https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

Hydrologic Model Parameter Estimation in Snow-Dominated Headwater Catchments Using Multiple Observation Datasets

Lauren H. North[1], Adrienne M. Marshall[1], Glenn A. Tootle[2], Lisa Davis[3], Andy W. Wood[1,4], Eric J. Anderson[1]

[1]Hydrologic Science and Engineering, Colorado School of Mines, Golden, 80401, United States of America
[2]Civil and Environmental Engineering, University of Alabama, Tuscaloosa, 35401, United States of America
[3]Department of Geography, University of Alabama, Tuscaloosa, 35401, United States of America
[4]National Center for Atmospheric Research, Boulder, 80305, United States of America

Correspondence to: Lauren H. North (north@mines.edu)

Abstract. Hydrologic models are often calibrated only using streamflow, but increasing availability of in situ and satellite based observations provide numerous opportunities to constrain model outputs and improve process representation. However, as new observation data emerges, it is often unclear whether calibration with additional data would inform or misinform streamflow prediction. Here, we carry out a multi-observational sensitivity and uncertainty analysis using the U.S. Geological Survey's National Hydrologic Model (NHM) in four headwater catchments in the Upper Colorado River Basin. We use seven different observational data products that pertain to discharge, snow water equivalent, snow-covered area, soil moisture, and evapotranspiration. Informative model parameters are identified using the Morris screening method across all data sets, followed by parameter estimation and streamflow performance assessment using a Latin Hypercube Sample Monte-Carlo filtering approach. Results show that an increased number of informative parameters are determined through the screening process with the use of observation data representing terms beyond streamflow, and that forcing corrections and rain-snow partitioning parameters are particularly impactful to the model fit to observations. Multi-objective Monte Carlo filtering reduces the number of behavioral parameter sets, and estimated parameter values can depend strongly on the observation data criteria. Evapotranspiration is informative for streamflow prediction across all catchments included in this study, but snow and soil moisture datasets are only informative in some. These results provide new insight into the variable value of alternative observation data for streamflow prediction and highlight challenges related to model/observation scale mismatches, compensating errors, and misinformative data.

1 Introduction

Improved scientific understanding of hydrologic processes, the growth of observational data, and advancements in computational power have led to the development of complex, spatially distributed hydrologic models (Beven, 1996; Gupta

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

et al., 1998). These models help provide essential services to the public, such as water supply forecasting (Gorski et al., 2025), flood forecasting (Emerton et al., 2016; Hogue et al., 2000), drought monitoring (Hao et al., 2017; Pendergrass et al., 2020), and analysis of the impact of climate variability (Christensen et al., 2004). In the western United States, the Colorado River Basin is of particular interest because it provides an invaluable resource for 40 million inhabitants across several metropolitan and agricultural areas, and it is particularly vulnerable to climate variability and drought (Nash & Gleick, 1991; Wheeler et al., 2022). Mountainous headwater catchments provide up to 92% of the total annual runoff for the entire Upper Colorado River Basin (UCRB) (Lukas & Payton, 2020); therefore, long term forecasting and climate related modeling in this domain is of particular interest to government agencies, water managers, and water rights holders. The value of hydrologic forecasts depends largely on how well the model performs with respect to observations; generally, model evaluation is carried out retrospectively to assess this. Long term discharge records (Q) are often the primary, if not only, observation against which hydrologic models are calibrated (Gupta et al., 1998; Mei et al., 2023). In recent decades, several different data products have emerged from satellite based or airborne missions, measuring or estimating variables such as soil moisture (SM), actual evapotranspiration (AET), snow-covered area (SCA), and snow water equivalent (SWE). These alternative observations have the potential to improve modeled streamflow performance when used to enhance model calibration; however, their inclusion has uncertain outcomes (Herrera et al., 2022).

Process based hydrologic models simulate several hydrologic processes and output time series of the state variables, which provides an avenue to compare with alternative observations. It has been widely reported that the use of gridded AET products improves streamflow performance when used in calibration (Dembélé et al., 2020; Huang et al., 2020; X. Liu et al., 2022; Livneh & Lettenmaier, 2012; Mei et al., 2023). Soil moisture (Mei et al., 2023; Oubeidillah et al., 2019) and terrestrial water storage (Hasan et al., 2025; Rakovec et al., 2016) have also been found to be informative. However, some studies have shown that in other cases, SM and AET can be misinformative or require bias correction depending on the product used (Kunnath-Poovakka et al., 2016; Széles et al., 2020). While in situ SWE observations have been shown to be informative for streamflow modeling in snow-dominated catchments (Livneh & Badger, 2020), there are several challenges associated with snowpack modeling, especially when semi-lumped model outputs are compared to point-scale measurements (Cho et al., 2022; Gelfan et al., 2004; Lundquist et al., 2013; Mazzotti et al., 2023). Remotely sensed SWE from the Airborne Snow Observatory (ASO) program has emerged within the last decade (Painter et al., 2016), with spatial patterning results that suggest in situ SWE observations are often not representative of the surrounding landscape (Herbert et al., 2024).

Multi-observational calibration studies have shown that streamflow prediction performance varies between datasets or combinations of multiple datasets (McCabe et al., 2005; Mei et al., 2023), and the results can depend on the catchment scale (Livneh & Lettenmaier, 2012). Uncertainties associated with calibration data can result in vastly different parameter estimates (Bárdossy & Singh, 2008), resulting in deleterious effects in flood hazard forecasting (Balbi & Lallemant, 2023), peak flow estimates (Bárdossy & Anwar, 2023), and climate change studies (Marshall et al., 2021). It has long been noted

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

that observation quality is an important source of uncertainty in multi-objective calibration (Gupta et al., 1998). These challenges underscore the importance of developing multi-objective calibration, uncertainty quantification, and diagnostic procedures for models (Gupta et al., 2008).

Calibration and uncertainty analysis requires sampling parameters at a high density; however, this can become computationally expensive for models with many parameters and long run times (Razavi et al., 2021). Thus, a type of sensitivity analysis, parameter screening, often precedes further analysis to identify informative/highly sensitive parameters (Pianosi et al., 2016; Saltelli et al., 2019). Screening out non-informative parameters reduces the dimensions of the parameter space, which reduces sample sizes and computational demand. Keeping in mind the goal of calibration or uncertainty quantification, objective functions can be used in a sensitivity analysis, referred to as "identifiability analysis" when sensitivity is assessed relative to an objective function, following the terminology of Gupta & Razavi (2018). The value of alternative observations in an identifiability analysis context is relatively unexplored, and multi-objective methods are only briefly discussed in the most recent reviews (Pianosi et al., 2016; Song et al., 2015).

Model structures and spatiotemporal simplifications rely on parameters to account for unresolved or unobserved physics when calibrated to observations (Pathiraja et al., 2016). This introduces uncertainty since many parameter sets may return simulations with acceptable performance, known as equifinality, or getting the "right answer for the wrong reason" (Beven & Freer, 2001; Kirchner, 2006). The advent of remotely sensed observation products allows the conditioning of multiple model state variables rather than just streamflow, and various multi-objective approaches have demonstrated improvements in streamflow performance and reduced model uncertainty (Choi & Beven, 2007; Dembélé et al., 2020; Y. Liu et al., 2012; Shafii et al., 2015; Vrugt et al., 2005). Computational resources in recent decades have permitted the development of large domain, physically based modeling infrastructure such as the North American Land Data Assimilation System (Mitchell et al., 2004), National Oceanic and Atmospheric Administration national water model (J. M. Johnson et al., 2023) and the U.S. Geological Survey (USGS) national hydrologic model (NHM) (Regan et al., 2019). These models are built to support nationwide water prediction initiatives and have been subject to extensive calibration (Hay et al., 2023; Nassar et al., 2025). However, there is a remaining need to assess whether alternative observations are informative to streamflow prediction, especially as more data products become available. For example, gridded soil moisture observations from the Soil Moisture Active Passive (SMAP) satellite mission, lidar based snow depth observations, and new evapotranspiration products could be used to assess the USGS NHM (Hay et al., 2023), but have not previously been applied in this context.

In this study, we present a multi-observational sensitivity and uncertainty analysis that leverages seven publicly available observation datasets. We employ a newly developed python package by the USGS, pywatershed, to run a process based, semi-distributed hydrologic model in four headwater catchments in the UCRB. Our approach begins with a screening method to identify informative parameters for each dataset. The informative parameters are carried into a Latin Hypercube

Sampling (LHS) design to assess parameter estimation and streamflow performance using a Monte-Carlo filtering approach. The remotely sensed products used for model evaluation are SM from the soil moisture active passive (SMAP) mission, airborne lidar SWE from ASO, snow-covered area from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite based product, and evapotranspiration from the OpenET project. We also include in situ measurements of SM, SWE, and Q to examine the outcomes of in situ versus remotely sensed observations. The SMAP, ASO, and OpenET products are relatively new, and studies incorporating both multi-variable and multi-dataset objectives remain rare. Three research questions guide this work:

1. How does the use of alternative observations in sensitivity analysis impact the outcomes of parameter screening?
2. How does the use of alternative observations in model calibration affect parameter estimation and streamflow prediction?
3. Are the findings consistent among different UCRB headwater catchments?

2 Methods

2.1 Selected hydrologic model

The distributed parameter hydrologic model used in this study is the US Geologic Survey's pywatershed. It is the successor of the Precipitation Runoff Modeling System (PRMS) (Leavesley et al., 1983; Markstrom et al., 2015). Pywatershed is a python package with the goal of modernizing legacy software and increasing flexibility. As is the case for PRMS, pywatershed is a deterministic, distributed parameter, physical process based hydrologic model. The modeling domain is discretized into hydrologic response units (HRUs) that are delineated through a variety of topographic, geologic, and climatologic factors. Each HRU is modeled as a homogenous unit, where energy and mass balances are computed at 12 hr and 24 hr timesteps, respectively. The simulated hydrologic response is conceptualized through a series of storage reservoirs (such as snowpack or the soil zone), stream segments, lakes, and fluxes between them.

PRMS is the primary component of the USGS National Hydrologic Model, a modeling system over the conterminous United States (CONUS) that includes a database of parameters and climate inputs (Regan et al., 2018, 2019). Extracts of the NHM are used to provide baseline parameter and climate input files specific to our watersheds of interest. Pywatershed requires three daily climatologic inputs for each HRU: minimum temperature, maximum temperature, and precipitation (Markstrom et al., 2015). The climate-by-HRU files sourced from the NHM are developed using the 1 km Daymet product (Thornton et al., 2016).

Pywatershed contains 145 parameters pertaining to hydrologic processes and HRU attributes. Parameters that represent geographic information, boundary conditions, or model configurations are considered "non-calibration" parameters and are

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

not modified from their initial values (Viger, 2014). Based on four recent PRMS sensitivity and calibration studies, we selected 51 calibration parameters (Douglas-Mankin & Moeser, 2019; Hay et al., 2023; Markstrom et al., 2016; Mei et al., 2023). Of these 51 parameters, four snow albedo parameters are included only in the present study and are marked in table S1. These parameters are included because radiative forcing is an important process in snowpack modeling, especially for wildfire related studies (Gleason et al., 2019; Maxwell & St Clair, 2019; Skiles et al., 2018). In this study, pywatershed was run from water years 1982 through 2022 (41 years) for the sensitivity analysis, and from 2013 through 2022 (10 years) for the Monte Carlo filtering.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

2.2 Study catchments in the UCRB

The Blue, Dolores, East, and Taylor River catchments each contain 1 or 2 USGS stream gages and between 1 and 5 in situ SWE and SM measurement sites (Figure 1). At least two lidar based SWE acquisitions from the ASO program are available in each catchment during our study period (up to WY 2022). The East and Taylor River share a catchment boundary. The four altogether represent a range of climatologic and geographic attributes in the Upper Colorado headwaters (Table S2, Figure 1).

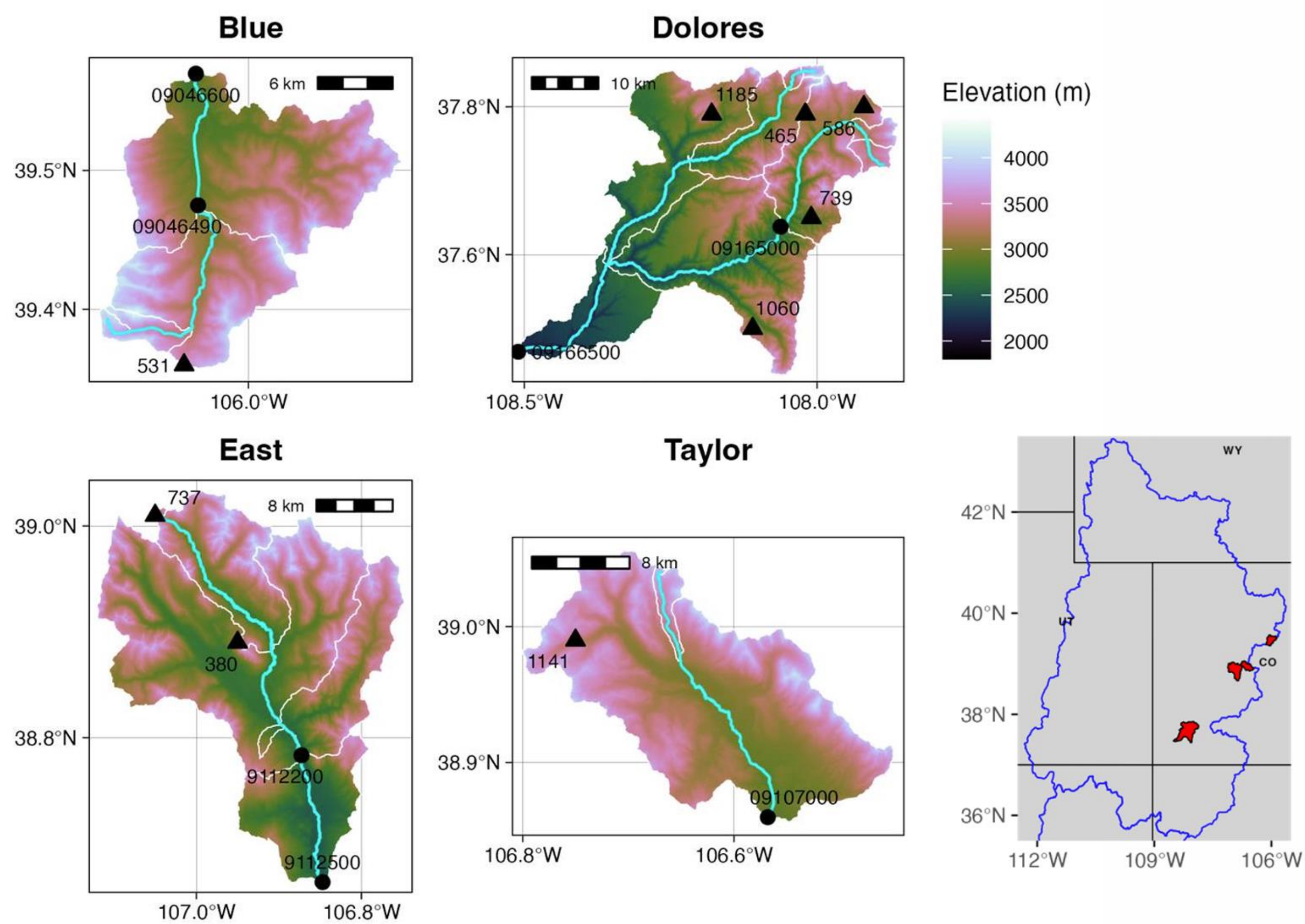

Figure 1. Digital elevation models of the four headwater catchments included in this study. White lines indicate HRU boundaries. Black dots represent USGS stream gages and black triangles represent NRCS SNOTEL stations.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

2.3 Observation data for model evaluation

We use seven observation data products for model sensitivity and calibration analysis in this study, consisting of both in-situ and remotely sensed measurements (Table 1). Daily mean discharge observations were obtained through the USGS National Water Information System dataRetrieval package in R (DeCicco et al., 2024). The locations of USGS stream gauges correspond to downstream points of stream segments in the model. Daily mean in situ SWE and SM was obtained from the NRCS SNOTEL (Snow Telemetry) online report generator. It is common for these observation stations to fall near HRU boundaries (Figure 1), which likely reduces their representativeness of a homogenous HRU.

High spatial resolution (50m) SWE rasters from ASO are derived from airborne LiDAR measurements where the aircraft flies over the catchment of interest (Painter et al., 2016). These acquisitions pertain to specific catchments and are requisitioned by water managers one to several times throughout a water year. The low temporal resolution is unique to this dataset; however, the spatial completeness allows for a more robust estimate of HRU mean SWE compared to SNOTEL point observations. The MODIS snow covered area product used in this study (MOD10A1.061) was used to provide an estimate of fractional snow-covered area (fSCA) and required additional screening and transformations before it could be directly compared to the model output (Supplementary Text S1).

Table 1. Observations used in the parameter sensitivity and estimation workflow. Start years are approximate for in situ observations as each station began recording in different years. All outputs from the model area in daily timesteps.

Variable		Simulated		Observed			
		Output	**Spatial**	**Source**	**Temp.**	**Start**	**Spatial**
Discharge	Q	seg_outflow	Seg.	USGS	Daily	>1980	Point
Snow Water Equivalent	SWE	pkwater_equiv	HRU	SNOTEL	Daily	>1980	Point
				ASO	Intermittent	>2021	50 m
Soil moisture	SM	soil_rechr	HRU	SNOTEL	Daily	>2000	Point
				SMAP	Daily	>2015	9 km
Snow Covered Area	SCA	snowcov_area	HRU	MODIS	Daily	>2000	500 m
Actual evapo-transpiration	AET	hru_actet	HRU	OpenET Ensemble	Monthly	>2013	30 m

The soil moisture observations also required pre-screening and transformations to be suitable for model evaluation (Supplementary Text S2). The soil moisture sensors at NRCS SNOTEL stations are at 2 cm, 8 cm, and 20 cm depths. Of the three in situ depths, the 2 cm depth observation had the best fit to the model and was used in our analysis. The level 4 SMAP product obtained from GEE (SPL4MGP.007) provides measures of saturation at the surface, rootzone, and soil profile at a 9 km spatial resolution, every 3 hours from March 31st, 2015 to present. Between the SMAP surface and rootzone wetness measurements, the surface zone had the best fit to the model and was used in the final analysis. It is aggregated by HRU and as a daily mean. The model simulates soil moisture storage in a conceptual reservoir, which does not have a physical depth in the soil column. This poses a challenge since it is not directly comparable to observations. Since the simulated and observed values do not match in magnitude, all are normalized between 0 and 1 to compare temporal variability (Hay et al., 2023). We note that there is considerable misalignment between the simulated and observed soil moisture (Figure S1), which limits the realization of behavioral models in this respect.

The OpenET product provides AET at a 30 meter resolution using an ensemble mean of multiple satellite based observations and models (Melton et al., 2022). Monthly AET was area-averaged to the HRU scale for comparison with the modeled AET output. The satellite remotely sensed observations (SCA, SMAP, AET) were obtained using Google Earth Engine (GEE) and HRU geometry files from the NHM.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

2.4 Selected performance metric

Summarizing a time series of the model's error or behavior into a statistic is a necessary step in model diagnostics and has strong implications for the results of sensitivity analysis and Monte Carlo filtering. Here, we use the normalized root mean squared error (NRMSE). It is calculated by normalizing the root mean squared error by the standard deviation of the observations σ_O, as shown in equation 1:

$$NRMSE = \frac{RMSE}{\sigma_O} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(S_t - O_t)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(O_t - \overline{O})^2}} \tag{1}$$

where T is the total number of t timesteps in the evaluation, S_t and O_t are simulated and observed values at each timestep, and $\overline{O}$ is the mean of the observations. This metric allows for comparisons between different catchments and observations because it is not in absolute units, and it accounts for the inherent variability of the location and data. Its value can be interpreted as a proportion: for example, a NRMSE of 0.5 means that the error is half of the variability in the observations. Squared error based metrics suffer limitations such as sensitivity to outliers and errors during high flows (Gupta et al., 2009); however, its interpretability is favorable for comparisons between observations and the rejection of non-behavioral models. NRMSE was used in a recent multi-observational, CONUS-wide NHM calibration study led by the USGS (Hay et al., 2023), and using the same metric makes this work relevant to current agency procedures. For remotely sensed observations, NRMSE is reported as a catchment-wide HRU area-weighted mean, computed over the entire available time series. In the case of ASO, which does not have a time series, NRMSE is reported on a by-acquisition basis.

2.5 Identifying sensitive model parameters using the Morris Method

To identify parameters to be used in the model calibration, we first conduct a type of parameter sensitivity analysis known as screening (Pianosi et al., 2016). This is typically done for models with a large number of calibration parameters (51 in this study) - an important outcome being a reduction in the number of parameters for further analysis. Here, we use the Morris Elementary Effects method, which coarsely samples the parameter space using a one-at-a-time (OAT) approach and is a relatively computationally efficient screening method (Herman et al., 2013b; Morris, 1991; Pianosi et al., 2016). We discuss two sensitivity measures: μ^* to describe the magnitude of parameter sensitivity (Campolongo et al., 2007), and a normalized metric η^* for the screening process (Cuntz et al., 2015). We use 51 trajectories in our sampling design (Cuntz et al., 2015; Gan et al., 2014), and use 1000 bootstrap replicates to identify type I (false positive) and type II (false negative) statical errors in our screening approach (Supplementary Text S3) (Campolongo & Saltelli, 1997; Saltelli et al., 2007). The Morris sampling algorithm and analysis for calculating sensitivity indices was carried out using the sensitivity package in R (Iooss et al., 2024).

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

The spatial and temporal variability of the sensitivity measures are also presented in this study. When observations are available, we calculate the sensitivity measures in 3 ways: (1) the full period in which forcings and observations are available, (2) at annual intervals between, and (3) for 10 year moving windows stepped in one year increments. Since discharge observations return a high number of sensitive parameters and have the longest period of record, we limit our scope to these observations for temporal analysis. In terms of spatial analysis, we leverage the spatially distributed remotely sensed observations to assess the influence of HRU attributes on parameter sensitivity based on Spearman Rank correlations.

2.6 Monte Carlo filtering to assess parameter equifinality

Following the identification of informative parameters, we employ a simple uncertainty analysis technique known as Monte-Carlo filtering to evaluate performance and parameter estimation. The method involves choosing an objective function (NRMSE in this case) and setting a threshold for behavioral ("good") or non-behavioral ("poor") performance. The parameter space is stochastically sampled with a high number of replicates, and the model is run with each parameter set. Depending on the performance criteria and observation data, usually several parameter sets will return as behavioral. The non-behavioral models are filtered out, and the behavioral simulations are used to assess model parameter value uncertainty (equifinality) and performance relationships (Shafii et al., 2015). In this work we select our behavioral threshold as NRMSE < 1.0, where the model error is less than the inherent variability in the observations. This threshold is a common benchmark and is analogous to a Nash-Sutcliffe Efficiency of 0.0, as these two metrics are related (Althoff & Rodrigues, 2021; Manikanta & Vema, 2022; Ritter & Muñoz-Carpena, 2013). To assess how the inclusion of alternative observations affects streamflow calibration, we define multi-objective criteria as joint constraints where the model performance is behavioral with respect to discharge *and* alternative observations. This criterion leads to relatively few behavioral models for the intersection of discharge and SCA, SMS2, and SMAP - leading us to relax the threshold to an NRMSE of 1.5 for these three alternative datasets only.

Latin Hypercube Sampling (LHS) is a sampling approach commonly applied in sensitivity and uncertainty analysis of complex models with a high number of parameters (Helton & Davis, 2003; Sheikholeslami & Razavi, 2017; Shields & Zhang, 2016). It is a suggested approach for generating parameter sets in Monte Carlo filtering based frameworks so the parameter space is uniformly sampled and equifinality can be assessed (Beven & Freer, 2001). In this study, we use maximinLHS function from the R lhs package (Carnell, 2024), which iteratively solves statistical criteria to maximize the minimum distance between sampling points (M. E. Johnson et al., 1990). This method is recognized for producing well-distributed, space-filled samples (Chen et al., 2017; Santner et al., 2018). Following recommendations in existing literature, we use 1000 trajectories per parameter in the LHS design (Pianosi et al., 2016). Due to the higher number of trajectories and computational limitations, the simulation period is reduced to one decade (2013-2022) in this experiment. Since pywatershed

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

contains several non-scalar parameters that are spatially distributed by HRUs, temporally distributed by month, or both, we preserved a priori spatiotemporal distributions from the NHM during the Morris and LHS experiments. We apply the "use the mean" procedure from previous PRMS analysis to address the complexity associated with non-scalar parameters (Hay et al., 2006; Hay & Makiko, 2007).

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

3 Results

3.1 Sensitivity analysis

3.1.1 Identifying sensitive parameters for calibration

Using the Morris method with seven different observation datasets, we find that both the number and type of parameters identified as informative are considerably influenced by the target observation (Figure 2). The 51 parameters are binned into their process representation or "module" in pywatershed (right-hand labels). Several parameters emerge as informative across all observations, particularly in the climate module. These include *tmax_allsnow* (rain-snow partitioning), *tmax_cbh_adj,* and *tmin_cbh_adj* (forcing corrections), which highlight the strong effect of meteorological forcings on multiple model processes. The Jenz-Haize potential evapotranspiration coefficient (*jh_coef*), which acts as an empirical multiplier for potential evapotranspiration, is highly sensitive for all non-snow related outputs. Two parameters governing groundwater flow, *gwflow_coef* and *soil2gw_max,* are only identified by discharge data, and the latter is the only standalone type II error across all observations.

The model fit to snow data is sensitive to parameters in the snow module, and as expected in snow-dominated headwaters. Of the four albedo parameters added to our study, none were identified as informative. Simply based on count, discharge observations consistently elicit the largest number of sensitive parameters, with between 10-14 parameters identified among the four catchments (Table S3). The SNOTEL, ASO, and SMS2 observations identify comparable numbers of parameters (10 to 16), but the precise number of identified parameters varies among catchments. Across all observation datasets, 18-22 informative parameters are identified (22-25 including type II errors). Many of the SNOTEL and ASO parameter identifications have a high number of type I and type II errors, shown by black squares and triangles in Figure 2. The frequency of these errors suggests that parameter sensitivity with respect to SWE has high variability. This is supported by visualizations of the fitted logistic function, where the error bars for the informative parameters are considerably larger (Figure S2-S5). While uncertain, the SWE related observations contribute one to six parameters in addition to what is identified by discharge, and the other observations generally contribute relatively fewer unique identifications. While SCA also assesses model performance with respect to snowpack simulation, it returns a consistent, yet, smaller number of identifications and does not exhibit the same extent of statistical errors as the SNOTEL and ASO data.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

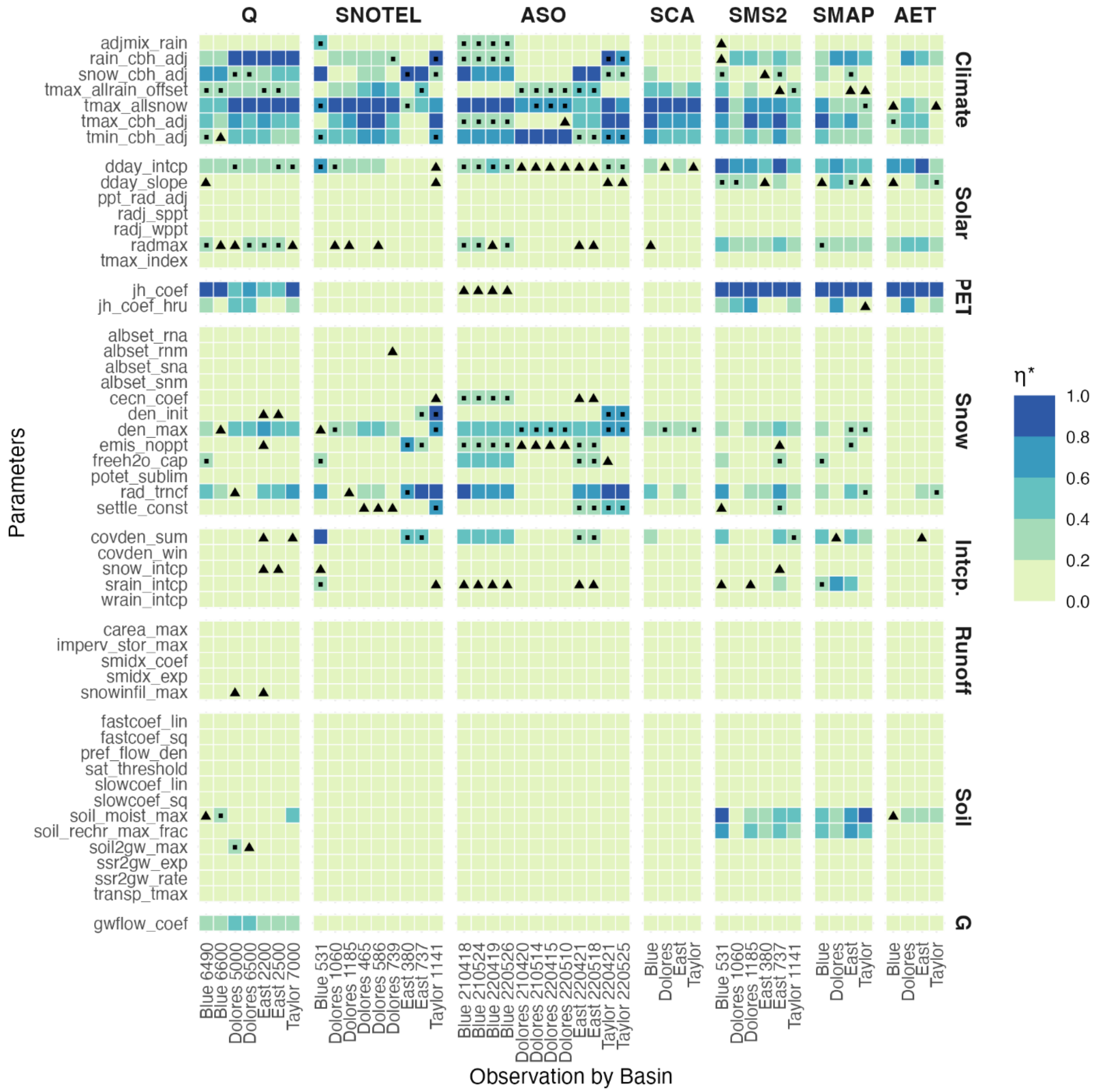

Figure 2. Normalized sensitivity metric η^* for all catchments and target calibration observations available from WY 1982 through 2022. The black squares denote type I errors and the triangles denote type II errors identified via bootstrapping, using an uncertainty bound of $\pm 1.0 \times SD(\mu_i^*)$. In the x axis text, the numbers following the catchment name correspond to the last four digits of the USGS gage ID for Q, the NRCS site ID for SNOTEL and SMS2, or the ASO acquisition date in yymmdd format.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

3.1.2 Parameter sensitivity to annual forcing anomalies

There is considerable interannual variability in the magnitude of parameter sensitivity with some influence of annual temperature and precipitation anomalies. For example, *jh_coef* was relatively sensitive in warm, dry conditions (Figure 3). We also see differences in the magnitude of sensitivity between catchments, such as *rad_trncf* being less sensitive in the Dolores compared to the others. Across the four most sensitive parameters, we find relative increased sensitivity in dry conditions (with respect to discharge observations). For this analysis, we use μ^* rather than η^* to assess the overall magnitude of sensitivity.

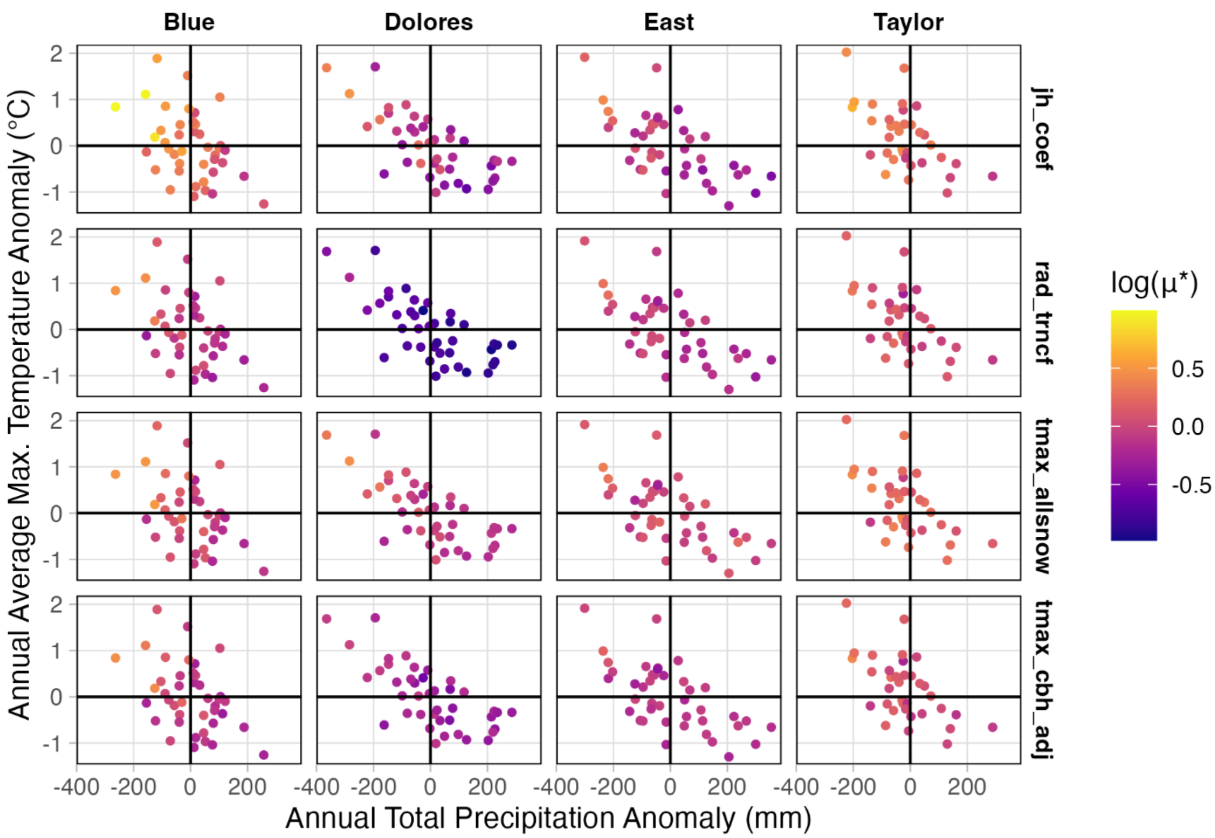

Figure 3. Annual sensitivity measures μ^* for select parameters with respect to discharge observations at the most downstream stream gage point. The sensitivity indices were logged for visual interpretation.

Time series analysis of parameter sensitivity illustrates these modest climate sensitivities across all parameters (Figures S6-S9). On an annual scale, the number of identified parameters ranges from 12 to 17, with the largest variability in snow and interception parameters (Figure S6). In terms of sensitivity magnitude, the climate, PET, and snow parameter groups show the most variability (S7). When the sensitivity measures are computed over a 10 year moving window, the screening results are far more stable (Figure S8). Parameters that are near the screening threshold fluctuate between being identified as

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

sensitive versus not, such as *radmax*. This is likely due to changes in the most sensitive climate parameters (S9), since their sensitivity indices are used to produce the normalized metric for screening.

3.1.3 Relationship of geographic attributes and parameter sensitivity

Among the highly sensitive parameters, relationships between HRU attributes and parameter sensitivity range from weak to moderate (Figure 4). Correlations among HRU average annual precipitation and μ^* were generally statistically significant but with limited explanatory power. The strongest correlation is a negative correlation between *rad_trncf* (solar radiation transmission through the canopy) and winter maximum temperature (Tmax); *rad_trncf* is also more sensitive at high elevations and relatively high precipitations. Similarly, *jh_coef*, *tmax_allsnow,* and *tmax_cbh_adj* are each generally more sensitive at higher elevations and cooler temperatures. While *jh_coef* is less sensitive to snow-related observational data, it shows a similar relationship with the climatic attributes across each observation type. Overall, the magnitude of sensitivity for these select parameters is positively correlated with precipitation and elevation and negatively correlated with temperature. We expect these relationships due to the strong covariance between climate forcings and elevation (Figure S10), but the differences in explanatory power suggest there are other confounding factors. These results demonstrate that while HRU attributes have relatively low predictability of precise sensitivity measures, the most sensitive parameters are moderately associated with temperature and elevation in particular.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

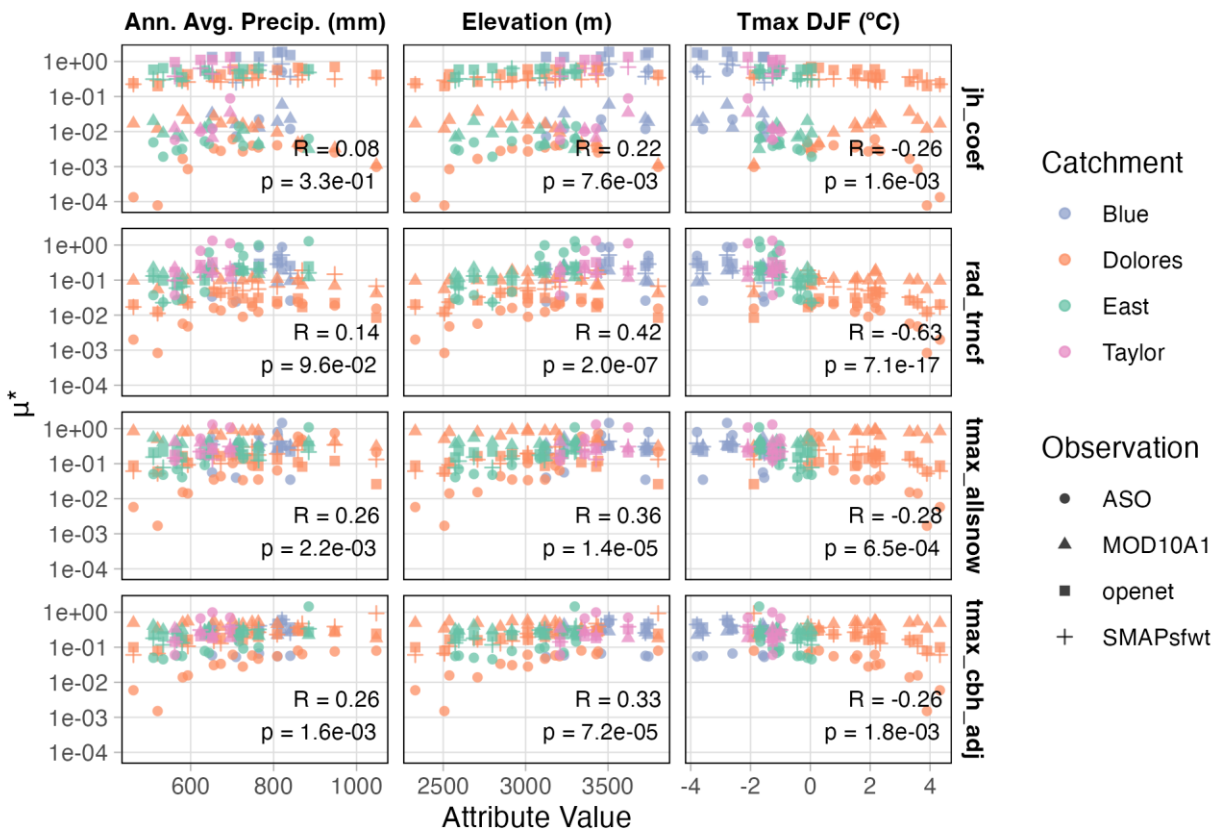

Figure 4. Morris bootstrapped sensitivity metric μ^* for selected parameters versus geographic attributes (annual average precipitation, elevation, and maximum temperature in DJF). Spearman rank correlation values are denoted by R with an associated p-value. Each point represents a by-HRU sensitivity measure for a specific observation dataset, catchment, and HRU attribute. The y-axis is logged for improved interpretation of the absolute sensitivity measures.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

3.2 Monte Carlo filtering calibrations

3.2.1 Constraining equifinality with multiple observations

With NRMSE of daily discharge as the governing objective, we demonstrate how the inclusion of alternative observations can constrain equifinality. Some catchments yield a much higher number of behavioral simulations than others with respect to discharge NRMSE alone (Table 2). For example, out of 22,000 simulations for the Dolores, over 20% are returned as behavioral, while less than 3% of 24,000 simulations are returned from the Blue. The Dolores and the East are the larger of the four catchments and return greater proportions of behavioral streamflow simulations, suggesting that model may be more representative over larger scales compared to headwater catchments.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

Table 2. Number of behavioral models for where the filtering threshold is set to NRMSE < 1.0. The percentage of total simulations is shown in parentheses. Where there are multiple rows, each row denotes a specific in situ observation station or ASO acquisition. The numbers in bold font indicate the greatest overall improvement in NRMSE(Q) of that criterion.

Criteria		Number of behavioral parameter sets			
		Blue	**Dolores**	**East**	**Taylor**
Q	Upstream	846 (3.53%)	5116 (23.3%)	4159 (16.6%)	1543 (7.01%)
	Downstream	2887 (12.0%)	6098 (27.7%)	3642 (14.6%)	-
	Both	654 (2.73%)	4621 (21.0%)	3577 (14.3%)	-
Q ∩ SNOTEL		561 (2.34%)	4047 (18.4%)	2525 (10.1%)	1265 (5.75%)
		-	4267 (19.4%)	**3260** (13.0%)	-
		-	1894 (8.61%)	-	-
		-	**3814** (17.3%)	-	-
		-	1314 (5.97%)	-	-
Q ∩ ASO	Apr. 21	13 (0.05%)	3858 (17.5%)	-	-
	May 21	186 (0.78%)	**3828** (17.4%)	-	-
	Apr. 22	**57** (0.24%)	3783 (17.2%)	**2282** (9.13%)	610 (2.77%)
	May 22	95 (0.40%)	2548 (11.5%)	889(3.56%)	**158** (0.72%)
Q ∩ SCA*		17 (0.07%)	569 (2.59%)	357 (1.43%)	58 (0.26%)
Q ∩ SMS2*		413 (1.72%)	**3084** (14.0%)	2226 (8.90%)	451 (2.05%)
		-	6 (0.03%)	**1691** (6.76%)	-
		-	0	**-**	-
Q ∩ SMAP*		4 (0.02%)	0	133 (0.53%)	45 (0.20%)
Q ∩ AET		187 (0.78%)	4056 (18.4%)	1540 (6.16%)	735 (3.34%)

* Denotes where the NRMSE threshold for the alternative observations is 1.5

Multiple ASO acquisitions show that behavioral SWE simulation is highly dependent on the catchment and date of acquisition. The larger catchments have overall a greater number of behavioral ASO simulations. Results in the Blue suggest that the modeled SWE is less erroneous in May than April. However, the other catchments do not provide a clear indication whether the model better represents April versus May SWE.

While hundreds to thousands of simulations are behavioral for the intersection of Q and SWE, the same filtering threshold (NRMSE < 1.0) resulted in zero to a few dozen parameter sets for SCA, SMS2, and SMAP. Instead of discarding the

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

information, we relaxed the threshold to 1.5 for these calibrations to permit further interpretation. Even then, very few behavioral parameter sets are yielded for the soil moisture observations in the Dolores catchment (Table 2). We note that this performance level is within the range of output from the National Hydrologic Model (Table S4). This finding suggests potential issues with model-to-observation alignment. The AET performance was calculated using a monthly mean, which yields behavioral models under the stricter threshold (NRMSE < 1.0). This is expected due to the model-to-observation congruency and the suppression of daily variability.

3.2.2 The effect of multiple criteria on streamflow performance

Of the alternative observations used in this study, AET is the only one to consistently improve discharge performance (Figure 5). The empirical cumulative distribution functions of NRMSE(Q) in Figure 5 show that the intersection with AET yields a distribution that is better than Q alone (green line is left of the black line in Figure 5). The observations that induce a worse distribution in discharge performance can be considered misinformative. However, whether these observations improve or reduce discharge performance is dependent on the catchment. For example, ASO shows slight reductions in performance for the Dolores and East (right of the black line in Figure 5), but some of the best performance in the Blue and Taylor. The latter have fewer behavioral discharge simulations to a begin with - thus, the inherently reduced parameter space may affect how new observations inform the model. The effect of ASO observations on streamflow performance also varies among individual acquisition dates and sites (Figures S11-S14). Similarly, some SNOTEL locations induce performance improvements while others induce reductions. These results suggest that introducing alternative observations does not always lead to positive streamflow performance outcomes. However, in the case of pywatershed, monthly mean AET observations may be useful in this respect.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

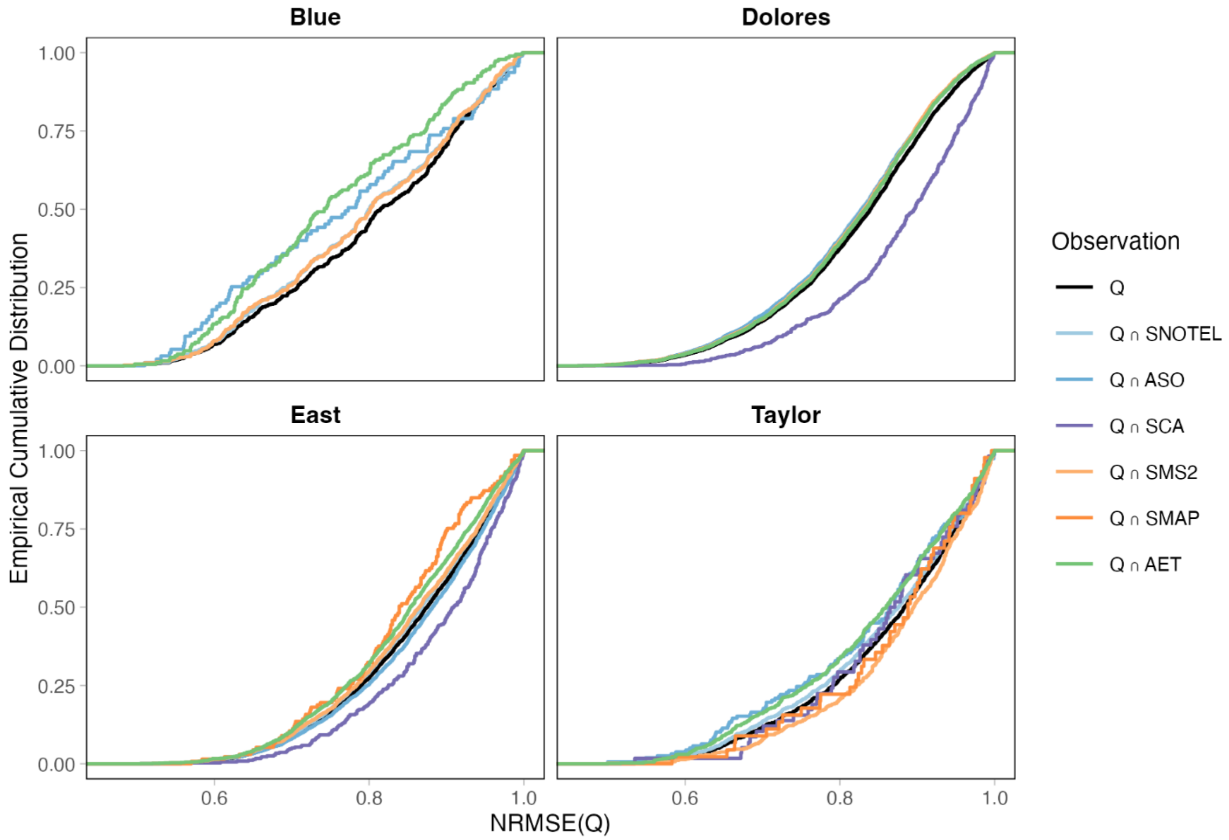

Figure 5. The cumulative distribution of streamflow performance, NRMSE(Q), is influenced by intersections with alternative observations. The black line represents the distribution when filtering with Q only. A distribution that is closer to zero (left) is considered informative in the case of NRMSE.

365

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

3.2.3 Parameter estimation

The effect of Monte Carlo filtering with alternative observations on behavioral parameter estimation ranges from marginal to pronounced. In most observation and catchment combinations, the parameter values span the entire possible range (Figure 6). This result alludes to issues with equifinality and the use of prescribed parameter ranges from the PRMS documentation. In only a few cases (such as the use of SMAP for *jh_coef* estimation in the East River), the introduction of additional observational data reduces the extent of behavioral parameter ranges. Parameter estimates from the three snow-related observations (SNOTEL, ASO, and SCA) show variable agreement. Overall, ASO is relatively consistent among acquisitions (Figure S15). The highly sensitive potential evapotranspiration parameter *jh_coef* is generally shifted to a lesser value when intersecting with AET observations

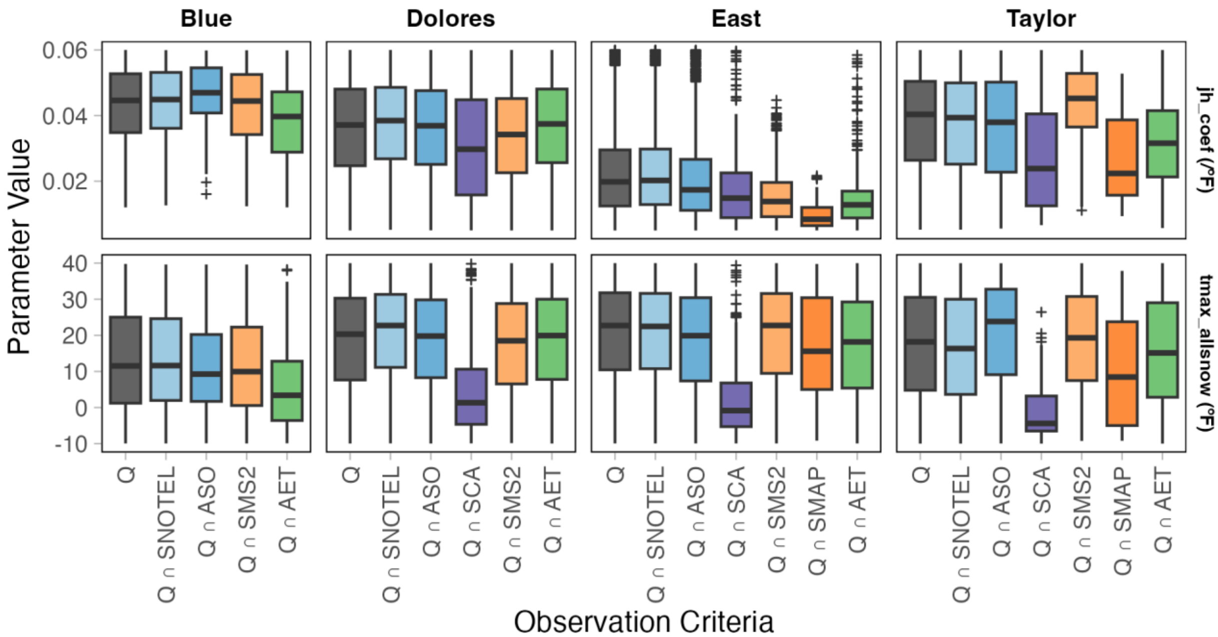

Figure 6. The inter-quartile ranges of behavioral parameter values fluctuate across observation intersection criteria. For catchments that have more than one stream gauge, the behavioral intersection of both stream gauges is used. For SNOTEL , SMS2, and ASO, the best performing site/acquisition is shown. Datasets that returned an insufficient number of behavioral parameter sets are excluded.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

4 Discussion

4.1 Model to observation alignment

A core challenge of using additional observational constraints in process based hydrologic modeling is that the simulated variables may not be physically well aligned with what is observed (McCabe et al., 2017). This becomes particularly evident in lumped or semi-lumped models (such as pywatershed) as well as coarsely gridded models (Ehlers et al., 2019; Motovilov et al., 1999), where *in-situ* point observations are compared to a much larger simulated area. This is the case for SNOTEL SWE and SMS2. In our case study watersheds, these observation points are near topographic high points and therefore are often near catchment or HRU boundaries (Figure 1). Observation points in these locations are likely not well-aligned with the simulated areal average snowpack or soil moisture over an HRU. We see from the Monte Carlo filtering that streamflow performance can be both hindered and improved by these point based SWE and SM measurements, but it is largely dependent on the station (Figures S10 – S13). Where the *in-situ* observations degraded streamflow performance, we attribute this to poor spatial representation of the HRU.

Aside from the spatial misalignment, the conceptual or physical representations may not fit observed quantities either. In the case of pywatershed, this issue was apparent when comparing simulated and observed soil moisture. The model uses a conceptual framework in terms of storage and fluxes, representing SM storage as a series of conceptual reservoirs. However, in situ measurements of SM are at discrete depths in the soil column, and similarly, satellite remotely sensed soil moisture from SMAP retrieves soil moisture at discrete depth ranges. To address these discrepancies, the model output and observations were normalized between 0 and 1 before comparison, as done in previous work (Hay et al., 2023). Yet, the conceptual misalignment seemed to persist during the Monte Carlo simulation, since the model failed to yield behavioral parameter sets when soil moisture was included in the performance criteria (depending on the catchment, Table 2). Similarly, the baseline parameter sets from the extensively calibrated NHM also perform poorly for soil moisture (Table S4, Figure A1). This suggests that alternative approaches for addressing the model-to-observation alignment may be needed. In Mei et al. (2023) where PRMS was calibrated with SM observations, both the simulated and observed datasets were treated as anomalies, which assesses timing rather than magnitude or variability. Brocca et al. (2014) showed that temporal SM anomalies show lesser spatial variability than absolute magnitude, and other SM calibration studies employ adjustments to in-situ and remotely sensed SM data to remove biases (Draper et al., 2009; Rajib et al., 2016). In another hydrologic calibration with SMAP, temporal correlations were used to assess performance (Koster et al., 2018). In light of the normalization technique used in this study, additional bias corrections or temporal relationships should be explored in future work.

While remotely sensed observation products largely address the point-to-HRU challenges of in situ observations, they are still subject to model-to-observation challenges or uncertainty in the observations themselves. Area averaged ASO SWE,

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

MODIS SCA, and OpenET AET are theoretically well-aligned with the HRU based output of pywatershed. However, uncertainty in remotely sensed data is complex, stemming from the sensors, cloud conditions, surface conditions, spatial sampling, and post-processing (Povey & Grainger, 2015). For example, the ASO SWE product is based on airborne lidar retrievals of snow depth and the snow density is modeled post-hoc, which inherently introduces uncertainty (Painter et al., 2016). The OpenET product is based on satellite optical data, weather data, and an ensemble of models, but yields a single estimate of AET (Melton et al., 2022). The MODIS SCA product and SMAP SM product also yield a singular estimate, which has well documented uncertainties (P.-W. Liu et al., 2021; Stillinger et al., 2023). Despite the well-known challenges of remotely sensed observations, in this paper they are not explicitly addressed, and we instead opt for "out of the box" implementations.

4.2 On identifying sensitive parameters in headwater catchments

Other PRMS sensitivity analysis studies agree with our overall findings. The most recent and comprehensive sensitivity analysis by (Markstrom et al., 2016) encounters similar results: the sensitivity in mountainous headwater catchments is largely driven by a select few parameters. They show that parameters such as *jh_coef* and *tmax_allsnow* explain the majority of parameter sensitivity. Other studies with different models also show that rain-snow partitioning and forcing corrections are highly sensitive components in modeled streamflow in mountainous headwaters (Mai et al., 2022; Singh et al., 2024). To the extent that the most sensitive parameters represent corrections of errors in forcing inputs, these results corroborate arguments that forcing input uncertainty is generally greater than model errors (Lundquist et al., 2019). Across two large scale studies, evapotranspiration emerges as the primary component of model sensitivity, or "dominant process" in the UCRB region (Mai et al., 2022; Markstrom et al., 2016).

In the present study, few to no runoff parameters were identified as informative. We posit two reasons for this: (1) that snowmelt has a much greater implication to runoff timing and volume, and (2) that pywatershed has a large emphasis on forcing data adjustments. In support of the first line of reasoning, the runoff parameter *snowinfil_max* is identified as a type II error in the East River (Figure 2), which suggests the importance of snowmelt. Snowpacks are the primary contributor to runoff volume in high elevation, snow-dominated catchments and rainfall has marginal contributions to runoff volumes (with the exception of rain on snow events) (Hammond & Kampf, 2020; Li et al., 2017). Since squared error based objective functions (NRMSE in this case) strongly penalize errors at high flows (Gupta et al., 2009), where the model inaccurately simulates the timing or magnitude of the spring snowmelt driven streamflow pulse, the parameters driving that inaccuracy would be deemed sensitive. Snowmelt also contributes to seasonal soil moisture regimes in mountainous catchments (Harpold et al., 2015), which could explain why we may not see sensitive runoff parameters for the fit to soil moisture observations either.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

Secondly, the forcing adjustment parameters and their ranges from the PRMS documentation influence parameter identification. Complex models with a large number of parameters often exhibit nonlinear sensitivities and strong parameter interactions (Saltelli et al., 2019), and previous work has shown that the selection of parameter ranges impacts the outcomes of sensitivity analysis (Shin et al., 2013). For example, *tmax_allsnow* has a wide range (Figure 6) and represents the monthly maximum temperature where precipitation is assumed to be snow (Table S1). Rain to snow partitioning parameters often are constrained within a few degrees of freezing on shorter timescales (Jennings et al., 2018), but this representation in the model makes it more of a tuning parameter with a less clear physical basis. The σ indices yielded from the Morris experiment indicate that this parameter has relatively high nonlinear effects (Figure S16) - if the range were narrowed, the sensitivity index μ^* would likely change. Since the sensitivity index for screening (η^*) is normalized by the maximum μ^*, this has implications for which parameters are identified as informative. Additionally, the Morris method may not provide as reliable indices for highly sensitive parameters when compared to quantitative methods, such as Sobol (Herman et al., 2013b; Sobol′, 2001). However, in the context of model calibration and parameter estimation, it is important to prescribe parameter ranges that cover the optimal space while remaining efficient (Bárdossy & Singh, 2008; Mai, 2023). The effects of *a priori* parameter ranges are not addressed in this study, which provides an opportunity for improved pywatershed analysis in future work.

We clarify that the objective of this work is to assess the impact of observational data selection on parameter identifiability, rather than conventional sensitivity analysis. This distinction is made by Gupta & Razavi (2018), where an identifiability analysis focuses on model sensitivity with respect to observations (by using an objective function) versus sensitivity to the output itself. These methods are fundamentally distinct from each other. Choosing an objective function to summarize the model responses limits the interpretation of process importance because the "sensitivity" is influenced by how well the model tracks observations. Given that the choice of objective function has a pronounced impact on how model residuals are penalized, it therefore influences what parameters are considered informative. Our approach therefore cannot support the identification of dominant processes; however, it holds particular utility in parameter screening with the aim of calibrating a model to a suite of observations (Pianosi et al., 2016).

4.3 Model selection impacts on sensitivity and uncertainty analysis

We sought to explore some of the numerous choices that a modeler faces during a calibration experiment. The primary focus of this study was on the choice of calibration target data, as well as simulation period and catchment. Other important choices include the model itself, the calibration algorithm, forcing inputs, and the objective function. These choices were controlled for in our study by using a single model, a uniform sampling LHS design, one forcing dataset, and a grounded objective function threshold. A vast body of work discusses the nuances in inter-model comparison (Mendoza et al., 2015), advancement of calibration techniques (Mai, 2023), the uncertainty in forcing inputs (Tang et al., 2023), and the implications

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

of objective function choice (Lamontagne et al., 2020). While there are limitations to using a single objective function for model evaluation (Clark et al., 2021; Legates & McCabe Jr., 1999), for simplicity and scope we find NRMSE to be appropriate for a multi-observation framework (Gupta et al., 2008, 2009). There is also subjectivity in the design of the Morris and LHS experiments (Gan et al., 2014); we made these choices by following recommended values for discretization levels, trajectories, and rejection criteria (Cuntz et al., 2015; Pianosi et al., 2016).

Previous works have found that multi-observational calibration leads to better representation of hydrologic process and improved streamflow simulations (Finger et al., 2015; Smyth et al., 2020; Wongchuig et al., 2024; Zhou et al., 2020). Other multi-observational PRMS based studies have found that AET and soil moisture can improve streamflow performance (Mei et al., 2023). Our results only partially support this notion. While these findings are promising, a recent review poses the question “Increasing amount of collected data: to use or not to use?” (Herrera et al., 2022). We provide a conflicting answer to this question, as some observations constrained behavioral simulations into worse performing areas. One prior study found that the inclusion of ASO has positive implications for streamflow prediction in one California catchment with a large number of acquisitions (Lahmers et al., 2022). Our results agree with this finding in the Blue and Taylor but disagree in the Dolores and East.

We suggest a few possible reasons for the instances of poorer distributions of model performance: first, simulations that fail to adequately simulate intermediate state variables (such as SWE, SM, or AET) may have had structural compensating errors that ultimately yield good streamflow performance, even if for the wrong reasons. For example, high precipitation biases could be compensated for by high soil moisture storage when soil moisture is not used as a calibration target, but these simulations would be removed when soil moisture observations are included. Second, the approach to the multi-objective problem influences the way equifinality is assessed. Our use of joint constraints is clear cut, demanding that performance criteria is met for more than one set of observations. However, there are numerous alternative approaches, such as adaptive data assimilation techniques (Y. Liu et al., 2012), pareto optimization (Madsen, 2003), or informal Bayesian methods (Choi & Beven, 2007). Each approach is unique in its integration of alternative observations and assessment of parameter uncertainty/equifinality. Our logical framework inherently reduces the equifinal space as more observational constraints are introduced, but pareto optimization or fuzzy logical constraints may expand it. Lastly, the model-to-observation alignment as discussed in section 4.1 plays a significant role in constraining parameter values.

Additionally, the modeler must make decisions on the spatial and temporal resolution for model evaluation. While streamflow observations are commonly used in daily timesteps, the modeler may consider using monthly or annual averages to assess performance (Hay et al., 2023). Notably, the OpenET dataset from Google Earth Engine is only available as monthly averages and was identified as the most informative alternative observation dataset in this study. Future work should include the use of different temporal resolutions in model evaluation. Errors at daily timesteps may result in harsh

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

penalization, while longer term trends could be adequately represented. In a similar vein, results from sensitivity analysis depend in part on the calibration period and window size (Herman et al., 2013a; Massmann et al., 2014; van Werkhoven et al., 2008). Previous literature notes that sensitivity measures become stable for five year windows or greater, muting the effects of interannual variability (Shin et al., 2013). Our assessment of streamflow over multiple 10 year rolling windows (Figures S8, S9) corroborates this finding. However, the user must be cautious of type II errors, as they arise for parameters that straddle the identifiability threshold in the Morris experiment (Figures S2-S5). We continue the recommendation of bootstrapping the elementary effects to identify these errors (Campolongo & Saltelli, 1997).

4.4 Implications for water resource operations and forecasting

The parameter identification and estimation results presented in this study help inform operational modeling practices and could be extended to forecasting frameworks. Multi-observational modelling techniques have been a subject of increasing attention in the last decade (Y. Liu et al., 2012) and are a promising tool for the improvement of ensemble forecasting (Troin et al., 2021). However, our results suggest that the integration of alternative observations have spatially heterogenous implications to streamflow performance, despite the four case study catchments being within a similar geographic region. In the catchments that border each other (the East and Taylor River), we see that the augmentation of streamflow performance for each dataset is different (Figure 5, S12, S13). This may result from uncertainty in the model, since the smaller headwater catchments yielded fewer behavioral parameter sets in the Monte Carlo Filtering step (Table 2). These differences between catchments could also be explained by differences in hydroclimatic variables, such as runoff ratio or aridity (Elkouk et al., 2024; van Werkhoven et al., 2008). At broader scales, the simulation of streamflow would likely be improved by different datasets, as the dominant hydrologic processes vary by ecologic and physiographic characteristics (Mai et al., 2022; Markstrom et al., 2016). The methods used in this study, Morris screening and LHS Monte-Carlo filtering, are relatively straightforward analytical techniques that were completed on a laptop computer. Future work to accomplish this assessment at broader scales in a computationally parsimonious way would be valuable.

5 Conclusions

A multi-observation sensitivity and uncertainty analysis of the pywatershed hydrologic model is presented in this study. In four headwater catchments in the UCRB, we obtained seven observation datasets pertaining to discharge, snow water equivalent, snow-covered area, soil moisture, and evapotranspiration to use as objective targets in a Morris parameter screening and Monte-Carlo filtering analysis. Results show that the use of alternative observations allows for the identification of more informative parameters in the screening analysis. The outcomes of streamflow performance and parameter estimation vary considerably across catchments and observation data criterion.

Starting with 51 model parameters, the Morris screening method identifies nearly twice as many informative parameters when including alternative observations versus discharge alone. Bootstrapping of the sensitivity metrics allows for the identification of type I and type II statistical errors, which avoids the exclusion of parameters that have sensitivities near the identification threshold. Across the four catchments, forcing corrections and rain-snow partitioning parameters have a high impact on the model fit to observations. The identification of informative parameters is highly variable over annual timescales, but over decadal timescales it is relatively stable due to the suppression of interannual variability. Spearman rank correlations between parameter sensitivity and catchment attributes such as precipitation, temperature, and elevation are weak to moderate.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

With the informative parameters carried into the maximin LHS Monte-Carlo filtering analysis, we find that multi-observation criteria considerably reduce equifinality. However, by reducing the number of acceptable parameter sets, streamflow performance may be either positively or negatively constrained. Our results suggest that AET is consistently useful for the improvement of streamflow simulations in UCRB catchments, but the value of other alternative datasets needs to be assessed on a case-by-case basis. Snow and soil moisture datasets yield both increased and decreased performance depending on the catchments. Additionally, the observation criterion has strong impacts on the range of estimated parameter values

The use of alternative observations is found to be informative in parameter screening but has uncertain and spatially heterogenous outcomes in terms of streamflow performance and parameter estimation. We note that observation quality and model-to-observation alignment are important aspects of the analytical framework used in this study. Given the nuance of observations such as SMAP, ASO, and OpenET, these findings may be considered in future work where multi-objective calibration is of interest.

Code Availability

The pywatershed model used for hydrologic modeling in this paper is publicly available on Github under a Creative Commons Zero v1.0 license (https://github.com/EC-USGS/pywatershed). The model input files, geometry files, and software developed for the analysis presented in this paper are publicly available (https://doi.org/10.5281/zenodo.17180693).

Data Availability

The United States Geological Survey streamflow datasets are retrieved through the dataRetrieval package in the R programming language (DeCicco et al., 2024). The in situ snow water equivalent and soil moisture datasets from the SNOTEL observation network are available through the U.S. Department of Agriculture, National Water and Climate Center online report generator https://wcc.sc.egov.usda.gov/reportGenerator/. The ASO remotely sensed snow water equivalent datasets are available online https://www.airbornesnowobservatories.com/. The MODIS snow covered area dataset (MOD10A1.061) is retrieved from Google Earth Engine at https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD10A1 .The SMAP Level 4 surface wetness dataset (SPL4SMGP.007) is retrieved from Google Earth Engine at https://developers.google.com/earth-engine/datasets/catalog/NASA_SMAP_SPL4SMGP_007?hl=en. The OpenET, Inc. actual evapotranspiration ensemble dataset is retrieved from Google Earth Engine at https://developers.google.com/earthengine/datasets/catalog/OpenET_ENSEMBLE_CONUS_GRIDMET_MONTHLY_v2_0

Author Contributions

LHN: data curation, formal analysis, methodology, software, visualization, and writing - original draft preparation. AMM: conceptualization, funding acquisition, methodology, project administration, supervision, and writing - review and editing. GAT and LD: conceptualization, funding acquisition, and writing - review and editing. AWW and EJA: writing - review and editing.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

We acknowledge the support from Parker Norton who provided configuration files from the National Hydrologic Model, and James McCreight who assisted with the model implementation.

Financial Support

L.H.N., A.M.M., G.A.T., L.D., A.W.W., and E.J.A. were supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) with joint funding under award NA22NWS4320003 from the National Oceanic and Atmospheric Administration (NOAA) Cooperative Institute Program and the U.S. Geological Survey (USGS). The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the opinions of NOAA or USGS.

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

References

Althoff, D. and Rodrigues, L. N.: Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment, Journal of Hydrology, 600, 126674, https://doi.org/10.1016/j.jhydrol.2021.126674, 2021.

Balbi, M. and Lallemant, D. C. B.: The Cost of Imperfect Knowledge: How Epistemic Uncertainties Influence Flood Hazard Assessments, Water Resources Research, 59, e2023WR035685, https://doi.org/10.1029/2023WR035685, 2023.

Bárdossy, A. and Anwar, F.: Why do our rainfall–runoff models keep underestimating the peak flows?, Hydrology and Earth System Sciences, 27, 1987–2000, https://doi.org/10.5194/hess-27-1987-2023, 2023.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Sciences, 12, 1273–1283, https://doi.org/10.5194/hess-12-1273-2008, 2008.

Beven, K.: A Discussion of Distributed Hydrological Modelling, in: Distributed Hydrological Modelling, edited by: Abbott, M. B. and Refsgaard, J. C., Springer Netherlands, Dordrecht, 255–278, https://doi.org/10.1007/978-94-009-0257-2_13, 1996.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11–29, https://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

Brocca, L., Zucco, G., Mittelbach, H., Moramarco, T., and Seneviratne, S. I.: Absolute versus temporal anomaly and percent of saturation soil moisture spatial variability for six networks worldwide, Water Resources Research, 50, 5560–5576, https://doi.org/10.1002/2014WR015684, 2014.

Campolongo, F. and Saltelli, A.: Sensitivity analysis of an environmental model: an application of different analysis methods, Reliability Engineering & System Safety, 57, 49–69, https://doi.org/10.1016/S0951-8320(97)00021-5, 1997.

Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software, 22, 1509–1518, https://doi.org/10.1016/j.envsoft.2006.10.004, 2007.

Carnell, R.: lhs: Latin Hypercube Samples, 2024.

Chen, Y., Steinberg, D. M., and Qian, P.: Maximin Sliced Latin Hypercube Designs with Application to Cross Validating Prediction Error, in: Handbook of Uncertainty Quantification, edited by: Ghanem, R., Higdon, D., and Owhadi, H., Springer International Publishing, Cham, 289–309, https://doi.org/10.1007/978-3-319-12385-1_6, 2017.

Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., and Jacobs, J. M.: Precipitation biases and snow physics limitations drive the uncertainties in macroscale modeled snow water equivalent, Hydrology and Earth System Sciences, 26, 5721–5735, https://doi.org/10.5194/hess-26-5721-2022, 2022.

Choi, H. T. and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, Journal of Hydrology, 332, 316–336, https://doi.org/10.1016/j.jhydrol.2006.07.012, 2007.

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., and Palmer, R. N.: The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin, Climatic Change, 62, 337–363, https://doi.org/10.1023/B:CLIM.0000013684.13621.1f, 2004.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The Abuse of Popular Performance Metrics in Hydrologic Modeling, Water Resources Research, 57, e2020WR029001, https://doi.org/10.1029/2020WR029001, 2021.

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S., and Samaniego, L.: Computationally inexpensive identification of noninformative model parameters by sequential screening, Water Resources Research, 51, 6417–6441, https://doi.org/10.1002/2015WR016907, 2015.

DeCicco, L., Hirsch, R., Lorenz, D., Read, J., Walker, J., Carr, L., Watkins, D., Blodgett, D., Johnson, M., and Krall, A.: dataRetrieval: Retrieval Functions for USGS and EPA Hydrology and Water Quality Data, 2024.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resources Research, 56, e2019WR026085, https://doi.org/10.1029/2019WR026085, 2020.

Douglas-Mankin, K. and Moeser, D.: Calibration of Precipiation-Runoff Modeling System (PRMS) to Simulate Prefire and Postfire Hydrologic Response in the Upper Rio Hondo Basin, New Mexico, U.S. Geological Survey, 2019.

Draper, C. S., Walker, J. P., Steinle, P. J., de Jeu, R. A. M., and Holmes, T. R. H.: An evaluation of AMSR–E derived soil moisture over Australia, Remote Sensing of Environment, 113, 703–710, https://doi.org/10.1016/j.rse.2008.11.011, 2009.

Ehlers, L. B., Sonnenborg, T. O., and Refsgaard, J. C.: Observational and predictive uncertainties for multiple variables in a spatially distributed hydrological model, Hydrological Processes, 33, 833–848, https://doi.org/10.1002/hyp.13367, 2019.

Elkouk, A., Pokhrel, Y., Livneh, B., Payton, E., Luo, L., Cheng, Y., Dagon, K., Swenson, S., Wood, A. W., Lawrence, D. M., and Thiery, W.: Toward Understanding Parametric Controls on Runoff Sensitivity to Climate in the Community Land Model: A Case Study Over the Colorado River Headwaters, Water Resources Research, 60, e2024WR037718, https://doi.org/10.1029/2024WR037718, 2024.

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., and Cloke, H. L.: Continental and global scale flood forecasting systems, WIREs Water, 3, 391–418, https://doi.org/10.1002/wat2.1137, 2016.

Finger, D., Vis, M., Huss, M., and Seibert, J.: The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments, Water Resources Research, 51, 1939–1958, https://doi.org/10.1002/2014WR015712, 2015.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., and Di, Z.: A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model, Environmental Modelling & Software, 51, 269–285, https://doi.org/10.1016/j.envsoft.2013.09.031, 2014.

Gelfan, A. N., Pomeroy, J. W., and Kuchment, L. S.: Modeling Forest Cover Influences on Snow Accumulation, Sublimation, and Melt, 2004.

Gleason, K. E., McConnell, J. R., Arienzo, M. M., Chellman, N., and Calvin, W. M.: Four-fold increase in solar forcing on snow in western U.S. burned forests since 1999, Nat Commun, 10, 2026, https://doi.org/10.1038/s41467-019-09935-y, 2019.

Gorski, G., Stets, E. G., Scholl, M. A., Degnan, J. R., Mullaney, J. R., Galanter, A. E., Martinez, A. J., Padilla, J., LaFontaine, J. H., Corson-Dosch, H. R., and Shapiro, A.: Water supply in the conterminous United States, Alaska, Hawaii, and Puerto Rico, water years 2010–20, Professional Paper, U.S. Geological Survey, https://doi.org/10.3133/pp1894B, 2025.

Gupta, H. and Razavi, S.: Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models, Water Resources Research, 54, 8692–8717, https://doi.org/10.1029/2018WR022668, 2018.

Gupta, H., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resources Research, 34, 751–763, https://doi.org/10.1029/97WR03495, 1998.

Gupta, H., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Gupta, H., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hammond, J. C. and Kampf, S. K.: Subannual Streamflow Responses to Rainfall and Snowmelt Inputs in Snow-Dominated Watersheds of the Western United States, Water Resources Research, 56, e2019WR026132, https://doi.org/10.1029/2019WR026132, 2020.

Hao, Z., Yuan, X., Xia, Y., Hao, F., and Singh, V. P.: An Overview of Drought Monitoring and Prediction Systems at Regional and Global Scales, Bulletin of the American Meteorological Society, 98, 1879–1896, https://doi.org/10.1175/BAMS-D-15-00149.1, 2017.

Harpold, A. A., Molotch, N. P., Musselman, K. N., Bales, R. C., Kirchner, P. B., Litvak, M., and Brooks, P. D.: Soil moisture response to snowmelt timing in mixed-conifer subalpine forests, Hydrological Processes, 29, 2782–2798, https://doi.org/10.1002/hyp.10400, 2015.

Hasan, H. M. M., Döll, P., Hosseini-Moghari, S.-M., Papa, F., and Güntner, A.: The benefits and trade-offs of multi-variable calibration of the WaterGAP global hydrological model (WGHM) in the Ganges and Brahmaputra basins, Hydrology and Earth System Sciences, 29, 567–596, https://doi.org/10.5194/hess-29-567-2025, 2025.

Hay, L. E. and Makiko, U.: USGS Open-File Report 2006-1323: Multiple-Objective Stepwise Calibration Using Luca, 2007.

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., and Umemoto, M.: Step Wise, Multiple Objective Calibration of a Hydrologic Model for a Snowmelt Dominated Basin, JAWRA Journal of the American Water Resources Association, 42, 877–890, https://doi.org/10.1111/j.1752-1688.2006.tb04501.x, 2006.

Hay, L. E., LaFontaine, J. H., Beusekom, A. E. V., Norton, P. A., Farmer, W. H., Regan, R. S., Markstrom, S. L., and Dickinson, J. E.: Parameter estimation at the conterminous United States scale and streamflow routing enhancements for the National Hydrologic Model infrastructure application of the Precipitation-Runoff Modeling System (NHM-PRMS), Techniques and Methods, U.S. Geological Survey, https://doi.org/10.3133/tm6B10, 2023.

Helton, J. C. and Davis, F. J.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliability Engineering & System Safety, 81, 23–69, https://doi.org/10.1016/S0951-8320(03)00058-9, 2003.

Herbert, J. N., Raleigh, M. S., and Small, E. E.: Reanalyzing the spatial representativeness of snow depth at automated monitoring stations using airborne lidar data, The Cryosphere, 18, 3495–3512, https://doi.org/10.5194/tc-18-3495-2024, 2024.

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T.: From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models, Hydrology and Earth System Sciences, 17, 5109–5125, https://doi.org/10.5194/hess-17-5109-2013, 2013a.

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T.: Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, Hydrology and Earth System Sciences, 17, 2893–2903, https://doi.org/10.5194/hess-17-2893-2013, 2013b.

Herrera, P. A., Marazuela, M. A., and Hofmann, T.: Parameter estimation and uncertainty analysis in hydrological modeling, WIREs Water, 9, e1569, https://doi.org/10.1002/wat2.1569, 2022.

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A., and Braatz, D.: A Multistep Automatic Calibration Scheme for River Forecasting Models, Journal of Hydrometeorology, 1, 524–542, https://doi.org/10.1175/1525-7541(2000)001%253C0524:AMACSF%253E2.0.CO;2, 2000.

Huang, Q., Qin, G., Zhang, Y., Tang, Q., Liu, C., Xia, J., Chiew, F. H. S., and Post, D.: Using Remote Sensing Data-Based Hydrological Model Calibrations for Predicting Runoff in Ungauged or Poorly Gauged Catchments, Water Resources Research, 56, e2020WR028205, https://doi.org/10.1029/2020WR028205, 2020.

Iooss, B., Veiga, S. D., Janon, A., and Pujol, G.: sensitivity: Global Sensitivity Analysis of Model Outputs and Importance Measures, 2024.

Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere, Nat Commun, 9, 1148, https://doi.org/10.1038/s41467-018-03629-7, 2018.

Johnson, J. M., Fang, S., Sankarasubramanian, A., Rad, A. M., Kindl da Cunha, L., Jennings, K. S., Clarke, K. C., Mazrooei, A., and Yeghiazarian, L.: Comprehensive Analysis of the NOAA National Water Model: A Call for Heterogeneous Formulations and Diagnostic Model Selection, Journal of Geophysical Research: Atmospheres, 128, e2023JD038534, https://doi.org/10.1029/2023JD038534, 2023.

Johnson, M. E., Moore, L. M., and Ylvisaker, D.: Minimax and maximin distance designs, Journal of Statistical Planning and Inference, 26, 131–148, https://doi.org/10.1016/0378-3758(90)90122-B, 1990.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, https://doi.org/10.1029/2005WR004362, 2006.

Koster, R. D., Liu, Q., Mahanama, S. P. P., and Reichle, R. H.: Improved Hydrological Simulation Using SMAP Data: Relative Impacts of Model Calibration and Data Assimilation, https://doi.org/10.1175/JHM-D-17-0228.1, 2018.

Kunnath-Poovakka, A., Ryu, D., Renzullo, L. J., and George, B.: The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction, Journal of Hydrology, 535, 509–524, https://doi.org/10.1016/j.jhydrol.2016.02.018, 2016.

Lahmers, T. M., Kumar, S. V., Rosen, D., Dugger, A., Gochis, D. J., Santanello, J. A., Gangodagamage, C., and Dunlap, R.: Assimilation of NASA's Airborne Snow Observatory Snow Measurements for Improved Hydrological Modeling: A Case

https://doi.org/10.5194/egusphere-2025-5815
Preprint. Discussion started: 4 January 2026

Study Enabled by the Coupled LIS/WRF-Hydro System, Water Resources Research, 58, e2021WR029867, https://doi.org/10.1029/2021WR029867, 2022.

Lamontagne, J. R., Barber, C. A., and Vogel, R. M.: Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data, Water Resources Research, 56, e2020WR027101, https://doi.org/10.1029/2020WR027101, 2020.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G.: Precipitation-runoff modeling system; user's manual, Water-Resources Investigations Report, U.S. Geological Survey, Water Resources Division, https://doi.org/10.3133/wri834238, 1983.

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233–241, https://doi.org/10.1029/1998WR900018, 1999.

Li, D., Wrzesien, M. L., Durand, M., Adam, J., and Lettenmaier, D. P.: How much runoff originates as snow in the western United States, and how will that change in the future?, Geophysical Research Letters, 44, 6163–6172, https://doi.org/10.1002/2017GL073551, 2017.

Liu, P.-W., Bindlish, R., Fang, B., Lakshmi, V., O'Neill, P. E., Yang, Z., Cosh, M. H., Bongiovanni, T., Bosch, D. D., Collins, C. H., Starks, P. J., Prueger, J., Seyfried, M., and Livingston, S.: Assessing Disaggregated SMAP Soil Moisture Products in the United States, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2577–2592, https://doi.org/10.1109/JSTARS.2021.3056001, 2021.

Liu, X., Yang, K., Ferreira, V. G., and Bai, P.: Hydrologic Model Calibration With Remote Sensing Data Products in Global Large Basins, Water Resources Research, 58, e2022WR032929, https://doi.org/10.1029/2022WR032929, 2022.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrology and Earth System Sciences, 16, 3863–3887, https://doi.org/10.5194/hess-16-3863-2012, 2012.

Livneh, B. and Badger, A. M.: Drought less predictable under declining future snowpack, Nat. Clim. Chang., 10, 452–458, https://doi.org/10.1038/s41558-020-0754-8, 2020.

Livneh, B. and Lettenmaier, D. P.: Multi-criteria parameter estimation for the Unified Land Model, Hydrology and Earth System Sciences, 16, 3029–3048, https://doi.org/10.5194/hess-16-3029-2012, 2012.

Lukas, J. and Payton, E.: Colorado River Basin Climate and Hydrology: State of the Science, https://doi.org/10.25810/3HCV-W477, 2020.

Lundquist, J. D., Dickerson-Lange, S. E., Lutz, J. A., and Cristea, N. C.: Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling, Water Resources Research, 49, 6356–6370, https://doi.org/10.1002/wrcr.20504, 2013.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Advances in Water Resources, 26, 205–216, https://doi.org/10.1016/S0309-1708(02)00092-1, 2003.

Mai, J.: Ten strategies towards successful calibration of environmental models, Journal of Hydrology, 620, 129414, https://doi.org/10.1016/j.jhydrol.2023.129414, 2023.

Mai, J., Craig, J. R., Tolson, B. A., and Arsenault, R.: The sensitivity of simulated streamflow to individual hydrologic processes across North America, Nat Commun, 13, 455, https://doi.org/10.1038/s41467-022-28010-7, 2022.

Manikanta, V. and Vema, V. K.: Formulation of Wavelet Based Multi-Scale Multi-Objective Performance Evaluation (WMMPE) Metric for Improved Calibration of Hydrological Models, Water Resources Research, 58, e2020WR029355, https://doi.org/10.1029/2020WR029355, 2022.

Markstrom, S. L., Regan, R. S., Hay, L. E., Viger, R. J., Webb, R. M., Payn, R. A., and LaFontaine, J. H.: PRMS-IV, the precipitation-runoff modeling system, version 4, Techniques and Methods, U.S. Geological Survey, https://doi.org/10.3133/tm6B7, 2015.

Markstrom, S. L., Hay, L. E., and Clark, M. P.: Towards simplification of hydrologic modeling: identification of dominant processes, Hydrology and Earth System Sciences, 20, 4655–4671, https://doi.org/10.5194/hess-20-4655-2016, 2016.

Marshall, A. M., Link, T. E., Flerchinger, G. N., and Lucash, M. S.: Importance of Parameter and Climate Data Uncertainty for Future Changes in Boreal Hydrology, Water Resources Research, 57, e2021WR029911, https://doi.org/10.1029/2021WR029911, 2021.

Massmann, C., Wagener, T., and Holzmann, H.: A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales, Environmental Modelling & Software, 51, 190–194, https://doi.org/10.1016/j.envsoft.2013.09.033, 2014.

Maxwell, J. and St Clair, S. B.: Snowpack properties vary in response to burn severity gradients in montane forests, Environ. Res. Lett., 14, 124094, https://doi.org/10.1088/1748-9326/ab5de8, 2019.

Mazzotti, G., Webster, C., Quéno, L., Cluzet, B., and Jonas, T.: Canopy structure, topography, and weather are equally important drivers of small-scale snow cover dynamics in sub-alpine forests, Hydrology and Earth System Sciences, 27, 2099–2121, https://doi.org/10.5194/hess-27-2099-2023, 2023.

McCabe, M. F., Franks, S. W., and Kalma, J. D.: Calibration of a land surface model using multiple data sets, Journal of Hydrology, 302, 209–222, https://doi.org/10.1016/j.jhydrol.2004.07.002, 2005.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The future of Earth observation in hydrology, Hydrology and Earth System Sciences, 21, 3879–3914, https://doi.org/10.5194/hess-21-3879-2017, 2017.

Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H., Eberts, S., Niswonger, R., Regan, R. S., and Hunt, R. J.: Can Hydrological Models Benefit From Using Global Soil Moisture, Evapotranspiration, and Runoff Products as Calibration Targets?, Water Resources Research, 59, e2022WR032064, https://doi.org/10.1029/2022WR032064, 2023.

Melton, F. S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Senay, G. B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., Daudert, B., Doherty, C., Dunkerly, C., Friedrichs, M., Guzman, A., Halverson, G., Hansen, J., Harding, J., Kang, Y., Ketchum, D., Minor, B., Morton, C., Ortega-Salazar, S., Ott, T., Ozdogan, M., ReVelle, P. M., Schull, M., Wang, C., Yang, Y., and Anderson, R. G.: OpenET: Filling a Critical Data Gap in Water Management for the Western United States, JAWRA Journal of the American Water Resources Association, 58, 971–994, https://doi.org/10.1111/1752-1688.12956, 2022.

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., and Arnold, J. R.: Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts, Journal of Hydrometeorology, 16, 762–780, https://doi.org/10.1175/JHM-D-14-0104.1, 2015.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2003JD003823, 2004.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33, 161–174, https://doi.org/10.1080/00401706.1991.10484804, 1991.

Motovilov, Y. G., Gottschalk, L., Engeland, K., and Rodhe, A.: Validation of a distributed hydrological model against spatial observations, Agricultural and Forest Meteorology, 98–99, 257–277, https://doi.org/10.1016/S0168-1923(99)00102-1, 1999.

Nash, L. L. and Gleick, P. H.: Sensitivity of streamflow in the Colorado Basin to climatic changes, Journal of Hydrology, 125, 221–241, https://doi.org/10.1016/0022-1694(91)90030-L, 1991.

Nassar, A., Tarboton, D., Anderson, M., Yang, Y., Fisher, J. B., Purdy, A. J., Baig, F., He, C., Gochis, D., Melton, F., and Volk, J.: Intercomparison of the U.S. National water model with OpenET over the Bear River Basin, U.S, Journal of Hydrology, 132826, https://doi.org/10.1016/j.jhydrol.2025.132826, 2025.

Oubeidillah, A., Tootle, G., and Piechota, T.: Incorporating Antecedent Soil Moisture into Streamflow Forecasting, Hydrology, 6, 50, https://doi.org/10.3390/hydrology6020050, 2019.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., and Winstral, A.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sensing of Environment, 184, 139–152, https://doi.org/10.1016/j.rse.2016.06.018, 2016.

Pathiraja, S., Marshall, L., Sharma, A., and Moradkhani, H.: Hydrologic modeling in dynamic catchments: A data assimilation approach, Water Resources Research, 52, 3350–3372, https://doi.org/10.1002/2015WR017192, 2016.

Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., Bonfils, C. J. W., Gallant, A. J. E., Hoerling, M., Hoffmann, D., Kaatz, L., Lehner, F., Llewellyn, D., Mote, P., Neale, R. B., Overpeck, J. T., Sheffield, A., Stahl, K., Svoboda, M., Wheeler, M. C., Wood, A. W., and Woodhouse, C. A.: Flash droughts present a new challenge for subseasonal-to-seasonal prediction, Nat. Clim. Chang., 10, 191–199, https://doi.org/10.1038/s41558-020-0709-0, 2020.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environmental Modelling & Software, 79, 214–232, https://doi.org/10.1016/j.envsoft.2016.02.008, 2016.

Povey, A. C. and Grainger, R. G.: Known and unknown unknowns: uncertainty estimation in satellite remote sensing, Atmospheric Measurement Techniques, 8, 4699–4718, https://doi.org/10.5194/amt-8-4699-2015, 2015.

Rajib, M. A., Merwade, V., and Yu, Z.: Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture, Journal of Hydrology, 536, 192–207, https://doi.org/10.1016/j.jhydrol.2016.02.037, 2016.

Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resources Research, 52, 7779–7792, https://doi.org/10.1002/2016WR019430, 2016.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., and Maier, H. R.: The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support, Environmental Modelling & Software, 137, 104954, https://doi.org/10.1016/j.envsoft.2020.104954, 2021.

Regan, R. S., Markstrom, S. L., Hay, L. E., Viger, R. J., Norton, P. A., Driscoll, J. M., and LaFontaine, J. H.: Description of the National Hydrologic Model for use with the Precipitation-Runoff Modeling System (PRMS), Techniques and Methods, U.S. Geological Survey, https://doi.org/10.3133/tm6B9, 2018.

Regan, R. S., Juracek, K. E., Hay, L. E., Markstrom, S. L., Viger, R. J., Driscoll, J. M., LaFontaine, J. H., and Norton, P. A.: The U. S. Geological Survey National Hydrologic Model infrastructure: Rationale, description, and application of a watershed-scale model for the conterminous United States, Environmental Modelling & Software, 111, 192–203, https://doi.org/10.1016/j.envsoft.2018.09.023, 2019.

Ritter, A. and Muñoz-Carpena, R.: Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, Journal of Hydrology, 480, 33–45, https://doi.org/10.1016/j.jhydrol.2012.12.004, 2013.

Saltelli, A., Ratto, M., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Elementary Effects Method, in: Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, 109–154, https://doi.org/10.1002/9780470725184.ch3, 2007.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q.: Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices, Environmental Modelling & Software, 114, 29–39, https://doi.org/10.1016/j.envsoft.2019.01.012, 2019.

Santner, T. J., Williams, B. J., and Notz, W. I.: The Design and Analysis of Computer Experiments, Springer, New York, NY, https://doi.org/10.1007/978-1-4939-8847-1, 2018.

Shafii, M., Tolson, B., and Shawn Matott, L.: Addressing subjective decision-making inherent in GLUE-based multi-criteria rainfall–runoff model calibration, Journal of Hydrology, 523, 693–705, https://doi.org/10.1016/j.jhydrol.2015.01.051, 2015.

Sheikholeslami, R. and Razavi, S.: Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models, Environmental Modelling & Software, 93, 109–126, https://doi.org/10.1016/j.envsoft.2017.03.010, 2017.

Shields, M. D. and Zhang, J.: The generalization of Latin hypercube sampling, Reliability Engineering & System Safety, 148, 96–108, https://doi.org/10.1016/j.ress.2015.12.002, 2016.

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., and Jakeman, A. J.: Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R, Journal of Hydrology, 503, 135–152, https://doi.org/10.1016/j.jhydrol.2013.08.047, 2013.

Singh, B., Ferdousi, T., Abatzoglou, J. T., Swarup, S., Adam, J. C., and Rajagopalan, K.: Sensitivity of snow magnitude and duration to hydrology model parameters, Journal of Hydrology, 645, 132193, https://doi.org/10.1016/j.jhydrol.2024.132193, 2024.

Skiles, S. M., Flanner, M., Cook, J. M., Dumont, M., and Painter, T. H.: Radiative forcing by light-absorbing particles in snow, Nature Clim Change, 8, 964–971, https://doi.org/10.1038/s41558-018-0296-5, 2018.

Smyth, E. J., Raleigh, M. S., and Small, E. E.: Improving SWE Estimation With Data Assimilation: The Influence of Snow Depth Observation Timing and Uncertainty, Water Resources Research, 56, e2019WR026853, https://doi.org/10.1029/2019WR026853, 2020.

Sobol′, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, Journal of Hydrology, 523, 739–757, https://doi.org/10.1016/j.jhydrol.2015.02.013, 2015.

Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., and Bair, E. H.: Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets, The Cryosphere, 17, 567–590, https://doi.org/10.5194/tc-17-567-2023, 2023.

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., and Blöschl, G.: The Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a Small Catchment, Water Resources Research, 56, e2019WR026153, https://doi.org/10.1029/2019WR026153, 2020.

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., Beck, H. E., Wood, A. W., Newman, A. J., and Papalexiou, S. M.: The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins, Water Resources Research, 59, e2022WR033767, https://doi.org/10.1029/2022WR033767, 2023.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wei, Y., Devarakonda, R., Vose, R. S., and Cook, R. B.: DaymetDaymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3, https://doi.org/10.3334/ORNLDAAC/1328, 15 July 2016.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years, Water Resources Research, 57, e2020WR028392, https://doi.org/10.1029/2020WR028392, 2021.

Viger, R. J.: Preliminary spatial parameters for PRMS based on the Geospatial Fabric, NLCD2001, and SSURGO, https://doi.org/10.5066/F7WM1BF7, 2014.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resources Research, 41, https://doi.org/10.1029/2004WR003059, 2005.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resources Research, 44, https://doi.org/10.1029/2007WR006271, 2008.

Wheeler, K. G., Udall, B., Wang, J., Kuhn, E., Salehabadi, H., and Schmidt, J. C.: What will it take to stabilize the Colorado River?, Science, 377, 373–375, https://doi.org/10.1126/science.abo4452, 2022.

Wongchuig, S., Paiva, R., Siqueira, V., Papa, F., Fleischmann, A., Biancamaria, S., Paris, A., Parrens, M., and Al Bitar, A.: Multi-Satellite Data Assimilation for Large-Scale Hydrological-Hydrodynamic Prediction: Proof of Concept in the Amazon Basin, Water Resources Research, 60, e2024WR037155, https://doi.org/10.1029/2024WR037155, 2024.

Zhou, J., Wu, Z., Crow, W. T., Dong, J., and He, H.: Improving Spatial Patterns Prior to Land Surface Data Assimilation via Model Calibration Using SMAP Surface Soil Moisture Data, Water Resources Research, 56, e2020WR027770, https://doi.org/10.1029/2020WR027770, 2020.