% and the RMSE decreasing by 6.44% and 5.22% relative to the $M_{Climate}$ models, for MAX30/area and (Q10-Q95)/area, respectively (Table 6). The (Q10-Q95)/area M_{All} model added stream density and the proportion of inundation that was seasonally inundated and occurred within the floodplain. The landscape-based variables added for MAX30/area included the amount of seasonally inundated area on the floodplain, stream density, and geologic permeability (Table 7). The inundation variables were both found to be significant, and their potential exclusion had a projected increase in model MSE of 15.44% and 8.32% for MAX30/area and (Q10-Q95)/area, respectively (Table 6). Further, like the flashiness signatures, the selected inundation variables were consistent with the inundation variables identified as significant in the correlation analysis (Table 6).

3.3 Low flow signatures

The DryMonth/area and baseflow index were highest within the East watersheds, on average, and lowest within the Southwest watersheds (Table A3, Fig. 4). Watersheds were also regionally variable. For example, the DryMonth/area signature graded west (lower) to east (higher) within the North Central region (Fig. 4), concurrent with the aridity gradient within the region (Fig. 1). The two low flow signatures had a significant, but weaker correlation with one another (R = 071, p<0.01). The DryMonth/area was significantly correlated with many more independent variables than the baseflow index. Like the peak flow signatures, DryMonth/area was positively correlated with greater annual precipitation and water demand (P-ET) and negatively correlated with greater aridity. The DryMonth/area was also positively correlated with total inundation within the floodplain, seasonally inundated area within the floodplain, and temporarily inundated area outside of the floodplain. No significant correlations for DryMonth/area, in contrast, were found with topographic or wetland variables (Table 4). The DryMonth/area had the greatest model explanatory power, relative to the other hydrologic signature models (Table 5, 6). However, despite

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

significant (p<0.01) correlations with remotely sensed inundation dynamics, there was no model improvement as landscape variables were added between the $M_{Climate}$ and M_{All} models (Table 5). The DryMonth/area was best explained by watershed aridity and the precipitation CV.

The baseflow index was negatively significantly (p<0.01) correlated with precipitation CV, evapotranspiration, and fraction of clay (Table 4). Adding landscape variables, unlike DryMonth/area, showed a limited amount of improvement with the adjusted R^2, RMSE and MSE improving by 2.95%, 2.90%, and 3.43%, respectively (Table 6) and improved the relationship between the observed and predicted baseflow index values (Fig. 6b). While the precipitation CV was the most important variable in both the baseflow index M_{All} and $M_{Climate}$ models, the M_{All} model's improvement was attributable to the inclusion of stream density, clay fraction, and the amount of non-floodplain area classified as semi-permanent to permanent (i.e., large wetlands and lakes outside of the floodplain) (Table 7). The selected inundation variable was significant (p<0.01) within the model and had a projected 15.47% increase in MSE with the potential exclusion of the variable within the M_{All} model (Table 6).

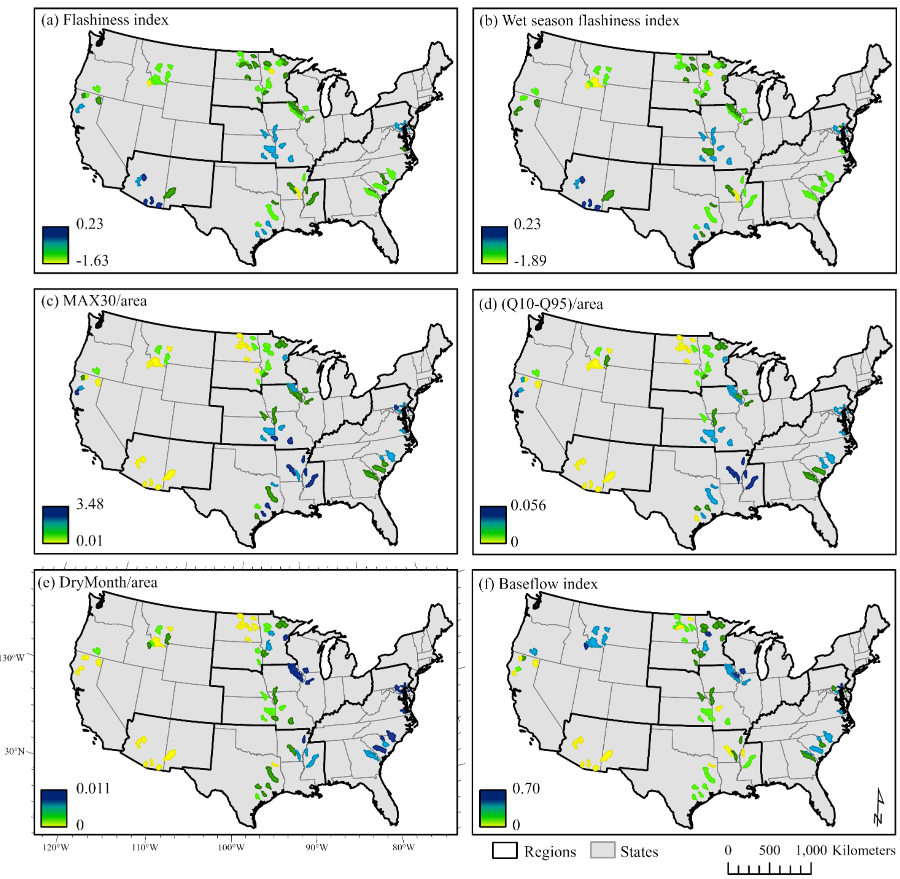


Figure 4. Hydrological signature values by watershed including (a) flashiness index, (b) wet season flashiness index, (c) MAX30/area ($m^3/sec/km^2$), (d) (Q10-Q95)/area ($m^3/sec/km^2$), (e) DryMonth/area ($m^3/sec/km^2$), and (f) baseflow index. Greater flashiness (a, b), higher peak flows (c, d), and greater flows during low flow periods (e, f) are shown in blue.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

423 **Table 4**. Correlation values between hydrologic signatures and variables. Significant (*p*<0.01) correlations, after
424 Bonferroni correction was applied, are shown in shaded gray. CV: coefficient of variation, FP: floodplain, NFP: non-
425 floodplain, Prop: proportion, MAX: maximum, SP+P: semi-permanent and permanent, Proport.: proportion, GIW:
426 geographically isolated wetlands, NWI: National Wetland Inventory

Variable Type	Variable	Flashiness Index	Flashiness (wet season)	MAX 30/area	(Q10-Q95)/area	DryMonth /area	Baseflow index
Climate	Precipitation	0.06	0.01	0.86	0.87	0.68	0.16
	Evapotranspiration	0.43	0.32	0.18	0.14	-0.15	-0.47
	Aridity index	-0.02	-0.04	-0.77	-0.80	-0.86	-0.42
	Water demand	-0.03	-0.02	0.78	0.83	0.82	0.41
	Precipitation seasonality	0.17	0.26	0.01	-0.04	0.21	0.20
	Precipitation CV	0.18	0.17	-0.59	-0.65	-0.85	-0.59
	Rainfall intensity	0.20	0.16	0.78	0.75	0.55	-0.02
	Max monthly precipitation	0.30	0.25	0.85	0.82	0.48	-0.10
	Temperature seasonality	-0.29	-0.18	-0.30	-0.28	-0.05	0.24
	Temperature CV	-0.38	-0.29	-0.35	-0.30	-0.08	0.28
Land cover	Forest	-0.14	-0.17	0.28	0.32	0.18	0.15
	Developed	0.22	0.18	0.62	0.60	0.62	0.18
	Cultivated crops	-0.16	-0.13	0.03	0.06	0.30	0.27
	Stream density	0.36	0.29	0.37	0.35	-0.06	-0.33
Soil and Geology	Clay fraction	0.40	0.37	0.25	0.15	-0.12	-0.44
	Sand fraction	-0.23	-0.27	-0.32	-0.26	-0.07	0.16
	Silt fraction	0.04	0.10	0.18	0.17	0.12	0.06
	Depth to bedrock	-0.29	-0.30	0.14	0.18	0.32	0.20
	Water table depth	0.12	0.13	-0.54	-0.55	-0.45	-0.09
	Geological permeability	-0.42	-0.40	-0.25	-0.20	0.16	0.43
Topography	Slope	0.13	0.13	-0.23	-0.22	-0.27	0.00
	Elevation range	0.12	0.03	0.24	0.24	0.14	-0.04
	Topographic diversity	0.11	0.11	-0.17	-0.15	-0.20	0.04
Inundation Dynamics	Temporarily flooded, FP	0.27	0.23	0.42	0.40	0.24	-0.05
	Temporarily inundated, NFP	-0.06	-0.03	0.49	0.51	0.58	0.30
	Seasonally inundated, FP	-0.12	-0.15	0.66	0.69	0.59	0.15
	Seasonally inundated, NFP	-0.21	-0.19	0.36	0.37	0.39	0.14
	SP+P inundated, FP	-0.44	-0.46	0.24	0.33	0.33	0.14
	SP+P, inundated, NFP	-0.34	-0.32	0.13	0.11	0.13	0.04
	Total inundation, FP	-0.12	-0.15	0.60	0.63	0.52	0.12
	Total inundation, NFP	-0.19	-0.17	0.37	0.37	0.41	0.17
	Proport. Seasonally inundated, FP	-0.17	-0.20	0.41	0.46	0.36	0.17
	Proport. Seasonally inundated, NFP	-0.20	-0.16	0.06	0.04	0.11	0.09
Wetland	GIW	-0.31	-0.29	0.07	0.08	0.13	0.04
	Prop. of wetland area that is GIW	-0.08	-0.05	0.08	0.03	0.06	0.01
	Floodplain	-0.02	-0.07	0.49	0.51	0.39	0.00
	NWI wetlands	-0.44	-0.44	0.12	0.19	0.28	0.19

427
428

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Table 5. Model statistics for each hydrologic signature and version of the model including (1) climate variables only ($M_{Climate}$) and (2) all variables including wetland and surface water variables (M_{All}). All models were significant at p<0.0001. RMSE: root mean square error, adj: adjusted, MSE: mean square error, AIC: Akaike information criterion, MAX: maximum

Signature	Model	R^2	R^2 adj.	RMSE	MSE	AIC	Trees	Max. tree depth	Variable count
Flashiness index	$M_{Climate}$	0.54	0.51	0.242	0.191	-196.04	700	4	4
	M_{All}	0.61	0.57	0.225	0.172	-202.74	700	4	5
Flashiness index (wet season)	$M_{Climate}$	0.47	0.43	0.273	0.209	-178.78	500	4	4
	M_{All}	0.51	0.47	0.262	0.198	-182.64	300	4	5
MAX30/ area	$M_{Climate}$	0.65	0.63	0.475	0.327	-101.11	1000	4	3
	M_{All}	0.69	0.66	0.445	0.311	-106.70	700	4	5
(Q10-Q95)/area	$M_{Climate}$	0.76	0.74	0.007	0.005	-712.87	1000	3	3
	M_{All}	0.78	0.76	0.006	0.005	-714.59	700	4	6
DryMonth/ area	$M_{Climate}$	0.80	0.78	0.001	0.001	-952.93	700	4	5
	M_{All}	0.80	0.78	0.001	0.001	-952.93	700	4	5
Baseflow index	$M_{Climate}$	0.60	0.57	0.114	0.085	-306.80	700	4	4
	M_{All}	0.62	0.59	0.111	0.082	-307.04	700	4	5

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
Discussions

Table 6. Difference in model performance between $M_{Climate}$, model in which only climate variables were considered, and M_{All}, model where all variables were considered, where positive values for R^2, and R^2 adjusted indicate model improvement, and negative values for RMSE (root mean square error), MSE (mean square error), and AIC (Akaike information criterion) indicate model improvement. Inundation variables selected as well as significant correlations with inundation variables are also shown. Chg: change, SP+P: semi-permanent and permanent inundation, Seas: seasonally inundated, Temp: temporary inundation, FP: floodplain, NFP: non-floodplain, Prop: proportion of inundation, FP (%): 100-year floodplain, NWI: National Wetland Inventory

Metric	**Flashiness index**	**Flashiness index (wet season)**	**MAX30/ area**	**(Q10-Q95)/area**	**DryMonth/ area**	**Baseflow index**
Comparison of M_{All} to $M_{Climate}$ model metrics						
R^2 (% chg)	11.57	8.72	6.78	3.28	0.00	3.77
R^2 adjusted (% chg)	11.69	7.96	5.44	1.75	0.00	2.95
RMSE (% chg)	-7.16	-3.99	-6.44	-5.22	0.00	-2.90
MSE (% chg)	-10.40	-5.13	-5.07	-0.58	0.00	-3.43
AIC (chg)	-6.69	-3.86	-5.59	-1.72	0.00	-0.24
Inundation variables selected for M_{All} model and their significance						
Selected inundation variables	SP+P (FP)	SP+P (FP)	Seas (FP)	Prop Seas (FP)	~	SP+P (NFP)
Increase in MSE with inundation variable exclusion (%) (*p*-value)	18.61% (p<0.01)	18.81% (p<0.01)	15.44% (p<0.01)	8.32% (p<0.01)	~	15.47% (p<0.01)
Significant correlations (with Bonferonni correction) between signatures and inundation variables						
Positive correlations (p<0.01)	~	~	Temp (NFP), Seas (FP), FP (%)	Prop Seas (FP), Inun (FP), Temp (NFP), Seas (FP), FP (%)	Inun (FP), Temp (NFP), Seas (FP)	~
Negative correlations (p<0.01)	SP+P (FP), NWI (%)	SP+P (FP), NWI (%)	~	~	~	~

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Table 7. Variable permutation importance of variables selected for $M_{Climate}$: model in which only climate variables were considered, and M_{All}: all variables were considered. CV: coefficient of variation, min.: minimum, FP: floodplain, NFP: non-floodplain, SP+P: semi-permanent and permanent, Q10 and Q95: discharge at 10th and 95th percentiles, MAX: maximum

Variable Type	Variable	Flashiness index		Flashiness index (wet season)		MAX30/area		(Q10-Q95)/area		DryMonth/area		Baseflow index	
		$M_{Climate}$	M_{All}	$M_{Climate}$	M_{All}	$M_{Climate}$	M_{All}	$M_{Climate}$	M_{All}	$M_{Climate}$	M_{All}	$M_{Climate}$	M_{All}
Climate	Precipitation			0.23		0.44		0.44	0.38				
	Evapo-transpiration	0.41	0.31	0.34	0.24							0.26	0.18
	Aridity index	0.23					0.25			0.39	0.39		
	Water demand					0.33		0.38	0.30				
	Precipitation seasonality	0.15	0.11	0.21				0.18	0.09	0.07	0.07	0.13	
	Precipitation CV		0.18	0.21						0.29	0.29	0.40	0.30
	Rainfall intensity	0.21	0.15		0.18					0.19	0.19	0.21	
	Max Month Precip						0.44						
	Temperature seasonality									0.05	0.05		
	Temperature CV				0.13	0.23							
Land Cover	Stream density						0.12		0.14				0.19
Soil and Geology	Clay fraction												0.18
	Silt fraction												
	Water table depth				0.19								
	Geologic permeability						0.05						
Inundation Dynamics	SP+P inundated, FP		0.24		0.26								
	SP+P inundated, NFP												0.16
	Seasonally inundated, FP						0.14						
	Prop. Seasonally inundated, FP								0.09				
Other	Residual autocovariate								0.00				
	Color Legend:	(0-25%)		(26-50%)		(51-75%)		(76-100%)					

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

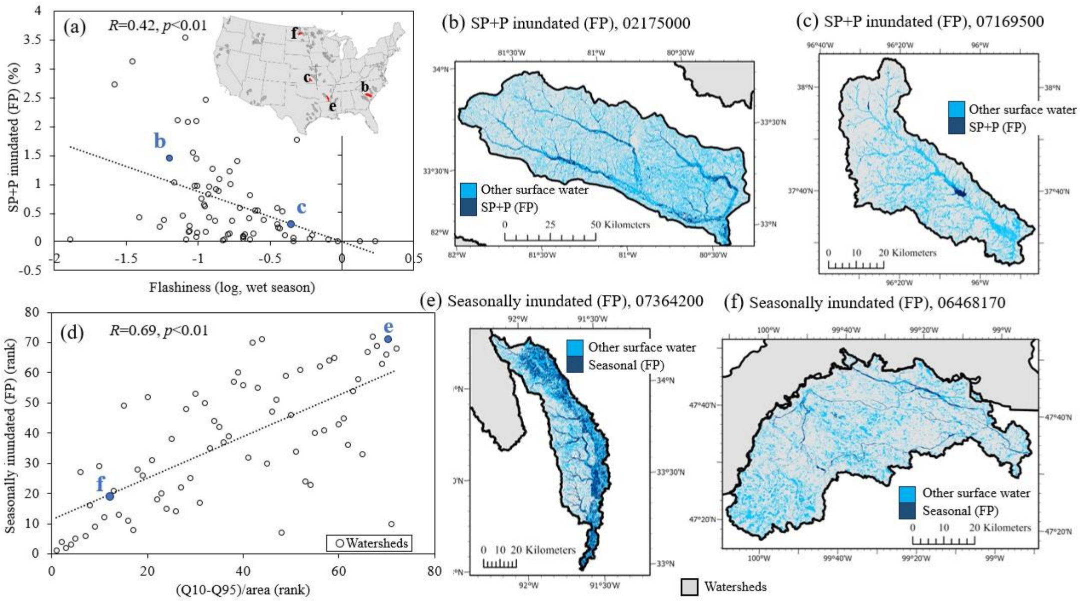


Figure 5. Scatter plot of (a) wet season flashiness versus the percent of semi-permanent and permanent (SP+P) floodplain (FP) inundation, which was included in the M_{All}, with corresponding examples (b, c) and (d) (Q10-Q95)/area in relation to the percent of seasonally inundated area (FP), with corresponding examples (e, f). To match the Spearman correlation analysis, both variables in panel d were converted to rank.

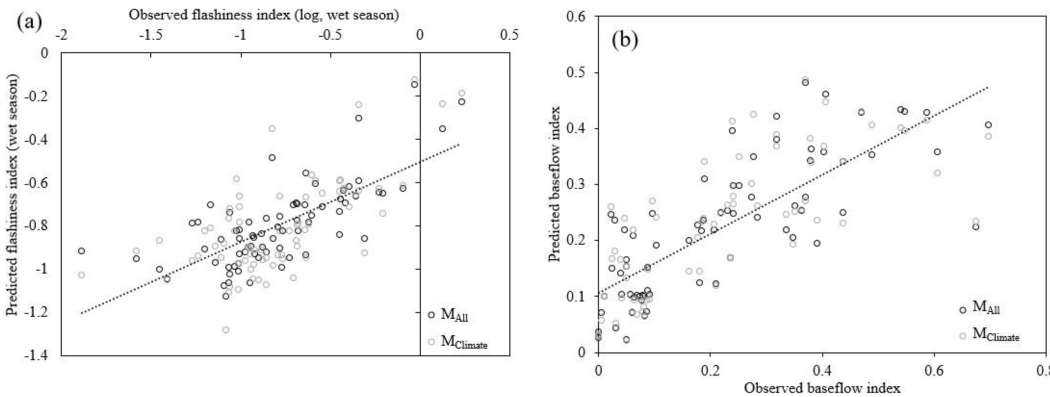


Figure 6. Scatter plots showing observed versus predicted with the $M_{climate}$ and M_{all} models for (a) flashiness (wet season, unitless) and (b) baseflow index (unitless).

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

4. Discussion

4.1 Contributions of climate and inundation variables to model hydrologic signatures

Hydrologic signatures can facilitate rapid comparison of river hydrology between diverse watersheds (McMillan, 2019). Climate variables often provide the highest predictive power for many hydrologic signatures (Beck et al., 2015; McMillan et al., 2021). Similarly, we found climate variables had the greatest variable importance in five of the six M_{All} models, with annual precipitation, evapotranspiration, aridity, and water demand dominating variable importance across the signatures tested (Table 7). Processes generating discharge are variable across the United States (Berghujs et al., 2016), but consistent with the variables selected in our analysis, rainfall, and for more northern watersheds, snowmelt (Jiang et al., 2022), as well as aridity (Sauqet et al., 2021) commonly account for variability in discharge. However, there are still opportunities to incorporate new watershed descriptors that may improve the characterization of flow signatures (Gnann et al., 2020). Specifically, McMillan et al. (2021) argued that novel relationships may be discovered where hydrology is more important than climate. For example, flood signatures have been predicted using watershed drainage patterns (Oppel & Schumann, 2020), and surface waterbodies have been found to help predict baseflow signatures (Beck et al., 2013). More generally, the influence of a watershed's landscape, including vegetation type (Trancoso et al., 2016; Addor et al., 2018), topography (Beck et al., 2015, and geology (Kuentz et al., 2017), on discharge has been well established. Likewise, in our analysis five of the six hydrologic signatures, showed an improvement in the M_{All} model performance, relative to relying on climate variables alone. Improvement, however, was moderate at best, with improvement in model metrics ranging from 0.24% to 11.57% (Table 6), making it difficult to interpret from model performance, alone, if the contributions were meaningful for hydrological processes.

More convincing was that inundation variables were selected for inclusion and were found to be significant in all hydrologic signature M_{All} models except DryMonth/area. The variables selected for inclusion were also highly consistent with the inundation variables that showed significant correlations with the corresponding signatures (Table 6), but further, and most importantly, the selected inundation variables were also consistent with our current understanding of watershed hydrology. The flashiness signatures, which reflect the rate that streamflow rises and falls in response to high rainfall and snowmelt events (Hannaford and March, 2008), selected the amount of SP+P inundation in the floodplain as the most and second most important variable for the wet season and all season flashiness, respectively (Table 7), where reduced flashiness was associated with greater SP+P floodplain inundation. This finding is consistent with stream-connected lakes or large wetlands moderating peak flows (Kuppel et al., 2015; Fritz et al., 2018). The signatures, MAX30/area and (Q10-Q95)/area, representing peak flows, relied instead on the amount and proportion of seasonal inundation within the floodplain, respectively. Greater peak flows in our analysis were positively correlated with greater seasonal floodplain inundation, consistent with our understanding of seasonal flooding coinciding with peak flow conditions (Blanchette et al., 2019; Wohl, 2021). However, the association of greater seasonal flooding with greater peak flows in our analysis does not allow us to understand how seasonal flooding extent may change the timing or magnitude of peak discharge conditions. Lastly, the baseflow index selected the amount of SP+P inundation outside of the floodplain. Differences in specific yield between uplands and non-floodplain wetlands leads to frequent reversals in hydraulic gradients, meaning that non-floodplain wetlands can

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

act as both groundwater sinks and sources (McLaughlin et al., 2014), contributing to baseflow (Evenson et al., 2015) and stabilizing low flow conditions (Ameli and Creed, 2017; Blanchette et al., 2019). Model selection of remotely sensed inundation dynamic variables over existing wetland and floodplain dataset variables suggests that consideration of surface water hydroperiod, alongside landscape position, was more helpful in explaining variability in hydrologic signatures then static datasets representing the spatial extent of wetlands (e.g., NWI, GIW) and floodplains (e.g., 100-year floodplain).

However, even when incorporating novel, remotely sensed inundation data, characterizing the potential influence of surface water storage on river discharge is challenging (Golden et al., 2021). In cases where variables can be isolated (e.g., basins with tile drainage, compared to basins without tile drainage), significant differences between models can be an appropriate mechanism to help quantify the impact of a variable (Rainio et al. 2024). But both discharge and surface water extent tend to be a function of climate inputs and catchment characteristics (Heimhuber et al., 2016; Vanderhoof et al., 2018). Consequently, our inundation variables were significantly correlated with not only catchment characteristics, such as depth to bedrock, slope, and topographic diversity, but also climate variables, including annual precipitation, aridity, and rainfall intensity (Table A4), with the highest correlation occurring between the amount of seasonal inundation on the floodplain and the watershed rainfall intensity (R=0.80, p<0.01) (Table A4). Because our inundation variables were significantly correlated with select climate variables, the $M_{Climate}$ cannot be considered a null model, relative to M_{All}, and therefore comparing variables selected, variable significance and importance as well as model improvement using evaluation metrics was seen as more appropriate than testing for significant differences between models. It is possible that an alternative methodological approach, for example integrating remotely sensed surface water into a process-based hydrologic model (e.g., Stacke and Hagemann, 2012; Rajib et al. 2020) or applying a deep learning approach to time series data (e.g., Kratzert et al. 2019), may help build upon these findings and provide additional clarity regarding the discrete role of surface water extent on diverse discharge regimes. However, process-based hydrologic models are most often developed for a single or series of nested watersheds (Jones et al., 2019), limiting our ability to compare geographically disparate watersheds. What was novel to this effort, conversely, was that we were able to explore the relevance of surface water storage variables from a spatial analysis across multiple, diverse watersheds, instead of from modelling temporal variability within a single watershed.

4.2 Sources of Uncertainty

Modelling hydrologic signatures to evaluate the relative influence of drivers on hydrologic responses has many potential sources of uncertainty. Our results, for instance, could depend on the hydrologic signatures included in the analysis (McMillan et al., 2021). It is possible that inundation has a greater or lesser influence on different aspects of the flow regime than those explored here. For the hydrologic signatures considered, such signatures can show substantial uncertainty, attributable to error in precipitation and discharge datasets (Westerberg and McMillan, 2015). To account for uncertainty, the hydrologic signatures were calculated annually, and then averaged across multiple years, while independent variables were averaged over multiple years and across each watershed, both steps that have been shown to reduce uncertainty (Westerberg and McMillan, 2015). Our findings may also depend on the variables included in the analysis. While we included diverse climate and catchment characteristics, it is possible that additional catchment variables, if available, such as data on aquifers (Bloomfield et al., 2021) or additional geologic characteristics, such as proportion sandstone (Carlier et al. 2018), could improve the explanatory power of certain hydrologic signatures, like baseflow index, and reduce our model uncertainty. However, included variables like clay fraction, crop cover, topography, aridity, bedrock depth, and precipitation have all previously been found to explain variability in baseflow (Aboelnour et al., 2021; Bloomfield et al., 2021; Briggs et al., 2022) Uncertainty can also be attributable to the watersheds selected (McMillan et al., 2021). While we limited the range of watershed sizes and sampled across diverse regions, we under-sampled certain regions including the northeastern U.S. and mountainous regions, where a high proportion of forest cover and steep slopes, respectively, tend to increase our uncertainty in mapping surface water. Generating the surface water variables was also computationally intensive and limited our feasible sample size, which also likely contributed uncertainty to the modelling effort. Further, while surface water extent was used to represent surface water storage, the two are distinct measurements, and in the future, conversion of surface water (2D) to storage (3D) will facilitate improved modelling of total water distribution. Lastly, uncertainty can be introduced by the statistical modelling approach itself. To minimize modelling-related uncertainty we applied hyper-parameter optimization and variable selection procedures. Random forest models have also previously been found to be an effective mechanism to model hydrologic signatures (Trancoso et al., 2016; Addor et al., 2018; Oppel and Schumann, 2020). Further exploration of how inundation impacts diverse components of flow regimes will be an important next step to reduce the uncertainty associated with this effort.

4.3 Management implications

Hydrologic signatures have been used to support watershed management. For example, signatures related to flow magnitude, high flow frequency and flow variability have applications for flood management (Mogollon et al., 2016), wildlife habitat condition (Lowe et al., 2019), and riparian vegetation (Richter et al., 1996). Further, changes in hydrologic signatures over time have been used to examine the impacts of management actions or to assess a watershed's vulnerability or resilience to change (Hannaford and Marsh, 2008; Mogollon et al., 2016; McMillan et al., 2021; Lane et al., 2023). Applying results linking different watershed characteristics (e.g., climate, land use, geology) to hydrologic signature variability can therefore help inform future watershed management actions. However, a challenge is how to synthesize this information in a useful way (Gnann et al., 2020). One approach would be to focus on managing watershed characteristics that are highly correlated with a pre-determined flow signature target, like those associated with flood risks. For example, in our analyses, the association of greater semi-permanent and permanent floodplain inundation with less flashiness suggests that protection and restoration of floodplains may be particularly important in watersheds with flashy discharge. On the other hand, we found that non-floodplain surface water inundation contributed to small improvements in modelling variability in the baseflow index, which describes the proportion of flow coming from groundwater, and by inference the relative potential vulnerabilities for drought and extreme low flow conditions. Results from our analyses, and other future analyses leveraging large satellite-based data sets against streamflow records, can therefore advance our ability to support improved watershed management in the face of future floods and drought (Winsemius et al., 2016; Stewart et al., 2020).

5. Conclusion

Hydrologic signatures are increasingly used to provide insights on process-based streamflow dynamics (Addor et al., 2018; McMillian, 2019). Most previous efforts that have modelled flow signatures have not tested inundation- related, observation-based variables. And conversely, most previous efforts that have tested the influence of inundation- related variables on discharge, have been limited in geographic extent. In this analysis we explored integrating novel remotely sensed surface water variables to help explain spatial variability in hydrologic signatures in watersheds across CONUS. While improvement in model performance between using only climate variables, and considering climate and catchment variables, was moderate at best, inundation variables were selected and significant in models for five of the six hydrologic signatures. Variables representing floodplain inundation dominated, with the amount of semi-permanent to permanent floodplain inundation supporting the greatest improvement in modelling the flashiness signatures, relative to the other signatures explored, and model improvement metrics ranging from 3.86% to 11.69% (Table 6). Enhancing our understanding of when and where surface water storage influences discharge regimes can help guide management of non-riverine surface water, including wetlands, lakes, and floodplains, and in turn, support greater watershed resilience against climate extremes and hydrologic disturbances (Lane et al., 2023).

Data Availability

The surface water data produced for this analysis are published and available (Vanderhoof et al., 2024).

Author Contribution

MV, PN, HG, CL and JC contributed to the work's conception. PN, WK, and MV contributed to data processing and analysis. MV, PW, HE, CL, JC, WK and WD contributed to the interpretation of the results as well as the writing and editing.

Competing Interests

The authors declare that they have no conflicts of interest.

Acknowledgements

This research was funded by the U.S. Geological Survey's National Land Imaging and Land Change Science Programs and the U.S. Environmental Protection Agency's, Office of Research and Development through an interagency agreement (DW-014-92569201-0, "Multisource remote sensing to enhance national mapping of aquatic resources"). We appreciate comments on earlier versions from Brent Johnson and Kyle McLean. We also appreciate support from Jeremy Havens and Kylen Solvik. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This publication represents the views of the authors and does not necessarily reflect the views or policies of the U.S. EPA.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
Discussions

References

Abatzoglou J. T.: Development of gridded surface meteorological data for ecological applications and modelling, Int. J. Climatol., 33(1), 121-131, 2013.

Abatzoglou J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C.: Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Sci. Data, 5:170191, 2018.

Aboelnour, M. A., Engel, B. A., Frisbee, M. D., Gitau, M. W., and Flanagan, D. C.: Impacts of Watershed Physical Properties and Land Use on Baseflow at Regional Scales, Journal of Hydrology: Regional Studies, 35, 100810, https://doi.org/10.1016/j.ejrh.2021.100810, 2021.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resour. Res., 54, 8792–8812, 2018.

Ameli, A. A. and Creed, I. F.: Quantifying hydrologic connectivity of wetlands to surface water systems, Hydrol. Earth Syst. Sc., 21, 1791–808, 2017.

Ameli, A. A. and Creed, I. F.: Does wetland location matter when managing wetlands for watershed-scale flood and drought resilience? J. Am. Water Resour. As., 55, 529–542, 2019.

Apurv, T. and Cai, X.: Regional drought risk in the contiguous United States, Geophys. Res. Lett., 48(5), e2020GL092200, 2021.

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A new flashiness index: Characteristics and applications to midwestern rivers and streams, J. Am. Water Resour. As., 40(2), 503-522, 2004.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, J. Hydrometeorol., 16, 1478-1501, 2015.

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., de Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., and Schellekens, J.: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour. Res., 49, 7843-4863, 2013.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., and Sivapalan, M.: Dominant flood generating mechanisms across the United States, Geophys. Res. Lett., 43, 4382-4390, 2016.

Betts, M. G., Diamond, A. W., Forbes, G. J., Villard, M. -A., and Gunn, J. S...: The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence, Ecol. Model., 191(2), 197-224, 2006.

Blanchette, M., Rousseau, A. N., Foulon, É., Savary, S., and Poulin, M.: What would have been the impacts of wetlands on low flow support and high flow attenuation under steady state land cover conditions? J. Environ. Manage., 234, 448–457, 2019.

Bloomfield, J. P., Gong, M., Marchant, B. P., Coxon, G., and Addor, N.: How is Baseflow Index (BFI) impacted by water resource management practices? Hydrol. Earth Syst. Sci., 25, 5355–5379, 2021.

Briggs, M. A., Goodling, P., Johnson, Z. C., Rogers, K. M., Hitt, N. P., Fair, J. B., and Snyder, C. D.: Bedrock depth influences spatial patterns of summer baseflow, temperature and flow disconnection for mountainous headwater streams, Hydrology and Earth System Sciences, 26, 3989–4011, https://doi.org/10.5194/hess-26-3989-2022, 2022.

Budyko, M.: *The Heat Balance of the Earth's Surface*. Washington, DC: Springer, 1958.

Bullock, A., Acreman, M.: The role of wetlands in the hydrological cycle, Hydrol. Earth Syst. Sc., 7, 358-389, 2003.

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., and Brunner, P.: Geology controls streamflow dynamics, J. Hydrol., 566, 756–769, 2018.

Cawley, G. C. and Talbot, N. L. C.: On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res., 11, 2079-2107, 2010.

Cowardin, L. M., Carter, and F. C., Golet, E. T.: Classification of wetlands and deepwater habitats of the United States. United States Department of the Interior, Fish and Wildlife Service, Washington, D.C., USA, 1979.

Daigle, A., St-Hilaire, A., Beveridge, D., Caissie, D., and Benyahya, L.: Multivariate analysis of the low-flow regimes in eastern Canadian rivers, Hydrolog. Sci. J., 56, 51-67, 2011.

Donnelly, J. P., Naugle, D. E., Collins, D. P., Dugger, B. D., Allred, B. W., Tack, J. D., and Dreitz, V. J.: Synchronizing conservation to seasonal wetland hydrology and waterbird migration in semi-arid landscapes, Ecosphere, 10(6), e02758, 2019.

Eamus, D., Hatton, T., Cook, P., and Colvin, C.: Ecohydrology: vegetation function, water and resource management, CSIRO Publishing, Australia, 360 pp, 2006.

Emerson, R. W.: Bonferroni correction and type I error., *J. Vis. Impair. Blind*., 114(1), 77-78, 2020.

Evenson, G. R., Golden, H. E., Lane, C. R., and D'Amico, E.: Geographically isolated wetlands and watershed hydrology: A modified model analysis, J. Hydrol., 529, 240-256, 2015.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Evenson, G. R., Jones, C. N., McLaughlin, D. L., Golden, H. E., Lane, C. R. DeVries, B., Alexander, L. C., Lang, M. W., McCarty, G. W., and Sharifi, A.: A watershed-scale model for depressional wetland-rich landscapes, J. Hydrol. X. 1, 100002, 2018.

Falcone, J.: GAGES-II: Geospatial attributes of gages for evaluating streamflow. U.S. Geological Survey, Reston, Virginia, https://water.usgs.gov/lookup/getspatial?gagesII_Sept2011 (last accessed April 1, 2024), 2011.

Fritz, K. M., Schofield, K. A., Alexander, L. C., McManus, M. G., Golden, H. E., Lane, C. R., Kepner, W. G., LeDuc, S. D., DeMeester, J. E., and Pollard, A. I.: Physical and chemical connectivity of streams and riparian wetlands to downstream waters: a synthesis, J. Am. Water Resour. As., 54(2), 323-345, 2018.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D.: The national elevation dataset, Photogramm. Eng. Rem. S., 68(1), 5–11, 2002.

Gnann, S., McMillan, H., Woods, R., and Howden, N.: Including regional knowledge improves baseflow signature predictions in large sample hydrology, Water Resour. Res., 57(2), e2020WR028354, 2021.

Golden, H. E., Lane, C. R., Amatya, D. M., Bandilla, K. W., Raanan Kiperwas, H., Knightes, C. D., and Ssegane, H.: Hydrologic connectivity between geographically isolated wetlands and surface water systems: a review of select modeling methods, Environ. Modell. Softw., 53, 190–206, 2014.

Golden, H. E., Lane, C. R., Rajib, A., and Wu, Q.: Improving global flood and drought predictions: integrating non-floodplain wetlands into watershed hydrologic models, Environ. Res. Lett., 16, 091002, 2021.

Gregorutti, B., Michel, B., and Saint-Pierre, P.: Correlation and variable importance in random forests, Stat. Comput., 27, 659-678, 2016.

Hannaford, J. and Marsh, T.: High-flow and flood trends in a network of undisturbed catchments in the UK, Int. J. Climatol., 28, 1325-1338, 2008.

Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning, Springer, New York, 2009.

Heidari, H., Arabi, M., Warziniack, T., and Kao, S. C.: Assessing shifts in regional hydroclimatic conditions of U.S. river basins in response to climate change over the 21st century, Earth's Future, 8(10), e2020EF001657, 2020.

Heimhuber, V., Tulbure, M. G., and Broich, M.: Modeling 25 years of spatio-temporal surface water and inundation dynamics on large river basin scale using time series of Earth observation data, Hydrol. Earth Syst. Sc., 20(6), 2227-2250, 2016.

Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A., and Darby, S. E.: Assessing the characteristics and drivers of compound flooding events around the UK coast, Hydrol. Earth Syst. Sc., 23, 3117-3139, 2019.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B.: Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, PeerJ, e5518, 2018.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., and Riiters, K.: Conterminous United States land cover change patterns 2001-2016 from the 2016 National Land Cover Database, ISPRS J. Photogramm., 162, 184-199, 2020.

Jiang, S., Zheng, Y., Wang, C., and Babovic, V.: Uncovering flooding mechanisms across the contiguous United States through interpretive deep learning on representative catchments, Water Resour. Res., 58(1), e2021WR030185, 2022.

Jones, N. C., Ameli, A., Neff, B. P., Evenson, G. R., McLaughlin, D. L., Golden, H. E., and Lane, C. R.: Modeling connectivity of non-floodplain wetlands: insights, approaches, and recommendations, J. Am. Water Resour. As., 55, 559-577, 2019.

Kelly, V.J., and White, S.: A method for characterizing late-season low-flow regime in the upper Grand Ronde River Basin, Oregon. U.S. Geological Survey Scientific Investigations Report 2016-5041, 2016.

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D., and Marsh, N.: Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies, River Res. Appl., 26, 137–156, 2010.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sc., 21, 2863–2879, 2017.

Klute, D., Lovallo, M., and Tzilkowski, W.: Autologistic regression modeling of American woodcock habitat use with spatially dependent data. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Sampson, F.B. (Eds.), Predicting Species Occurrences, Issues of Accuracy and Scale. Island Press, Washington, pp. 335–343, 2002.

Kuppel, S., Houspanossian, J., Nosetto, M. D., and Jobbágy, E. G.: What does it take to flood the Pampas? Lessons from a decade of strong hydrological fluctuations, Water Resour. Res., 51, 2937–2950, 2015.

Lane, C. R. and D'Amico, E.: Identification of putative geographically isolated wetlands of the conterminous United States, J. Am. Water Resour. As., 52 705–22, 2016.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Lane, C. R., Leibowitz, S. G., Autrey, B. C., LeDuc, S. D., and Alexander, L. C.: Hydrological, physical, and chemical functions and connectivity of non-floodplain wetlands to downstream waters: a review, J. Am. Water Resour. As., 54(2), 346-371, 2018.

Lane, C. R., Creed, I. F., Golden, H. E., Leibowitz, S. G., Mushet, D. M., Rains, M. C., Wu, Q., D'Amico, E., Alexander, L. C., Ali, G. A., Basu, N. B., Bennett, M. G., Christensen, J. R., Cohen, M. J., Covino, T. P., DeVries, B., Hill, R. A., Jensco, K., Lang, M. W., McLaughlin, D., Rosenberry, D. O., Rover, J., and Vanderhoof, M. K.: Vulnerable waters are essential to watershed resilience, Ecology, 26, 1-28, 2022.

Leibowitz, S. G.: Isolated wetlands and their functions: An ecological perspective, Wetlands, 23, 517-531, 2003.

Lowe, W. H., Swartz, L. K., Addis, B. R., and Likens, G. E.: Hydrologic variability contributes to reduced survival through metamorphosis in a stream salamander, Proc. Natl. Acad. Sci., 116(39), 19563–19570, 2019.

McLaughlin, D. L., Kaplan, D. A., and Cohen, M. J.: A significant nexus: geographically isolated wetlands influence landscape hydrology, Water Resour. Res., 50, 7153–66, 2014.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: a review, Hydrol. Process., 34(6), 1393-1409, 2019.

McMillan, H. K.: A review of hydrologic signatures and their applications, WIREs Water, 8(1), doi: 10.1002/wat2.1499, 2021.

Mehdipoor, H., Zurita-Milla, R., Izquierdo-Verdiguier, E., and Betancourt, J. L.: Influence of source and scale of gridded temperature data on modelled spring onset patterns in the conterminous United States, Int. J. Climatol., 38(14), 5430-5440, 2018.

Mogollon, B., Frimpong, E. A., Hoegh, A. B., and Angermeier, P. L.: Recent changes in stream flashiness and flooding, and effects of flood management in North Carolina and Virginia, J. Am. Water Resour. As., 52, 561-577, 2016.

Murphy, M. A., Evans, J. S., and Storfer, A.: Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics, Ecology, 91, 252–261, 2010.

National Atlas of the United States: Major Dams of the United States, Puerto Rico and the US Virgin Islands. Delivered by ArcGIS online (last accessed September 6, 2022), 2006.

National River Flow Archive: Derived flow statistics. Available online: https://nrfa.ceh.ac.uk/derived-flow-statistics (last accessed April 1, 2024), 2024.

Newman, A., Sampson, K., Clark, M.P., Bock, A., Viger, R.J., Blodgett, D.: A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO:UCAR/NCAR, https://dx.doi.org/10.5065/D6MW2F4D, 2014.

NOAA: U.S. Billion-dollar weather and climate disasters, National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information, https://doi.org/10.25921/stkw-7w73, 2020.

Oppel, H., and Schumann, A. H.: Machine learning based identification of dominant controls on runoff dynamics, Hydrol. Process., 34, 2450–2465, 2020.

Oueslati, O., De Girolamo, A. M., Abouabdillah, A., Kjeldsen, T. R., and Lo Porto, A.: Classifying the flow regimes of Mediterranean streams using multivariate analysis, Hydrol. Process., 29, 4666-4682, 2015.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. and Dubourg, V.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Pilgrim, D. H., Cordery, I., Baron, B. C.: Effects of catchment size on runoff relationships, J. Hydrol., 58(3-4), 205-221, 1982.

Portet, S.: A primer on model selection using the Akaike Information Criterion, Infect. Dis. Model, 5, 111-128, 2020.

Rainio, O., Teuho, J., Klén, R.: Evaluation metrics and statistical tests for machine learning, Sci. Rep., 14, 6086, 2024.

Rains, M. C., Leibowitz, S. G., Cohen, M. J., Creed, I. F., Golden, H. E., Jawitz, J. W., Kalla, P., Lane, C. R., Lang, M. W., and McLaughlin, D. L.: Geographically isolated wetlands are part of the hydrological landscape, Hydrol. Process., 30(1), 153-160, 2016.

Rajib A, Golden H. E., Lane, C. R., and Wu, Q.: Surface depression and wetland water storage improves major river basin hydrologic predictions, Water Resour. Res., 56, e2019WR026561, 2020.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A method for assessing hydrologic alteration within ecosystems, Conserv. Biol., 10, 1163–1174, 1996.

Sauqet, E., Shanafield, M., Hammond, J. C., Sefton, C., Leigh, C., and Datry, T.: Classification and trends in intermittent river flow regimes in Australia, northwestern Europe, and USA: A global perspective, J. Hydrol., 597, 126170, 2021.

Scott, D. T., Gomez-Velez, J. D., Jones, C. N., and Harvey, J. W.: Floodplain inundation spectrum across the United States, Nat. Commun., 10, 5194, 2019.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Shaw, D. A., Vanderkamp, G., Conly, F. M., Pietroniro, A., and Martz, L.: The fill-spill hydrology of prairie wetland complexes during drought and deluge, Hydrol. Process., 26, 3147–3156, 2012.

Sherrouse, B.C. and Hawbaker, T.J.: HOPS: Hyperparameter optimization and predictor selection v1.0, U.S. Geological Survey Software Release, https://doi.org/10.5066/P9P81HUR, 2023.

Stacke, T. and Hagemann, S.: Development and evaluation of a global dynamical wetlands extent scheme, Hydrol. Earth Syst. Sc., 16, 2915-2933, 2012.

Stepchinski, L. M., Rains, M. C., Lee, L. C., Lis, R. A., Nutter, W. L., Rains, K. C., and Stewart, S. R.: Hydrologic connectivity and flow generation from California vernal pool, swale, and headwater stream complexes to downstream waters, Wetlands, 43, 34, 2023.

Stewart, I. T., Rogers, J., and Graham, A.: Water security under severe drought and climate change: Disparate impacts of the recent severe drought on environmental flows and water supplies in Central California, J. Hydrol. X, 7, 100054, 2020.

Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., and Albano, C. M.: Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning, PloS ONE, 10(12), e0143619, 2015.

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., Wilson, B. E.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1840, 2020.

Trancoso, R., Phinn, S., McVicar, T. R., Larsen, J. R., McAlpine, C. A.: Regional variation in streamflow drivers across a continental climatic gradient, Ecohydrology, 10, e1816, 2016.

USFS: U.S. Stream Flow Metric Dataset: Modeled metrics for stream segments in the United States under historical conditions and projected climate change scenarios. Data Guide. Boise, ID, U.S. Department of Agriculture, U.S. Forest Service (USFS), (Last accessed September 6, 2022), 2022.

USFWS: National Wetlands Inventory. U.S. Fish and Wildlife (USFWS) Service. https://www.fws.gov/program/national-wetlands-inventory. (Last accessed April 1, 2024), 2019.

USGS: High Resolution, National Hydrography Dataset, U.S. Geological Survey (USGS), The National Map, Hydrography, https://apps.nationalmap.gov/services/ (Last accessed August 4, 2022), 2022.

USGS: U.S. Geological Survey water data for the Nation: U.S. Geological Survey (USGS) National Water Information System database, https://doi.org/10.5066/F7P55KJN (Last accessed (Last accessed April 1, 2024), 2024.

Vabalas, A., Gowen, E., Poliakoff, E. and Casson, A. J.: Machine learning algorithm validation with a limited sample size, PLoS ONE, 14(11), e0224365, 2019.

van Dijk, A. I. J. M.: Climate and terrain factors explaining streamflow response and recession in Australian catchments, Hydrol. Earth Syst. Sc., 14, 159-169, 2010.

Vanderhoof, M. K., Alexander, L. C., and Todd, M. J.: Temporal and spatial patterns of wetland extent influence variability of surface water connectivity in the Prairie Pothole Region, United States, Landscape Ecol., 31(4), 805-824, 2016.

Vanderhoof, M. K., Lane, C. R., McManus, M. G., Alexander, L. C., and Christensen, J. R.: Wetlands inform how climate extremes influence surface water expansion and contraction, Hydrol. Earth Syst. Sc., 22(3), 1851-1873, 2018.

Vanderhoof, M. K., Alexander, L., Christensen, J., Solvik, K., Nieuwlandt, P. and Sagehorn, M.: High-frequency time series comparison of Sentinel-1 and Sentinel-2 for open and vegetated water across the United States (2017-2021), Remote Sens. Environ., 288, 113498, 2023.

Vanderhoof, M. K., Christensen, J. R., Alexander, L. C., Lane, C. R., and Golden, H. E.: Climate change will impact surface water extents across the central United States, Earth's Future, 12(2), e2023EF004106, 2024.

Vanderhoof, M.K., Nieuwlandt, P., Golden, H.E., Lane, C.R., Christensen, J.R., Keenan, W., and Dolan, W.: Data release for integrating remotely sensed surface water dynamics in hydrologic signature modelling, U.S. Geological Survey data release, https://doi.org/10.5066/P9RLFMEQ, 2024.

Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19, 3951–3968, 2015.

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P., and Ward, P. J.: Global drivers of future river flood risk, Nat. Clim. Change, 6, 381-385, 2016.

Wohl, E.: An integrative conceptualization of floodplain storage, Rev. Geophy., 59(2), e2020RG000724, 2021.

Woznicki, S. A., Baynes, J., Panlasigui, S., Mehaffey, M., and Neale, A.: Development of a spatially complete floodplain map of the conterminous United States using random forest, Sci. Total Environ., 647, 942-953, 2019.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Wu, G., Chen, J., Shi, X., Kim, J. S., Xia, J., and Zhang, L.: Impacts of global climate warming on meteorological and hydrological droughts and their propagations, Earth's Future, 10(3), e2021EF002542, 2022.

Wu, Y., Zhang, G., Rousseau, A. N., Xu, Y. J. and Foulon, E.: On how wetlands can provide flood resilience in a large river basin: a case study in Nenjiang river Basin, China, J. Hydrol., 587, 125012, 2020.

Zeng, L., Shao, J., and Chu, X.: Improved hydrologic modeling for depression-dominated areas, J. Hydrol., 590, 125269, 2020.

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

Appendix

Table A1. The 72 U.S. Geological survey gages and watersheds included in the analysis. The 2016-2023 period is shown relative to the Palmer Drought Severity Index (PDSI, 1980-2021). NHD: National hydrographic dataset, NWI: National Wetland Inventory. CC: cultivated crops, DF: deciduous forest, D: developed, HP: hay/pasture, EF: evergreen forest, WW: woody wetlands, MF: mixed forest, SS: shrub/scrub, H: herbaceous

Gage ID	Site ID	U.S. State(s)	Area (km^2)	NHD Density (m km^2)	NWI (% area)	PDSI (min, %)	PDSI (max, %)	PDSI (median, %)	Primary land cover
01491000	MD1	MD, DE	292	2030.7	28.6	32.6	100.0	66.5	CC (47%)
01578475	MD2	MD, PA	458	1069.2	2.6	6.1	100.0	72.4	CC (43%)
01580520	MD3	MD, PA	425	1130.1	2.1	8.9	100.0	67.4	DF (30%)
01594440	MD4	MD	907	1571.9	6.4	16.7	100.0	64.4	D (36%)
01643000	MD5	MD, PA	2112	1394.3	3.0	4.2	100.0	57.5	HP (27%)
02049500	VA1	VA	1583	1497.8	15.7	33.0	100.0	79.3	EF (29%)
02131500	SC1	SC, NC	1720	1451.6	10.3	18.3	100.0	53.7	EF (26%)
02135000	SC2	SC, NC	7256	1628.6	27.2	8.5	100.0	78.0	WW (31%), CC (31%)
02136000	SC3	SC	3211	1738.0	27.0	17.6	100.0	75.2	CC (32%), WW (31%)
02175000	SC4	SC	7077	1163.0	17.3	27.1	98.0	75.5	EF (25%), WW (24%)
02198000	GA1	GA	1676	1365.2	12.0	19.4	96.6	61.1	EF (26%)
02202500	GA2	GA	6887	1249.8	16.8	21.1	97.7	60.2	EF (26%)
05056000	ND1	ND	4862	283.9	10.6	1.2	100.0	56.3	CC (52%)
05057200	ND2	ND	1897	259.2	11.6	0.0	100.0	65.0	CC (67%)
05062500	MN1	MN	2407	745.9	23.9	3.4	100.0	58.6	CC (39%)
05066500	ND3	ND	3218	774.1	6.9	0.3	100.0	63.4	CC (81%)
05078500	MN2	MN	3518	862.3	23.5	1.2	100.0	54.5	CC (48%)
05090000	ND4	ND	1742	1068.9	3.7	1.5	100.0	51.0	CC (73%)
05123400	ND5	ND	3206	515.6	12.2	1.0	97.8	48.8	CC (48%)
05131500	MN3	MN	4384	608.9	42.4	4.5	100.0	84.4	WW (49%)
05132000	MN4	MN	3895	537.3	48.7	5.6	100.0	71.1	WW (49%)
05244000	MN5	MN	2683	471.2	23.8	0.9	100.0	52.3	DF (27%)
05300000	MN6	MN, SD	2468	1286.4	11.5	11.6	100.0	66.6	CC (68%)
05304500	MN7	MN	4899	733.6	17.0	4.8	100.0	62.5	CC (66%)
05313500	MN8	MN, SD	1801	1129.0	8.8	8.5	100.0	58.8	CC (80%)
05336700	MN9	MN	2252	676.5	34.1	17.0	100.0	87.8	WW (34%)
05388250	IA1	IA, MN	2010	1548.4	2.7	9.2	100.0	76.1	CC (61%)
05412500	IA2	IA	3858	1414.9	2.4	6.9	100.0	81.6	CC (66%)
05418500	IA3	IA	4019	1452.5	2.1	6.3	100.0	70.8	CC (69%)
05422000	IA4	IA, MN	6049	1248.5	4.6	5.9	99.7	70.2	CC (79%)
05434500	WI1	WI, IL	2677	1618.6	3.0	5.1	100.0	71.9	CC (44%)
05447500	IL1	IL	2576	1115.6	1.9	20.7	100.0	74.9	CC (85%)
06018500	MT1	MT	9373	1628.9	3.9	0.3	89.3	50.9	SS (47%)
06052500	MT2	MT, WY	4634	1376.2	2.9	1.2	97.4	61.8	EF (47%)
06076690	MT3	MT	2189	1695.3	4.3	1.4	98.1	62.7	H (35%)
06468170	ND6	ND	2809	302.6	7.4	1.0	100.0	66.3	CC (67%)
06471200	ND7	ND, SD	1869	627.2	11.2	1.2	100.0	70.8	CC (62%)
06479525	SD1	SD	2467	947.8	9.8	19.3	100.0	67.4	CC (59%)
06481500	SD2	SD	1604	1102.0	8.7	8.8	100.0	62.0	CC (72%)
06815000	NE1	NE, KS	3473	1688.2	1.8	4.1	99.2	52.8	CC (54%)
06821190	MO1	MO, IA	6179	1925.6	4.8	11.3	99.0	56.6	CC (50%)
06908000	MO2	MO	2895	1737.9	4.2	3.5	90.4	51.9	HP (38%)
06916600	KS2	KS, MO	8387	1685.9	3.8	12.3	100.0	57.5	HP (37%)
06918060	MO3	MO, KS	2773	1669.2	5.4	4.7	100.0	57.0	HP (56%)

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Gage ID	**Site ID**	**U.S. State(s)**	**Area (km^2)**	**NHD Density (m km^2)**	**NWI (% area)**	**PDSI (min, %)**	**PDSI (max, %)**	**PDSI (median, %)**	**Primary land cover**
06928000	MO4	MO	3275	1538.7	1.8	12.8	100.0	79.9	DF (45%), HP (43%)
07047950	AR1	AR	1985	1864.2	12.5	20.2	100.0	82.5	CC (73%)
07169500	KS3	KS	2098	1781.3	2.9	4.9	100.0	62.0	H (595)
07288500	MS1	MS	2009	1809.9	9.8	7.1	97.9	55.9	CCs (82%)
07290000	MS2	MS	7124	2565.5	10.0	13.3	100.0	74.8	EF (19%), MF (19%)
07346070	TX1	TX	1809	2010.3	9.3	6.5	100.0	70.4	HP (27%)
07363500	AR2	AR	5429	1762.5	3.0	28.8	99.1	83.0	EF (40%)
07364200	LA1	AR, LA	3138	1507.9	14.6	22.8	100.0	79.5	CC (31%)
08033500	TX2	TX	9406	1712.0	8.0	3.2	99.9	64.9	EF (29%)
08068090	TX4	TX	2539	1695.0	9.9	10.9	100.0	71.4	EF (32%)
08110000	TX5	TX	2616	1630.0	4.8	8.9	100.0	73.8	HP (55%)
08117500	TX6	TX	1869	1085.4	5.6	6.4	98.6	64.8	HP (43%)
08164000	TX7	TX	2124	1435.4	2.1	8.8	94.1	52.3	HP (59%)
09439000	AZ1	AZ, NM	9279	1679.3	1.2	1.2	98.1	40.2	SS (45%)
09485700	AZ2	AZ	2238	2347.0	2.1	0.0	95.4	48.3	SS (64%)
09487000	AZ3	AZ	2028	3229.6	2.3	0.0	87.7	42.4	SS (79%)
09512800	AZ4	AZ	2876	1639.6	1.3	0.1	88.1	47.1	SS (68%)
09517000	AZ5	AZ	3967	1664.7	1.7	0.2	90.8	50.6	SS (81%)
09537500	AZ6	AZ	2912	1392.5	1.1	0.0	96.6	46.0	SS (67%)
11348500	CA1	CA	3884	1469.4	8.0	0.0	84.1	55.6	SS (50%)
11376000	CA2	CA	2313	2450.2	1.9	0.0	89.1	29.9	SS (56%)
11473900	CA3	CA	1925	4181.6	1.2	0.0	88.2	35.5	EF (45%)
11501000	OR1	OR	4121	1028.4	8.2	0.0	83.3	43.4	EF (55%)
11517500	CA4	CA	2047	1495.8	5.6	0.0	94.6	17.6	EF (37%)
11519500	CA5	CA	1714	2381.7	3.8	0.0	97.6	26.3	EF (46%)
12324680	MT4	MT	4590	1287.2	3.5	1.4	97.7	46.4	EF (45%)
13302005	ID1	ID	2143	1615.5	1.2	0.5	97.8	51.2	SS (76%)
13305000	ID2	ID	2412	1443.0	1.3	0.5	93.6	48.6	SS (59%)
All (median)	~	~	**2647**	**1461.0**	**5.6**	**5.0**	**100.0**	**62.0**	~

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Discussions

Table A2. Thresholds selected from 5-year Sentinel-1 (S1) and Sentinel-2 (S2) based surface water percentiles to account for variable accuracy between sites, sensors, and classes (open water (OW) compared to vegetated water (VW). ~ indicates that this output was excluded from the allowable water mask.

Site ID	S1 OW (%)	S1 VW (%)	S2 OW (%)	S2 VW (%)	Site ID	S1 OW (%)	S1 VW (%)	S2 OW (%)	S2 VW (%)
AR1	15	30	15	30	MN7	10	35	5	25
AR2	5	20	10	25	MN8	10	25	5	15
AZ1	10	5	15	10	MN9	5	25	5	25
AZ2	5	15	10	15	MO1	10	20	10	20
AZ3	5	10	10	15	MO2	5	15	15	25
AZ4	5	20	15	20	MO3	5	30	10	30
AZ5	5	20	20	15	MO4	10	15	10	35
AZ6	10	15	10	~	MS1	10	30	10	35
CA1	5	15	10	20	MS2	5	10	5	30
CA2	10	10	15	15	MT1	25	~	10	30
CA3	10	~	15	15	MT2	25	~	15	40
CA4	10	15	20	20	MT3	30	~	10	40
CA5	10	10	20	15	MT4	30	~	10	30
GA1	5	5	5	20	ND1	15	20	5	10
GA2	5	10	10	15	ND2	20	20	10	20
IA1	~	15	10	15	ND3	15	~	5	20
IA2	~	10	10	20	ND4	15	35	5	30
IA3	10	10	10	20	ND5	20	30	5	25
IA4	10	20	10	20	ND6	15	30	5	25
ID1	30	~	15	35	ND7	20	30	5	20
ID2	30	~	20	35	NE1	15	15	10	20
IL1	10	30	10	15	OR1	10	20	25	25
KS2	10	20	10	30	SC1	5	20	10	35
KS3	~	15	10	20	SC2	5	25	5	25
LA1	10	25	15	35	SC3	5	30	5	30
MD1	5	~	10	20	SC4	5	30	10	35
MD2	5	15	10	20	SD1	10	25	5	25
MD3	5	10	10	15	SD2	15	25	5	25
MD4	5	10	10	15	TX1	5	10	10	30
MD5	10	30	15	~	TX2	5	30	10	35
MN1	15	20	5	20	TX4	5	30	10	35
MN2	10	20	5	30	TX5	5	20	10	~
MN3	10	30	10	30	TX6	10	~	10	35
MN4	5	30	5	30	TX7	5	35	10	30
MN5	10	30	5	25	VA1	5	30	10	45
MN6	15	25	5	20	WI1	~	15	10	20

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Table A3. Hydrologic signatures by watershed. The blue to red shading reflects the high to low values for each signature. The bold values indicate the average values for the watersheds within each region.

Region	ID	Gage	Flashiness Index	Flashiness index (wet season)	MAX30 /area	(Q10-Q95)/area	Dry Month /area	Baseflow index
East			**-0.74**	**-0.78**	**1.37**	**0.023**	**0.0065**	**0.38**
East	MD1	01491000	-0.48	-0.45	2.16	0.034	0.0072	0.28
East	MD2	01578475	-0.44	-0.43	1.52	0.024	0.0105	0.55
East	MD3	01580520	-0.52	-0.64	1.45	0.024	0.0112	0.54
East	MD4	01594440	-0.43	-0.42	1.38	0.021	0.0089	0.49
East	MD5	01643000	-0.35	-0.40	1.98	0.028	0.0058	0.24
East	VA1	02049500	-0.87	-1.01	1.27	0.028	0.0060	0.36
East	SC1	02131500	-0.66	-0.64	1.29	0.022	0.0059	0.39
East	SC2	02135000	-1.05	-1.07	1.55	0.025	0.0055	0.35
East	SC3	02136000	-0.91	-1.04	1.22	0.023	0.0039	0.28
East	SC4	02175000	-1.13	-1.20	0.89	0.017	0.0055	0.44
East	GA1	02198000	-0.90	-0.95	0.86	0.016	0.0043	0.37
East	GA2	02202500	-1.09	-1.17	0.92	0.017	0.0030	0.24
Gulf Coast			**-0.79**	**-0.83**	**1.88**	**0.032**	**0.0026**	**0.09**
Gulf Coast	AR1	07047950	-0.99	-1.01	3.48	0.050	0.0057	0.18
Gulf Coast	MS1	07288500	-0.79	-0.90	2.23	0.056	0.0035	0.04
Gulf Coast	MS2	07290000	-0.85	-0.93	2.22	0.046	0.0030	0.10
Gulf Coast	TX1	07346070	-0.74	-0.71	1.64	0.025	0.0006	0.02
Gulf Coast	AR2	07363500	-0.82	-0.86	2.46	0.050	0.0024	0.05
Gulf Coast	LA1	07364200	-1.45	-1.58	1.37	0.044	0.0030	0.16
Gulf Coast	TX2	08033500	-0.94	-1.01	1.19	0.024	0.0027	0.08
Gulf Coast	TX4	08068090	-0.35	-0.31	2.30	0.016	0.0022	0.09
Gulf Coast	TX5	08110000	-1.00	-1.02	0.54	0.020	0.0024	0.08
Gulf Coast	TX6	08117500	-0.51	-0.59	2.10	0.021	0.0019	0.08
Gulf Coast	TX7	08164000	-0.21	-0.23	1.13	0.003	0.0010	0.07
Midwest			**-0.62**	**-0.60**	**1.43**	**0.021**	**0.0042**	**0.28**
Midwest	IA1	05388250	-0.78	-0.68	1.51	0.025	0.0083	0.47
Midwest	IA2	05412500	-0.73	-0.62	1.53	0.024	0.0066	0.37
Midwest	IA3	05418500	-0.80	-0.69	1.11	0.016	0.0077	0.59
Midwest	IA4	05422000	-0.99	-1.06	1.14	0.023	0.0060	0.41
Midwest	WI1	05434500	-1.12	-1.01	0.96	0.014	0.0094	0.70
Midwest	IL1	05447500	-0.79	-0.78	1.03	0.018	0.0055	0.38
Midwest	NE1	06815000	-0.25	-0.21	0.96	0.007	0.0013	0.24
Midwest	MO1	06821190	-0.52	-0.55	1.14	0.016	0.0018	0.19
Midwest	MO2	06908000	-0.40	-0.44	1.61	0.022	0.0010	0.05
Midwest	KS2	06916600	-0.55	-0.60	1.48	0.023	0.0013	0.09
Midwest	MO3	06918060	-0.39	-0.45	2.13	0.030	0.0020	0.06
Midwest	MO4	06928000	-0.38	-0.34	2.24	0.026	0.0024	0.10
Midwest	KS3	07169500	-0.42	-0.36	1.69	0.032	0.0015	0.06
North-Central			**-0.93**	**-0.93**	**0.52**	**0.008**	**0.0016**	**0.19**
North-Central	ND1	05056000	-1.04	-0.98	0.11	0.002	0.0005	0.08
North-Central	ND2	05057200	-0.83	-0.88	0.21	0.004	0.0004	0.07
North-Central	MN1	05062500	-0.94	-0.92	0.48	0.007	0.0014	0.24
North-Central	ND3	05066500	-0.76	-0.79	0.54	0.006	0.0007	0.09
North-Central	MN2	05078500	-0.81	-0.77	0.54	0.006	0.0011	0.23
North-Central	ND4	05090000	-0.78	-0.82	0.34	0.004	0.0004	0.05
North-Central	ND5	05123400	-1.09	-1.11	0.10	0.002	0.0001	0.06

Region	**ID**	**Gage**	**Flashiness Index**	**Flashiness index (wet season)**	**MAX30 /area**	**(Q10-Q95)/area**	**Dry Month /area**	**Baseflow index**
	MN3	05131500	-0.90	-0.86	1.15	0.018	0.0028	0.19
	MN4	05132000	-1.01	-0.95	0.77	0.013	0.0020	0.27
	MN5	05244000	-1.45	-1.46	0.31	0.006	0.0038	0.68
	MN6	05300000	-0.99	-0.96	0.65	0.011	0.0019	0.21
	MN7	05304500	-1.16	-1.14	0.46	0.010	0.0029	0.34
	MN8	05313500	-0.90	-0.89	0.83	0.015	0.0025	0.18
	MN9	05336700	-0.77	-0.78	1.69	0.027	0.0055	0.25
	ND6	06468170	-0.93	-0.93	0.18	0.003	0.0001	0.04
	ND7	06471200	-0.68	-0.64	0.24	0.002	0.0002	0.09
	SD1	06479525	-1.00	-1.09	0.23	0.005	0.0010	0.22
	SD2	06481500	-0.73	-0.73	0.48	0.009	0.0016	0.18
	Southwest		**-0.12**	**-0.16**	**0.06**	**<0.001**	**<0.0001**	**0.01**
	AZ1	09439000	-0.61	-0.83	0.09	0.001	0.0000	0.03
	AZ2	09485700	0.07	0.12	0.08	0.000	0.0000	0.00
Southwest	AZ3	09487000	0.23	0.23	0.01	0.000	0.0000	0.00
	AZ4	09512800	-0.08	-0.09	0.16	0.001	0.0000	0.00
	AZ5	09517000	-0.30	-0.34	0.02	0.000	0.0001	0.05
	AZ6	09537500	-0.02	-0.03	0.01	0.000	0.0000	0.00
	West		**-1.03**	**-1.09**	**0.67**	**0.012**	**0.0009**	**0.26**
	MT1	06018500	-1.23	-1.41	0.04	0.001	0.0004	0.44
	MT2	06052500	-1.18	-1.06	0.69	0.013	0.0022	0.32
	MT3	06076690	-1.04	-1.01	0.18	0.004	0.0007	0.32
	CA1	11348500	-0.70	-0.77	0.22	0.004	0.0001	0.02
	CA2	11376000	-0.51	-0.69	1.62	0.023	0.0006	0.06
West	CA3	11473900	-0.51	-0.69	3.31	0.051	0.0003	0.01
	OR1	11501000	-1.25	-1.24	0.31	0.005	0.0011	0.35
	CA4	11517500	-1.13	-1.27	0.14	0.003	0.0005	0.21
	CA5	11519500	-0.82	-0.95	0.92	0.023	0.0003	0.03
	MT4	12324680	-1.16	-1.07	0.33	0.006	0.0015	0.38
	ID1	13302005	-1.63	-1.89	0.12	0.002	0.0019	0.61
	ID2	13305000	-1.22	-1.08	0.20	0.003	0.0012	0.40

https://doi.org/10.5194/hess-2024-298
Preprint. Discussion started: 8 November 2024

Table A4. Correlation values between remotely sensed surface water variables and other independent variables. Significant (p<0.01) correlations, after Bonferroni correction has been applied, are shown shaded in gray. Correlations between surface water variables ranged from 0 to 0.98 with a median correlation of 0.35. CV: coefficient of variation, FP: floodplain, NFP: non-floodplain, temp: temporarily, seas: seasonally, SP+P: semi-permanent to permanent, inun: inundation, Prop.: proportion, Geographically Isolated Wetlands: GIW

Variable Type	Variable	Temp. flooded, FP	Temp. inun., NFP	Seas. inun., FP	Seas. inun., NFP	SP+P inun., FP	SP+P inun., NFP	Total inun., FP	Total inun., NFP	Prop. seas. inun., FP	Prop. seas. inun., NFP
	Precipitation	0.39	0.52	0.75	0.44	0.41	0.21	0.69	0.45	0.47	0.06
	Evapo-transpiration	0.40	-0.12	0.19	-0.22	-0.10	-0.27	0.19	-0.23	0.22	-0.44
	Aridity index	-0.27	-0.69	-0.67	-0.55	-0.40	-0.26	-0.61	-0.59	-0.28	-0.20
	Water demand	0.22	0.61	0.61	0.46	0.34	0.16	0.53	0.50	0.31	0.16
	Precipitation seasonality	0.03	0.19	0.06	0.29	0.03	0.09	0.11	0.26	-0.20	0.32
Climate	Precipitation CV	-0.20	-0.52	-0.64	-0.45	-0.38	-0.25	-0.55	-0.46	-0.41	-0.17
	Rainfall intensity	0.49	0.47	0.80	0.47	0.41	0.26	0.77	0.45	0.43	0.03
	Max monthly precipitation	0.51	0.44	0.63	0.33	0.26	0.09	0.63	0.34	0.34	-0.06
	Temperature seasonality	-0.37	0.02	-0.25	0.19	0.06	0.23	-0.21	0.19	-0.31	0.46
	Temperature CV	-0.44	-0.01	-0.32	0.14	0.05	0.23	-0.28	0.15	-0.32	0.43
	Forest	0.00	0.30	0.10	-0.02	0.06	0.00	0.04	0.04	0.20	-0.15
Land Cover	Developed	0.39	0.37	0.63	0.28	0.28	0.04	0.58	0.28	0.44	-0.04
	Cultivated crops	0.07	0.05	0.21	0.28	0.17	0.16	0.23	0.25	0.04	0.30
	Stream density	0.43	-0.11	0.13	-0.32	-0.24	-0.48	0.13	-0.32	0.32	-0.45
	Clay fraction	0.39	-0.01	0.27	0.00	0.00	-0.10	0.27	-0.06	0.20	-0.14
	Sand fraction	-0.35	0.05	-0.17	0.08	0.10	0.22	-0.18	0.09	-0.25	0.11
	Silt Fraction	0.22	-0.06	0.02	-0.12	-0.15	-0.25	0.04	-0.11	0.19	-0.07
Soil and Geology	Depth to bedrock	-0.12	0.33	0.49	0.71	0.66	0.69	0.51	0.68	-0.03	0.52
	Water table depth	-0.18	-0.51	-0.68	-0.78	-0.67	-0.64	-0.72	-0.73	-0.12	-0.47
	Geological permeability	-0.36	0.16	-0.06	0.21	0.18	0.31	-0.09	0.21	-0.17	0.27
	Slope	0.02	-0.30	-0.55	-0.77	-0.63	-0.76	-0.56	-0.71	0.05	-0.59
Topography	Elevation range	0.21	0.02	0.12	-0.22	-0.13	-0.23	0.04	-0.18	0.24	-0.35
	Topographic diversity	0.02	-0.22	-0.50	-0.71	-0.57	-0.70	-0.51	-0.65	0.05	-0.58
	GIW	-0.27	0.32	0.37	0.80	0.73	0.89	0.40	0.76	-0.21	0.73
	Prop. of wetland area that is GIW	-0.09	0.14	0.26	0.55	0.38	0.62	0.29	0.50	-0.12	0.59
Wetland	Floodplain	0.64	0.28	0.84	0.36	0.55	0.19	0.92	0.30	0.57	-0.17
	NWI wetlands	-0.27	0.48	0.45	0.81	0.86	0.85	0.46	0.80	-0.13	0.60