Improving Runoff Simulation in the Western United States with Noah-MP and VIC

Lu Su[a,b], Dennis P. Lettenmaier[b], Ming Pan[a], Benjamin Bass[c]

a Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, United States

b Department of Geography, University of California, Los Angeles, United States

c Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, United States

Correspondence: Dennis P. Lettenmaier (dlettenm@ucla.edu)

Abstract

Streamflow ~~forecasts~~ predictions are critical for managing water resources and for environmental conservation, especially in the water-short Western U.S. Land Surface Models (LSMs), such as the Variable Infiltration Capacity (VIC) model and the Noah-Multiparameterization (Noah-MP) play an essential role in providing comprehensive runoff ~~forecasts~~ predictions across the region. Virtually all LSMs require parameter estimation (calibration) to optimize their predictive capabilities. Here, we focus on the calibration of VIC and Noah-MP models at a 1/16° latitude-longitude resolution across the Western U.S. We first performed global optimal calibration of parameters for both models for 263 river basins in the region. We find that the calibration significantly improves~~d~~ the models' performance, with the median daily streamflow Kling-Gupta Efficiency (KGE) increasing from 0.37 to 0.70 for VIC, and from 0.22 to 0.54 for Noah-MP. In general, post-calibration model performance is for watersheds with relatively high precipitation and runoff ratios, and at lower elevations. At a second stage, we regionalized the river basin calibrations using the donor-basin method, which establishes~~d~~ transfer relationships for

Style Definition: Heading 1: Indent: Left: 0.63 cm, Space Before: 0 pt, After: 0 pt, No bullets or numbering, No widow/orphan control

Style Definition: Heading 2

Style Definition: Title

Formatted: French (France)

hydrologically similar basins, via which we extended our calibration parameters to 4,816 HUC-10 basins across the region. Using the regionalized parameters, we show that the models' capabilities to simulate high and low flow conditions were substantially improved following calibration and regionalization. The refined parameter sets we developed are intended to support regional hydrological studies and hydrological assessments of climate change impacts.

1. Introduction

Streamflow ~~forecasts~~ predictions play a key role in ~~various aspects of~~ water and environmental management, especially in the water-stressed Western U.S. (WUS). In the short term, these ~~forecasts~~ predictions provide early warnings for impending flood events, thereby enabling timely preparation and response to mitigate immediate flood risk and damages (Maidment, 2017). They also serve as crucial input for managing reservoirs effectively for water supply (Raff et al., 2013), hydroelectric power generation (Boucher & Ramos, 2018), and river navigation (by providing a basis for predicting water levels) (Federal Institute of Hydrology, 2020). In the longer term, streamflow ~~forecasts~~ predictions enable water utilities and agencies to plan water distribution within and across multiple uses—urban, agricultural, and industrial—which is especially vital during drought conditions when efficient water use becomes a necessity (Anghileri et al., 2016;). Streamflow ~~forecasts~~ predictions also aid in understanding and ~~predicting~~ foreseeing the impacts of climate change on water systems, thereby informing adaptive strategies for water resource management. Thus, in both short and longer-term contexts, streamflow ~~forecasts~~ predictions are an important tool for promoting sustainable water practices and resilience to water-related challenges.

Streamflow ~~forecasts~~ predictions are derived via a synthesis of hydrometeorological data, statistical methodologies, and computational modeling. Direct measurement of runoff is an important element of ~~streamflow forecasts~~this process, however it is only possible in river basins with well-developed observational infrastructure (Sharma and Machiwal, 2021). This limitation leaves vast areas, often critical to water resource management and climatology, without direct runoff observations on which to base streamflow ~~forecasts~~predictions. As an alternative, Land Surface Models (LSMs) can be used to simulate streamflow. LSMs typically are forced with air temperature, precipitation and other surface meteorological ~~forcings~~variables. By integrating climatic, topographic, and land-use information, they can fill streamflow observation gaps and provide comprehensive, spatially distributed runoff ~~forecasts~~ predictions (Fisher and Koven, 2020). The capabilities of LSMs equip us with the necessary tools to produce streamflow ~~forecasts~~ predictions that can be used to prepare for severe weather conditions, form the basis for water resource management, and inform water management associated with our evolving climate. These benefits hold true irrespective of the limitations associated with direct streamflow observations. Through off-line simulations and reconstructions, LSMs enable us to gain insights into land surface hydrology at various scales - regional, continental, and global.

~~The p~~Parameterizations of the underlying hydrological processes ~~varies~~ vary across different LSMs, but virtually all models require some level of parameter estimation based on historical observed streamflow data at forecast point, to ensure trustworthy predictions throughout the region (Beven,1989; Troy et al., 2008; Gong et al., 2015). In cases where observations don't exist, parameters can be transferred from river basins where they do (Arsenault and Brissette, 2014). In cases where observations do exist but aren't current, shorter records of historical streamflow data can be used for model calibration and subsequently streamflow ~~forecasts~~ predictions can be produced using meteorological forcings for more recent periods when streamflow data aren’t available.

Implementation of hydrological models for the above purposes usually involves calibration of model parameters using streamflow observations, which are more readily available than other model prognostic variables like soil moisture or evapotranspiration (Demaria et al., 2007; Gao et al., 2018; Troy et al., 2008; Yadav et al., 2007). Calibration has always been a critical and evolving component of hydrologic model application, and has been improved by advances in model parameterization, enhanced spatial resolution providing more detailed and accurate spatial information, improved soil/vegetation data, meteorological inputs, and training data. Furthermore, advances in calibration methods and computing power have facilitated regional approaches to model calibration, and inclusion of multiple hydrologic models. Previous~~,~~ studies ~~often~~mostly focus~~ed~~ on a single hydrologic model due to computational constraints (e.g., Mascaro et al. (2023), Sofokleous et al. (2023), and Gou et al. (2020)). However, we incorporate two models to address structural model uncertainty and to ensure broader applicability of the calibration methods we employ.

The Variable Infiltration Capacity (VIC, Liang et al. (1994)) model and Noah-Multiparameterization (Noah-MP, Niu et al. (2011)), which we use here, are widely used hydrologic models both in the U.S. and globally, as highlighted by Mendoza et al. (2015) and Tangdamrongsub (2023). Many previous implementations of VIC for the Western United States (WUS) have been based on the Livneh et al. (2013) data set, and its predecessor, Maurer et al. (2002), which performed initial calibrations across the region. In the case of Noah-MP, Bass et al. (2023) performed manual calibration across the region ~~efforts~~. Neither of these implementations, however, ~~sets~~ employs globally optimized calibration, as we do here.

Formatted: Font: Not Bold, No underline, Font color: Auto

The process of calibration can be computationally demanding, and prior research typically has focused on obtaining parameters appropriate to facilitating model simulations that match observations as closely as possible at ~~the observation point~~stream gauge locations (Duan et al,1992; Tolson and Shoemaker, 2007). Most previous studies have concentrated on a limited number of gauges/river basins ~~basins~~ and a single model (e.g. Mascaro et al.~~,~~ (2023); Sofokleous et al.~~,~~ (2023); and Gou et al.~~,~~ (2020)). Here, we aim to establish parameterizations for two LSMs -- ~~the Variable Infiltration Capacity (~~VIC~~)~~ ~~model~~ and ~~the Noah Multiparameterization (~~Noah-MP~~)~~ ~~LSM~~ across the entire WUS. ~~Both models have found extensive application both within the U.S. and internationally (Mendoza et al.,2015; Tangdamrongsub, 2023).~~ In doing so, we apply global optimization~~ed~~ ~~calibration~~ methods at the river basin level, followed by a second stage ~~and~~ regionalization~~, with the objective of facilitating these models to provide reliable runoff simulations~~.

~~Building on the context outlined earlier, o~~The work we report here ~~ur study~~ aims to develop ~~high-resolution, optimally~~ calibrat~~i~~ion~~ed~~ parameters for the VIC and Noah-MP models that can be implemented at the catchment (Hydrologic Unit Code or HUC) 10 level across the region~~in the WUS.~~ ~~In particular, w~~We explore and elucidate (i) the choice of physical parameterizations and calibration of land surface parameters, (ii) extension of these calibrated parameters to areas without gauges, and (iii) factors that influence calibration efficiency and LSM performance using regional parameter estimates. Following this introduction, Section 2 describes our calibration basins, the hydrologic models used, and the forcing dataset. The framework of our procedures is illustrated in Figure 1. Section 3 provides an in-depth exploration of the calibration process. In the case of Noah-MP, which offers multiple runoff generation (physics) options, our initial step involves choosing the most effective runoff parameterization option. Following this, we perform the calibration of land surface parameters. In the case of the VIC model, the runoff parameterization scheme is predetermined, so we commence immediately with calibration at 263 river basins across our region. Our second stage regionalization (section 4) extends the calibrated parameters to ungauged basins using the technique known as the donor basin method, as implemented by Bass et al. (2023). In Section 5, we evaluate both flood and low flow simulation skills both pre- and post-calibration, and following regionalization. Finally, following discussion and interpretation (section 6) section 7 presents conclusions, encapsulating the insights and implications of our study.

Formatted: Justified

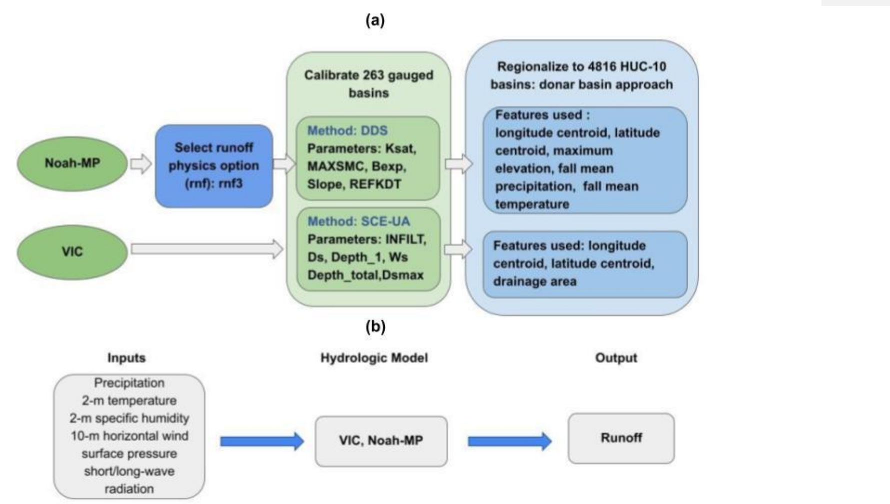


Figure 1 (a) framework of the calibration and regionalization processes adopted in this study. (b) model simulation inputs and output.

2. Study basins, land surface models and forcing dataset overview

2.1 Study Basins

We selected 263 river basins distributed across the WUS for calibration of the two models. Most of the basins were from USGS Gages II reference basins (Falcone 2011) which have minimum upstream anthropogenic effects such as dams and diversions. Among these basins, our selection criteria included having at least 20 years of record, and a minimum drainage area of 144 square kilometers, which is the size of four model grid cells. In addition to 250 Gages II reference stations, we included 13 basins located in California's Sierra Nevada for which naturalized flows (effects of upstream reservoir storage and/or diversions removed) are available from the California Department of Water Resources (2021). The locations of the 263 basins are shown in Figure 2~~1~~. ~~We focused on the hydrological models' calibration to full natural flow (the same as observed streamflow for GAGES II stations; estimated by DWR for the 13 Sierra Nevada sites), which indicates water flow conditions devoid of human interventions like reservoirs or diversions. Each basin was calibrated using~~ We used the most recent 20-year period of streamflow observations for calibration in each of the 263 basins~~when the observation is available~~.

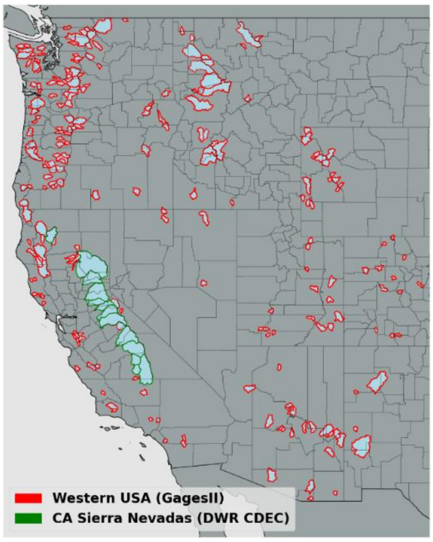


Figure. 2~~1~~. 263 river basins for which calibration was performed. The Gages II reference basins are delineated with red boundaries and the CA Sierra Nevada~~e~~ basins with green boundaries.

2.2 Land Surface Models

Formatted: Heading 2, Left, Indent: First line: 0 cm

The two models we used (VIC and Noah-MP) were chosen due to their broad application and proven effectiveness in hydrological simulations. The VIC model is renowned globally for its success in runoff simulation, as evidenced by studies such as Adam et al. (2003 & 2006), Livneh et al. (2013), and Schaperow et al. (2021). Conversely, Noah-MP, though relatively newer, forms the hydrologic core of the U.S. National Water Model (NWM) and is increasingly used both within the U.S. and abroad.

Our selection is further reinforced by a study conducted by Cai et al. (2014), which assessed the hydrologic performance of four LSMs in the United States using the North American Land Data Assimilation System (NLDAS) test bed. This study highlighted Noah-MP's proficiency in soil moisture simulation and its strong performance in Total Water Storage (TWS) simulations, while recognizing VIC's capabilities in streamflow simulations.

Our choice of models also was informed by the varying levels of complexity these two models offer in conceptualizing the effects of vegetation, soil, and seasonal snowpack on the land surface energy and water balances (refer to Table 1 for more details). VIC and Noah-MP employ different parameterizations for various hydrological processes, such as canopy water storage, base flow, and runoff. Noah-MP features four runoff physics options (see Table 1). It utilizes four soil layers, each with a fixed depth. In contrast, the VIC model, with its variable infiltration capacity approach (Liang et al., 1994), uses up to three soil layers per grid cell with variable depths, providing flexibility in modeling soil moisture dynamics. The unique runoff generation methodologies of each model are particularly pertinent for capturing the diverse hydrological characteristics of the WUS.

The calibrated parameters we develop here for both models will provide future researchers with essential tools for comprehensive hydrological analysis across the WUS. Utilizing these two distinct models, each with unique strengths and methods, will facilitate thorough exploration of the WUS's varied hydrological characteristics, and response of the watersheds in the region to climate change, as well as implementation of improved streamflow forecast methods. Our results will help to facilitate~~s~~ a deeper understanding of hydrological processes and spatial variability across the entire WUS region.

In our implementation of both models, we accumulated runoff over each of the calibration watersheds. We chose not to implement the channel routing schemes of either model since their~~its~~ impact on daily streamflow simulations ~~was~~ is small given the relatively small size of most of the basins. This aligns with earlier research (e.g., Li et al. 2019). However, in both the case of VIC and Noah-MP, the output of our simulations (runoff) could be used as input to routing models, such as those that are options in the implementation of both models. We describe below the particulars of the two models.

2.2.1 VIC

VIC is a macroscale, semi-distributed hydrologic model (described in detail by Liang et al 1994) that determines land surface moisture and energy states and fluxes by solving the surface water and energy balances. VIC is a research model and in its various forms it has been employed to study many major river basins worldwide (e.g. Adam et al 2003 & 2006; Livneh et al 2013; Schaperow et al 2021). This model enjoys a broad user community — as per the citation index Web of Science, the initial VIC paper has been referenced more than 2600 times, with contributing authors spanning at

Formatted: Justified, First line: 2 ch least 56 different countries (Schaperow et al 2021). We obtained initial VIC model parameters from Livneh et al 2013, who validated model discharges over major CONUS river basins. The origins of the soil and land cover data are outlined in Table 1. The version of the VIC model implemented here is 4.1.2, and it operates in energy balance mode. We selected VIC 4.1.2 for two key reasons: First, our initial parameters were based on Livneh et al. (2013), who validated model discharges over major CONUS river basins using this model version. Second, in a preliminary assessment of snow water equivalent (SWE) simulation skills at select SNOTEL sites across the WUS, we found that VIC 4.1.2 demonstrated superior performance compared to VIC 5 (see Figure S1). This finding, coupled with our research group's extensive experience and proven results with VIC 4.1.2, informed our decision to use this version.

Commented [DL1]: You need to show some of those results in the supplement.

2.2.2 Noah-MP

Noah-MP was originally designed as the land surface scheme for numerical weather prediction (NWP) models like the Weather Research and Forecasting (WRF) regional atmospheric model. Currently, it's being utilized for physically based, spatially-distributed hydrological simulations as a component of the National Water Model (NWM) (NOAA, 2016). It enhances the functionalities of the Noah LSM (as per Chen et al., 1996 and Chen and Dudhia, 2001) previously used in NOAA's suite of numerical weather prediction models by offering multiple options for key processes that control land-atmosphere transfers of moisture and energy. These include surface water infiltration, runoff, evapotranspiration, groundwater movement, and channel routing (see Niu et al., 2007; 2011). The model has been widely used for forecasting seasonal climate, weather, droughts, and floods not only across the continental United States (CONUS) but also globally (Zheng et al., 2019). We utilized the most current version (WRF-HYDRO 5.2.0)

Formatted: Justified

2.3 Forcing Dataset

We ran both models at a 3-hour time step and at 1/16° latitude–longitude spatial resolution. The forcings were the gridded observation dataset developed by Livneh et al (2013) and extended to 2018 by Su et al (2021) (hereafter referred to as L13). This data set spans the period from 1915 to 2018. For the VIC model, the L13 dataset provided daily values of precipitation, maximum and minimum temperatures, and wind speed (additional variables used by VIC including downward solar and longwave radiation, and specific humidity, are computed internally using MTCLIM algorithms as described by Bohn et al. (2013)). The Noah-MP model, on the other hand, necessitated additional meteorological data such as specific humidity, surface pressure, and downward solar and longwave radiation, in addition to precipitation, wind speed, and air temperature. We used the MTCLIM algorithms, as detailed by Bohn et al. (2013), to calculate specific humidity and downward solar radiation. We employed the Prata (1996) algorithm to compute the downward longwave radiation. Additionally, we deduced surface air pressure by considering the grid cell elevation in conjunction with standard global pressure lapse rates. Following this, we transitioned the daily data to hourly metrics using a cubic spline to interpolate between Tmax and Tmin, and derived other variables using the methods explained by Bohn et al. (2013). Lastly, we distributed the daily precipitation evenly across three hourly intervals.

We used a 3-hour simulation timestep given numerical considerations with Noah-MP (which don't affect VIC, however for consistency we used a 3-hour timestep for VIC as well. Despite the fact that precipitation in particular was available daily (and hence apportioned equally to 3-hour timesteps) resolving the diurnal cycle is sometimes important in the case of snow (accumulation and ablation) processes which vary diurnally.

Table 1. Overview of hydrologic model components and parameter data sources.

MODEL	SNOW ACCUMULATION AND MELT	MOISTURE IN THE SOIL AND COLUMN/SURFACE RUNOFF	BASE FLOW	CANOPY STORAGE	VEGETATION DATA	SOIL D
VIC (V4.1.2)	Two-layer energy–mass balance model	Infiltration capacity function. Vertical movement of moisture through soil follows 1D Richards equation.	A function of the soil moisture in the third layer. Linear below a soil moisture threshold and becomes nonlinear above that threshold. [Liang et al., 1994]	Mosaic representation of different vegetation coverages at each cell.	University of Maryland 1-km Global Land Cover Classification (Hansen et al. 2000)	1-km STATSGO database (Miller and White 1998).
NOAH-MP (WRF-HYDRO 5.2.0)	Three-layer energy–mass balance model that represents percolation, retention, and refreezing of meltwater within the snowpack.	(1) TOPMODEL-based runoff scheme (2) Simple TOPMODEL-based runoff scheme with an equilibrium water table (hereafter SIMTOP) (3) Infiltration-excess-based surface runoff scheme (4) BATS runoff scheme, which parameterized surface runoff as a 4th power function	Simple groundwater (hereafter SIMGM) [Niu et al., 2007]. Similar to SIMGM, but with a sealed bottom of the soil column [Niu et al., 2005] Gravitational free-drainage subsurface runoff scheme [Schaake et al., 1996] Gravitational free drainage [Dickinson et al.,1993]	Semi-tile approach for computing longwave, latent heat, sensible heat and ground heat fluxes	MODIS 30-second Modified IGBP 20-category land cover product	1-km STATSGO database (Miller and White 1998).

Formatted Table

Formatted: French (France)

		of the top 2 m soil wetness (degree of saturation)

3. Model calibration

3.1 Calibration methods

The initial step in our calibration effort was to optimize the land surface parameters of the two models for the 263 WUS basins. These parameters, primarily soil properties which can exhibit a substantial degree of uncertainty, were iteratively updated via hundreds of simulations to accurately reflect streamflow conditions in each basin.

Our focus on calibrating soil-related parameters was based on their critical role in runoff generation. In this respect, we focused on key processes including infiltration, soil moisture storage, and groundwater recharge. The calibration of parameters that control these processes was prioritized to improve the representation of soil-water interactions, a major driver of runoff variability in the region. Given the importance of snow processes across much of the region, we conducted snow simulation verification at 20 Snow Telemetry (SNOTEL) (Natural Resources Conservation Service, ~~,~~2023) sites across WUS. –Our assessment (see –Figure S1) indicated that the existing parameterizations for snow processes in both models reproduced observed SWE well across our study region.

-Prior to calibration, we conducted a sensitivity analysis to identify the most influential parameters for streamflow simulation in both models. We also drew ~~Drawing~~ on insights from previous research in this respect (Mendoza et al. 2015; Hussein 2020; Shi et al. 2008; Holtzman et al., 2020; Bass et al., 2023; Schaperow et al., 2023)~~.~~. We then performed a sensitivity analysis, focusing on how variations in the most sensitive

~~se~~ parameters impacted Kling-Gupta Efficiency (KGE; Gupta et al., 2009)~~KGE outcomes~~. Based on these analyses, we chose to calibrate six parameters for the VIC model and five for the Noah-MP model (Table 2). For each parameter, we defined a physically viable range (refer to Table 2), drawing from values utilized in prior studies (Cai et al. 2014; Mendoza et al. 2015; Hussein 2020; Shi et al. 2008; Gochis et al., 2019; Holtzman et al., 2020; Lahmers et al. 2021; Bass et al., 2023; Schaperow et al., 2023). ~~Through our iterative calibration method, each subsequent simulation learns from the previous ones using algorithms designed to reduce the discrepancy between the simulated and observed streamflow.~~

In recent years, the development of hydrologic model calibration has evolved from manual, trial-and-error approaches to advanced automated techniques. This has included a shift towards global optimization methods, notably the Shuffled Complex Evolution algorithm (SCE-UA; Duan et al., 1992). Typically, SCE-UA has been applied to computationally efficient models (simulation time often on the order of a few minutes or less; see e.g., Franchini et al., (1998)). However, its application becomes less practical with more recent distributed hydrologic models such as the Noah-MP which require longer simulation times. To address these computational challenges, Tolson and Shoemaker (2007) introduced the Dynamically Dimensioned Search (DDS) algorithm, tailored for complex, high-dimensional problems. DDS is more computationally efficiency than SCE-UA, and we therefore used it for our Noah-MP calibrations.

Formatted: First line: 0 ch

Formatted: Font: (Asian) +Body Asian (新細明體), (Asian) Chinese (Traditional, Taiwan)

~~Our selection of calibration algorithms was guided by the need to balance computational efficiency with robustness. In the calibration of the VIC model, we employed SCE-UA. This method is a global optimization method widely used in hydrology and environmental modeling, owing to its robustness and efficiency when addressing complex, non-linear, and multi-modal objective functions (Naeini et al.,~~

~~2015). The SCE-UA is well-established and widely recognized for its efficacy with this particular model (and with which we have considerable experience). The SCE-UA has been a benchmark in calibrating VIC for decades (see Naeini et al., 2019). The computational efficiency of VIC made SCE-UA a suitable choice, despite its requirement for a higher number of iterations. In practical terms, iterating a 20-year simulation in VIC takes about 2 minutes for a mid-sized basin, which we found manageable in terms of the computer resources available to us.~~

~~For VIC parameter estimation, we employed the Shuffled Complex Evolution algorithm developed at the University of Arizona (SCE-UA, Duan et al. 1992). This method is a global optimization method widely used in hydrology and environmental modeling, owing to its robustness and efficiency when addressing complex, non-linear, and multi-modal objective functions (Naeini et al., 2015).~~

~~For the Noah-MP model, which requires more computational core-hours per simulation, we used the DDS method. NOAA employs the DDS algorithm for their CONUS implementation of NWM, which is grounded in Noah-MP (Gochis et al. 2019). Although we had not used DDS previously, the fact that we had available to us a computational structure which embedded Noah-MP, in addition to its computational efficiency, was a deciding factor.~~

To assure that the parameter sets we estimated weren't dependent on the optimization method, we conducted a comparison between SCE-UA and DDS for calibrating VIC across 20 randomly chosen basins. We found that the DDS algorithm achieved optimal calibration with fewer iterations (typically around 3000 iterations vs only about 250 for DDS). The parameter sets identified were nearly identical, affirming our decision to use distinct algorithms tailored to the computational demands of each model. ~~We evaluated both calibration methods (DDS and SCE-UA) for VIC for 20 randomly chosen basins, and obtained similar results. For VIC, we chose SCE-UA due~~

Formatted: Justified

~~to its inherent compatibility with the model and because the additional computation (relative to DDS) was less important given that the inherent computation required for VIC is considerably less than for Noah-MP.~~ ~~While it was possible to calibrate all basins using DDS, the similar outcomes for the chosen basins with both algorithms did not necessitate this approach, (we note that we performed calibration on all of the basins for VIC before starting with the Noah-MP calibration). While using the same calibration method for both models could simplify comparisons, as noted above, the two methods produce essentially the same results, the only difference being computational efficiency.~~

~~In our application of SCE-UA, we performed a maximum of 3000 iterations for each basin, while the DDS method employed 250 iterations for each basin for Noah-MP. Each basin was calibrated using the most recent 20 years of streamflow data.~~ For both models, our objective function was the ~~Kling-Gupta Efficiency (~~KGE~~, Gupta et al., 2009)~~ metric for daily streamflow. KGE is a widely used performance measure because of its advantages in orthogonally considering bias, correlation and variability (Knoben et al., 2019). KGE = 1 indicates perfect agreement between simulations and observations; KGE values greater than -0.41 indicate that a model improves upon the mean flow benchmark (Konben et al., 2019).

TABLE 2. Calibration methods, parameters and modifications to their initial default values evaluated in the calibration.

Model	**VIC**		**Noah-MP**	
Calibration Method	SCE-UA		DDS	
Iterations	3000		250	
Calibrated Parameter	Variable Infiltration Curve Parameter (INFILT)	0.001 – 0.4 (Shi et al.,2008)	Saturated Hydraulic Conductivity (Ksat)	$2 \times 10^{-9} to$ 0.07(Cai et al.,2014)
	Baseflow parameter (Ds)	0.001 – 1.0 (Shi et al.,2008)	Saturation soil moisture content (MAXSMC)	0.1 to 0.71 (Cai et al.,2014)

Formatted: Font: (Asian) +Body Asian (新細明體), (Asian) Chinese (Traditional, Taiwan)

Formatted Table

Formatted: French (France)

Thickness of Soil in Layer 1 (Depth_1)	0.01 – 0.2 (Shi et al.,2008)	Pore size distribution index (Bexp)	1.12 to 22 (Cai et al.,2014; Gochis et al.,2019)
Total thickness of soil column (Depth_total)	0.6 – 3.5 (Shi et al.,2008)	Linear scaling of "openness" of bottom drainage boundary (Slope)	0.1-1 (Lahmers et al 2021)
Max velocity parameter of baseflow (Dsmax)	0.001 – 30 (Schaperow et al.,2023)	Parameter in surface runoff (REFKDT)	0.1-10 (Lahmers et al 2021)
Fraction of max soil moisture where nonlinear baseflow occurs (Ws)	0.001 – 1 (Shi et al.,2008)		

Formatted: French (France)

3.2 Noah-MP parameterization

As specified in Table 1, Noah-MP has four runoff and groundwater physics options (rnf). Initially, we adopted the options that are incorporated in the NWM, as elaborated in Gochis et al. (2020). Before we could proceed with calibrating Noah-MP for all the WUS basins, it was necessary to determine suitable rnfs. To streamline computational time, we initially selected 50 basins randomly from the total of 263 from which we created four experimental groups. Each group employed a different rnf option. We applied the DDS method to these groups and compared the cumulative distribution functions (CDF) of their baseline and calibrated KGEs (Figure 3~~2~~). From this figure, it's apparent that the KGE improved post-calibration for all four rnfs. Notably, rnf3, also known as free drainage, exhibited the most substantial performance enhancement after calibration. As a result, we chose to continue using this option which is incorporated in the NWM. Nonetheless, it's worth noting that the use of different options for different basins—a feature currently not utilized in Noah-MP or WRF-Hydro—could potentially result in improved overall model performance.

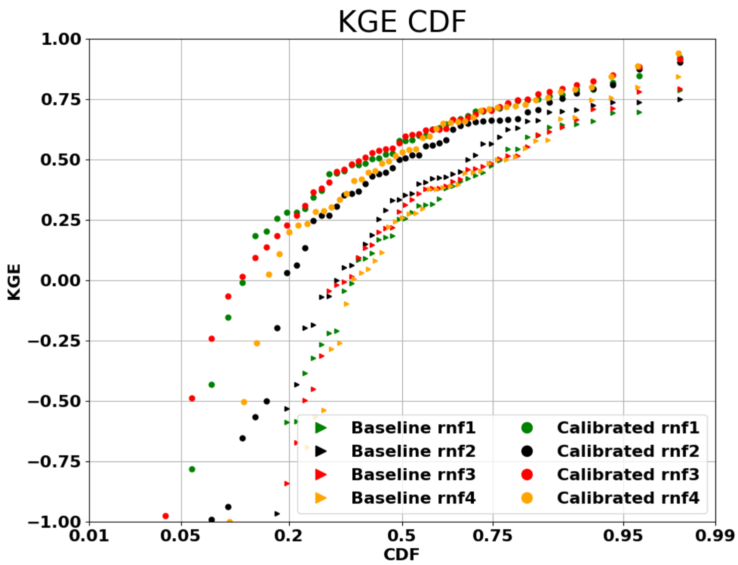


Figure 3~~2~~. Streamflow performance (KGE of daily streamflow simulations) of different Noah-MP runoff generation options across 50 (of 263) randomly selected basins. The performances are shown for both baseline and calibrated simulations.

3.3 Calibration of gauged basins

Following the selection of the most effective set of runoff generation options across the domain, we estimated model parameters for all 263 basins. The comparative performance of the models, before and after calibration, is shown in Figure 4~~3~~. It's apparent from the figure that both Noah-MP and VIC have significantly enhanced their daily streamflow simulation skills post-calibration. After calibration, the median KGE of Noah-MP improved from 0.22 to 0.54, and the VIC's median

KGE increased from 0.37 to 0.70. When contrasting the two models, we observed that VIC outperformed Noah-MP both pre- and post-calibration. One possible explanation could be that the baseline VIC parameters were taken from Livneh et al. (2013), and these parameters had already been validated and adjusted for major U.S. basins (although not for our 263 basins specifically), while the Noah-MP parameters are default values from NWM. Another possibility is inherent differences in the physics of streamflow simulation between the two models (VIC primarily generates runoff via the saturation excess mechanism), although that isn't the main focus of our research.

Following the calibration with data from the past 20 years, we performed a test where we calibrated the streamflow using the first 10 years of data and validated with the subsequent 10 years of data. This test revealed that the KGE distribution from the 10-year calibration is similar to that from the 20-year data. The median KGE values for VIC and Noah-MP after calibration with 10 years of observations were 0.52 and 0.69, respectively. Correspondingly, the median KGEs during the validation period were 0.50 and 0.68, respectively, which are only slightly lower. These comparisons demonstrate general consistency over time in the performance of the calibrated parameters.

To validate the robustness of our calibration methodology, we calculated alternative (to KGE) performance metrics , specifically Nash-Sutcliffe Efficiency (NSE) and bias. Our analyses, detailed in Figures S2&3, revealed significant enhancements in model performance ~~in~~ as measured by these metrics. The observed improvements across multiple evaluation criteria affirm the efficacy of our calibration process, and in particular that the performance of our procedures is not contingent upon the choice of evaluation metrics.

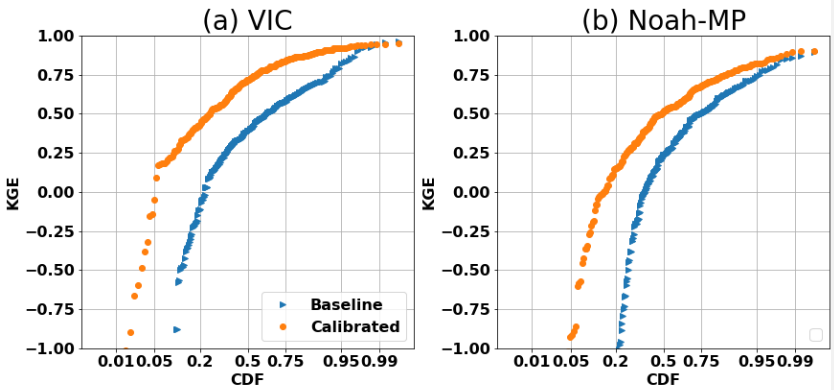


Figure 4~~3~~. Cumulative Distribution Function (CDF) plot of the daily streamflow KGE for (a) VIC and (b) Noah-MP, comparing baseline and calibrated runs across all 263 basins.

We examined the spatial variability of daily streamflow KGE for Noah-MP and VIC, both before and after the calibration (see Figure 5~~4~~). The highest baseline KGEs are along the Pacific Coast, in central to northern CA for both models. VIC's baseline KGE generally is high in the Pacific Northwest. Post-calibration improvements occurred for both models in most areas, especially in regions where the baseline KGE was low, such as southern CA and the southeastern part of the study region. Median improvements after calibration were 0.27 for Noah-MP and 0.30 for VIC.

We observed that basins displaying higher KGE values typically were more humid than those with lower KGE. To further delve into the relationship between KGE and basin characteristics, we explored correlations between KGE and 21 different characteristics, including drainage area, elevation, seasonal/annual average temperature and precipitation, annual maximum precipitation, and seasonal/annual runoff ratio. Of these, 12 characteristics were statistically significantly correlated with the VIC KGE, including four seasonal and annual runoff ratios; mean precipitation in winter, spring, and fall; annual maximum precipitation; and minimum elevation.

Figure 6~~5~~ shows scatterplots of eight representative characteristics. Apart from minimum elevation and mean summer temperature, all other characteristics were positively correlated with KGE. Typically, spring runoff ratio, annual runoff ratio, mean annual max precipitation, and mean winter precipitation exhibited the highest correlations with KGE. This implies that basins with higher runoff ratios (particularly in spring), higher precipitation (especially maximum precipitation), lower summer temperature, and lower elevation are more likely to exhibit strong VIC performance. The same applies to Noah-MP, as indicated in Figure 7~~6~~, although Noah-MP showed relatively weaker correlations. Correlations between mean summer temperature and mean fall precipitation and Noah-MP KGE weren't statistically significant.

The spatial distribution of the eight characteristics is qualitatively similar with the KGE spatial distribution, as shown in Figure 8~~7~~. Generally, basins with higher KGE have higher characteristic values when the correlation is positive, and lower characteristic values when the correlation is negative. As noted above, both models show good baseline performance along the Pacific Coast, and in central to northern CA (Figure 5). Those areas have high runoff ratios (specifically spring and annual) and high mean winter precipitation. These features generally lead to runoff physics that ~~have infiltration~~are dominated by the -saturation-excess mechanism, which is well represented by both ~~thus both~~ VIC and Noah-MP. VIC's baseline KGE generally is high in the inland Northwest which has somewhat lower ~~mean annual max daily precipitation~~ runoff ratios and (relatively) deeper groundwater tables. VIC's superior performance relative to Noah-MP may also be because of its variable rather than fixed soil moisture depths (as is the case for Noah-MP).

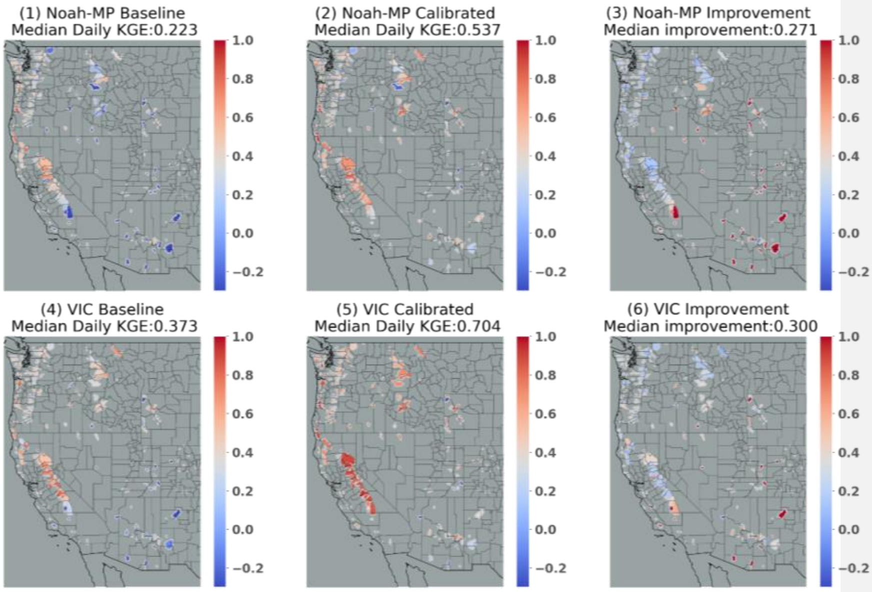


Figure 54. Spatial distribution of daily streamflow KGE for Noah-MP baseline (1); calibrated Noah-MP (2); difference between calibrated and baseline Noah-MP (3); VIC baseline (4); calibrated VIC (5); difference between calibrated and baseline VIC (6).

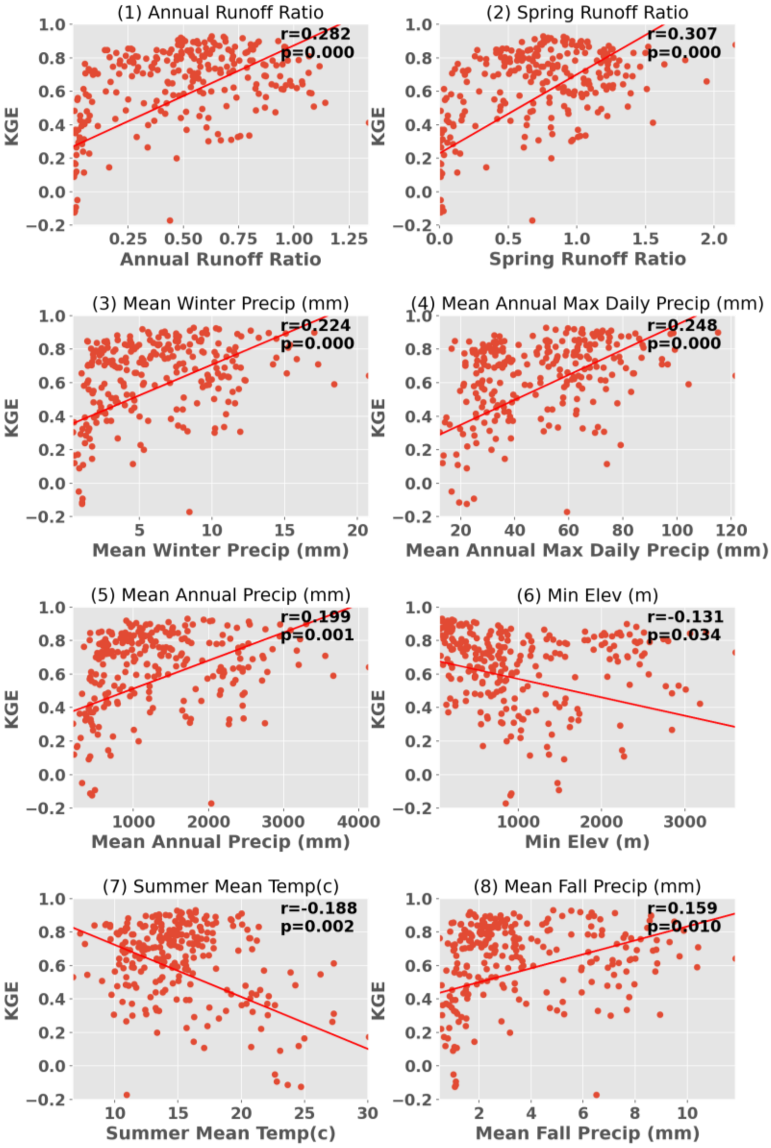


Figure 6~~5~~. Scatterplots of VIC KGE in relation to significantly correlated characteristics. Each subplot indicates the corresponding Pearson correlation coefficients and the P-value.

Formatted: Indent: First line: 2 ch

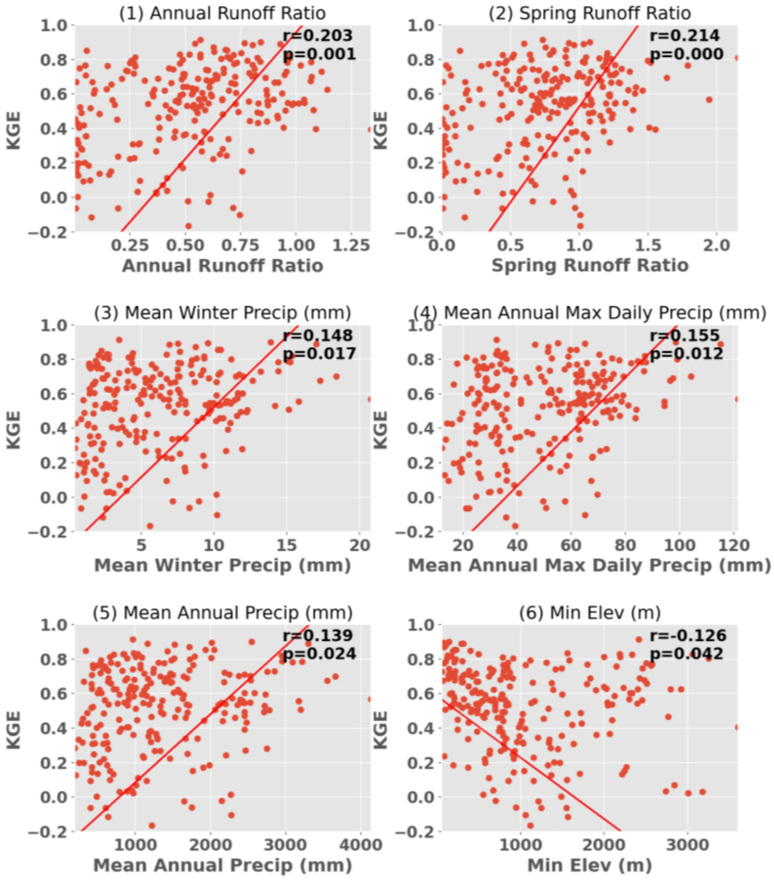


Figure 7~~6~~. Scatterplot of Noah-MP KGE in relation to significantly correlated characteristics. Each subplot indicates the corresponding Pearson correlation coefficients and the P-value.

Formatted: Indent: First line: 2 ch

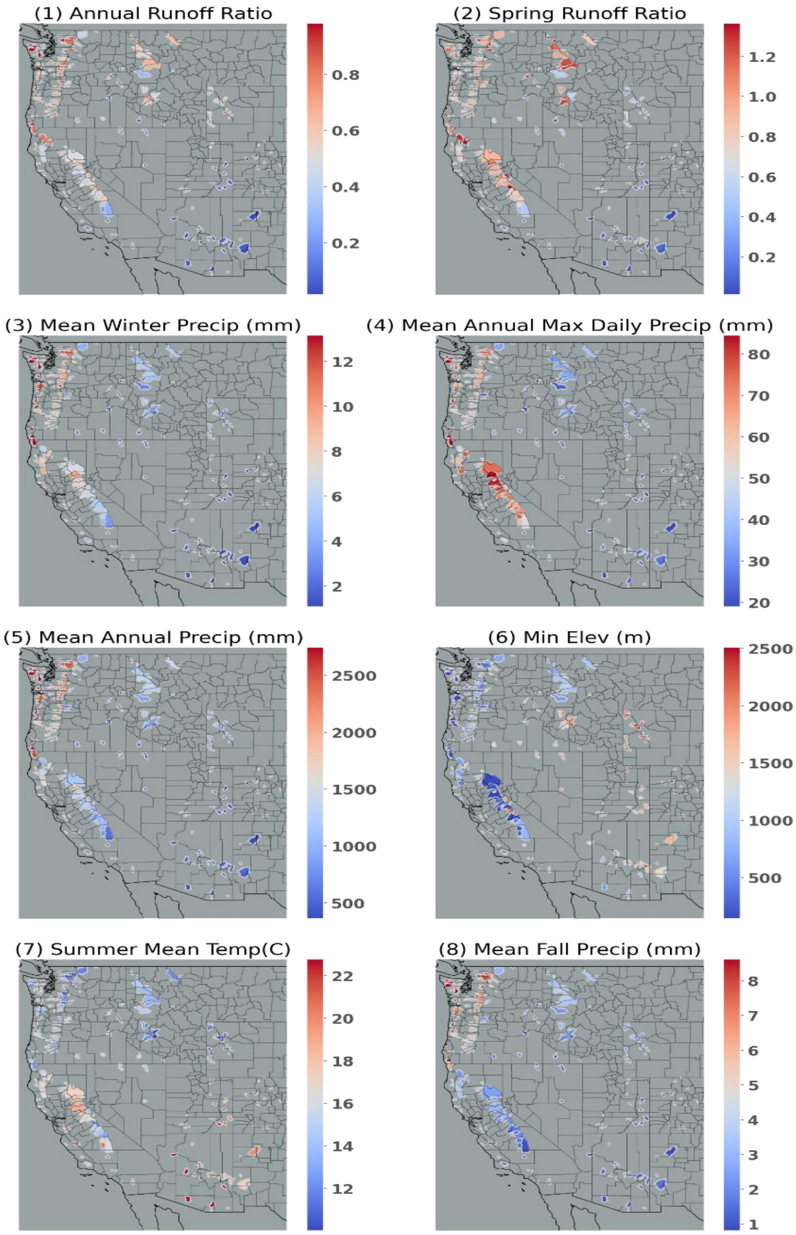


Figure 8~~7~~. Spatial distribution of characteristics that are statistically significantly correlated with KGE. Note that all characteristics are significantly correlated with VIC KGE whereas only (1)-(6) are significantly correlated with Noah-MP KGE.

4. Regionalization

To distribute parameters from the calibration basins to the entire region~~,~~, we used the donor-basin method ~~, where ungauged basins inherit parameters from similar gauged basins. This choice is supported by its successful application in NOAA/NCAR's National Water Model and proven effectiveness in various studies (e.g., ~~as implemented in numerous previous studies (e.g., Arsenault and Brissette (2014); Poissant et al. (2017); Razavi and Coulibaly (2017); Gochis et al. (2019); Qi et al. (2021) and ~~;~~Bass et al.~~,~~ (2023). ~~The donor-basin approach allows for more accurate and detailed application to ungauged basins, making it a suitable choice for our study.~~

Following the calibration process, we regionalized the parameters from gauged to ungauged basins based on a mathematical assessment of the spatial and physical proximity between the gauged and ungauged basins. ~~; following previous studies by Arsenault and Brissette (2014), and Razavi and Coulibaly (2017). We opted for this method over an alternate approach that first regionalizes the streamflow attributes (such as runoff depth, high flow indicators) and then standardizes the model throughout (as proposed by Castiglioni et al., 2010; Oubeidillah et al., 2014; and Yang et al 2017). The reason for our choice is our interest in actual streamflow time series rather than metrics. We carried out the regionalization after calibrating to specific streamflow gauges, ensuring high precision for these gauged basins and facilitating high-quality regionalization in ungauged basins. In hydrology, ~~We considered two primary methods for implementing the donor basin approach. The first uses models calibrated to spatially continuous gridded runoff metrics (Beck et al. 2015; Yang et al. 2019). The second approach, which we ultimately adopted, calibrates models to individual gauges, then extends these parameters to ungauged basins, based either on a statistical or mathematical similarity measures (e.g., Arsenault and Brissette 2014;

Formatted: Indent: First line: 0 ch

Razavi and Coulibaly 2017). Our preference for the second method was guided by a key limitation of the first approach, specifically it is limited to calibrating against runoff metrics, such as long-term mean flow and flow percentiles, rather than streamflow time series. ~~Additionally, their usually coarser resolution compared to forcing datasets and hydrologic models is a constraint. Our focus on actual streamflow time series led us to this approach. Specifically, we employed a donor-basin approach which has demonstrated successful outcomes, as reported by Bass et al. (2023).where an ungauged basin adopts calibrated parameters from its most similar gauged basin(s). This method has been applied in many studies including Arsenault and Brissette (2014); Poissant et al. (2017); Razavi and Coulibaly (2017); Gochis et al. 2019; Qi et al. (2021); and Bass et al (2023).~~

In the donor-basin method, an ungauged basin inherits its land surface parameters from the most similar gauged basin(s) (or the ~~top 'x'~~ 'n' most similar gauged basins). Here, we evaluated the similarity or proximity between gauged and ungauged basins based on the similarity index SI as defined and used by Burn and Boorman (1993) and Poissant et al. (2017):

$$SI = \sum_{i=1}^{k} \frac{|X_i^G - X_i^U|}{\Delta X_i} \qquad (1)$$

In Eq. 1~~this formula~~, k stands for the total number of features considered, X_i^G represents the ith feature of the gauged basin G, X_i^U is the ith feature of a specific ungauged basin, and ΔX_i is the range of potential values for the ith feature, grounded in the data from the gauged basins. This yields a unique value of SI for each gauged basin, contingent on the specific ungauged basin it is compared with. Typically, gauged basins that exhibit greater resemblance to the ungauged basin will have a smaller SI.

We assessed the donor-basin method's efficacy using a cross-validation approach, where each gauged basin was treated as ungauged one at a time. The pseudo-ungauged basin inherits its hydrological parameters from its three most similar gauged basins, determined by SI. The parameters inherited are a weighted average from the three donor basins. After testing one to five donor basins, we found that using three donors yielded the best results. Thus, every basin inherits parameters from the three most similar gauged basins in each simulation, offering a concise evaluation of the donor-basin method's regionalization performance.

We used 18 basin-specific features in the donor basin method, detailed in Table S1, calculated based on the forcings and parameters used in the study. For feature selection in the donor-basin method, we adopted an iterative approach, explained in detail in the following paragraph.~~.~~ ~~Each iteration added a single feature to the index, with the most beneficial feature (based on median KGE improvement) retained. This process was repeated until the median KGE no longer improved.~~ Only basins with a KGE exceeding 0.3 were considered, following previous studies suggesting that inclusion of poorly performing basins can lower regionalization performance. We found that a KGE threshold of 0.3 resulted in a median performance improvement of 0.08 larger than did a KGE threshold of 0, hence it was chosen. After screening, 223 basins were utilized in VIC regionalization and 194 in Noah-MP regionalization. We note that the parameters used for calibration and the features used to determine the similarity index in the regionalization process are different. The physics that control the key hydrological processes of the two models are different, so we explored their best regionalization features separately.

Formatted: Left, First line: 0 ch

To determine the most effective regionalization features from the 18 basin characteristics listed in Table S1, we employed a systematic iterative approach. The first iteration includes 18 simulations, each of which incorporates one of the 18 features. The feature that yielded the greatest increase in the median KGE across all basins, based

Formatted: Justified, First line: 2 ch

on leave-one-out cross validation, was then retained. In the second iteration, we conducted 17 simulations, each combining the retained feature from the first iteration with one of the remaining 17 features. This process was repeated iteratively, reducing the number of features considered in each subsequent round, until the addition of new features no longer resulted in an appreciable increase in median KGE. The sequence of features shown in Figure 9 (also shown in Table S1) indicated the importance of the features. This iterative approach ensured that each feature's individual and combined contribution to model performance was thoroughly assessed. It allowed us to identify a subset of features that, when used together, optimally improved model accuracy. We recognize the potential existence of inter-feature correlations that may exert a discernible influence on their collective efficacy when utilized in combination.

This procedure ~~, Wwe~~ resulted in ~~found~~ five features generated the best regionalization performance for VIC (longitude centroid, latitude centroid, maximum elevation, fall mean precipitation, and fall mean temperature). ~~and t~~Three features were found to be best for Noah-MP (latitude centroid, longitude centroid, and drainage area) (see Figure 9~~8~~). Among them, latitude and longitude are the common features that contribute the most to regionalization when using the similarity index method. This suggests that geographical similarities are the most important factor in parameter information transfer from gauged to ungauged basins.

Upon evaluating the performance of baseline, calibrated, and regionalized simulations, the respective median daily KGEs for the VIC model were found to be 0.41, 0.71, and 0.49. For the Noah-MP, these values were 0.38, 0.60, and 0.49 (refer to Figures 9~~8~~ & S4~~9~~). These metrics are for basins that have a calibrated KGE greater than 0.3 only, resulting in higher median KGEs than for all 263 basins (See Figure 4~~3~~). The KGE distribution also improved overall. It's noteworthy that the regionalization improvement relative to baseline is higher for Noah-MP than for VIC.

Formatted: Not Highlight

While VIC's baseline and calibrated KGE skill distribution outperforms Noah-MP's, the regionalized skills of Noah-MP and VIC are quite comparable. This observation might be attributable to the constraints of the regionalization setup and could warrant future investigation.

After optimizing the features and specific design of the donor-basin method, parameters were regionalized to 4816 ungauged USGS Hydrologic Unit Code (HUC) -10 basins across the WUS. HUCs are delineated and quality controlled by USGS using high-resolution DEMs. For each of the 4816 HUC-10 basins, we calculated a similarity index with the calibrated basins using the selected features. The three most similar basins were identified as donor basins, and their weighted average parameters were then adopted by the target HUC-10 basin. The final hydrologic parameters for both VIC and Noah-MP for all WUS HUC-10 basins are shown in Figures S5~~1~~&S6~~2~~. The baseline HUC-10 parameters are shown in Figures S7~~3~~&S8~~4~~.

Comparison of Figures S4-5 to Figures S6-7 makes it clear that the baseline model parameters lack accuracy, and exhibit significant spatial uniformity where large geographical regions share identical parameter values. For example, parameters such as Ds and Soil_Depth1 in VIC show this uniformity. Furthermore, certain parameters, such as SLOPE and REFKDT in ~~the~~ Noah-MP, remain~~ed~~ invariant across all spatial domains, and don't reflect real-world conditions. Regionalization, improved the parameters, leading to increased accuracy and strengthening of region-specific characteristics.

Formatted: Indent: First line: 0 ch

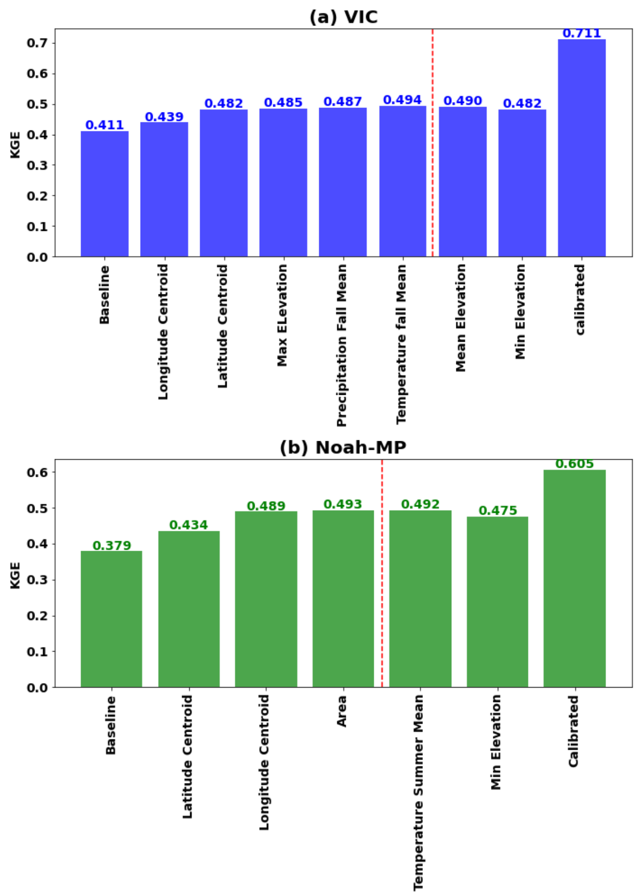
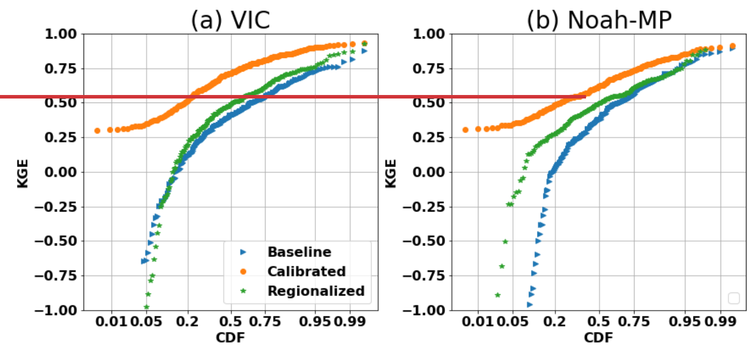


Figure 9~~8~~. Best regionalization features for (a) VIC and (b) Noah-MP. The final regionalization to ungauged basins of the WUS incorporated all features up to the point marked by the red line since the addition of further features doesn't improve KGE.



~~**Figure 9. CDF of daily KGE for (a) VIC and (b) Noah-MP, comparing baseline and calibrated runs across selected regionalized basins within the WUS.**~~

Formatted: Heading 1, Indent: First line: 0 cm

5. Evaluation of high and low flow simulation skill

Our primary calibration objective was to enhance the accuracy of daily streamflow simulations. However, to ensure the versatility of our parameter sets for research related to both floods and dry conditions, we also evaluate the models' capabilities in reproducing high and low streamflow. To understand the capabilities of the two models in reconstructing high and low streamflow, we assessed their performance across baseline, calibrated, and regionalized settings.

(a) Evaluation of high flow performance

We used the peaks-over-threshold (POT) method (Lang et al. 1999) to identify extreme streamflow events as in Su et al (2023) and Cao et al. (2019, 2020). We first applied the event independence criteria from USWRC (1982) to daily streamflow data to identify independent events. We set thresholds at each basin that resulted in 3 extreme events per year on average. After selecting the flood events over the study period based on the observation, we sorted the floods based on the return period and then calculated the KGE of baseline, calibrated and regionalized floods. Figure S9~~10~~ displays the associated CDF plots. The median KGE for baseline floods in Noah-MP

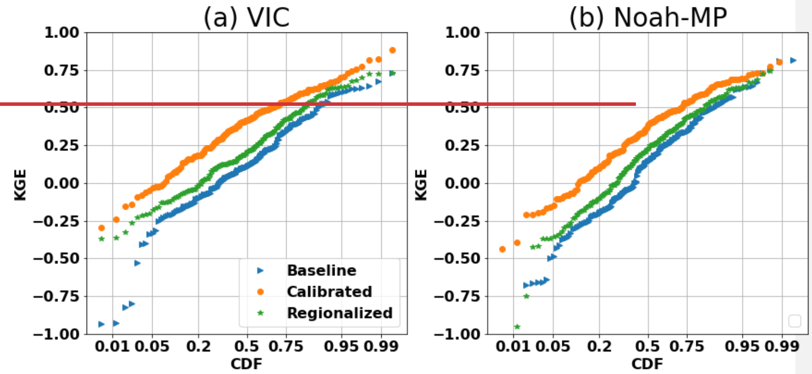
Formatted: Not Highlight

was 0.14, which rose to 0.37 post-calibration, and receded to 0.22 after regionalization. For VIC, the flood KGE started at 0.11, increased to 0.41 after calibration, and declined to 0.20 post-regionalization. As anticipated, these numbers are lower than (all) daily streamflow skill due to our calibration target being daily streamflow. Still, flood competencies experienced considerable enhancement, surpassing the Noah-MP KGE benchmark of -0.41 found by Knoben et al. (2019).

(b) Evaluation of low flow performance

To assess low flow performance, we utilized the 7q10 metric. This hydrological statistic, commonly adopted in water resources management and environmental engineering, is the lowest 7-day average flow that occurs (on average) once every 10 years (EPA,2018). Scatterplots of 7q10 (Figure S10~~11~~) showed high correlation between our model's simulated low flows and the observed data. Post-calibration, this alignment intensified. The VIC model tended to underestimate the low flows. After calibration, the median bias improved from -23.6% to -9.9%, and with regionalization, it was -11.7%. In contrast, Noah-MP began with an 11.20% overestimation in the baseline, improved to 0.61% post-calibration, and was -9.5% after regionalization. The outcomes underline the proficiency of both models for low flow prediction, exhibiting enhanced competencies post-calibration and commendable performance after regionalization.

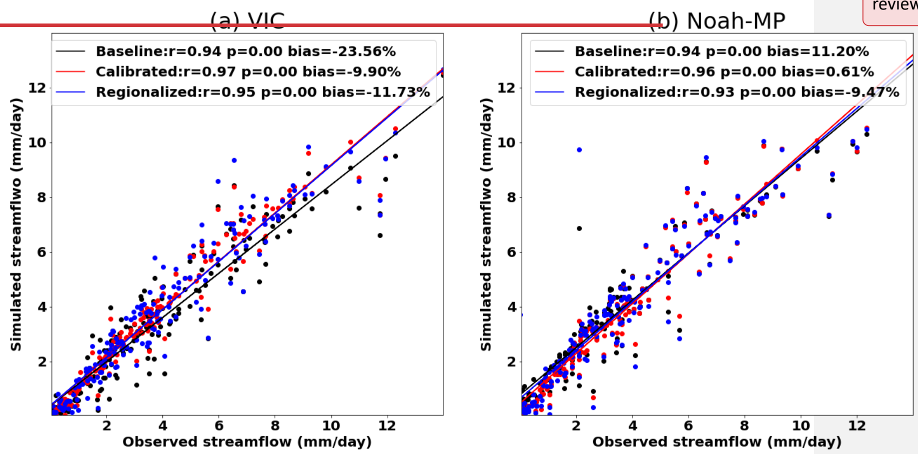
Formatted: Not Highlight



~~Figure 10. CDF of high flow KGE for (a) VIC and (b) Noah-MP, comparing baseline and calibrated runs across selected regionalized basins within the WUS.~~

Commented [DL2]: Where are figure 10-11 titles?

Commented [LS3R2]: Moved to supplement as suggested by reviewer



~~Figure 11. Scatterplot of 7q10 low flows (the lowest 7-day average flow that occurs (on average) once every 10 years) for the baseline and calibrated and regionalized runs for (a) VIC model and (b) Noah-MP. The correlation coefficients, P-values and percentage bias are denoted in the upper section of the figures. The x axis is observed low flow and the y axis is simulated low flow.~~

6. Discussion

In this discussion, we summarize our key accomplishments in calibrating the two hydrological models, examine our approach to choosing calibration objective

Formatted: Normal, Indent: First line: 0.85 cm functions and metrics, and we consider lessons learned in model regionalization.

(a) Improved parameter sets

We generated calibrated parameter sets for the VIC and Noah-MP hydrological models at 1/16° latitude-longitude scale across WUS. These calibrated parameter sets are intended to facilitate the use of the two models for climate change and water investigations across the region, among other applications. Our focus on calibrating daily streamflow aligns with common practice in hydrology, providing a comprehensive representation of catchment hydrology dynamics which should enhance future understanding of hydrological phenomena and their spatial variations across the region.

(b) Selection of calibration objective function

We used objective functions based on streamflow observations. We chose this approach due to its applicability elsewhere, given the widespread accessibility of streamflow observations as compared to alternative metrics such as soil moisture or evapotranspiration (Demaria et al., 2007; Gao et al., 2018; Troy et al., 2008; Yadav et al., 2007). While we acknowledge the potential of remote sensing products like MODIS, SMAP, SMOS, ESA, and ALEXI to improve calibration efforts, especially for variables like actual evapotranspiration (AET) and soil moisture (SM), we were limited by the scarcity of observations for these variables. Future studies could, nonetheless, leverage from the methods we've employed to incorporate additional variables into the objective functions we used.

(c) Selection of calibration metric

We used the KGE metric applied to daily streamflow, which we chose for its ability to address bias, correlation, and variability simultaneously (Knoben et al., 2019). We also evaluated NSE and BIAS metrics, and found substantial improvements in both models' performance after calibration when these metrics were

Formatted: Font: (Asian) +Body Asian (新細明體)

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Asian) DengXian, (Asian) Chinese (Simplified, Mainland China)

Formatted: Indent: Left: 0.85 cm, First line: 0 ch

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: a, b, c, … + Start at: 1 + Alignment: Left + Aligned at: 0.85 cm + Indent at: 1.48 cm used in place of KGE (See Figures S2-3) . Our assessment of high and low flow reconstruction in Section 5 further validated our generated parameter sets. While we used a single objective function due to data and computing constraints, incorporating multiple objective functions is feasible in principle.

(d) Regionalization possibilities

We calibrated model parameters directly for individual basins, considering their unique hydrological features, and then transferred these calibrated parameters to similar basins based on similarity assessments. Alternative parameter transfer strategies could be used within the same framework we employed (e.g., pedo-transfer functions, e.g. Imhoff et al.,2020) or multiscale parameter regionalization (e.g. Schweppe et al.,2022). We do note that our regionalization approach facilitates the transfer of calibrated parameters to comparable regions, which could be explored in future research.

7. Conclusions

Our intent was to develop a regional parameter estimation strategy for the VIC and Noah-MP land surface schemes, and to apply it across the WUS region at the HUC-10 catchment scale. We've described what we believe is a robust framework that can be applied in future hydrological and climate change studies across the WUS, and is applicable to other regions as well. Our key findings and conclusions are:

a) Our catchment scale calibration of the two models to 263 sites across WUS resulted in major improvements in the performance of both models relative to a priori parameters, but performance improvement was greatest for Noah-MP – although this may be in part because VIC a priori parameters benefitted from prior calibration and hence resulted in better baseline performance than did a priori Noah-MP.

b) Both models performed best in more humid basins, mainly in the Pacific

Formatted: Normal, Indent: First line: 0.85 cm

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Northwest and central to northern CA where runoff ratios are high. This is consistent with previous results (e.g. Bass et al.,2023).

Formatted: Font: (Default) Times New Roman

c) Post-calibration regional model performance improved for both models in most areas, especially where the baseline KGE was low, such as southern CA and the southeastern part of the study region.

Formatted: Font: (Default) Times New Roman

d) VIC performance across all calibration basins generally was better than for Noah-MP. However, Noah-MP performance benefitted more from regionalization than did VIC, and ultimately post-regionalization VIC performance was only slightly superior to that of Noah-MP.

~~The calibrated parameters, derived from global optimal calibration methods, are set to enhance hydrological simulations and forecasting in the WUS, providing valuable support for regional hydrologic and river hydrodynamic modeling studies. The improved model predictions are expected to benefit water management practices in the region and contribute to a better understanding of climate change impacts. Furthermore, the methodologies employed in this study have broader applicability and can be adapted to other geographic areas, extending the significance of our work beyond the WUS. This research thus contributes to the global effort to improve hydrological system understanding and management.~~

~~While this study has yielded valuable insights and contributions, we acknowledge certain limitations, including the use of a single objective function and the exclusive focus on streamflow calibration. In future research, we intend to explore additional objective functions and broaden the scope of calibration to include other hydrological variables, thereby further enriching the robustness and applicability of our methodology.~~

~~Our objective was to produce parameter sets for VIC and Noah-MP over WUS~~

~~that could be used in regional studies, and would result in better model performance than default or other "off the shelf" parameters. We identified preferred runoff generation options for Noah-MP (physics options are fixed in VIC) using a subset of our WUS basins (50 in total) for which we evaluated all four Noah-MP runoff generation options. Once we identified the optimal runoff generation options for Noah-MP, we identified (calibrated) parameters for both Noah-MP and VIC for each of our 263 basins across WUS using the most recently available 20-years of streamflow observations. Following calibration, the Noah-MP median KGE increased from 0.22 to 0.54, while the median VIC KGE rose from 0.37 to 0.70. VIC KGEs were higher than Noah-MP's both before and after calibration across the 263 basins, possibly because the initial VIC parameters had the benefit of some previous calibration, albeit for much larger river basins across WUS (in the case of post-calibration KGE, it's unclear whether and how they might have been affected by the choice of initial parameters). Other possible cause of the differences could be inherent differences in streamflow simulation physics between the two models. We also conducted a test using the initial 10 years of data for calibration and the following 10 years for validation, and found results that were consistent with those we obtained using the entire 20 years for calibration.~~

~~Upon the selection of suitable parameterizations for Noah-MP and calibration of gauged basins for both VIC and Noah-MP, we extended the use of the calibrated parameters to ungauged basins across the WUS for both models. This extension was achieved through the donor-basin regionalization method, which allows ungauged basins to inherit parameters from gauged basins with similar hydroclimatic properties. We discovered that using a weighted combination of three similar basins yielded better regionalization results (in terms of KGE) compared to using the single most similar donor basin, as determined by a similarity index. Following regionalization,~~

~~the median KGE for VIC rose from 0.41 to 0.49, and for Noah-MP it increased from 0.38 to 0.49 over the selected basins. Interestingly, even though the pre-regionalization KGE for VIC was considerably higher than for Noah-MP, the post-regionalization values for the two models were nearly identical. Stated otherwise, the regionalization enhancement was considerably greater for Noah-MP than for VIC. We further evaluated high and low flow simulation skills and found the skill significantly improved after calibration for both VIC and Noah-MP and improvements remained after regionalization. Following calibration and regionalization, we developed gridded parameter sets for both models at 1/16° latitude-longitude resolution for all 4816 HUC-10 basins across the WUS. These parameter sets should be useful for regional hydrologic and river hydrodynamic modeling studies over all or parts of the WUS domain. Improving the accuracy of the models' predictions should have benefits for water management across the region, and more and more generally for understanding the potential impacts of climate change across the region. Moreover, the methods and procedures we utilized are not restricted to our current research domain; they could be transferred readily to other geographic regions. In effect, our research contributes to both local and global efforts to understand and manage our critical hydrological systems better, demonstrating its broader relevance and utility.~~

Data Availability statement

The Livneh (2013) forcings are available at http://livnehpublicstorage.colorado.edu:81/Livneh.2013.CONUS.Dataset/. The extended forcings used in this study are available at ftp://livnehpublicstorage. colorado.edu/public/sulu. The results are available online at https://figshare.com/s/66fe8305bff516e80f6f .

Author contribution

LS and DL conceptualized the study. LS generated the dataset and analysis with support of DL, MP and BB. LS drafted the manuscript with support of DL.

Formatted: Font: (Asian) +Body Asian (新細明體)

Competing interests. The contact author has declared that none of the authors has any competing interests.

References

Adam, J.C. and Lettenmaier, D.P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108(D9), 1-14, doi:10.1029/2002JD002499, 2003.

Adam, J.C., Clark, E.A. , Lettenmaier, D.P. and Wood, E.F.: Correction of Global Precipitation Products for Orographic Effects, J. Clim., 19(1), 15-38, doi: 10.1175/JCLI3604.1, 2006.

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F. , Nijssen, B. and Lettenmaier, D.P.: Value of Long-Term Streamflow Forecasts to Reservoir Operations for Water Supply in Snow-Dominated River Catchments. Water Resources Research 52: 4209–25, 2016.

Arsenault, R., and Brissette, F. P.: Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. Water Resour. Res., 50, 6135–6153, https://doi.org/ 10.1002/2013WR014898, 2014.

Bass, B., Rahimi, S., Goldenson, N., Hall, A., Norris, J. and Lebow, Z.J.: Achieving Realistic Runoff in the Western United States with a Land Surface Model Forced by Dynamically Downscaled Meteorology. Journal of Hydrometeorology, 24(2), 269-283, 2023.

Beck, H. E., Roo, A. de and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments. J. Hydrometeor., 16, 1478–1501, https://doi.org/10.1175/JHM-D-14-0155.1, 2015.

Bennett, A. R., Hamman, J. J. and Nijssen, B.: MetSim: A Python package for estimation and disaggregation of meteorological data. J. Open Source Software, 5, 2042, https://doi.org/10.21105/joss.02042, 2020.

Beven, K.: Changing ideas in hydrology-the case of physically-based models. Journal of Hydrology, 105(1-2), 157–172. https://doi. org/10.1016/0022-1694(89)90101-7, 1989.

Bohn, T. J., Livneh, B., Oyler, J. W., Running, S. W., Nijssen, B. and Lettenmaier, D. P.: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. Agric. For. Meteor., 176, 38–49, https://doi.org/10.1016/j.agrformet.2013.03.003, 2013.

Boucher, M.-A., and Ramos, M.-H.: Ensemble Streamflow Forecasts for Hydropower Systems. In Handbook of Hydrometeorological Ensemble Forecasting, edited by Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H.L. Cloke, and J.C. Schaake, 1–19. Berlin Heidelberg: Springer, 2018.

Burn, D. H., and Boorman, D. B.: Estimation of hydrological parameters at ungauged catchments. J. Hydrol., 143,429454, https://doi.org/10.1016/0022-1694(93)90203-L, 1993.

Cai, X., Yang, Z.-L. , David, C. H., Niu, G.-Y. and Rodell, M.: Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. J. Geophys. Res. Atmos., 119, 23–38, https://doi.org/10.1002/2013JD020792, 2014.

California Department of Water Resources: California data exchange center: Daily full natural flow for December 2022. California Department of Water Resources, accessed 1 October 2021, https://cdec.water.ca.gov/reportapp/javareports?name= FNF, 2021.

Cao, Q., Mehran, A. , Ralph, F. M. and Lettenmaier, D. P.: The role of hydrological initial conditions on atmospheric river floods in the Russian River basin. J. Hydrometeor., 20, 16671686, https://doi.org/10.1175/JHM-D-19-0030.1, 2019.

Cao, Q., Gershunov, A., Shulgina, T., Ralph, F. M. , Sun, N. and Lettenmaier, D. P.: Floods due to atmospheric rivers along the U.S. West Coast: The role of antecedent soil moisture in a warming climate. J. Hydrometeor., 21, 1827–1845, https:// doi.org/10.1175/JHM-D-19-0242.1, 2020.

Castiglioni, S., Lombardi, L., Toth, E. , Castellarin, A. and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. Advances in Water Resources, 33(10), 1235–1242. https://doi.org/10.1016/j.advwatres.2010.04.009, 2010.

Chen, F., and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Wea. Rev., 129, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Chen, F., and Coauthors: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. J. Geophys. Res., 101, 7251–7268, https://doi.org/10.1029/95JD02165, 1996.

Cosby, B.J., Hornberger, G.M., Clapp, R.B. and Ginn, T.: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water resources research, 20(6), 682-690, 1984.

Demaria, E. M., Nijssen, B., & Wagener, T.: Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. Journal of Geophysical Research, 112, D11113. https://doi.org/10.1029/2006JD007534, 2007.

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., and Stisen, S.: Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, Hydrol. Earth Syst. Sci., 22, 1299–1315, https://doi.org/10.5194/hess-22-1299-2018, 2018.

Dickinson, R. E., Henderson-Sellers, A. & Kennedy, P. J.: Biosphere–Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community

Climate Model. NCAR Tech. Note TN383+STR, NCAR, 1993.

Duan, Q., Sorooshian, S. and Gupta, V. : Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res., 28, 1015–1031, https://doi.org/ 10.1029/91WR02985, 1992.

Environmental Protection Agency (EPA) Office of Water: Low Flow Statistics Tools: A How-To Handbook for NPDES Permit Writers. EPA-833-B-18-001, 2018.

Falcone, J.: GAGES-II: Geospatial attributes of gages for evaluating streamflow. U.S. Geological Survey, accessed 1 April 2021, https://water.usgs.gov/GIS/metadata/usgswrd/XML/ gagesII_Sept2011.xml, 2011.

Federal Institute of Hydrology: "SOSRHINE." http://sosrhine.euporias.eu/en/sosrhine_overview, 2020.

Fisher, R.A. and Koven, C.D.: Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. Journal of Advances in Modeling Earth Systems, 12(4), p.e2018MS001453, 2020.

Franchini, M., Galeati, G. and Berra S.: Global optimization techniques for the calibration of conceptual rainfall-runoff models, Hydrol. Sci. J., 43, 443 – 458, 1998.

Gao, H., Birkel, C., Hrachowitz, M., Tetzlaff, D., Soulsby, C., & Savenije, H. H. (2018). A simple topography-driven, calibration-free runoff generation model. Hydrology and earth system sciences discussions.,1–42. https://doi.org/10.5194/hess-2018-141

Gochis, D. and Coauthors: Overview of National Water Model Calibration: General strategy and optimization. National Center for Atmospheric Research, accessed 1 January 2023, 30 pp., https://ral.ucar.edu/sites/default/files/public/9_RafieeiNasab_CalibOverview_CU

AHSI_Fall019_0.pdf, 2019.

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., et al.: Multi-objective parameter optimization of common land model using adaptive surrogate modeling. Hydrology and Earth System Sciences, 19(5), 2409–2425. https://doi.org/10.5194/hess-19-2409-2015, 2015.

Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J. and Zhou, R.: Sensitivity analysis-based automatic parameter calibration of the VIC model for streamflow simulations over China. Water Resources Research, 56(1), e2019WR025968, 2020.

Gupta, H. V., et al.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377, 80-91,2009.

Holtzman, N.M., Pavelsky, T.M., Cohen, J.S., Wrzesien, M.L. and Herman, J.D.: Tailoring WRF and Noah-MP to improve process representation of Sierra Nevada runoff: Diagnostic evaluation and applications. Journal of Advances in Modeling Earth Systems, 12(3), p.e2019MS001832, , 2020.

Hussein, A.: Process-based calibration of WRF-hydro model in unregulated mountainous basin in Central Arizona. M.S. thesis, Ira A. Fulton Schools of Engineering, Arizona State University, 110 pp., https://keep.lib.asu.edu/_flysystem/fedora/c7/224690/Hussein_asu_0010N_19985.pdf, 2020.

Imhoff, R.O., Van Verseveld, W.J., Van Osnabrugge, B. and Weerts, A.H.: Scaling point-scale (pedo) transfer functions to seamless large-domain parameter estimates for high-resolution distributed hydrologic modeling: An example for the Rhine River. Water Resources Research, 56(4), p.e2019WR026807,2020.

Kimball, J. S., Running, S. W. and Nemani, R. R.: An improved method for estimating surface humidity from daily minimum temperature. Agric. For. Meteor., 85, 87–98, https:// doi.org/10.1016/S0168-1923(96)02366-0, 1997.

Lahmers, T.M., et al.: Evaluation of NOAA national water model parameter calibration in semiarid environments prone to channel infiltration. Journal of Hydrometeorology, 22(11), 2939-2969, 2021.

Li, D., Lettenmaier, D. P., Margulis, S. A. and Andreadis, K.: Theroleofrain-on-snowinflooding over the conterminous United States. Water Resour. Res., 55, 8492–8513, https://doi. org/10.1029/2019WR024950, 2019.

Liang, X., Lettenmaier, D. P. , Wood, E. F. and Burges S. J. : A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415–14428, doi:10.1029/94JD00483, 1994.

Livneh B, Rosenberg, E.A. , Lin, C. , Nijssen, B. , Mishra, V. , Andreadis, K. , Maurer, E.P. and Lettenmaier, D.P.: A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States: Updates and extensions, Journal of Climate, doi:10.1175/JCLI-D-12-00508.1, 2013.

Maidment, D.R.: Conceptual Framework for the National Flood Interoperability Experiment. Journal of the American Water Resources Association 53: 245–57, 2017.

Mascaro, G., Hussein, A., Dugger, A. and Gochis, D.J.: Process-based calibration of WRF-Hydro in a mountainous basin in southwestern US. Journal of the American Water Resources Association, 59(1), 49-70, 2023.

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(2), 762-780, 2015.

Miller, D.A. and White, R.A.: A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. Earth interactions, 2(2), pp.1-26, 1998.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B. , Rakovec, O. and Samaniego, L. :Towards seamless large-domain parameter estimation for hydrologic models. Water Resour. Res., 53, 8020–8040, https://doi.org/10.1002/2017WR020401, 2017.

Naeini, M.R., Analui, B., Gupta, H.V., Duan, Q. and Sorooshian, S.. Three decades of the Shuffled Complex Evolution (SCE-UA) optimization algorithm: Review and applications. Scientia Iranica, 26(4), pp.2015-2031, 2019.

Natural Resources Conservation Service: SNOTEL (Snow Telemetry) Data. USDA. https://www.nrcs.usda.gov/wps/portal/wcc/home/, 2023.

Niu, G.-Y., Yang, Z.-L. , Dickinson, R. E. , Gulden, L. E. and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. J. Geophys. Res., 112, D07103, https://doi.org/10.1029/2006JD007522, 2007.

Niu, G. Y., Yang, Z. L., Dickinson, R. E., & Gulden, L. E.: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. Journal of Geophysical Research: Atmospheres, 110(D21), 2005.

Niu, G.-Y., and Coauthors: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res., 116, D12109, https://doi.org/10.1029/2010JD015139, 2011.

NOAA (National Oceanic and Atmospheric Administration): National Water Model: Improving NOAA's Water Prediction Services, 2016.

Oubeidillah, A. A., Kao, S.-C., Ashfaq, M. , Naz, B. S. and Tootle, G.: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. Hydrology and Earth System Sciences, 18(1), 67–84. https://doi.org/ 10.5194/hess-18-67-2014, 2014.

Prata, A.J.: A new long-wave formula for estimating downward clear-sky radiation at the surface. Quarterly Journal of the Royal Meteorological Society, 122(533), 1127-1151, 1996.

Poissant, D., Arsenault, A. and Brissette, F. : Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. J. Hydrol. Reg. Stud., 12,220–237, https://doi.org/10.1016/j.ejrh.2017.05.005, 2017.

Qi, W.Y., Chen, J. , Li, L. , Xu, C.-Y. , Xiang, Y.-h. , Zhang, S.-B. and Wang, H.-M.: Impact of the number of donor catchments and the efficiency threshold on regionalization performance of hydrological models. J. Hydrol., 601, 126680, https://doi.org/10. 1016/j.jhydrol.2021.126680, 2021.

Raff, D., Brekke, L. , Werner, K. , Wood, A. and White. K.: Short-Term Water Management Decisions: User Needs for Improved Climate, Weather, and Hydrologic Information. U.S. Bureau of Reclamation. https://www.usbr.gov/research/st/roadmaps/WaterSupply.pdf, 2013.

Razavi, T., and Coulibaly, P.: An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. Can. Water Resour. J., 42,2–20, https://doi.org/10.1080/07011784.2016. 1184590, 2017.

Rajsekhar, D., Singh, V.P. and Mishra, A.K., 2015. Hydrologic drought atlas for Texas. Journal of Hydrologic Engineering, 20(7), p.05014023.

Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., & Chen, F.: Simple water balance model for estimating runoff at different spatial and temporal scales. Journal of Geophysical Research, 101(D3), 7461–7475.

https://doi.org/10.1029/95JD02892, 1996.

Schaperow J.R, Li, D., Margulis, S.A., Lettenmaier D.P. :A near-global, high resolution land surface parameter dataset for the variable infiltration capacity model. Scientific Data. Aug 11;8(1):216, 2021.

Schweppe, R., Thober, S., Müller, S., Kelbling, M., Kumar, R., Attinger, S., and Samaniego, L.: MPR 1.0: a stand-alone multiscale parameter regionalization tool for improved parameter estimation of land surface models, Geosci. Model Dev., 15, 859–882, https://doi.org/10.5194/gmd-15-859-2022, 2022.

Sharma, P. and Machiwal, D.: Streamflow forecasting: overview of advances in data-driven techniques. Advances in Streamflow Forecasting,1-50. https://doi.org/10.1016/B978-0-12-820673-7.00013-5, 2021

Shi, X., Wood, A.W. and Lettenmaier, D.P. : How essential is hydrologic model calibration to seasonal streamflow forecasting? Journal of Hydrometeorology, 9(6), 1350-1363, 2008.

Sofokleous, I., Bruggeman, A., Camera, C. and Eliades, M.: Grid-based calibration of the WRF-Hydro with Noah-MP model with improved groundwater and transpiration process equations. Journal of Hydrology, 617, 128991 , 2023

Su, L., Cao, Q. , Xiao, M., Mocko, D. M., Barlage, M. , Li, D. , Peters-Lidard, C. D. and Lettenmaier, D. P.: Drought variability over the conterminous United States for the past century. J. Hydrometeor., 22, 1153–1168, https://doi.org/10.1175/JHM-D-20-0158.1, 2021.

Su, L., Cao, Q. , Shukla, S., Pan, M. and Lettenmaier, D.P.: Evaluation of Subseasonal Drought Forecast Skill over the Coastal Western United States. Journal of Hydrometeorology, 24(4), 709-726, 2023.

Tangdamrongsub, N.: Comparative Analysis of Global Terrestrial Water Storage Simulations: Assessing CABLE, Noah-MP, PCR-GLOBWB, and GLDAS

Performances during the GRACE and GRACE-FO Era. Water, 15(13), p.2456, 2023.

Thornton, P. E., and Running, S. W.: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agric. For. Meteor., 93, 211–228, https://doi.org/10.1016/S0168-1923(98) 00126-9, 1999.

Tolson, B. A., and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res., 43, W01413, https://doi.org/10.1029/2005WR004723, 2007.

Troy, T. J., Wood, E. F. and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling. Water Resources Research, 44, W09411. https://doi.org/10.1029/2007WR006513, 2008

USWRC: Guidelines for determining flood flow frequency. Bulletin 17B of the Hydrology Subcommittee, 183 pp., https:// water.usgs.gov/osw/bulletin17b/dl_flow.pdf, 1982.

Yang, Y., Pan, M., Beck, H.E. , Fisher, C.K., Beighley, R.E. , Kao, S.C. , Hong, Y. and Wood, E.F.: In quest of calibration density and consistency in hydrologic modeling: Distributed parameter calibration against streamflow characteristics. Water Resources Research, 55(9), 7784-7803, 2019.

~~Yang, Z. L., and Dickinson R. E. : Description of the BiosphereAtmosphere Transfer Scheme (BATS) for the soil moisture workshop and evaluation of its performance, Global Planet. Change, 13, 117–134, doi:10.1016/0921-8181(95)00041-0, 1996.~~

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged

Formatted: No underline, Font color: Auto

basins. Advances in Water Resources, 30(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

Zheng, H., Yang, Z.-L. , Lin, P. , Wei, J. , Wu, W.-Y., Li, L. , Zhao, L. and Wang, S.: On the sensitivity of the precipitation partitioning into evapotranspiration and runoff in land surface parameterizations. Water Resour. Res., 55, 95–111, https://doi.org/10.1029/2017WR022236, 2019.

Zink M, Mai J, Cuntz M, Samaniego L. Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. Water Resources Research. 2018 Apr;54(4):2976-98.