

# Water Resources Research

## RESEARCH ARTICLE

10.1029/2020WR027639

### Key Points:

- Uncertainty in simulated groundwater storage change is a result of topographic redistribution of uncertainty in precipitation forcing
- Subsurface flow pathways in both soil and deeper bedrock control watershed response to variable precipitation input
- Merging multiple precipitation data sets provides better results than merging simulated fluxes, due to high model sensitivity to changes in precipitation

### Supporting Information:

- Supporting Information S1

### Correspondence to:

A. P. Schreiner-McGraw,  
adampschreiner@gmail.com

### Citation:

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of uncertainty in precipitation forcing data sets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. *Water Resources Research*, 56, e2020WR027639. <https://doi.org/10.1029/2020WR027639>

Received 1 APR 2020

Accepted 25 OCT 2020

Accepted article online 29 OCT 2020

## Impact of Uncertainty in Precipitation Forcing Data Sets on the Hydrologic Budget of an Integrated Hydrologic Model in Mountainous Terrain

Adam P. Schreiner-McGraw<sup>1</sup>  and Hoori Ajami<sup>1</sup> 

<sup>1</sup>Department of Environmental Sciences, University of California, Riverside, CA, USA

**Abstract** Precipitation is a key input variable in distributed surface water-groundwater models, and its spatial variability is expected to impact watershed hydrologic response via changes in subsurface flow dynamics. Gridded precipitation data sets based on gauge observations, however, are plagued by uncertainty, especially in mountainous terrain where gauge networks are sparse. To examine the mechanisms via which uncertainty in precipitation data propagates through a watershed, we perform a series of numerical experiments using an integrated surface water-groundwater hydrologic model, ParFlow-CLM. The Kaweah River watershed in California, USA, is used as our virtual catchment laboratory to characterize watershed response to variable precipitation forcing from headwaters to groundwaters. By applying the three-cornered hat method, we quantify the spatially distributed uncertainty in four publically available precipitation forcing data sets and their simulated hydrology. Simulations demonstrate that uncertainty in the simulated groundwater storage is primarily a result of topographic redistribution of uncertainty in precipitation forcing. Soil water redistribution is the primary pathway that redistributes uncertainty downslope. We also find that topography exerts a larger impact than variable subsurface parameters on propagating uncertainty in simulated fluxes. Finally, we find that improvement in model performance metrics is higher for a single simulation forced with the mean precipitation from the available data sets than the averaged simulated results of separate simulations forced with each data set. Results from this study highlight the importance of topography-moderated flow through the critical zone in shaping the groundwater response to climate variability.

## 1. Introduction

Precipitation is the primary driver of the terrestrial hydrologic cycle. Historically, precipitation estimates have been derived from ground-based gauges at point locations. However, watershed scale studies require spatially distributed precipitation inputs over large regions at high temporal resolution (minutes to daily). Several techniques to extrapolate precipitation from gauge-based observation networks to gridded data sets at continental scales have been proposed and widely adopted (Daly et al., 1994; Thornton et al., 1997). It is well known that point-scale gauge measurements are prone to uncertainty, which is most commonly caused by gauge undercatch (Groisman & Legates, 1994; Pollock et al., 2018). Additional uncertainty is introduced by interpolation algorithms that extrapolate ground-based precipitation observations to gridded data sets covering large domains. As a result, precipitation uncertainty increases from 5–16% to 33–65% of rainfall rates with increasing scales from gauges to interpolated gridded data sets (scales of  $10^3$  to  $10^4$  m) (McMillan et al., 2012).

Advances in remotely sensed precipitation products can overcome the observational gap of in situ precipitation data sets. However, data quality challenges still remain, particularly in the mountain catchments that provide 20–90% of discharge (Viviroli & Weingartner, 2004). While satellite-based remote sensing products are being improved with higher resolutions, they currently lack sufficient spatiotemporal resolution and have limited accuracy in quantifying solid-phase precipitation (Lettenmaier et al., 2015). Ground-based radar measurements are also limited by terrain blockage in mountainous regions (Westrick et al., 1999). As a result, precipitation observations in mountainous terrain are largely limited to in situ precipitation gauges despite uncertainty in gauge-based gridded data sets.

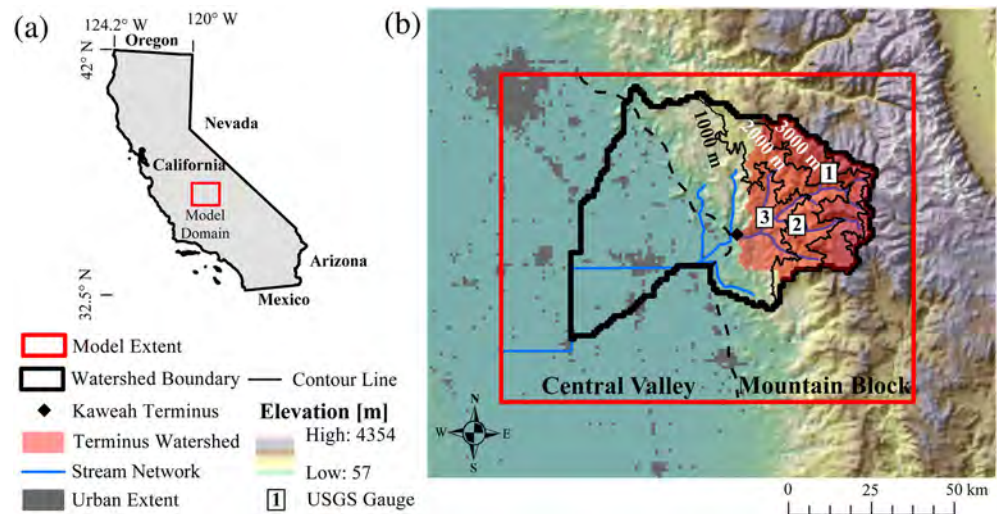
Previous evaluations of gridded precipitation data sets found significant uncertainty in precipitation estimates at high elevations and that uncertainty estimates from data intercomparisons are lower than the

true uncertainty (Henn et al., 2018; Lundquist et al., 2015). These shortcomings are caused by sparse gauge networks, due to limited accessibility and infrastructure (Lundquist et al., 2003), or significant underestimation of snow (Rasmussen et al., 2012). In addition to precipitation measurement challenges at a single location, accuracy of gridded data sets depends on extrapolation algorithms across complex terrain, the use of different input gauges, and how snow measurements are considered (Daly et al., 1994; Thornton et al., 1997). Thus, there is considerable uncertainty in the volume and spatial distributions of precipitation data sets built from gauge observations, especially in the high-elevation regions where greater precipitation occurs. This uncertainty in precipitation data sets creates challenges for simulating water balance partitioning of mountain catchments where precipitation forcing has traditionally been treated as truth.

Despite the substantial uncertainty in precipitation inputs, hydrologic models have historically been run with deterministic inputs that do not account for the uncertainties in precipitation (Krzysztofowicz, 2001). With the development of Markov chain Monte Carlo (MCMC) methods and advances in computing power, Bayesian inference or ensemble-based approaches have become popular methods to quantify uncertainty in hydrologic modeling caused by uncertain model inputs, parameters, or structure (Forman et al., 2008; Hong et al., 2006; Kavetski et al., 2006; Tang et al., 2019). The computational time of MCMC methods limits their application, however, when distributed process-based hydrologic models are utilized at large catchment scales. Process-based models are often preferable to empirical or semidistributed models in simulating watershed response to changes in climate or land cover as they capture nonstationarity (e.g., Deb et al., 2019). Although assessing the impact of uncertainty in precipitation on groundwater resources has received less attention, a plethora of studies have investigated the impact of precipitation inputs on streamflow (Forman et al., 2008; Fu et al., 2011; Lobligois et al., 2014; Mizukami et al., 2014, 2016; Moreno et al., 2012; Nicótina et al., 2008) with a common goal of finding an optimum grid resolution to represent streamflow generation processes and routing. Previous work suggests that using the mean of multiple precipitation data sets is the best approach to accurately represent spatially distributed precipitation (Clark & Slater, 2006; Cornes et al., 2018). It has been suggested that running separate simulations with each data set and averaging the simulated streamflow may improve results (Jiang et al., 2012; Zhu et al., 2019). However, there is limited evidence to confirm that these approaches improve simulated subsurface hydrology.

The response of rainfall-runoff models to spatial precipitation variability is dependent on the runoff generating mechanism. Infiltration-excess-type models are seen to be more sensitive to precipitation inputs than saturation-excess-type models (Koren et al., 1999; Milly & Eagleson, 1988; Winchell et al., 1998). This response suggests that the subsurface hydrologic processes play an important role in determining model response to variable precipitation inputs. It has been shown that precipitation input resolution can impact groundwater levels at local scales without any impact on streamflow (Fu et al., 2011). The response of a hydrologic system to variable precipitation input is complex, and in order to disentangle this complexity, consideration of the integrated hydrologic system from land surface to the subsurface is necessary. Integrated hydrologic models simulate both the surface water system and the groundwater system, as well as their interactions. Few of the mentioned modeling studies examined the impact of precipitation uncertainty on the vadose zone and groundwater system.

In this study, we use an integrated groundwater-land surface model, ParFlow.CLM (Kollet & Maxwell, 2008), to investigate how uncertainty in precipitation forcings propagates from precipitation to groundwater. Due to the lack of a dense precipitation gauge network in the mountainous study domain, we quantify the uncertainty between four common gridded precipitation data sets, Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al., 1994), Daymet (Thornton et al., 1997), Gridmet (Abatzoglou, 2013), and a downscaled version of the North American Land Data Assimilation System (NLDAS-2) (Pan et al., 2016). Then we force a hydrologic model with these data sets to examine how this uncertainty propagates through the hydrologic system, with a focus on changes to groundwater. This study is guided by the following fundamental questions: (1) How does uncertainty in forcing data sets propagate from precipitation to groundwater? (2) Does the subsurface parameterization alter the propagation of uncertainty from precipitation to groundwater? and (3) Does a single simulation with the mean of the uncertain precipitation data sets perform better than running a separate simulation with each data set and averaging the simulated results?



**Figure 1.** (a) Study site location. (b) The 30-m digital elevation model of the Kaweah River Watershed model domain, the watershed boundary, and stream network as formed by ParFlow.CLM simulations. The Central Valley and Mountain Block provinces are defined from a geologic map of the region. The location of the Kaweah Terminus full natural flow estimates is shown as well as the Terminus Watershed draining to this point. USGS stream gauge locations are also shown. The station IDs are as follows: 11208730, 11206820, and 11208600 for Sites 1, 2, and 3, respectively.

To address the research questions, we perform three numerical experiments using ParFlow.CLM. These experiments assess the propagation of uncertainty from precipitation to groundwater by forcing the model with four gridded precipitation products (Experiment 1) or changing the subsurface structure to assess precipitation uncertainty impacts (Experiment 2). In Experiment 3, the impact of averaging the precipitation forcings of the four input data sets for each grid cell at a daily scale is compared against the simulated hydrology from each individual data set in Experiment 1.

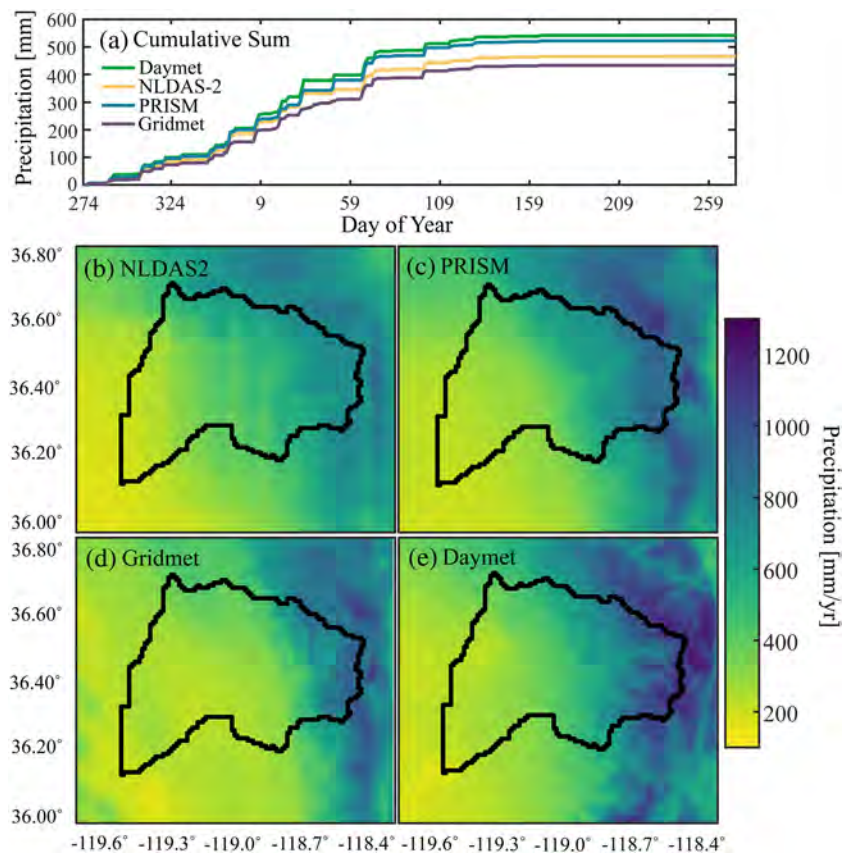
## 2. Methods

### 2.1. Study Site

The study site is the Kaweah River watershed located in the southern part of the Sierra Nevada Range and Central Valley of California (Figure 1a). The regional topography is dominated by the north-south trending granitic mountain block of the Sierra Nevada (Jennings, 1977). The highest elevations, exceeding 4,000 m, are in the eastern portion of the range where the mountain block was uplifted the highest. A complex assemblage of landforms combines to create the alluvial fan that erodes from the mountain block (Olmsted & Davis, 1961). This piedmont slope is composed of loosely consolidated sediments and gradually transitions to the flat basin floor of the Central Valley (Davis et al., 1959). The Central Valley portion of the study domain is primarily composed of interbedded sand and silt layers and is a highly productive groundwater aquifer (Faunt, 2009). The low-elevation valley floor has been subject to extensive development and human settlement (Figure 1b), and the natural flow of water is highly disrupted by a series of dams, canals, and wells. The climate is a Mediterranean climate with warm, dry summers and cool, wet winters. The mean annual precipitation ranges from 193 to 1,443 mm, roughly following the elevation gradient with orographic effects increasing precipitation at higher elevations (Figure 2). The land cover in the model domain contains a sequence of grassland and oak savannah grading into shrublands then conifer forests as elevation increases and finally becoming grassland and bare ground at elevations above the tree line (USDA-NASS, 2017).

### 2.2. Model Forcing and Validation Data Sets

For model forcing, we use four precipitation data sets with at least daily temporal resolution and a spatial resolution of 4 km or higher across the study domain, Daymet, Gridmet, PRISM, and a downscaled version of NLDAS-2. All of the data sets are based on gauge measurements that are extrapolated to a uniform grid. The PRISM data set uses over 13,000 observation stations across the continental United States and builds a



**Figure 2.** (a) The cumulative sum of the WY2016 daily mean precipitation for the model domain for each of the four gridded data sets. Spatial maps of the total annual precipitation applied to each model grid cell for (b) downscaled NLDAS-2, (c) PRISM, (d) Gridmet, and (e) Daymet forcing data sets. The Kaweah River watershed boundary is displayed with a black line.

spatial data set based on a climate-elevation regression for each DEM grid cell at a 4-km resolution (Daly et al., 2008). The observational stations used to build this regression for each grid cell are determined based on the location, elevation, coastal proximity, topographic facet orientation, vertical atmospheric layer, topographic position, and orographic effectiveness of the terrain (Daly et al., 1994). The Daymet data set is constructed in a similar manner with a focus on extrapolating precipitation estimates from gauges across mountainous topography. Daymet is built using fewer rain gauges than PRISM, but it includes data from the SNOTEL network. The extrapolation algorithm to create the gridded data set is based on the original MTCLIM logic and weights each station by its distance from the target grid cell resulting in precipitation estimates at 1-km resolution (Thornton et al., 1997). The NLDAS-2 and Gridmet data sets both use the PRISM algorithm to perform topographic adjustments to daily gauge data. The NLDAS-2 data set is built from the Climate Prediction Center gauge data that has been adjusted to topography using the PRISM algorithm at a  $(1/8)^{\circ}$  spatial scale (Higgins et al., 2000). Temporal downscaling is performed by matching the daily precipitation patterns in the National Weather Service Stage II, 4-km hourly precipitation analyses (Fulton et al., 1998). In this study, we use the hourly Princeton CONUS Forcing data set, which provides a  $(1/32)^{\circ}$  ( $\sim 3$ -km) downsampled version of NLDAS-2 meteorological product over CONUS using 4-km Stage IV and Stage II radar/gauge products (Beck et al., 2019; Pan et al., 2016). We use the downsampled version of NLDAS-2 because the original NLDAS-2 resolution is too coarse to represent precipitation patterns in the complex topography of the Sierra Nevada. Similar to the downsampled NLDAS-2 data set, the Gridmet data set blends the temporal advantages of the original NLDAS-2 with the spatial advantages of PRISM to downscale the original NLDAS-2 data to a 4-km grid resolution using a climatically aided interpolation method (Abatzoglou, 2013; Abatzoglou et al., 2018).



ParFlow.CLM is run at an hourly time step using hourly inputs of precipitation, air temperature, relative humidity, net shortwave radiation, net longwave radiation, atmospheric pressure, and wind speed in the  $x$  and  $y$  directions. Hourly precipitation forcings are generated for each of the three precipitation data sets (PRISM, Daymet, and Gridmet) by applying the temporal downscaling method of NLDAS-2, that is, we apply the diurnal pattern of NLDAS-2 hourly precipitation to the other daily data sets. We obtain publically available data from each of the four data sets for the water year (WY) 2016, 1 October 2015 to 30 September 2016. There is considerable variability in gridded precipitation estimates generated with these four methods (Figure 2). The domain average precipitation for the year varies by more than 100 mm (Gridmet—433 mm, NLDAS-2—466 mm, PRISM—522 mm, and Daymet—542 mm). The temporal distribution throughout the year is relatively consistent between the data sets meaning that the discrepancy in annual totals is primarily a result of differences in the volume of precipitation events at the domain average scale (Figure 2a).

For other meteorological forcings than precipitation, we use the Princeton CONUS Forcing data set at 3-km, 1-hourly resolution. We further downscale the hourly air temperature estimates to 1-km resolution by applying a constant  $-6.5^{\circ}\text{C}/\text{km}$  lapse rate. Watershed average forcing values are shown in the supporting information (Figure S3).

We use several data sets to validate model performance. Streamflow is validated with United States Geological Survey (USGS) stream gauge observations from stream reaches that have been minimally impacted by humans (Falcone, 2011). Because most rivers in the Kaweah River watershed have been extensively modified from their natural state, the California Department of Water Resources estimates unimpaired or “full natural flow” for various rivers in California. Daily full natural flow data are calculated for the Kaweah river at the Terminus dam (Figure 1b) by adjusting for gains and losses associated with retention and diversion.

In mountainous terrain, snow is an important component of the hydrologic cycle. To validate simulated snow processes, we use a reanalysis gridded snow water equivalent (SWE) and snow covered area (SCA) data set covering the Sierra Nevada (Margulis et al., 2016). This 90-m, daily resolution reanalysis data set is generated with a Bayesian data assimilation technique using remotely sensed SCA (Margulis et al., 2016). Finally, we obtain evapotranspiration (ET) estimates from the remotely sensed MODIS MOD16 global ET product at  $1\text{-km}^2$  spatial and 8-day temporal resolutions (Mu et al., 2011). Model performance is assessed using the mean relative error (MRE; Mean absolute error/mean of observed values), bias (B), and Nash-Sutcliffe Efficiency (NSE) between simulated fluxes and the aforementioned observational data sets.

### 2.3. Integrated Hydrologic Model Development

#### 2.3.1. ParFlow.CLM Model Framework

We use the ParFlow.CLM integrated hydrologic model code (Kollet & Maxwell, 2006; Maxwell, 2013; Maxwell & Miller, 2005) to simulate the variably saturated subsurface flow, streamflow, and land surface processes such as the energy balance. The ParFlow model solves the Richards' equation in three dimensions for variably saturated flow in the subsurface and simultaneously solves the kinematic wave approximation of the shallow water equations at the land surface. Predefined river networks with user defined geometries are not used in ParFlow, instead streamflow develops according to a uniform application of the kinematic wave approximation to every cell in the domain. In this representation, horizontal grid resolution and elevation values determine geometry of channel cross sections. ParFlow has been fully coupled with the Common Land Model (CLM) 3.0 (Dai et al., 2003) to simulate the water and energy fluxes at the land surface. The CLM portion of the code is coupled via the pressure and saturation in the top model layers of the ParFlow grid, and ParFlow replaces the CLM soil moisture and runoff generation processes. We apply the terrain-following grid formulation of ParFlow, which is best suited for mountainous terrain with high relief (Maxwell, 2013).

#### 2.3.2. Conceptual Model of the Kaweah River Watershed

The study domain encompasses an area of  $12,276\text{ km}^2$  of highly variable terrain in the Kaweah River watershed (Figure 1). The total relief in the study basin exceeds 4,000 m and includes the relatively humid Sierra Nevada Mountains and the semiarid Central Valley. We conceptualize these geographic regions as the two primary hydrologic systems. Across both systems, the surface soils are considered to be 2-m thick and are divided into six layers with variable thicknesses consistently defined across the entire domain. The soil overlies the 620 m thick bedrock aquifers that consist of nine computational layers.

Parameterization of these layers is different for the mountain and valley aquifer systems according to our conceptual geological model. The mountain system is characterized by higher precipitation, steep slopes, shallow soils, and fractured bedrock geology. We conceptualize the subsurface as three distinct geological layers with variable thickness and hydrologic properties: weathered saprolite, fractured bedrock, and less fractured bedrock at depth. Thicknesses of subsurface layering, particularly the saprolite layer are uncertain (Banks et al., 2009; Graham et al., 1997; Mohr et al., 2012). Our base model uses a 15-m thick saprolite layer consistent with observations in similar granitic rock in California (Graham et al., 1997). The valley aquifer system is a sedimentary basin with highly productive aquifers, relatively low precipitation, and a flat land surface. While the total depth of the model is consistent for the two aquifers, we conceptualize the Central Valley aquifer as a series of nine aquifer layers of variable thickness based on well core observations and consistent with the Faunt (2009) geologic model of the Central Valley (supporting information Text S2). The mountain and valley aquifers are considered to be hydrologically interconnected surface and ground-water systems that respond differently to forcing and model parameters. In this view, the apparent differences in hydrologic responses are emergent properties of consistent fundamental physical processes that operate across a varied terrain.

We focus simulations on the relatively shallow groundwater that is most likely to interact with surface processes and use a total model thickness of 622 m. The bottom boundary as well as the north, east, and southern boundaries of the model are considered no-flow boundaries. As this study domain is part of a larger regional system, we allow a small flux (0.01 m/day) of water out of the western boundary to prevent the artificial accumulation of water in the Central Valley. Finally, to isolate the impact of precipitation data set uncertainty on simulated hydrologic processes, we conceptualize the system as a quasi-predevelopment state where the influence of groundwater pumping, stream impoundments and reservoirs, or surface diversions is not considered. This is due to the paucity of water transfer data and the inability of the model to simulate reservoir processes.

### 2.3.3. Summary of Model Parameterization

The Kaweah River watershed model has a horizontal grid resolution of 1 km<sup>2</sup> covering the area shown in Figure 1. This grid size results in a domain with 124 columns and 99 rows and 15 variable thickness layers (six soil and nine aquifer layers; in the Sierra Nevada region some adjacent computational layers are assigned the same hydraulic parameters resulting in three types of bedrock layers—Table S1) covering the top 622 m of the subsurface. At the land surface, slopes are derived from a 30-m digital elevation model obtained from the National Elevation Data Set (Gesch et al., 2018) and Manning's roughness values are assigned based on reference tables (Chow, 2009). Vegetation types are based on the USDA CropScape data set and are aggregated to the IGBP classification scheme. The vertical discretization represents the surface soils with high resolution (0.05-, 0.1-, 0.15-, 0.3-, 0.4-, and 1-m layers) where hydraulic conductivity and porosity are obtained from the POLARIS data set (Chaney et al., 2016). Model Layers 7 through 15 represent key features of the aquifer region where their hydraulic conductivity and porosity parameters within the Central Valley are assigned from drill core observations compiled by Faunt (2009). In the Sierra Nevada region, bedrock was parameterized based on a geologic map and reference table values (Jennings, 1977; Welch & Allen, 2014) due to the lack of drill core observations of bedrock. To help decrease model run time, the van Genuchten parameters are assigned as constant values of  $n = 2$  and  $\alpha = 3.5 \text{ m}^{-1}$  over the domain. A more detailed description of the inputs and data processing steps are provided in supporting information Text S2.

### 2.3.4. Numerical Experiments and Model Spin-Up

In this study, we perform three numerical experiments. Experiment 1 assesses the propagation of uncertainty from precipitation to groundwater and includes four scenarios with constant physical parameterizations and consistent meteorological forcings except for hourly precipitation. In these scenarios, each precipitation data set is resampled to 4-km spatial resolution while preserving the total precipitation volume from its native resolution. For Experiment 2, we assess the impact of uncertainty in the subsurface properties by changing the saprolite layer thickness. We create subsurface models with saprolite layers of 2-, 8-, and 15-m thickness and run each of the four precipitation data sets at 4-km resolution with each subsurface parameterization, for a total of 12 model scenarios. Experiment 3 consists of a single scenario generated by averaging the precipitation forcings of the four input data sets for each grid cell at a daily scale. We compare results of Experiment 3 to the simulated hydrology from each individual data set

(Experiment 1) as well as the mean of the simulated streamflow, ET, and SWE from the four simulations with the four precipitation forcing data sets, all using the 15-m saprolite subsurface parameterization.

We initialize the hydrologic model through a two-step spin-up process to bring the subsurface storage into dynamic equilibrium with each set of meteorological forcings. The first step applies a constant in time net precipitation flux ( $P - ET$ ) to the initially dry model system to fill the underground storage and create a rough approximation of the flow network (Livneh et al., 2013). From this point, each ParFlow.CLM model scenario is individually spun up to dynamic equilibrium by iteratively performing 1-year simulations using the hourly meteorological forcing of WY2016 for that scenario (Ajami et al., 2014). WY2016 approximately represents average precipitation and air temperature conditions and is recent enough to include several available remote sensing and reanalysis products for model performance evaluation. Dynamic equilibrium for each scenario is defined as the absolute percent change in subsurface storage value becoming less than 0.1% in consecutive simulations (Ajami et al., 2015). The ParFlow.CLM model is optimized for massively parallel computer systems (Kollet et al., 2010), and simulations were performed on a high-performance parallel computing cluster using 35 processors. One year of simulation required ~850 service units, with a service unit equivalent to 1 hr of time used by one processor.

#### 2.4. Analysis of Uncertainty in Precipitation and Simulated Hydrology

We assess the uncertainty of precipitation data sets using the three-cornered hat (3CH) method (Gray & Allan, 1974). The 3CH method estimates the variance from the ensemble of each data set by computing the differences between multiple combinations of three data sets. In this article, we present the square root of the variance from ensemble ( $\sigma$ ), analogous to the standard deviation with units of mm/day, and we refer to this as the uncertainty in a particular data set in the rest of the text. The 3CH method is based on the assumption that the errors are normally distributed, and the theory is described in the following. A given set of observations,  $obs_i$ , contains two components, the true value ( $x$ ) and the measurement error ( $e_i$ ):  $obs_i = x + e_i$  (Long et al., 2014). When three sets of observations ( $i, j, k$ ) exist, the difference between a pair of observations can be written as follows:

$$obs_i - obs_j = x + e_i - (x + e_j) = e_i - e_j \quad (1)$$

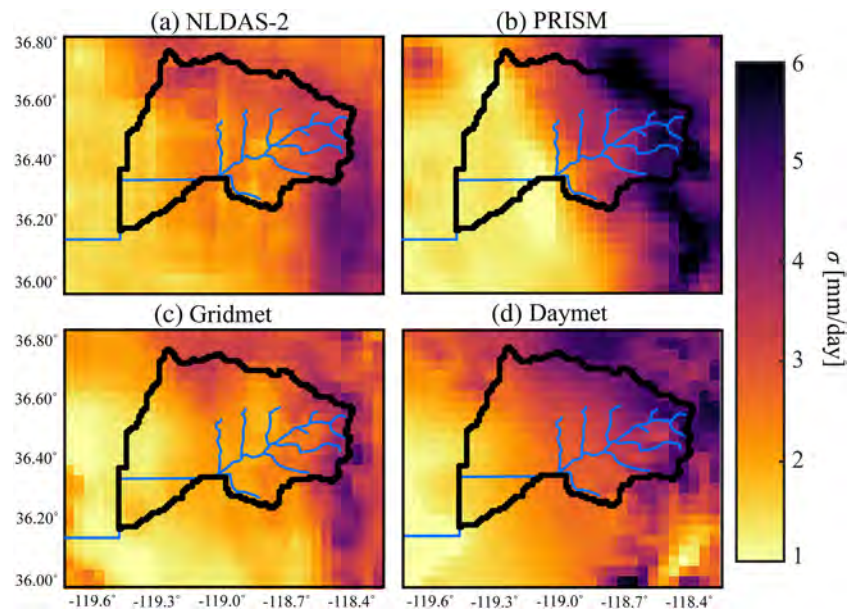
The variance of the differences between observations ( $\sigma_{ij}^2$ ) can be written as follows:

$$\sigma_{ij}^2 = \sigma_{ei}^2 + \sigma_{ej}^2 - 2\text{cov}(e_i, e_j) \quad (2)$$

Finally, the individual error variances  $\sigma_{ei}^2$  can be obtained through substitution and the variance of differences between two data sets ( $\sigma_{ij}^2$ ).

A basic hypothesis of the original method is that all considered data sets are uncorrelated, and the covariance in Equation 2 is zero. Given that our precipitation data sets are generated using many of the same input gauges and assumptions, we use a modified version of the 3CH method that does not assume uncorrelated data sets (Premoli & Tavella, 1993) and apply it to the four data sets (Anthes & Rieckh, 2018). By introducing the covariances of the measured data and utilizing an optimization approach to solve the linear equations (Premoli & Tavella, 1993), the complete covariance matrix between three or more measurements is estimated, allowing for the uncertainty estimation with correlated samples.

We apply the 3CH method to the 4-km resolution precipitation data sets at two scales: a daily watershed average scale where the uncertainty of the time series in Figure 2a is computed and a grid scale where the uncertainty of a data set is calculated in a spatially distributed manner. The uncertainty values are not true estimates of the data set's error because each data set has unquantified bias due to issues such as rain gauge undercatch and sparse gauge networks. We also apply the 3CH method to several model output variables including  $ET$ , potential recharge ( $R$ ), soil moisture ( $\Theta$ ), and subsurface storage ( $S$ ). We recognize that uncertainty of output variables calculated with the 3CH method is not representative of a true uncertainty introduced by model structure and parameters. We believe that these uncertainties are illustrative of the uncertainty in simulated hydrologic variables caused by the uncertainty in precipitation inputs. The 3CH method was selected because it considers covariances of correlated data sets in estimating uncertainty for both spatially distributed or basin-averaged values of input forcings and simulated hydrology using a



**Figure 3.** Uncertainty as predicted from the 3CH method for daily precipitation forcing data sets at 4-km spatial resolution. The uncertainty ( $\sigma$ ) is presented for the (a) downscaled NLDAS-2, (b) PRISM, (c) Gridmet, and (d) Daymet data sets.

single consistent approach. This method is advantageous over using a metric-like variance that does not consider covariances between multiple data sets or the MCMC approaches that require a large number of simulations for uncertainty quantification.

To analyze the propagation of uncertainty, we utilize the simulated grid cell saturation and pressure head to calculate surface and subsurface fluxes for certain regions in the domain. We compute the net vertical and lateral soil water fluxes from the soil zone (top 2 m of domain) at each grid cell using the simulated pressure head and the Richards' equation (Maxwell & Miller, 2005). Additionally, we calculate run-on infiltration, the infiltration of excess precipitated water from upslope that has reached a grid cell as overland flow and will contribute to soil water storage, at an annual scale. Run-on is calculated as the total infiltration at the land surface minus the precipitation. By setting negative values (representing runoff + interception) to zero, we identify locations where infiltration is greater than precipitation indicating run-on infiltration. Finally, we calculate the net groundwater flux at the watershed boundary ( $GW$ ) as a residual of the water balance,  $GW = P - ET - Q - dS$ , where  $dS$  is the change in total subsurface storage in the watershed and  $GW = GW_{in} - GW_{out}$ . While this definition does not explicitly define  $GW$  flux as recharge to the Central Valley aquifer, the elevation gradients in the model domain would result in  $GW$  influx to the Central Valley.

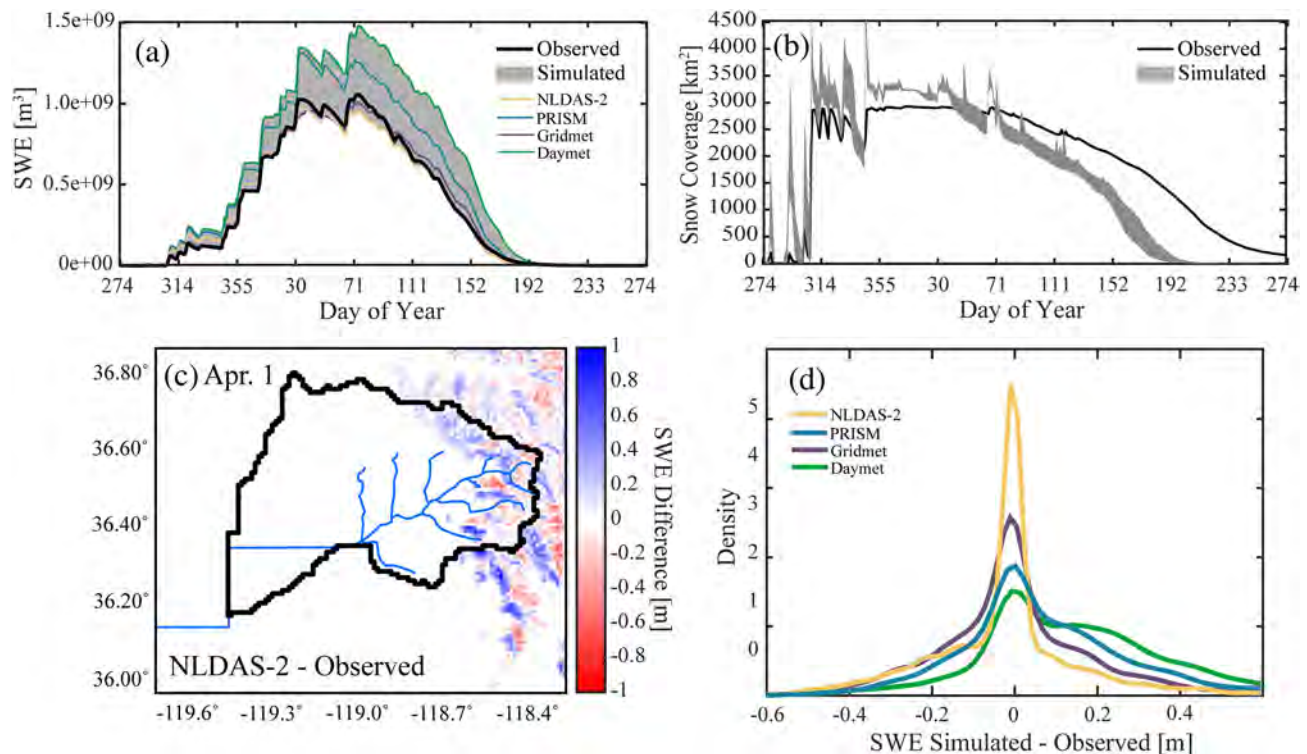
Finally, we apply the Diebold-Mariano (DM) test for predictive accuracy to test if there are statistically significant differences in the simulation performance of the two scenarios from Experiment 3 (Gilleland & Roux, 2015). The DM test is a statistical test for the null hypothesis that the forecast accuracy between two simulated time series is equal compared to observations. We choose the DM test because it makes no assumptions on the distribution of errors between simulated and observed values, and it incorporates temporal autocorrelations.

### 3. Results and Discussion

#### 3.1. Uncertainty of Precipitation Data Sets

The 3CH method was applied to every grid cell of the four daily precipitation data sets (NLDAS-2, PRISM, Gridmet, and Daymet) at 4-km resolution to provide spatially distributed estimates of daily precipitation uncertainty. Figure 3 presents the uncertainty for each data set. Across all data sets, the precipitation uncertainty is highest in the Sierra Nevada mountains covering the eastern portion of the study domain. The locations of maximum uncertainty coincide with the locations of maximum annual precipitation in the domain





**Figure 4.** Daily time series of simulated and observed snow water equivalent (a) and snow covered area (b). (c) Spatial distribution of annual differences between simulated and observed SWE using the downscaled NLDAS-2 precipitation forcing. (d) Probability density functions of the difference between simulated and observed SWE for the four precipitation forcing data sets from Experiment 1 using April 1, 2016 data.

(Correlation Coefficients: Daymet = 0.86; Gridmet = 0.89; NLDAS = 0.87; PRISM = 0.85). Uncertainties reach 6 mm/day in some locations and highlight the importance of considering precipitation uncertainty in hydrologic modeling studies. The spatial patterns and magnitudes of uncertainties are consistent with the literature characterizing errors in gridded precipitation data sets who found the highest uncertainty of ~100–200 mm/year in the mountain regions (Henn et al., 2018; Lundquist et al., 2015). The uncertainty values presented in Figure 3 are larger than previously reported values because uncertainty calculated with the 3CH method includes the differences in the annual total, as well as day-to-day variability between data sets.

### 3.2. Evaluation of Model Performance in Complex Terrain

We build confidence in the Kaweah River watershed model by comparing simulated hydrologic fluxes to available observational equivalents during the WY2016 study period. We compare the simulated hydrology to (1) estimates of daily unimpaired streamflow at USGS gauge locations, (2) data assimilation and remote sensing derived daily snowpack properties, and (3) remotely sensed 8-day *ET*. The historical record of observation coincides with widespread human manipulation of the hydrologic system making observations of dynamic steady-state rare. As a result, we focus our comparisons primarily on the Sierra Nevada region where human impacts on the hydrologic system are limited. The majority of the Kaweah River watershed is located in this less-impacted zone allowing reasonable comparison of simulated and observed variables, particularly their spatial patterns and variation over the annual cycle. We compare simulations from Experiment 1, forced with the four 4-km resolution precipitation data sets, to observed data sets. A summary of performance metrics is presented in Table 2.

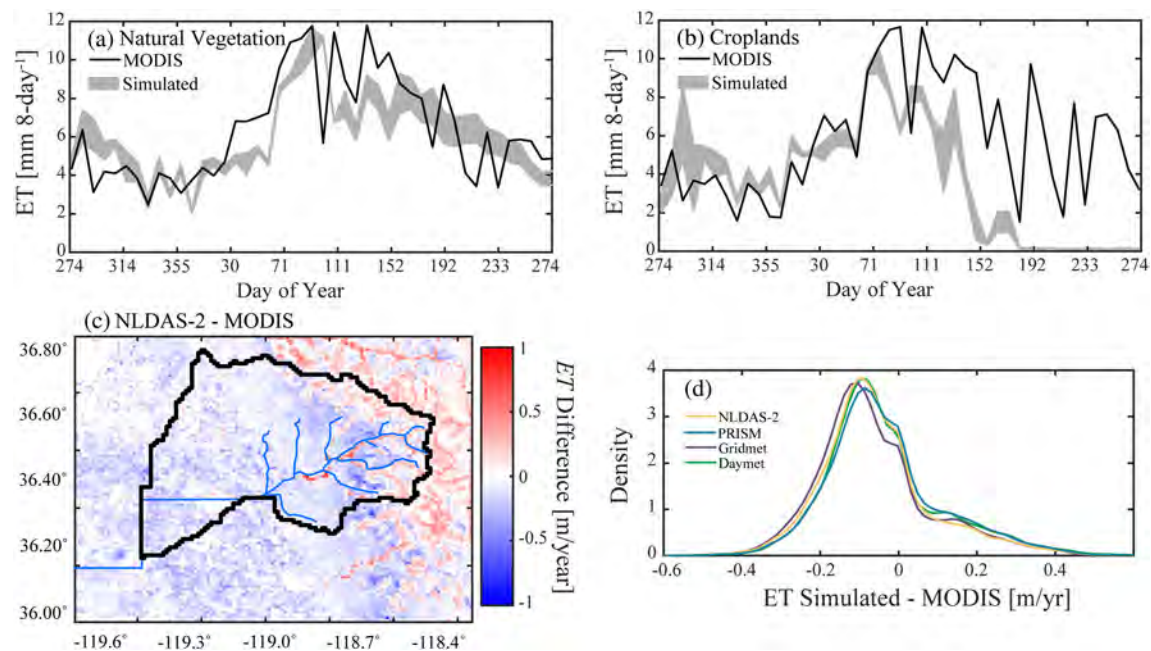
Comparisons between simulated and observed snowpack can provide assessment of the model's ability to capture the energy budget and hydrologic processes in the snow-dominated Sierra Nevada mountains. Figure 4 presents comparisons between the simulated *SWE* and *SCA* and estimates of the actual quantities from the reanalysis data set (Margulis et al., 2016). As the Margulis et al. (2016) data set is masked at an

elevation of 1,500 m, below which SWE is set to 0, we apply the same mask to the simulated SWE for comparisons. The annual cycle of snowpack accumulation and melting is represented in Figure 4a where the solid black line is the reanalysis SWE values and the gray shaded region represents the range of simulated values from Scenarios 1–4. The simulated annual pattern of daily SWE matches the reanalysis values with bias and NSE values close to 1 for all data sets. The estimated actual SWE values from Margulis et al. (2016) lie within the range of simulated values at all times throughout the year. The estimated actual SCA does not fall within the range of simulated SCA values during the peak snowpack season (DOY 340–60) when the model overestimates the snowpack (Figure 4b). Additionally, small amounts of estimated actual SCA remain throughout the summer period that are not preserved in the simulations. We attribute these differences primarily to the coarse resolution topography and air temperature inputs to our model. Small ravines can cause shading effects that are not captured by the coarse resolution model.

We compare the spatial patterns of simulated snowpack to the reanalysis data set in Figure 4c using the downscaled NLDAS-2 scenario. The simulated SWE from the downscaled NLDAS-2 scenario does not contain systematic bias, and the locations where snow is under or overestimated are not correlated with elevation, air temperature, or vegetation type ( $r^2$  and correlation coefficient values are  $<0.3$  in all cases). Comparison of the reanalysis SWE data set to snow pillow observations found an average RMSE of approximately 0.1 m (Margulis et al., 2016), which contributes to the differences between our simulated SWE and the reanalysis estimates. Figure 4d presents probability density functions (PDFs) of the differences between simulated and reanalysis SWE on 1 April 2016 for each of the 4-km resolution precipitation data set scenarios. The PDFs are consistent with results from Figure 4a in that the downscaled NLDAS-2 scenario has relatively low bias compared to observations, whereas the Gridmet, PRISM, and Daymet scenarios have successively higher bias in estimating SWE. This result is influenced by the SWE reanalysis data set generated by forcing a land surface model by topographically downscaled NLDAS-2 data set (Margulis et al., 2016). Although we use another downscaled version of the original NLDAS-2 data set, finding that our simulated SWE from the NLDAS-2 scenario provides the best match with the reanalysis data set does not necessarily mean that NLDAS-2 is the best data set to use.

We compare simulated *ET* fluxes to remote sensing estimates from the MOD16 product (Mu et al., 2011). As our model represents predevelopment conditions and does not incorporate irrigation, simulated *ET* is assessed separately over irrigated croplands and “natural” vegetation (all land cover types except croplands and urban areas). The 8-day annual pattern of *ET* from natural vegetation is shown in Figure 5a where the solid black line is the MODIS observed value and the shaded region is the range of simulated values from Experiment 1. The range of simulated *ET* encompasses the observed *ET* throughout the year in natural vegetation. For croplands, simulated *ET* matches the observations during the wet season, but it is lower than the observed *ET* that is sourced from irrigation in summer. Figure 5c presents the differences between the total annual values of simulated *ET* from the downscaled NLDAS-2 scenario and the MODIS data. In the natural vegetation, the simulated *ET* is overestimated at the highest elevations and within riparian areas at all elevations. The riparian *ET* overestimation is likely due to the river channel width of 1 km, equal to the model horizontal resolution as ParFlow.CLM does not consider local grid refinement. The PDFs of spatial differences between simulated and observed *ET* show that the peak differences in *ET* occur at negative values (Figure 5d), an artifact of the irrigation impacting the observed *ET*.

In order to demonstrate streamflow (*Q*) performance from multiple simulations in a clear manner, Figure 6 presents flow duration curves (FDC) of simulated *Q* at three USGS gauge locations as well as the observed FDC. A time series comparison is presented in Figure S4. We also present the simulated and observed *Q* at the Kaweah Terminus dam where the observed *Q* is the full natural flow *Q* product from the CA Department of Water Resources. The range of simulated *Q* from Experiment 1, at Gauges 1 and 2, encompasses the observations. Simulations underestimate daily *Q* at lower elevation sites, Site 3 and the Kaweah Terminus dam, during the high flow regimes (exceedance probability  $< \sim 40\%$ ). We primarily attribute this difference to unrealistic stream widths (1-km resolution) in the model that affects surface-groundwater interactions along streambeds. ParFlow.CLM users often calibrate Manning's *n* values to account for this problem (Bhaskar et al., 2015). We did not perform any calibration which might bias the model performance under certain precipitation conditions as we are investigating the impacts of forcing data on simulated hydrology. As an uncalibrated model, the *Q* simulations are not intended for streamflow or flood prediction, and the model performance is reasonable for investigating the precipitation uncertainty propagation into the

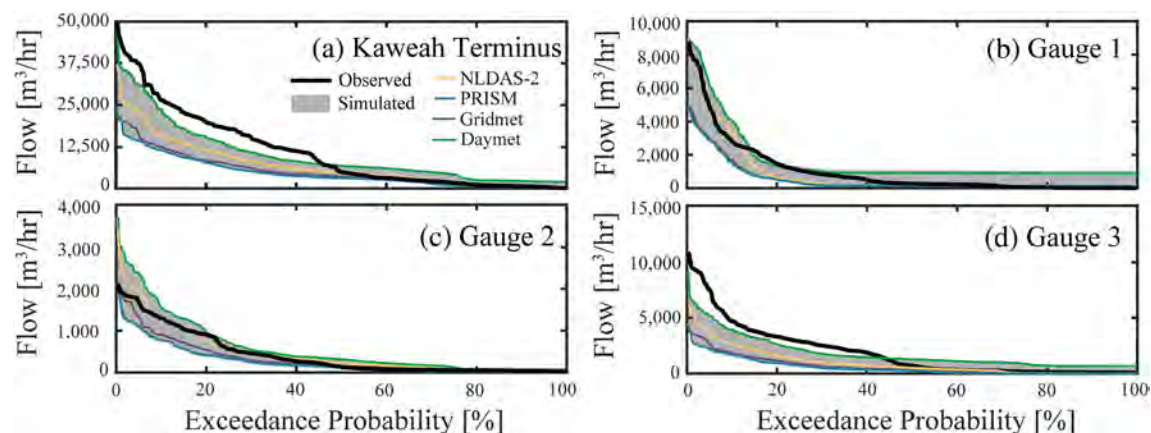


**Figure 5.** Time series of spatially averaged 8-day  $ET$  values as observed by MODIS or simulated in Experiment 1 for (a) Natural vegetation and (b) Croplands. The gray shaded regions represent the variability in simulated values from four precipitation data sets in Experiment 1. (c) Spatial difference in annual values of simulated and MODIS observed  $ET$  using the simulation forced with downscaled NLDAS-2 precipitation. (d) Probability density functions illustrate the differences between spatially distributed annual  $ET$  observed from MODIS and model simulations in Experiment 1.

groundwater. We focus primarily on understanding how precipitation uncertainty propagates into groundwater within the mountain system where the model is best validated. We recognize, however, that model parameter choice will impact the hydrologic fluxes and states and may impact the propagation of uncertainty into the groundwater system. The variation in simulated streamflow highlights how the surface-groundwater interactions are impacted by both model parameters and precipitation forcing and disentangling the effects of each merits future research.

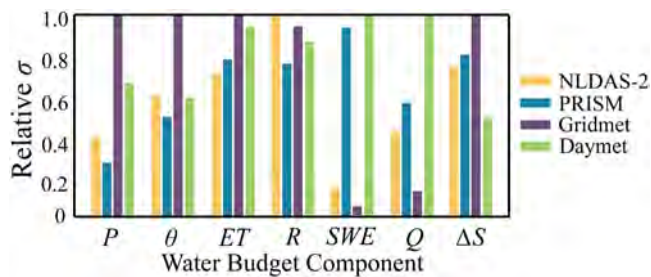
### 3.3. Uncertainty Propagation From Precipitation to Groundwater

We applied the 3CH method to the simulated water budget components to estimate the uncertainty of outputs caused by variability in precipitation forcing. We summarize overall uncertainty results at the watershed scale for the Kaweah Terminus watershed (located entirely in the Sierra Nevada portion of the study domain, Figure 1b) and select elevation intervals across the domain. The calculated uncertainties



**Figure 6.** Simulated and observed flow duration curves for four validation locations. Gauge locations are shown in Figure 1.





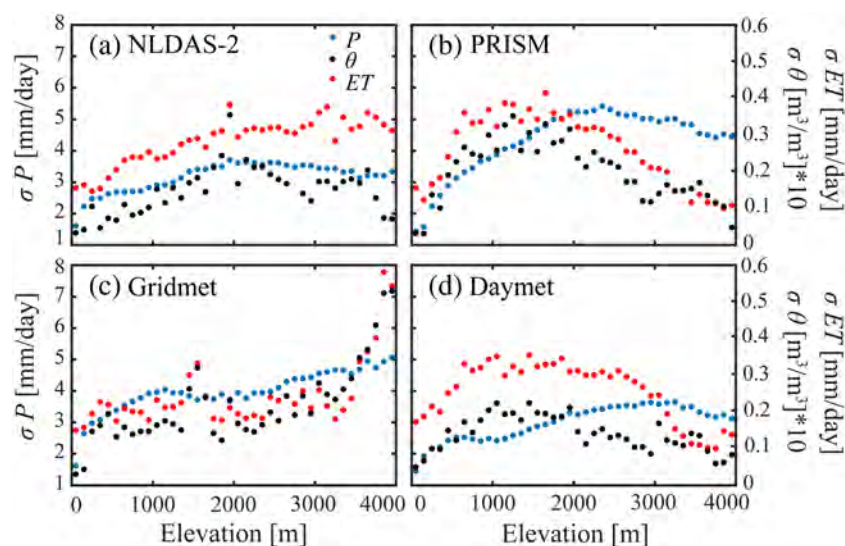
**Figure 7.** Relative uncertainty ( $\sigma$ ) calculated for the Kaweah River Watershed using the daily water budget components of models in Experiment 1: precipitation ( $P$ ), root zone soil moisture ( $\theta$ ), evapotranspiration ( $ET$ ), potential recharge ( $R$ ), snow water equivalent ( $SWE$ ), streamflow at Kaweah Terminus ( $Q$ ), and change in total subsurface storage ( $\Delta S$ ). In order to display several metrics on one chart, uncertainties are normalized to the precipitation forcing scenario with the highest value of  $\sigma$  for that variable.

represent the uncertainty of daily fluxes and states caused by variability in precipitation forcing. In Figure 7, we present  $\sigma$  of the  $P$  forcing and simulated hydrology from Experiment 1 scenarios. To ease comparison,  $\sigma$  is presented as relative uncertainty, normalized to the maximum  $\sigma$  of that variable. At the watershed average scale, the Gridmet precipitation data set has the highest  $\sigma$  even though the uncertainty in  $P$  is lower than other data sets at a grid scale (Figure 3). This result is possible when individual grid cells are not dramatically different from the other data sets, but the watershed total  $P$  is very different relative to the other data sets (Figure 2a). The PRISM data set has locations with considerably higher and lower  $P$  than the ensemble average that cancel out in the watershed average. This result highlights the difficulty in selecting the “best” precipitation forcing data set because the strengths of each data set vary, and what is best depends on the objectives.

An important question in hydrologic modeling is whether high uncertainty in precipitation forcing magnifies uncertainty of simulated water budget components. Results of Experiment 1 demonstrate that it does

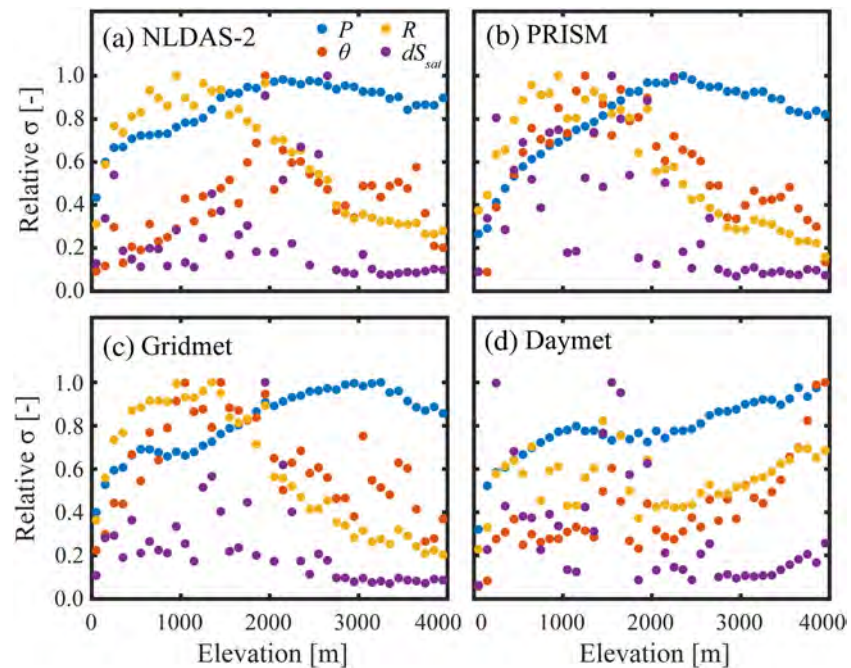
not (Figure 7). For example, the PRISM data set with the smallest  $\sigma$  for  $P$  does not result in smaller  $\sigma$  for other fluxes or storages. The Gridmet data set with the highest  $\sigma$  for  $P$ ,  $\theta$ ,  $ET$ , and  $\Delta S$  has the lowest  $\sigma$  for  $SWE$  and  $Q$ , which are important metrics for water management. In mountainous watersheds, the  $SWE$  magnitude depends on the spatial patterns of precipitation more than the watershed average values. For snow accumulation, precipitation must fall in the high-elevation regions where air temperatures are below the freezing point (Figure S4). Because streamflow in the Sierra Nevada region depends on snowmelt, the lower  $\sigma$  for  $SWE$  in the Gridmet scenario results in a lower  $\sigma$  for  $Q$ . This result highlights the importance of accurate spatial patterns in precipitation forcing for hydrologic simulation.

To demonstrate how uncertainty in precipitation forcing propagates from precipitation to groundwater, we apply the 3CH method in a spatially distributed manner to the 1-km scale simulated outputs. Figure 8 presents the absolute uncertainty in  $P$ ,  $\theta$ , and  $ET$  binned across the elevation gradient in the study domain. The spatial patterns of  $\sigma$  in soil moisture do not match the spatial patterns of  $\sigma$  in precipitation, and there are no significant relationships between the  $\sigma$  of precipitation and soil moisture at either the grid scale or the elevation binned values ( $p$  values  $> 0.05$ ). Spatial patterns of  $\sigma$  in soil moisture are related to topographic



**Figure 8.** Uncertainty ( $\sigma$ ) in surface water budget components averaged across 100-m elevation intervals. Precipitation ( $P$ ), root zone soil moisture ( $\theta$ ), and evapotranspiration ( $ET$ ) are from models in Experiment 1 using downscaled NLDAS-2 (a), PRISM (b), Gridmet (c), and Daymet (d) precipitation forcings.

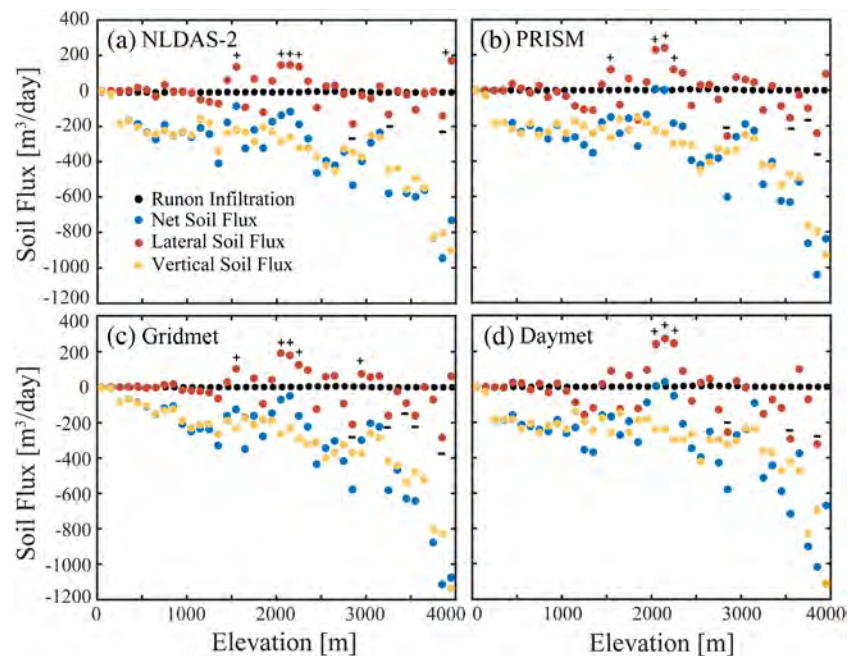




**Figure 9.** Relative uncertainty ( $\sigma$ ) in water budget components averaged across 100-m elevation intervals. Precipitation ( $P$ ), root zone soil moisture ( $\theta$ ), potential recharge ( $R$ ), and change in groundwater storage ( $dS_{sat}$ ) are from models in Experiment 1 using downscaled NLDAS-2 (a), PRISM (b), Gridmet (c), and Daymet (d) precipitation forcings.

properties, particularly the slope of the grid cell, with  $p$  values  $\ll 0.05$ , although correlation coefficient values are low ( $r = 0.12, 0.18, 0.15$ , and  $0.14$  for the downscaled NLDAS-2, PRISM, Gridmet, and Daymet scenarios, respectively). The uncertainty in simulated  $ET$  is tightly correlated to the uncertainty in  $\theta$  with correlation coefficients of  $0.70, 0.91, 0.74$ , and  $0.92$  for the downscaled NLDAS-2, PRISM, Gridmet, and Daymet scenarios, respectively.  $ET$  is the largest component of the water budget and by extracting water from the soil reduces the water available for potential recharge.

Precipitation is connected to the subsurface water budget via groundwater recharge. Figure 9 presents the binned average values of relative  $\sigma$  in  $P$ ,  $\theta$ ,  $R$ , and daily change in total subsurface storage in the saturated zone ( $dS_{sat}$ ) across the elevation gradient. Generally, the  $\sigma$  for precipitation increases with elevation, except at the highest elevations ( $>3,000$  m) where Gridmet is the only data set with increasing  $\sigma$ . This is not the case, however, for water stored underground. As discussed previously, the uncertainty in  $\theta$  is related to the topographic conditions. For  $R$  to occur, sufficient  $\theta$  is needed to increase unsaturated hydraulic conductivity and facilitate downward water flux. Although high-elevation regions during the snowmelt have the largest  $R$  (Figure S6), the uncertainty patterns in  $R$  across the elevation gradient typically mimic the uncertainty patterns of  $\theta$  ( $r = 0.08, 0.77, 0.74$ , and  $0.60$  for the downscaled NLDAS-2, PRISM, Gridmet, and Daymet simulations, respectively). Higher precipitation uncertainty in the semiarid Central Valley region (Figure 3) of the downscaled NLDAS-2 data set causes the low correlation between  $\theta$  and  $R$  for the downscaled NLDAS-2 data set. Groundwater recharge in semiarid regions is typically not related to antecedent  $\theta$  (Scanlon et al., 2006; Schreiner-McGraw & Vivoni, 2017), so the low correlation for the NLDAS-2 data set is expected. The uncertainty in  $dS_{sat}$  follows a similar trend with elevation to that of  $R$  but is complicated by the geologic structure and subsurface hydraulic properties. The uncertainty in  $dS_{sat}$  is high at the mountain front where sharp contrasts in subsurface permeability create large head gradients. The average elevation bin across the four scenarios of Experiment 1 for the maximum  $\sigma$  of  $P$  ranges from 2,800–2,900 m, while the maximum  $\sigma$  of  $\theta$  is in 2,000- to 2,100-m bin,  $R$  is in 1,200- to 1,300-m bin, and  $dS_{sat}$  is in 1,300 to 1,400-m bin. For the model structure applied in this study, as water flows downslope, topography-driven flow processes and convergence move the locations of uncertainty in simulated hydrology downslope.



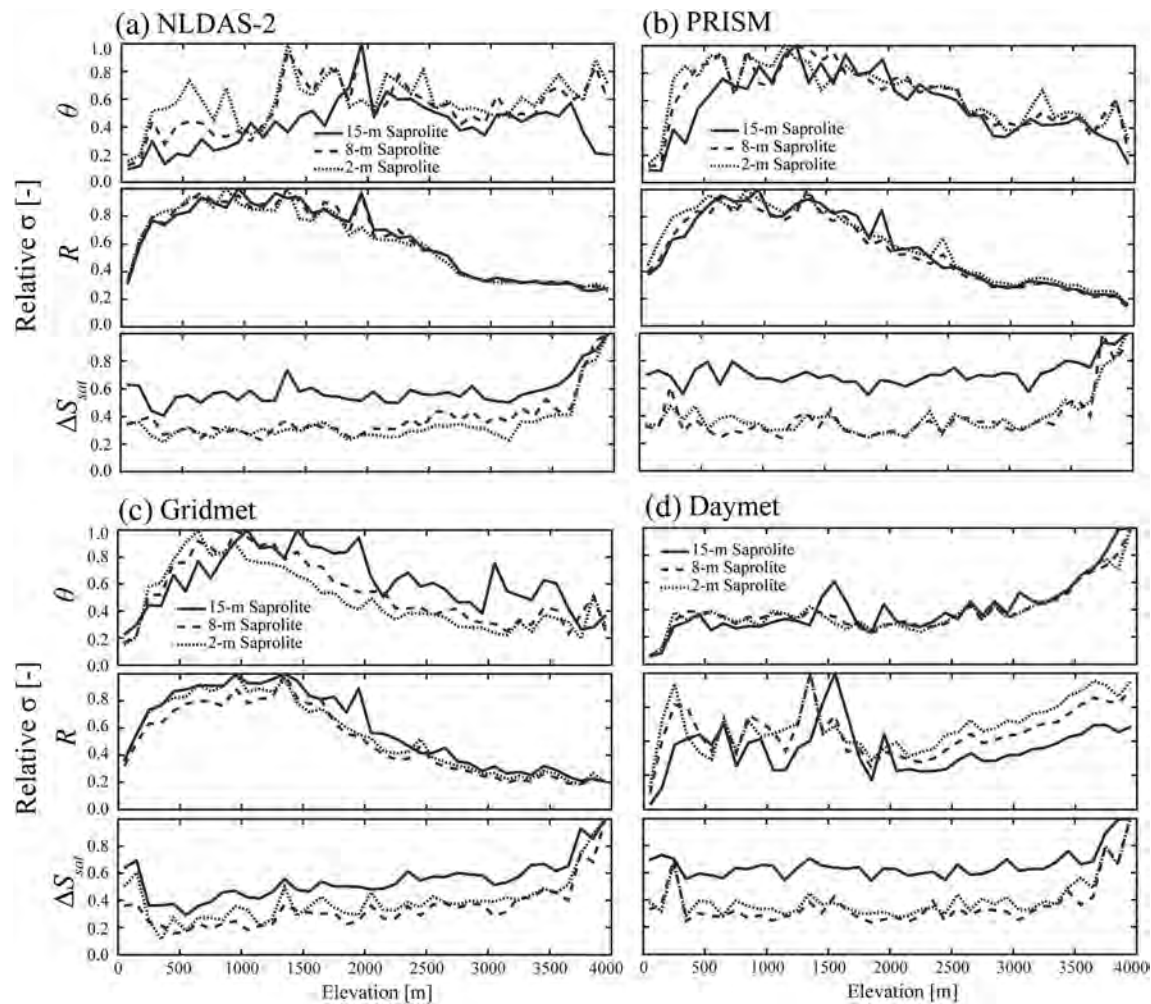
**Figure 10.** Elevation binned values of mean net, lateral, and vertical soil water fluxes as well as run-on infiltration simulated from models in Experiment 1 using downscaled NLDAS-2 (a), PRISM (b), Gridmet (c), and Daymet (d) precipitation forcing. Symbols indicate bin values of lateral soil flux that a one-way ANOVA determined to be significantly greater (+) or lower (−) than 0.

For further evidence that topography exerts control on the water budget, we present subsurface water fluxes as average values binned by elevation for each of the four simulated precipitation scenarios (Figure 10). For all the elevation bins and all  $P$  forcing scenarios, the soil net vertical flux is negative indicating net downward flow into the saprolite layer. The downward soil vertical fluxes become larger as elevation increases causing higher potential  $R$ , potentially a result of lower  $\theta$  and  $ET$  in these regions. Although average vertical soil fluxes across all elevation bands are downward, at elevations below 3,000 m, many soil locations (14.8%, 12%, 13.5%, and 15.1% of the Kaweah River watershed area in the Daymet, Gridmet, downscaled NLDAS-2, and PRISM scenarios, respectively) in local topographic lows receive positive upward fluxes from the saprolite layer (Figure S6).

High-elevation (>3,000 m) soil water contents are further depleted by lateral water flux out of soils. For elevation bands between 1,500 and 2,500 m, there is typically a positive lateral flux of up to 250 m<sup>3</sup>/day per grid cell into the soil. To compare the importance of subsurface and overland flow processes, we compute the annual run-on infiltration to each grid cell. Figure 10 demonstrates that run-on infiltration is negligible across all elevation gradients. These analyses suggest that the high  $\theta$  and  $ET$  uncertainty in the middle-elevation range (1,500 and 2,500 m) is primarily controlled by lateral soil water redistribution from the high elevations rather than moisture contributions from overland flow or deeper flow paths involving saprolite layers. These results may be biased by the relatively coarse resolution of the model. Although streamflow is generated in the small, upland portions of the watershed, lower-order watersheds are not accurately represented at 1-km resolution. This representation likely impacts the surface-groundwater interactions in these regions.

### 3.4. Impacts of Uncertain Subsurface Parameters

Given that subsurface flow patterns are governed by the conceptual geological model and subsurface parameters in integrated hydrologic models, we assess the impact of subsurface properties on the conclusions drawn from Experiment 1 (sections 3.1–3.3). An important uncertainty in the subsurface parameterization is the thickness of the highly weathered saprolite zone, which has been found to vary between 2- and 20-m thickness in regions with similar geology (Banks et al., 2009; Graham et al., 1997). Previous research has

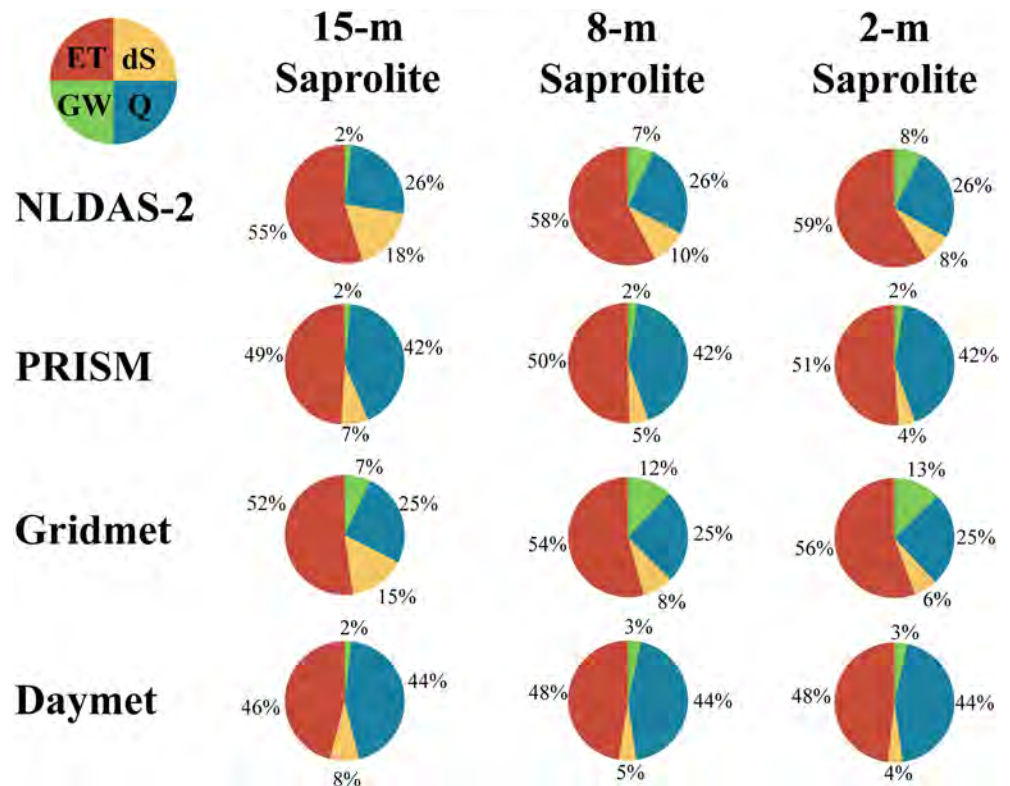


**Figure 11.** (a–d) Elevation binned values of relative uncertainty ( $\sigma$ ) in root zone soil moisture ( $\theta$ ), potential recharge ( $R$ ), and daily change in total saturated subsurface storage ( $\Delta S_{sat}$ ) for each of the scenarios in Experiment 2.

shown that this layer of higher permeability plays an important role in governing surface water–groundwater interactions (Foster & Maxwell, 2019). We applied the 3CH method to the simulated subsurface water budget, consisting of  $\theta$ ,  $R$ , and  $\Delta S_{sat}$ , from each of the model scenarios in Experiment 2. Figure 11 presents the relative ( $\sigma$ ) ( $\sigma$  normalized by the maximum of that variable) for each scenario across elevation bins, similar to Figure 9. The elevation binned spatial patterns of  $\sigma$  in simulated subsurface water budget components are consistent across the three different saprolite layer thicknesses. This result indicates that the saprolite thickness does not impact the propagation of uncertainty from precipitation to groundwater in this watershed. Additionally, across all three saprolite scenarios, positive soil water fluxes occur via lateral soil water flux from elevations  $>3,000$  m to elevations between 1,500 and 2,500 m (data not shown). These results help build confidence in the role of topographic controls on the propagation of uncertainty from precipitation to groundwater for different geologic structures and not a single realization of the watershed.

Although changes to the saprolite layer do not systematically impact the propagation of uncertainty in precipitation to groundwater, they do impact regional flow pathways in the groundwater system. Figure 12 presents the partitioning of the precipitation into water budget components for the Kaweah Terminus watershed where each flux is presented as a percentage of the  $P$ . Across all four precipitation forcing data sets, a decrease in the saprolite layer thickness reduces total subsurface storage and increases annual  $ET$  and  $ET/P$  ratio. A thinner saprolite layer restricts potential recharge beneath the 2-m soil layer,





**Figure 12.** Water budget partitioning as a fraction of the input precipitation for the Kaweah Terminus watershed using the 12 scenarios from Experiment 2. Water budget components include evapotranspiration ( $ET$ ), streamflow ( $Q$ ), groundwater flux out of the watershed ( $GW$ ), and change in total subsurface storage over the watershed ( $dS$ ).

which elevates soil moisture values and increases available water for  $ET$ . Across all Experiment 2 scenarios, a decrease in the saprolite layer thickness increases  $GW$  flux out of the Kaweah Terminus watershed particularly for the NLDAS-2 and Gridmet data sets where the differences are up to 4 times the value from the 15-m saprolite layer. For the PRISM and Daymet scenarios, however, the differences in  $GW$  flux out are less than 1% of the precipitation. Increases in  $GW$  outflux are caused by the shallower water table in the mountain regions of the 2-m saprolite layer scenario, as well as larger potential vertical recharge into fractured bedrock layers. The larger values of potential vertical recharge into bedrock are a result of less lateral flow within the thin saprolite layer compared to thicker saprolite layers. It should be noted, however, that vertical layer thickness can impact the accuracy of model simulations, so by changing the model layer thickness small differences in simulated fluxes may be observed.

For all saprolite layer thicknesses, there are consistent patterns in the water budget partitioning using different precipitation forcings. The  $ET$  ( $Q$  flux) is lowest (highest) in scenarios where more precipitation occurs at higher elevations. For example, the Daymet precipitation data set has the highest fraction of precipitation at high elevations (Table 1) where lower vegetation coverage and higher snow coverage lead to higher potential vertical recharge. This higher vertical recharge results in more water entering the mountain flow paths where it eventually discharges as  $Q$  flux to the Central Valley.

### 3.5. Merging Multiple Forcings or Merging Their Simulated Fluxes

Table 2 presents the performance metrics obtained for each of the precipitation forcing scenarios, including the mean of the four data sets run with the base case (15-m saprolite layer) model structure and the mean of the

**Table 1**  
Number of Precipitation Grid Cells per Annual Precipitation Quantile Obtained From Spatially Distributed Total Annual Precipitation Data Sets

Quantile	Range (mm/year)	NLDAS-2	PRISM	Gridmet	Daymet
0–10%	148–217	1,842	989	922	1,145
10–20%	217–248	664	1,189	1,979	1,091
20–30%	248–286	592	981	<b>2,340</b>	997
30–40%	286–349	992	1,163	1,495	1,253
40–50%	349–427	1,518	1,073	1,129	1,198
50–60%	427–536	1,695	1,232	827	1,156
60–70%	536–651	1,940	1,235	639	1,097
70–80%	651–737	<b>2,036</b>	1,331	714	829
80–90%	737–855	961	1,490	1,355	1,105
90–100%	855–1258	36	<b>1,593</b>	876	<b>2,405</b>

*Note.* Quantile ranges are derived from the combination of all four precipitation data sets. Bold values indicate the quantile with the highest number of cells for each dataset.



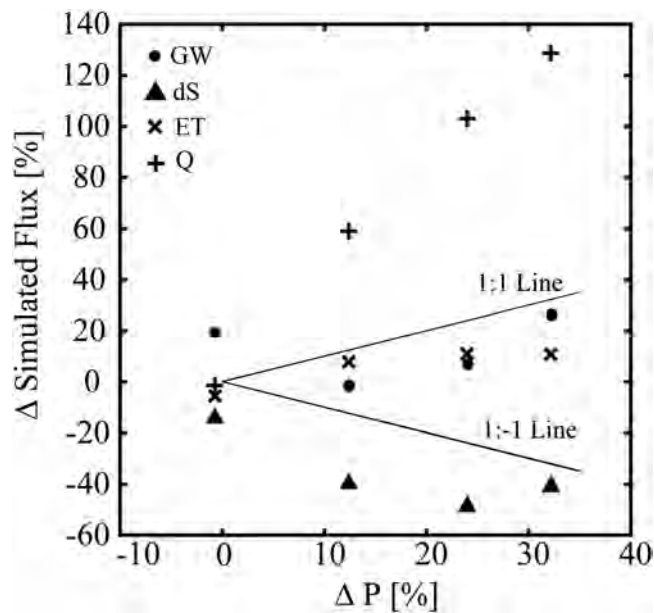
**Table 2**  
Performance Metrics Comparing Simulated Fluxes to Observed Values for the Model Scenarios Comprising Experiment 3

Metric			Daymet	Gridmet	NLDAS	PRISM	Mean of inputs	Mean of outputs	
Temporal	ET (mm/8 days)	Bias	0.91	0.80	0.87	<b>0.95</b>	0.92	0.88	
		MRE	0.32	0.32	0.31	<b>0.31</b>	0.32	0.31	
	SWE (mm)	Bias	1.60	0.99	<b>0.99</b>	1.39	1.20	1.24	
		MRE	0.60	0.09	<b>0.08</b>	0.57	0.19	0.28	
	Q1 (m <sup>3</sup> /hr)	Bias	1.48	0.83	0.72	1.21	<b>1.01</b>	1.06	
		NSE	−0.35	<b>0.54</b>	0.35	0.13	0.44	0.41	
	Q2 (m <sup>3</sup> /hr)	MRE	0.90	0.59	0.64	0.69	<b>0.58</b>	0.60	
		Bias	<b>0.87</b>	0.54	0.53	0.86	0.69	0.70	
	Q3 (m <sup>3</sup> /hr)	NSE	0.72	0.61	0.59	0.76	<b>0.80</b>	0.79	
		MRE	0.51	0.55	0.56	0.48	<b>0.43</b>	0.43	
	Q Terminus (m <sup>3</sup> /hr)	Bias	<b>0.71</b>	0.40	0.36	0.58	0.49	0.51	
		NSE	0.36	0.23	0.08	<b>0.37</b>	0.31	0.31	
	Spatial	ET (mm/year)	MRE	0.63	0.66	0.72	0.58	<b>0.56</b>	0.61
			Bias	1.45	0.63	0.63	1.29	1.01	<b>1.00</b>
		SWE (mm/year)	NSE	−0.57	<b>0.32</b>	0.24	−0.21	0.30	0.29
Spatial	ET (mm/year)	MRE	0.82	0.54	0.58	0.69	<b>0.52</b>	0.58	
		Bias	0.88	0.78	0.85	<b>0.92</b>	0.89	0.86	
	SWE (mm/year)	MRE	0.14	0.15	0.14	0.13	0.13	<b>0.13</b>	
		Bias	1.52	<b>1.01</b>	0.98	1.29	1.17	1.20	
		MRE	0.70	<b>0.42</b>	0.48	0.57	0.48	0.49	

Note. Bold values indicate the scenario with the best performance for that metric.

simulated hydrologic fluxes. We compare simulated hydrology with the limited available observational data sets by calculating performance metrics (MRE and B) for the Kaweah Terminus watershed using watershed averaged values consistent with the temporal scale of observations, as well as spatially distributed annual values. Consistent with discussion in previous sections, no individual forcing data set (Daymet, Gridmet, downscaled NLDAS-2, or PRISM) consistently outperforms the others. For example, the downscaled NLDAS-2 scenario has the best performance in simulating temporal patterns of *SWE* ( $B = 0.99$ ,  $MRE = 0.08$ ), whereas the PRISM scenario performs best in simulating spatial patterns of annual *ET* ( $B = 0.92$ ,  $MRE = 0.13$ ). The mean of simulated fluxes from the four precipitation forcing scenarios (Experiment 1) generally improves model performance where  $B$  is closer to 1, and a low MRE indicates improved prediction of observed fluxes. In the study watershed, however, running a single simulation with the mean of the precipitation forcing data sets results in the best performance metrics compared to performing four simulations and averaging the simulated results. Results from the DM test of equal forecast accuracy reveal that the mean of inputs scenario has statistically significant improvements in performance ( $p < 0.05$ ) for the time series of *SWE* and streamflow at Gauges 1 and 2. While the performance improvement gained by taking the mean of the forcing data sets is small, reduction in the computation time makes this strategy much more appealing. This finding is contrary to the previous work that showed that merging the simulated hydrologic fluxes improves model performance (Zhu et al., 2019).

We attribute this result, in part, to the sensitivity of simulated hydrologic fluxes to changes in  $P$ . Figure 13 illustrates the sensitivity of water budget components ( $Q$ ,  $ET$ ,  $dS$ , and  $GW$ ) to changes in mean annual  $P$  relative to the downscaled NLDAS-2 scenario for the Daymet, Gridmet, PRISM, and mean forcing scenarios. A 1% change in watershed annual  $P$  results in a greater than 1% change in simulated  $Q$ ,  $dS$ , and sometimes  $GW$ , while  $ET$  and  $GW$  flux from some scenarios are less sensitive and fall below the 1:1 line. Uncertainties in the  $P$  forcing magnify the sensitivity of the simulated fluxes more than linear increases or decreases in the  $P$  data. Additionally, the mean of multimodel ensemble approach works best when each model provides an independent estimate of a predicted output (e.g., streamflow) (Abramowitz et al., 2018). In our experiments, the forcing data are more independent than the simulations from four ParFlow.CLM models that are likely have similar biases due to the same model structure and parameters. Therefore, using the mean of the available precipitation data sets as the model forcing results in the best performance metrics.



**Figure 13.** The sensitivity of simulated annual fluxes to changes in annual precipitation using model scenarios from Experiment 3. The percent change in watershed average precipitation between the base case downscaled NLDAS-2 scenario and the Gridmet, Mean of Inputs, PRISM, and Daymet is displayed on the x axis from left to right, respectively, and the corresponding percent change in hydrologic fluxes relative to the downscaled NLDAS-2 scenario is displayed on the y axis.

#### 4. Conclusions

This study explores the sensitivity of water budget partitioning to variable precipitation forcings as simulated by an integrated surface water-groundwater model. We present a hydrologic model of the Kaweah River watershed conceptualized to represent a dynamic steady-state condition without human modification. The model captures important hydrologic processes in the study region such as seasonal variations in snowpack, streamflow duration curves, and spatially distributed *ET* and *SWE*. By applying the 3CH method, we quantify spatial patterns of the uncertainty in precipitation data sets and find that uncertainty reaches 6 mm/day in some mountainous locations. We also apply the 3CH method to the simulated hydrology. Results of the 3CH method applied to the watershed average values indicate that high uncertainty in precipitation forcing results in high uncertainty in soil moisture and *ET*, but not *SWE* or streamflow (Figure 7). The contrasting responses of several simulated variables highlight the difficulties in selecting a single best precipitation forcing data set.

By using an integrated hydrologic model, we quantify the propagation of uncertainty in precipitation forcing to uncertainty in the simulated groundwater flow and storage. Soil moisture is a key hydrologic variable that controls both *ET* and potential recharge. We find that the uncertainty in simulated root zone soil moisture is not correlated to the precipitation uncertainty. Rather, it is correlated to topographic properties including the slope and elevation. An analysis of the soil water flux demonstrates that topographically enhanced lateral flows in the soil layer are responsible for redistributing the uncertainty in precipitation forcing downslope.

As a result, the uncertainty in both potential recharge and groundwater storage changes is decoupled from the uncertainty in the precipitation forcing (Figure 9). Indeed, although majority of the potential groundwater recharge occurs at the highest elevations in response to snowmelt, the largest uncertainty in potential recharge is located downslope in response to topography.

We find that modifying the critical zone parameterization does not alter the role of topography in propagating uncertainty to groundwater. This is demonstrated with consistent results between Experiment 1 and Experiment 2 simulations where the thickness of saprolite layer is changed. Assessing the patterns of uncertainty in hydrologic variables across the elevation gradient shows that the uncertainty in precipitation is highest for the 2,800- to 2,900-m elevation bin, while soil moisture, *ET*, and potential recharge have the highest uncertainties in the 2,000- to 2,100-, 1,700- to 1,800-, and 1,200- to 1,300-m elevation bins, respectively. This finding remains true for all realizations of the subsurface tested because soil water redistribution in the soil layer (top 2 m of model domain) is the primary mechanism redistributing the uncertainty in soil water balance processes downslope.

When precipitation uncertainty is prevalent in available forcing data sets, a common approach is to consider merging forcing data sets or merging simulated outputs using an ensemble approach. We assess which approach works best for an integrated hydrologic model in the study domain. By comparing the simulated hydrology to available observational data sets, we quantify the performance of different modeling approaches. We find that generally, merging either the input precipitation data sets or the simulated hydrology from multiple simulations with different precipitation forcing results in improved overall model performance. In the study domain, an unexpected pattern emerges where the simulation with merged input forcings shows better model performance than the mean of simulated hydrologic fluxes from multiple scenarios. We attribute this to the high sensitivity of hydrologic fluxes and states to changes in precipitation. A 1% change in mean annual precipitation in the study domain results in a greater than 1% response in streamflow, subsurface storage changes, and in some cases, groundwater recharge from the Kaweah Terminus watershed to the Central Valley aquifer. Furthermore, the difference in performance statistics is related to the model independence in each ensemble (i.e., forcings or outputs). As a result of

independence in the model forcings and using a consistent model structure, merging the multiple precipitation data sets provides the best model performance with shorter simulation time. Therefore, averaging of multiple precipitation data sets should be considered when developing a hydrologic model when these data sets are available.

## Data Availability Statement

All data sets used and model code are publically available. Model outputs will be deposited at UC Riverside Dryad Digital Repository (<https://doi.org/10.6086/D14T2B>).

## Acknowledgments

We would like to acknowledge funding from the California Energy Commission grant “Advanced Statistical-Dynamical Downscaling Methods and Products for California Electrical System” (Award EPC-16-063). We thank Brian Henn and two anonymous reviewers for providing constructive comments to further improve the manuscript.

## References

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 1–12. <https://doi.org/10.1038/sdata.2017.191>
- Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., et al. (2018). Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth System Dynamics Discussions*, 6, 1–20. <https://doi.org/10.5194/esd-2018-51>
- Ajami, H., McCabe, M. F., & Evans, J. P. (2015). Impacts of model initialization on an integrated surface water-groundwater model. *Hydrological Processes*, 29(17), 3790–3801. <https://doi.org/10.1002/hyp.10478>
- Ajami, H., McCabe, M. F., Evans, J. P., & Stisen, S. (2014). Assessing the impact of model spin-up on surface water-groundwater interactions using an integrated hydrologic model. *Water Resources Research*, 50, 2636–2656. <https://doi.org/10.1002/2013WR014258>
- Anthes, R., & Rieckh, T. (2018). Estimating observation and model error variances using multiple data sets. *Atmospheric Measurement Techniques*, 11(7), 4239–4260. <https://doi.org/10.5194/amt-11-4239-2018>
- Banks, E. W., Simmons, C. T., Love, A. J., Cranswick, R., Werner, A. D., Bestland, E. A., et al. (2009). Fractured bedrock and saprolite hydrogeologic controls on groundwater/surface-water interaction: A conceptual model (Australia). *Hydrogeology Journal*, 17(8), 1969–1989. <https://doi.org/10.1007/s10040-009-0490-7>
- Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I. J. M., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23(1), 207–224. <https://doi.org/10.5194/hess-23-207-2019>
- Bhaskar, A. S., Welty, C., Maxwell, R. M., & Miller, A. J. (2015). Untangling the effects of urban development on subsurface storage in Baltimore. *Water Resources Research*, 51, 1158–1181. <https://doi.org/10.1002/2014WR016039>
- Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, 274, 54–67. <https://doi.org/10.1016/j.geoderma.2016.03.025>
- Chow, V. T. (2009). *Open-channel hydraulics*. Caldwell, NJ: The Blackburn Press.
- Clark, M. P., & Slater, A. G. (2006). Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, 7(1), 3–22. <https://doi.org/10.1175/JHM474.1>
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123, 9391–9409. <https://doi.org/10.1029/2017JD028200>
- Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The common land model. *Bulletin of the American Meteorological Society*, 84(8), 1013–1023. <https://doi.org/10.1175/BAMS-84-8-1013>
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15), 2031–2064. <https://doi.org/10.1002/joc.1688>
- Daly, C., Neilson, R., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33(2), 140–158. [https://doi.org/10.1175/1520-0450\(1994\)033<0140:astmf>2.0.co;2](https://doi.org/10.1175/1520-0450(1994)033<0140:astmf>2.0.co;2)
- Davis, G., Green, J. H., Olmsted, F. H., & Brown, D. W. (1959). *Ground-Water Conditions and Storage Capacity in the San Joaquin Valley California, Geological Survey Water-Supply Paper* (Vol. 1469). Washington, DC: United States Government Printing Office.
- Deb, P., Kiem, A. S., & Willgoose, G. (2019). A linked surface water-groundwater modelling approach to more realistically simulate rainfall-runoff non-stationarity in semi-arid regions. *Journal of Hydrology*, 575, 273–291. <https://doi.org/10.1016/J.JHYDROL.2019.05.039>
- Falcone, J. A. (2011). *GAGES-II: Geospatial attributes of gages for evaluating streamflow*. Reston, VA: US Geological Survey. <https://doi.org/10.3133/70046617>
- Faunt, C. C. (Ed.) (2009). *Groundwater availability of the central valley aquifer, California, U.S. Geological Survey Professional Paper 1766* (p. 255). Reston, Virginia: United States Geological Survey. Retrieved from <http://pubs.usgs.gov/pp/1766/>
- Forman, B. A., Vivoni, E. R., & Margulis, S. A. (2008). Evaluation of ensemble-based distributed hydrologic model response with disaggregated precipitation products. *Water Resources Research*, 44, W12409. <https://doi.org/10.1029/2008WR006827>
- Foster, L. M., & Maxwell, R. M. (2019). Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions. *Hydrological Processes*, 33(3), 332–349. <https://doi.org/10.1002/hyp.13327>
- Fu, S., Sonnenborg, T. O., Jensen, K. H., & He, X. (2011). Impact of precipitation spatial resolution on the hydrological response of an integrated distributed water resources model. *Vadose Zone Journal*, 10(1), 25–36. <https://doi.org/10.2136/vzj2009.0186>
- Fulton, R. A., Breidenbach, J. P., Seo, D. J., Miller, D. A., & O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and Forecasting*, 13(2), 377–395. [https://doi.org/10.1175/1520-0434\(1998\)013<0377:TWRA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2)
- Gesch, D. B., Evans, G. A., Oimoen, M. J., & Arundel, S. (2018). *The national elevation dataset* (pp. 83–110). American society for photogrammetry and remote sensing. Retrieved from <http://pubs.er.usgs.gov/publication/70201572>

- Gilleland, E., & Roux, G. (2015). A new approach to testing forecast predictive accuracy. *Meteorological Applications*, 22(3), 534–543. <https://doi.org/10.1002/met.1485>
- Graham, R. C., Anderson, M. A., Sternberg, P. D., Tice, K. R., & Schoeneberger, P. J. (1997). Morphology, porosity, and hydraulic conductivity of weathered granitic bedrock and overlying soils. *Soil Science Society of America Journal*, 61, 516–522. <https://doi.org/10.2136/sssaj1997.03615995006100020021x>
- Gray, J. E., & Allan, D. W. (1974). A method for estimating the frequency stability of an individual oscillator. In *28th Annual Symposium on Frequency Control* (pp. 243–246). Atlantic City, NJ: IEEE. <https://doi.org/10.1109/FREQ.1974.200027>
- Groisman, P., & Legates, D. (1994). The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society*, 75(2), 215–227. [https://doi.org/10.1175/1520-0477\(1994\)0752.0.CO;2](https://doi.org/10.1175/1520-0477(1994)0752.0.CO;2)
- Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2018). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, 556, 1205–1219. <https://doi.org/10.1016/j.jhydrol.2017.03.008>
- Higgins, R. W., Shi, W., Yarosh, E., & Joyce, R. (2000). Improved United States precipitation quality control system and analysis. NCEP/Climate prediction center ATLAS No. 7.
- Hong, Y., Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resources Research*, 42, W08421. <https://doi.org/10.1029/2005WR004398>
- Jennings, C. W. (1977). *Geologic map of California*. Sacramento: California Division of Mines and Geology.
- Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., & Ma, M. (2012). Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *Journal of Hydrology*, 452–453, 213–225. <https://doi.org/10.1016/j.jhydrol.2012.05.055>
- Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1 Theory. *Water Resources Research*, 42, W03407. <https://doi.org/10.1029/2005WR004368>
- Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945–958. <https://doi.org/10.1016/j.advwatres.2005.08.006>
- Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, 44, W02402. <https://doi.org/10.1029/2007WR006004>
- Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., & Simmer, C. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, 46, W04201. <https://doi.org/10.1029/2009WR008730>
- Koren, V. I., Finnerty, B. D., Schaake, J. C., Smith, M. B., Seo, D. J., & Duan, Q. Y. (1999). Scale dependencies of hydrologic models to spatial variability of precipitation. *Journal of Hydrology*, 217(3–4), 285–302. [https://doi.org/10.1016/S0022-1694\(98\)00231-5](https://doi.org/10.1016/S0022-1694(98)00231-5)
- Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1–4), 2–9. [https://doi.org/10.1016/S0022-1694\(01\)00420-6](https://doi.org/10.1016/S0022-1694(01)00420-6)
- Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51, 7309–7342. <https://doi.org/10.1002/2015WR017616>
- Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., et al. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, 26(23), 9384–9392. <https://doi.org/10.1175/JCLI-D-12-00508.1>
- Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, 18(2), 575–594. <https://doi.org/10.5194/hess-18-575-2014>
- Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, 50, 1131–1151. <https://doi.org/10.1002/2013WR014581>
- Lundquist, J. D., Cayan, D. R., & Dettinger, M. D. (2003). In F. Zhao & L. Guibas (Eds.), *Meteorology and hydrology in Yosemite national park: A sensor network application BT-information processing in sensor networks* (pp. 518–528). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lundquist, J. D., Hughes, M., Henn, B., Gutmann, E. D., Livneh, B., Dozier, J., & Neiman, P. (2015). High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada California. *Journal of Hydrometeorology*, 16(4), 1773–1792. <https://doi.org/10.1175/JHM-D-15-0019.1>
- Margulis, S. A., Cortés, G., Giroto, M., & Durand, M. (2016). A landsat-era Sierra Nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, 17(4), 1203–1221. <https://doi.org/10.1175/JHM-D-15-0177.1>
- Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale integrated hydrologic modeling. *Advances in Water Resources*, 53, 109–117. <https://doi.org/10.1016/j.advwatres.2012.10.001>
- Maxwell, R. M., & Miller, N. L. (2005). On the development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*, 6, 233–247. <https://doi.org/10.1175/JHM422.1>
- McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078–4111. <https://doi.org/10.1002/hyp.9384>
- Milly, P. C. D., & Eagleson, P. S. (1988). Effect of storm scale on surface runoff volume. *Water Resources Research*, 24(4), 620–624. <https://doi.org/10.1029/WR024i004p00620>
- Mizukami, N., Clark, M. P., Gutmann, E. D., Mendoza, P. A., Newman, A. J., Nijssen, B., et al. (2016). Implications of the methodological choices for hydrologic portrayals of climate change over the contiguous United States: Statistically downscaled forcing data and hydrologic models. *Journal of Hydrometeorology*, 17(1), 73–98. <https://doi.org/10.1175/JHM-D-14-0187.1>
- Mizukami, N., Clark, M. P., Slater, A. G., Brekke, L. D., Elsnor, M. M., Arnold, J. R., & Gangopadhyay, S. (2014). Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *Journal of Hydrometeorology*, 15(1), 474–488. <https://doi.org/10.1175/JHM-D-13-036.1>
- Mohr, C. H., Montgomery, D. R., Huber, A., Bronstert, A., & Iroumé, A. (2012). Streamflow response in small upland catchments in the Chilean coastal range to the MW 8.8 Maule earthquake on 27 February 2010. *Journal of Geophysical Research*, 117, F02032. <https://doi.org/10.1029/2011JF002138>
- Moreno, H. A., Vivoni, E. R., & Gochis, D. J. (2012). Utility of quantitative precipitation estimates for high resolution hydrologic forecasts in mountain watersheds of the Colorado front range. *Journal of Hydrology*, 438–439, 66–83. <https://doi.org/10.1016/j.jhydrol.2012.03.019>
- Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8), 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>



- Nicotina, L., Alessi Celegon, E., Rinaldo, A., & Marani, M. (2008). On the impact of rainfall patterns on the hydrologic response. *Water Resources Research*, 44, W12401. <https://doi.org/10.1029/2007WR006654>
- Olmsted, F. H., & Davis, G. H. (1961). *Geologic features and ground-water storage capacity of the Sacramento Valley, California, U.S. Geological Survey Water-Supply Paper* (Vol. 1497, p. 241). Washington, DC: United States Government Printing Office.
- Pan, M., Cai, X., Chaney, N. W., Entekhabi, D., & Wood, E. F. (2016). An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophysical Research Letters*, 43, 9662–9668. <https://doi.org/10.1002/2016GL069964>
- Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., et al. (2018). Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resources Research*, 54, 3863–3875. <https://doi.org/10.1029/2017WR022421>
- Premoli, A., & Tavella, P. (1993). A revisited three-cornered hat method for estimating frequency standard instability. *IEEE Transactions on Instrumentation and Measurement*, 42(1), 7–13. <https://doi.org/10.1109/19.206671>
- Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., et al. (2012). How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society*, 93(6), 811–829. <https://doi.org/10.1175/BAMS-D-11-00052.1>
- Scanlon, B. R., Keese, K. E., Flint, A. L., Flint, L. E., Gaye, C. B., Edmunds, W. M., & Simmers, I. (2006). Global synthesis of groundwater recharge in semiarid and arid regions. *Hydrological Processes*, 20, 3335–3370. <https://doi.org/10.1002/hyp>
- Schreiner-McGraw, A. P., & Vivoni, E. R. (2017). Percolation observations in an arid piedmont watershed and linkages to historical conditions in the Chihuahuan Desert. *Ecosphere*, 8, e02000. <https://doi.org/10.1002/ecs2.2000>
- Tang, Y., Marshall, L., Sharma, A., & Ajami, H. (2019). Modelling precipitation uncertainties in a multi-objective Bayesian ecohydrological setting. *Advances in Water Resources*, 123(September 2017), 12–22. <https://doi.org/10.1016/j.advwatres.2018.10.015>
- Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, 190(3–4), 214–251. [https://doi.org/10.1016/S0022-1694\(96\)03128-9](https://doi.org/10.1016/S0022-1694(96)03128-9)
- USDA-NASS (2017). USDA national agricultural statistics service cropland data layer. USDA-NASS. Washington, DC. Retrieved from <https://nassgeodata.gmu.edu/CropScape/>
- Viviroli, D., & Weingartner, R. (2004). The hydrological significance of mountains: From regional to global scale. *Hydrology and Earth System Sciences*, 8(6), 1017–1030. <https://doi.org/10.5194/hess-8-1017-2004>
- Welch, L. A., & Allen, D. M. (2014). Hydraulic conductivity characteristics in mountains and implications for conceptualizing bedrock groundwater flow. *Hydrogeology Journal*, 22(5), 1003–1026. <https://doi.org/10.1007/s10040-014-1121-5>
- Westrick, K. J., Mass, C. F., & Colle, B. A. (1999). The limitations of the WSR-88D radar network for quantitative precipitation measurement over the coastal western United States. *Bulletin of the American Meteorological Society*, 80(11), 2289–2298. [https://doi.org/10.1175/1520-0477\(1999\)080<2289:TLOTWR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2289:TLOTWR>2.0.CO;2)
- Winchell, M., Gupta, H. V., & Sorooshian, S. (1998). On the simulation of infiltration- and saturation-excess runoff using radar-based rainfall estimates: Effects of algorithm uncertainty and pixel aggregation. *Water Resources Research*, 34(10), 2655–2670. <https://doi.org/10.1029/98WR02009>
- Zhu, Q., Gao, X., Xu, Y. P., & Tian, Y. (2019). Merging multi-source precipitation products or merging their simulated hydrological flows to improve streamflow simulation. *Hydrological Sciences Journal*, 64(8), 910–920. <https://doi.org/10.1080/02626667.2019.1612522>