https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

A WRF-Hydro-based retrospective simulation of water resources for US integrated water availability assessment

Arezoo Rafieeinasab[1], Amir Mazrooei[1], Thomas Enzminger[1], Ishita Srivastava[1], Aubrey Dugger[1], Dave Gochis[1], Nina Omani[1], Joe Grim[1], Kevin Sampson[1], Yongxin Zhang[1], Jacob LaFontaine[2], Roland Viger[3], Yuqiong Liu[4], and Tim Schneider[1]

[1]NSF NCAR, Research Applications Laboratory (RAL), Boulder, Colorado, USA
[2]U.S. Geological Survey, Water Resources Mission Area, Norcross, Georgia, USA
[3]U.S. Geological Survey, Water Resources Mission Area, Boulder, Colorado, USA
[4]Earth Resources Technology, Inc., Laurel, Maryland, USA

Correspondence: Arezoo Rafieeinasab (arezoo@ucar.edu)

Abstract. A systematic and periodic evaluation of water supply across the United States is critical for gaining comprehensive insights into the present state of the nation's water resources and strategically planning for the future. The U.S. Geological Survey (USGS) Integrated Water Availability Assessments (IWAAs) is a national initiative designed to characterize past, present, and future water availability at selected basins in the United States. The Weather Research and Forecasting model hydrological modeling extension package (WRF-Hydro) is one of the selected hydrologic models used to generate an estimate of national hydrological fluxes and storage across the conterminous United States (CONUS). The WRF-Hydro application is being forced using the state-of-the-art CONUS404 dataset, a regional hydroclimate dataset over the CONUS, and evaluated over water years 2010–2021. Calibration leads to substantial improvements in simulated streamflow across most of the CONUS. Following parameter regionalization, streamflow performance is reasonable at USGS gages, particularly in the eastern and western regions. However, certain challenges arise in the central US, Arizona, and south Florida, where the model exhibits poor performance. The observed shortcomings in these regions can be attributed to a combination of deficiencies within the framework of the model code, its configuration and atmospheric forcing errors, with a specific emphasis on temporal accuracy issues.

Throughout the CONUS, WRF-Hydro IWAAs based simulations of snow water equivalent closely align with the Snow Data Assimilation System (SNODAS) during the snow accumulation season but show low biases during the snow ablation season. WRF–Hydro IWAAs based actual evapotranspiration (ET) simulations generally exhibit close agreement with Global Land Evaporation Amsterdam Model (GLEAM) ET estimates when comparing cumulative distribution functions across CONUS. Despite this overall agreement, simulated WRF-Hydro IWAAs ET is higher in parts of the central US and lower in parts of the northeast, southeast, and northwest regions of the US, and in urban areas when compared to GLEAM. There is a strong agreement between WRF-Hydro IWAAs based simulations and GLEAM surface soil moisture (top 10 cm) values, with the WRF-Hydro IWAAs model simulating some higher estimates particularly over the eastern US. Similarly, simulated WRF-Hydro IWAAs root-zone soil moisture is underestimated in the southeast US while there are positive biases observed in the western US, relative to the GLEAM simulations. These comparisons to independent datasets indicate the WRF-Hydro

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

application developed for the USGS IWAAs is producing reasonable simulations in many locations across CONUS but is over- or underestimating model variables in some regions.

1 Introduction

The U.S. Geological Survey (USGS) Integrated Water Availability Assessments (IWAAs) is a comprehensive national initiative designed to evaluate water availability in the United States (US) on a recurring basis. The inaugural cycle of this national water availability assessment has two primary objectives: firstly, to provide a status assessment of water availability for the period 2010 to 2020 on a national scale, and secondly to conduct a historical trend analysis exploring multi-decadal changes over time for the period 1980 to 2020. Subsequent USGS IWAAs will expand the assessment scope to include projections and undertake more focused regional studies (Miller et al., 2020). Although the IWAAs encompasses various facets of water availability, including quantity, quality, use, and aquatic ecosystems, this paper specifically concentrates on the water quantity aspect within the IWAAs.

The modeling applications used to support the first cycle of the IWAAs are forced by the state-of-the-art, CONUS404 dataset, a regional hydroclimate dataset over the conterminous United States (CONUS) developed through a collaborative initiative between the USGS and the National Center for Atmospheric Research (NCAR) (Rasmussen et al., 2023). The Weather Research and Forecasting (WRF) model hydrological modeling extension package (WRF-Hydro) (Gochis et al., 2020) is one of the hydrological model applications used in the first cycle of the IWAAs. This paper focuses on providing an in-depth account of the WRF-Hydro modeling effort within the IWAAs, specifically delving into the details of the WRF-Hydro model configuration and evaluating its performance.

WRF-Hydro is a community-based modeling framework originally designed to facilitate coupling between the WRF atmospheric model and components of terrestrial hydrologic models (Gochis et al., 2020). In the WRF model, as with most numerical weather models, the land surface is represented by a column land surface model (LSM). Runoff-infiltration partitioning is performed in a one-dimensional column without considering the lateral water movement. In WRF-Hydro, the column LSM is enhanced by accounting for horizontal water movement with overland, shallow subsurface, and channel flow routing options as well as the addition of a conceptual baseflow model. WRF-Hydro has been widely used in research and operations in configurations coupled to the atmosphere (e.g. Yucel et al., 2015; Fredj et al., 2015; Senatore et al., 2015; Arnault et al., 2016; Givati et al., 2016; Kerandi et al., 2018; Naabil et al., 2017; Verri et al., 2017; Varlas et al., 2018) and uncoupled applications (e.g. Xiang et al., 2017; Yin et al., 2022, 2021; Mehboob et al., 2022; Lee et al., 2022; Bao et al., 2022) where the model is forced by reanalysis or observational atmospheric data. Here, the model is driven by the CONUS404 dataset, making it an uncoupled application.

One of the most prominent applications of WRF-Hydro is the National Oceanic and Atmospheric Administration (NOAA) National Water Model (NWM). A particular instance of WRF-Hydro has been running operationally as the NWM since August of 2016 (Cosgrove et al., 2024; Read et al., 2023). Covering the CONUS along with parts of Canada and Mexico, the NWM has significantly enhanced both temporal and spatial simulation resolutions of operational hydrological forecasting across the

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

CONUS. The number of features for which forecasts are generated has increased from approximately 3,700 River Forecast Center prediction locations to over 2.7 million stream reaches derived from the National Hydrography Dataset NHDPlus version 2.1 (McKay et al., 2012). Subsequent updates to the NWM introduced additional model domains covering Hawaii, Puerto Rico, and Alaska. The WRF-Hydro instance employed in this study aligns with the hydrography specifications of the NWM version 3.0.

Initial parameter values used in the application were improved using established hydrologic model calibration practices. Traditional automated calibration methods in hydrology often involve numerous iterations (Duan et al., 1992, 1993; Vrugt and Ter Braak, 2011), rendering calibration computationally demanding, especially for national-scale applications of distributed models like WRF-Hydro. Addressing this challenge, Tolson and Shoemaker (2007) introduced the Dynamically Dimensioned Search (DDS) algorithm, presenting a more cost-effective alternative. The DDS algorithm scales the search strategy in the model parameter space based on the user-specified maximum number of iterations, enhancing computational efficiency, making it an attractive choice for hydrological calibration. This is achieved through scaling (decreasing) the number of perturbed parameters as it approaches the maximum number of iterations. The DDS method is used for calibration of the WRF-Hydro application described in this study.

The subsequent sections of this paper are structured as follows: Section 2 provides a brief overview of the WRF-Hydro model, delineating its physics components and static inputs. Section 3 details the CONUS404 bias-adjusted forcing dataset employed to drive the WRF-Hydro model. The calibration methodology and procedure are elaborated in Section 4, followed by an explanation of the parameter regionalization procedure. Section 5 presents our findings, encompassing the evaluation of various water budget components, such as streamflow, snow, soil moisture, and evapotranspiration. The paper concludes in Section 6, offering a summary of key insights and outlining potential for model enhancements.

2 Model Description

WRF-Hydro comprises a suite of modules that can be selectively activated or deactivated based on the specific requirements of the model application. Figure 1 illustrates the model components utilized in this study, along with their interdependencies. This particular setup of the WRF-Hydro model, denoted as the WRF-Hydro IWAAs configuration hereafter, encompasses components operating at distinct spatial and temporal scales.

WRF-Hydro supports two different land surface models (Noah and Noah-MP, Gochis et al., 2020). The land surface model (LSM) used in the WRF-Hydro IWAAs configuration is Noah-MP (Niu et al., 2011). Noah-MP provides multiple options for key land-atmosphere processes (Niu et al., 2011), such as canopy stomatal resistance and snowpack albedo. For further information regarding each module and available options, the reader is referred to (He et al., 2023). The application of the Noah-MP LSM within the WRF-Hydro IWAAs configuration has a spatial resolution of 1 kilometer (km) and temporal resolution of 1 hour.

After simulating the hourly response of the land using the LSM at a 1-km grid cell resolution, WRF-Hydro carries out terrain routing (Figure 1). The terrain routing transfers both subsurface and surface flow horizontally. In the WRF-Hydro

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

IWAAs configuration, terrain routing happens on a finer mesh at 250-meter resolution. The higher spatial resolution resolves finer details in the model such as local topographic features (e.g., depressions, floodplains, and riparian corridors) that are not readily resolved at the 1-km resolution of the LSM application. State variables (soil moisture, infiltration excess water) are disaggregated from the land surface model resolution (1 km) to the terrain routing resolution (250 m). Next, subsurface flow routing is executed to allow exfiltration from fully saturated soil columns. Finally, when the depth of ponded water within a grid cell surpasses a predetermined retention threshold, the overland flow routing is executed through the utilization of a fully unsteady, spatially explicit, diffusive wave formulation (refer to Gochis et al. (2020) for detailed information).

Similar to the NWM, the IWAAs configuration of WRF-Hydro also utilizes the geospatial framework provided by the USGS National Hydrography Dataset (NHD) Plus Version 2.1 (NHDPlusV2.1) medium-resolution dataset (McKay et al., 2012), which includes both streams and corresponding catchments. A custom extension of this dataset with similar resolution was developed to extend coverage into contributing Canadian drainage basins in the Great Lakes region (Mason et al., 2019) as well as contributing areas of Mexico. In the WRF-Hydro IWAAs configuration, channel routing is performed on the NHDPlusV2.1 streams while the conceptual baseflow routing is performed on the NHDPlusV2.1 catchments. Each NHDPlusV2.1 catchment has a single groundwater reservoir with a conceptual volumetric capacity.

The inflow to each groundwater reservoir is aggregated from the lower boundary of the soil column (1-km regular mesh) to the NHDPlusV2.1 catchments. Then, the conceptual groundwater component is operated as a non-linear reservoir with spill behavior where outflow is estimated using an exponential storage-discharge function. Should the reservoir fill, excess water is added to the groundwater reservoir outflow. The outflow is then combined with lateral channel inflows from the overland flow component and directly input to the NHDPlus stream (associated with the NHDPlus catchment). The in-channel water volumes are then routed through the reaches of the channel network using the Muskingum-Cunge routing method (Read et al., 2023). WRF-Hydro also includes options to represent lakes and reservoirs (i.e., waterbodies). However, waterbodies and water use are being represented in the IWAAs as a post-process, so the hydrologic models are estimating "natural" stream and waterbody inflows only.

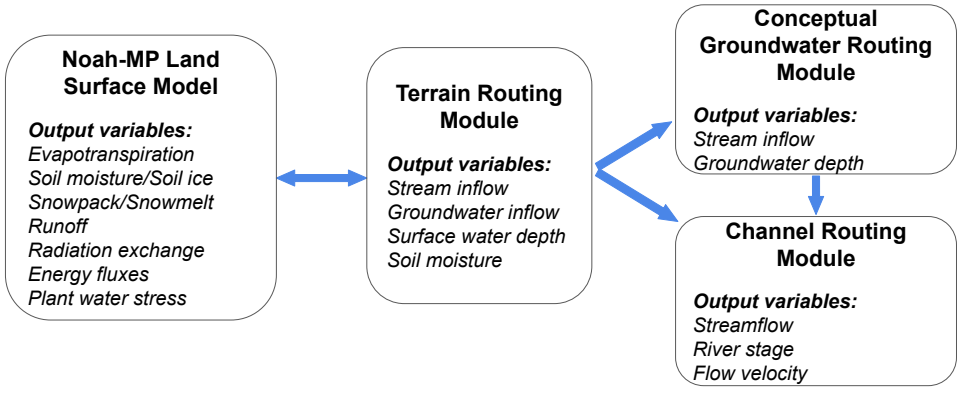


Figure 1. Existing model physics in WRF-Hydro used in IWAAs study. The arrows show the direction of information passing between the modules.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

3 Atmospheric Forcing

Errors in simulated hydrologic components such as streamflow are aggregated errors emerging from errors in initial states, deficiencies in model structure, and atmospheric forcing. Errors in the forcing dataset nonlinearly contribute to streamflow errors (Rafieeinasab et al., 2015) and, therefore, it is of great importance to choose the right forcing dataset for the application at hand. Ideally, one would like to force (and calibrate) the model using a dataset with an appropriate temporal and spatial resolution, a long-term data record, and physically consistent variables. In this study, the CONUS404 dataset (Rasmussen et al., 2023) with added air temperature and precipitation bias adjustment is used to force (and calibrate) the WRF-Hydro IWAAs model application.

CONUS404 is a mesoscale hydroclimate dataset over the CONUS produced under a collaborative initiative between USGS and NCAR. This publicly available dataset is at a 4-km horizontal spatial resolution with hourly temporal resolution, which are suitable resolutions for the IWAAs hydrologic modeling applications. The dataset is available from October 1979 to October 2022 (43 years) at the time of writing this paper, although only the water years of 2010–2021 were used in the IWAAs configuration of WRF-Hydro (Rafieeinasab et al., 2024). CONUS404 is produced using the WRF atmospheric model to dynamically downscale the fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) dataset (Hersbach et al., 2020). Through better representation of fine-scale weather phenomena, such as mesoscale convective systems and orographic precipitation, CONUS404 is able to produce a relatively accurate distribution of rainfall and temperature over a large area and a long period. The CONUS404 dataset provides an opportunity to study water-budget components at a relatively high spatial and temporal scale, which is of importance to hydroclimate studies.

The initial examination of CONUS404 data revealed notable regional biases in both its precipitation and temperature illustrated in Figure 2. These biases have the potential to introduce inaccuracies in the WRF-Hydro IWAAs modeling and calibration processes, likely resulting in calibrations with parameter compensation to address precipitation errors. In alignment with similar studies (e.g., Robertson et al., 2023; Grim et al., 2023), we implemented a bias correction approach to mitigate the most pronounced biases observed. While this adjustment does compromise the inherent consistency of CONUS404 variables to some extent, it prioritizes the necessity for the hydrological model to be calibrated using the most realistic hydrologic forcings, especially precipitation.

In order to perform bias correction to the CONUS404 dataset, we calculated the biases of two separate fields over the CONUS404 domain: 2-m air temperature and precipitation. For the observational dataset, several observation and observational analysis datasets (Automated Surface Observing Systems (ASOS, 1998), SNOpack TELemetry (SNOTEL, Serreze et al., 1999), Daymet (Thornton et al., 2016), and Parameter-elevation Regression on Independent Slopes Model (PRISM, Daly et al., 1994)) were compared with CONUS404. Of these, the Daymet dataset was selected as the datum from which to calculate bias in both temperature and precipitation because: 1) it is a spatially-continuous dataset, unlike ASOS and SNOTEL that only have values where there are stations; 2) its coverage seamlessly extends beyond the CONUS into Canada and Mexico, covering all US tributary water basins of interest; and 3) its accuracy is very similar to PRISM but does not have international boundary artifacts like PRISM.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

We calculated day-of-year bias for both fields (daily total precipitation and daily mean temperature) for every pixel in the CONUS404 domain over land using the time period Jan. 1, 1980–Dec. 31, 2017. This time frame was used since it covered the concurrent dates from the Daymet and CONUS404 datasets. The bias correction of CONUS404 was performed on the full dataset, and not just the time period used in the analysis of this manuscript. The bias was calculated as a percentage for precipitation and an absolute value offset (Kelvin or °C can be used interchangeably for bias corrections) for temperature. Figure 2 below shows the mean daily temperature and precipitation bias for all days and years of the period of analysis for each pixel in the modeling domain. The day-of-year biases were then calculated at every pixel by averaging the biases for each day of the year from the 38-year data set, applying a 31-day smoothing to both precipitation and temperature biases to remove anomalous day-to-day fluctuations in calculated biases. Figure 3 shows the domain-averaged temperature and precipitation biases. The day-of-year precipitation biases were corrected at every pixel for every day of the year by multiplying the CONUS404 precipitation amounts by the inverse of day-of-year precipitation bias ratio at each pixel; for example, a pixel with a +27% precipitation bias for that day of year would be multiplied by 1/1.27 to remove the bias. The day-of-year temperature biases were corrected at every pixel and each day of the year by subtracting the unique day-of-year biases for each pixel from the CONUS404 temperature for altitudes less than 2,000 m. Locations above 2,500 m elevation were not corrected because the Daymet values were influenced by biased SNOTEL stations (Oyler et al., 2015), and pixels between 2,000– 2,500 m elevation had their biases linearly corrected between no correction (2,500 m) to full correction (2,000 m), depending on their altitude (see Figure 4 for how SNOTEL temperature biases above 2,500 m also skew the biases for the Daymet and PRISM). The version of the CONUS404 dataset that includes bias adjustments to air temperature and precipitation is referred to as CONUS404BA (Zhang et al., 2024).

4 Model Calibration and Regionalization

Conducting regional calibration for distributed models like WRF-Hydro is computationally expensive. One strategy to minimize this cost is to calibrate a select subset of basins, prioritizing representative basins or those of particular importance to stakeholders, then extrapolating parameters to non-calibrated locations through a parameter regionalization process. We employ this strategy in calibrating the WRF-Hydro IWAAs configuration, choosing to directly calibrate GAGES-II reference basins (Falcone, 2017) possessing sufficient observational data. GAGES-II reference basins are selected for direct calibration because they have minimal human impacts and are generally considered mostly natural flow basins, consistent with the WRF-Hydro IWAAs configuration's exclusion of reservoirs, diversions, and other management. To address gaps in the spatial coverage of the GAGES-II reference basins, several lightly impacted USGS basins (limited to only basins with a hydrologic disturbance index < 21 and further filtered based on comments and reports), select gages managed by the California Department of Water Resources (CADWR), and several basins in Canada are also integrated into the calibration process. Only basins with an area smaller than 10,000 km^2 undergo calibration, with exceptions made for data-sparse regions to ensure comprehensive coverage. Additionally, calibration basins are limited to those with at least 50 percent data availability of hourly streamflow measurements during the calibration period (October 2016 to October 2021). This lenient threshold purposely allows inclusion of basins with

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

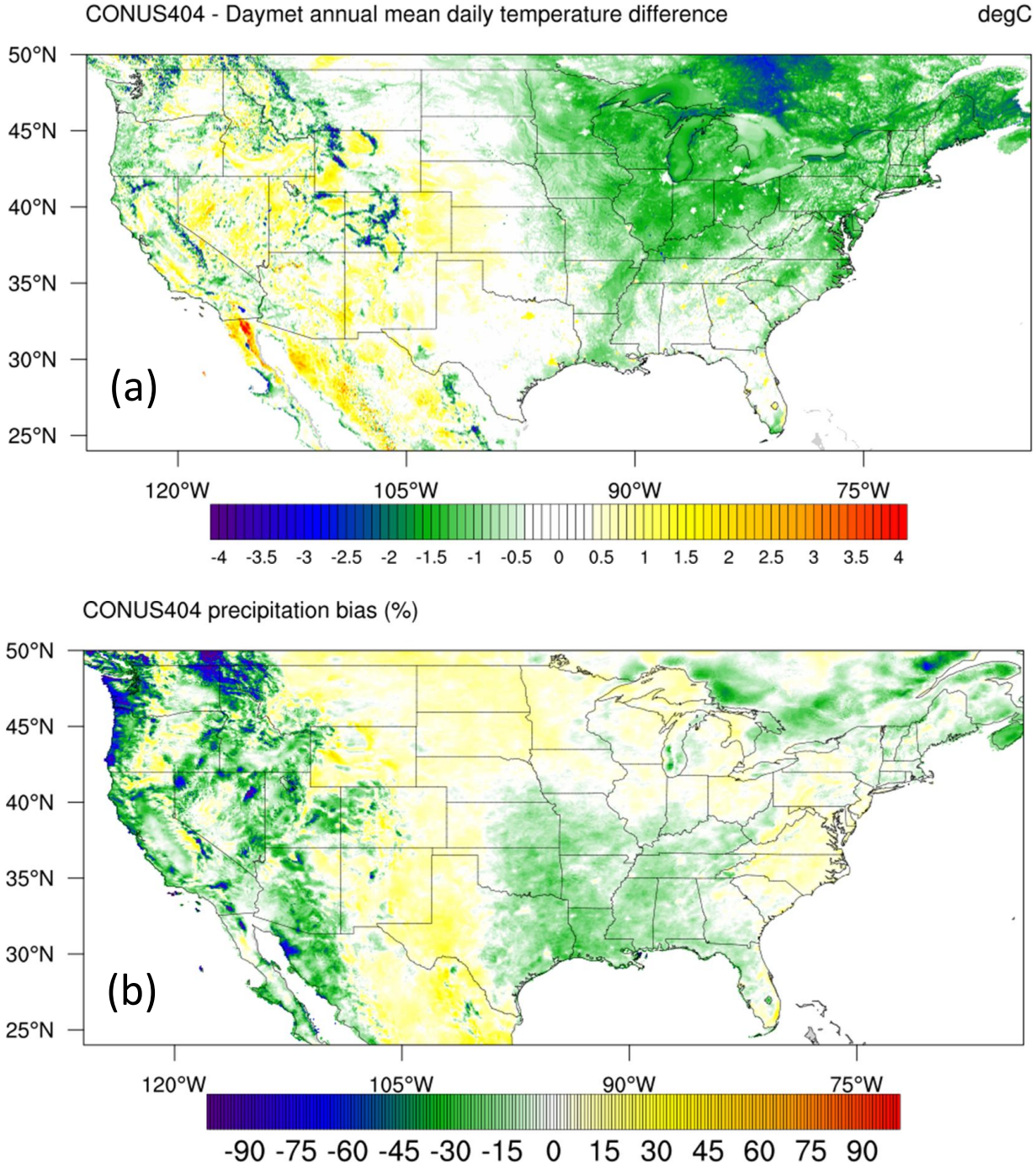


Figure 2. Annual mean daily (a) temperature difference (K) and (b) precipitation bias (%) over the modeling domain.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

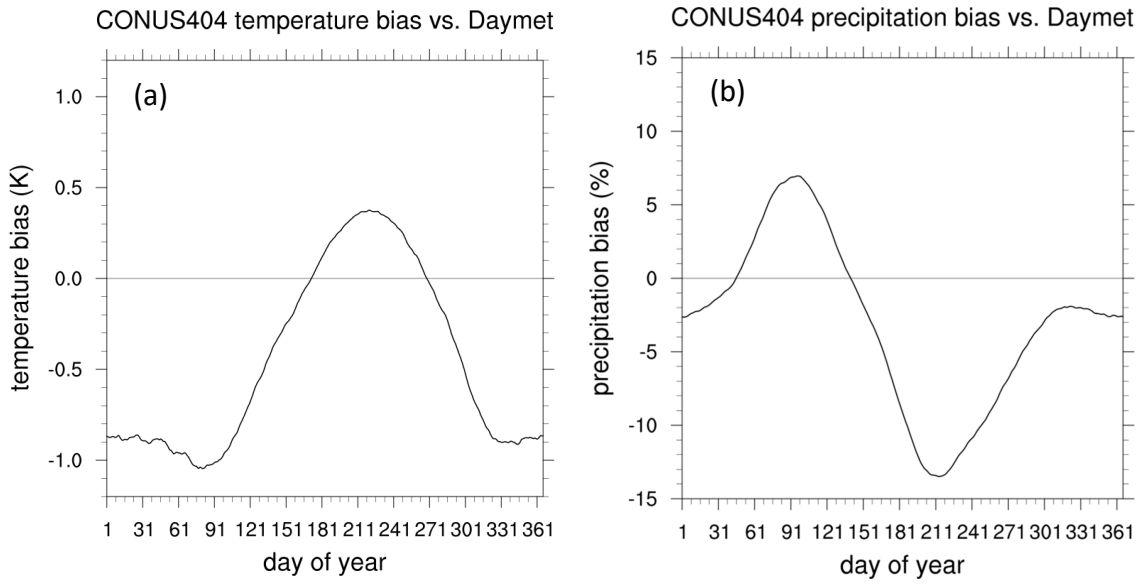


Figure 3. Annual cycle of domain-averaged (a) temperature bias and (b) precipitation bias.

only seasonal gages in the calibration process. In cases where gages fall short of the 50 percent data availability requirement for hourly streamflow over the calibration period, we check daily streamflow data availability, and if exceeding the 50 percent threshold, daily data are employed for calibration.

Overall, 1,522 basins are calibrated, including 1,470 USGS gages across CONUS (958 GAGESII reference basins), 27 gages in Canada, and 25 CADWR gages (Figure 5). For 10 of the USGS gages where natural flow time series were available, calibration is performed using the natural flow time series. The natural flow time series are downloaded from data publicly provided by the Bureau of Reclamation (https://www.usbr.gov/pn/hydromet/).

Calibration parameters: Hydrologic model parameters generally describe the properties of and relationships between various model components and are often presumed to be time invariant. Certain parameters, especially in conceptual models, lack direct physical interpretations, and even in semi- or fully-distributed physical models, some parameters are challenging to observe directly (particularly at the scales they are represented in the models). Consequently, model calibration, an iterative process aligning observed and modeled watershed behavior, is widely employed to deduce and fine-tune these parameters.

In total, 17 WRF-Hydro model parameters are calibrated for the IWAAs configuration, as shown in Table 1. The choice of these parameters is informed by a combination of pertinent scientific literature, expert opinion, and previous simulations and sensitivity analyses conducted on a subset of the selected calibration basins. A short description of each model parameter is provided in Table 1. The snow depletion curve melt factor (MFSNO) parameter is only calibrated for snow-dominated basins. Several of these parameters are a function of vegetation type or soil type and, therefore, we apply and calibrate a multiplier on these parameters to preserve the original spatial patterns. The remainder of the parameter values are calibrated directly.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

Table 1. Calibrated model parameters (x in the values denotes that the calibration parameter is a multiplier)

Name	Description	Units	Multiplier	Default value	Value Range: Min	Value Range: Max
SOIL PARAMETERS						
BEXP	Pore size distribution index in Brooks-Corey equation (Brooks and Corey, 1964).	dimensionless	Yes	x1	x0.4	x1.9
SMCMAX	Saturation soil moisture content (i.e., porosity)	volumetric fraction	Yes	x1	x0.8	x1.2
DKSAT	Saturated hydraulic conductivity	m/s	Yes	x1	x0.2	x10
RSURFEXP	Exponent in the resistance equation for soil evaporation	dimensionless	No	5	1	6
RUNOFF PARAMETERS						
AXAJ	Tension water distribution inflection parameter	unitless	Yes	x1	x-29	x29
BXAJ	Tension water distribution shape parameter	unitless	Yes	x1	x0.01	x10
XXAJ	Free water distribution shape parameter	unitless	Yes	x1	x0.01	x10
SLOPE	Linear scaling of "openness" of bottom drainage boundary	0-1	No	0.3	0	1
RETDEPRTFAC	Multiplier on retention depth limit	unitless	No	1	0.1	20,000
LKSATFAC	Multiplier on lateral hydraulic conductivity (controls anisotropy between vertical and lateral conductivity)	unitless	No	1,000	10	10,000
NEXP	Exponent in the decay function for lateral Ksat over depth	unitless	No	1	0.1	15
BASEFLOW PARAMETERS						
ZMAX	Maximum groundwater bucket depth	mm	No	50	10	250
EXPON	Exponent controlling rate of bucket drainage as a function of depth	dimensionless	No	3	1	8
VEGETATION PARAMETERS						
CWPVT	Canopy wind parameter for canopy wind profile formulation	1/m	Yes	x1	x0.5	x2
VCMX25	Maximum carboxylation at 25C	umol/m2/s	Yes	x1	x0.6	x1.4
MP	Slope of Ball-Berry conductance relationship	unitless	Yes	x1	x0.6	x1.4
SNOW PARAMETERS						
MFSNO	Melt factor for snow depletion curve	dimensionless	Yes	x1	x0.25	x2

.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

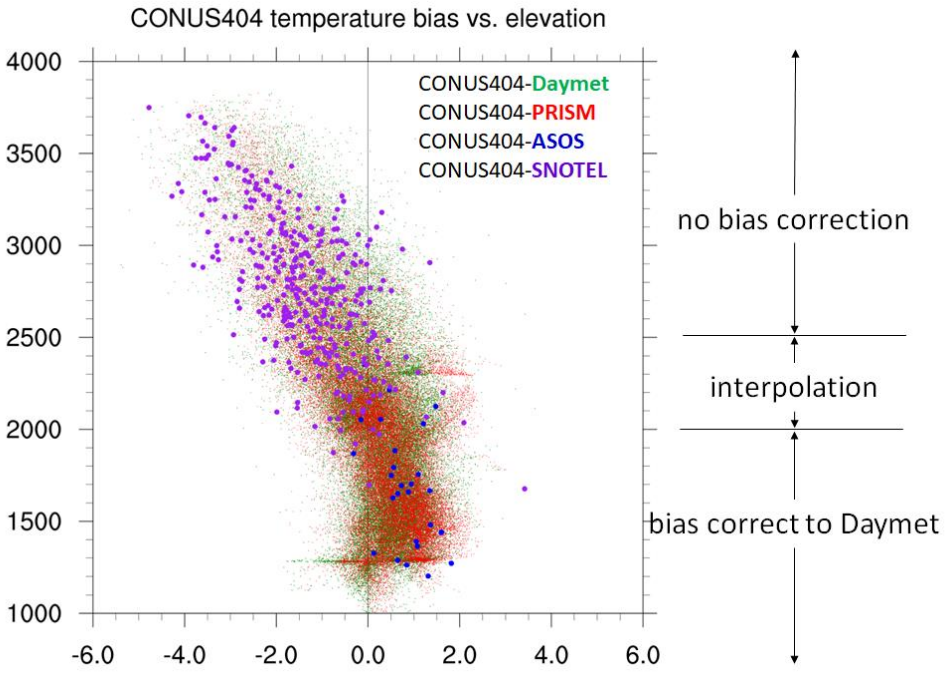


Figure 4. Temperature bias by elevation for four different products over the central Rockies.

Calibration methodology: The core optimization algorithm used is the Dynamically Dimensioned Search (DDS) algorithm introduced by Tolson and Shoemaker (2007). DDS was selected because, per Tolson and Shoemaker (2007), it is a computationally efficient optimization algorithm that provides comparable performance to algorithms like shuffled complex evolution (SCE, Duan et al., 1992, 1993). Additionally, the DDS algorithm is very simple, making it easy to code in any language.

In initial iterations the DDS algorithm searches globally and, as the procedure approaches the maximum user-defined number of iterations, the search transitions from a global to a local search. This transition from global to local search is achieved by dynamically and probabilistically reducing the search dimension, which is the subset of the calibration parameters that will be updated in a given iteration. Parameters selected in each iteration are perturbed within the defined parameter range. For the WRF-Hydro IWAAs calibration, the lower and upper limits for each parameter are given in Table 1. The limits are selected based on literature review and expert opinion. The number of iterations is set to 400 except for large domains (> 5,000 km^2), where only 200 iterations are used for computational tractability.

The optimization procedure exclusively employs streamflow observations, with the (minimized) calibration objective function defined as 1 minus the Kling-Gupta efficiency (KGE) of hourly streamflow, where KGE is as proposed by Gupta et al. (2009). KGE for daily streamflow is applied in instances where there are insufficient hourly flow measurements. While a multivariate calibration is preferable, constraints imposed by project timelines led us to focus solely on streamflow calibration.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

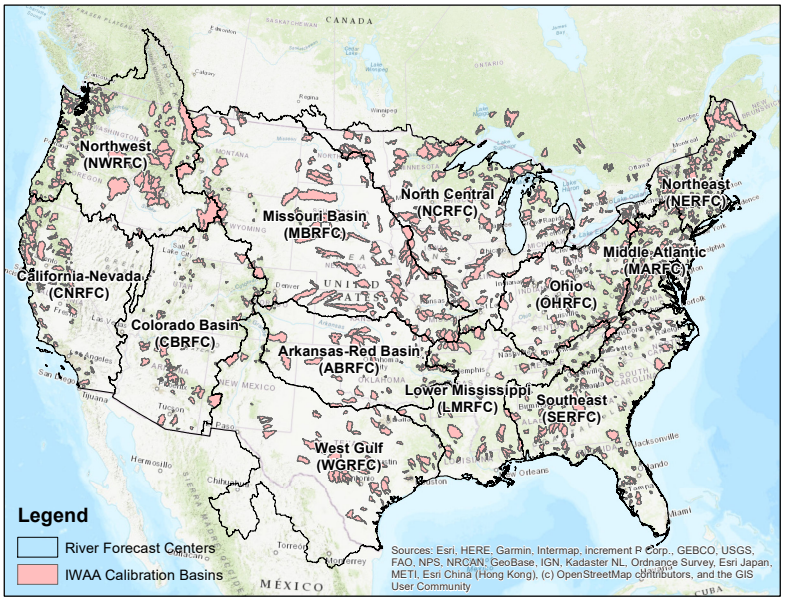


Figure 5. Spatial location of the calibration basins and their extent. The River Forecast Center (RFC) boundaries defined by National Weather Service (NWS) are also shown and labeled with the name of the RFC as it will be used frequently in the results and discussions. RFC boundaries were obtained from https://www.weather.gov/gis/RFCBounds

.

Before initiating the calibration process, a model run for each basin from October 2010 to October 2021 was spun up using default parameters. The default parameters are the initial guesses for each parameter distributed with the model code. Subsequently, the "warm" model states from October 2021 serve as initial conditions for the calibration model runs, commencing from October 2012. While it is recognized that conditions in 2021 may differ from those in 2012, we assume that the seasonality and regional climate are similar.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

In addition to the single spin-up run with the default parameter, each calibration cycle incorporates a distinct 1-year acclimation period (from October 2012 to October 2013) with updated model parameters. This is to mitigate instabilities that could arise from the parameter change. The calibration phase spans a total of five water years (from October 2013 to October 2018). This duration provides a compromise between available computational resources and a desire to calibrate the model to a range of hydroclimate conditions. Following evaluation of the objective function, the DDS algorithm tracks the "best" parameter set and identifies the next testable parameter set (Tolson and Shoemaker, 2007).

Concluding the calibration process, we implement the "best" parameter set identified by the DDS algorithm and conduct an additional model simulation spanning from October 2010 to October 2021. Disregarding the initial year as acclimation period, we calculate metrics for a validation period that includes 2 years preceding the calibration interval (October 2011 to October 2013) and 3 years succeeding the calibration period (October 2018 to October 2021). The error metrics of simulated streamflow for both calibration and validation periods are reported in the results section.

Regionalization Methodology: All the calibration basins collectively constitute less than 10% of the total land area within the model domain. To successfully execute the model application with spatially varying parameters across the CONUS, it is imperative to assign appropriate parameters to each grid cell within the model domain. For the WRF-Hydro IWAAs configuration, this task is accomplished through a parameter regionalization approach. The attributes of the cells in each calibration basin are summarized and compared to summaries of attributes of all (non-calibrated) cells in each USGS 10-digit hydrologic unit code (HUC10) of the Watershed Boundary Dataset (Jones et al., 2022). For each HUC10, the parameters from the calibration basin with the most similar characteristics are assigned to the cells within the HUC10. Similarity is determined using the Gower's distance metric (Gower, 1971) to compare relative positions of calibration basins and HUC10s in an attribute-based information space. Figure 6 provides an illustrative diagram of the regionalization procedure and its associated steps, which are further explained below.

In order to compute the Gower's distance metric to measure the physical similarity between donor and receiver basins, a number of attributes characterizing a basin's physio-climatic aspects are identified. In this study, we leverage two distinct basin attribute datasets to delineate physical similarity, undertaking two separate regionalization analyses. The initial approach mirrors the methodology employed in National Water Model v2.1 (Cosgrove et al., 2024; Liu et al., 2021), utilizing a limited set of basin attributes based on the Hydrological Landscape Region (HLR) framework (Winter, 2001; Wolock et al., 2004; Liu et al., 2008). The second methodology, as introduced in National Water Model v3.0 (Liu et al., 2021), uses a greater number of basin attributes based on the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) dataset (Addor et al., 2017). The main difference between the HLR and CAMELS frameworks is the use of a hydrologic signature category, which includes attributes like mean flow, runoff ratio, baseflow index, and flow quantiles. For more information on the attributes, please refer to Addor et al. (2017). In order to apply the Gower's distance metric properly, a principal component analysis (PCA) was conducted on the basin attribute data to remove potential correlation among them. Prior to applying the PCA, the raw basin attributes values are scaled to [-1,1] by subtracting the respective mean and then dividing by the respective standard deviation. The scores of these principal components and the percentages of the total variance explained by individual principal components were then used to calculate the Gower's distance metric.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

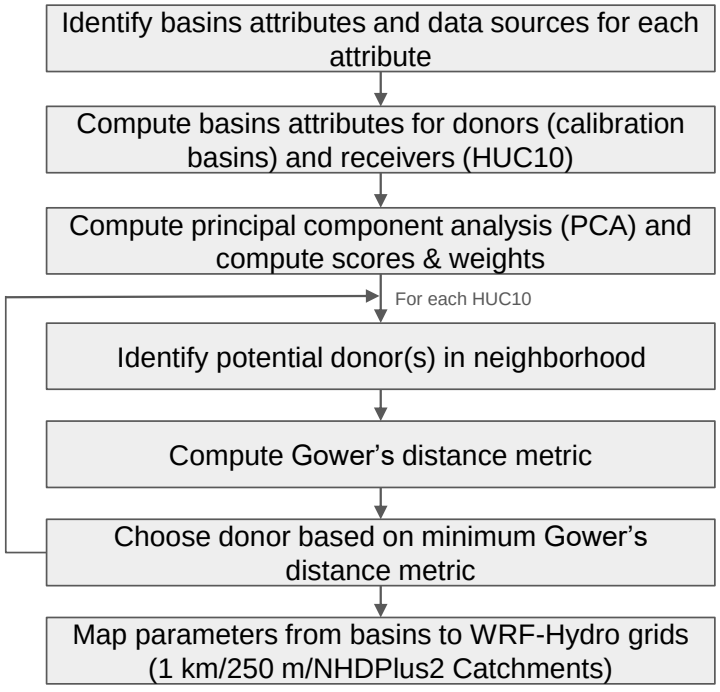


Figure 6. Workflow for regionalization

.

Donors are identified as calibration basin(s) that have a Gower's distance metric smaller than the predefined minimum distance threshold. If multiple donors are selected, then the median of the parameters are used. Once parameter values are transferred from the donor calibration basins to HUC10s, the final step of parameter regionalization consists of mapping the model parameters (for soil, vegetation, runoff, and baseflow) to the underlying grid cells used in the WRF-Hydro application. In addition, the optimal parameter values of the calibration basins are applied to the parameter grids. This last step ensures that, in the calibrated areas, optimal parameters from the calibration, rather than the regionalized parameters, are used.

Finally, since neither the HLR- or CAMELS-based regionalization approach exhibits universal superiority across all spatial contexts, we optimize the performance on a national scale across the CONUS by employing a mix-and-match strategy to select the better-performing approach (HLR or CAMELS). To do this, USGS 8-digit hydrologic unit codes (HUC8) are chosen as the spatial unit. For each HUC8 basin, we select the regionalization scenario that yields the best KGE calculated based on daily streamflow across the HUC8.

Verification metrics: During the calibration process a common set of metrics (Nash-Sutcliffe model efficiency (NSE, Nash and Sutcliffe, 1970), Log NSE, root mean square error (RMSE), KGE, percent bias, correlation coefficient, etc.) are calculated at each iteration using the hydroGOF library (Mauricio, 2017) in R (R Core Team, 2023) with the exception of percent bias which is calculated as follows:

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

$$PBIAS = \frac{\sum(Q_{sim} - Q_{obs})}{\sum Q_{obs}} * 100 \quad (1)$$

where Q_{sim} and Q_{obs} indicates the simulated and observed flow from the model, respectively. Here, we only present values of KGE, percent bias, correlation coefficient, and NSE for brevity. Percent bias is measuring the average tendency of the model simulation to overestimate or underestimate the observation with 100% being the optimal value. Pearson correlation coefficient ranges between -1 to 1, with 1 being the optimal value. Pearson correlation coefficient indicated the degree of collinearity between model simulations and observations. NSE ranges between -Inf to 1, with value of 1 being the optimal value. An efficiency less than zero (NSE < 0) means the observed mean is a better predictor than the model indicating unacceptable performance. In general, model simulation can be judged as satisfactory if NSE > 0.50 (Moriasi et al., 2007). KGE is ranging between -Inf to 1, with 1 being the perfect model simulations. Simulations with KGE values greater than -0.41 are performing better than the climatology (Knoben et al., 2019).

5 Result and Discussion

5.1 Evaluation of Calibration Basins

Upon the completion of 400 iterations, parameters from the optimal iteration, as identified by the objective function, are employed for a conclusive simulation spanning from October 2010 to October 2021. Disregarding the initial year as a spin-up period, the simulation incorporates five years from October 2013 to October 2018 as the calibration period, while an additional five years from October 2011 to October 2013 and October 2018 to October 2021 serve as an independent validation period. Metrics are computed over both the calibration and validation periods. As previously mentioned, the calibration is executed at an hourly time step utilizing hourly streamflow observations, except for a limited number of gages where hourly streamflow data are either unavailable or fail to meet the 50% completeness criteria. In these cases, daily observations are employed.

Figure 7 illustrates spatial maps of KGE and percent bias for hourly streamflow (daily in instances where hourly data are unavailable) in model simulations before (using default parameters) and after calibration (with the best-calibrated parameters) across both the calibration and validation periods for all 1,522 calibration basins in the CONUS. A visual comparison indicates substantial improvements across most of the US due to calibration, underscoring its meaningful benefits. In the default model simulations, a negative bias is evident in the Great Lakes region, Ohio Valley, and Mississippi Valley. However, this negative bias is notably mitigated during both the calibration and validation periods through the calibration process. Likewise, large positive biases observed in the southeast, central, and southwest US with default model parameters are also ameliorated during calibration. Figure 8 presents a summary of error metrics in the form of cumulative density functions, highlighting the considerable impact of calibration in reducing biases and enhancing other metrics such as KGE, NSE, and correlation coefficients for both the calibration and validation periods. While there is a general improvement in the validation period, the magnitude of improvements is less than the calibration period as expected.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

It is essential to highlight that in the calibration period, an enhancement in KGE is expected, given that it serves as the objective function for the calibration procedure. Nonetheless, it is noteworthy that approximately 150 sites across the US experienced a degradation in model performance during the validation period. This could potentially stem from overfitting during the calibration period or differences in climate regimes between the calibration and validation periods.

Ideally, the division of periods for calibration and validation should encompass the entire spectrum of climate conditions (normal, dry, wet, extremely dry, extremely wet) for each basin. Achieving such comprehensive coverage in both the calibration and validation periods is crucial for enhancing the model's predictive power, as emphasized by Zheng et al. (2018). However, given the scale and complexity of this study, a uniform data splitting strategy was employed, as incorporating a basin-specific approach for accommodating such variability is challenging over this large number of basins.

5.2 Regionalized Streamflow Verification

A subset of calibration basins exhibited improvement during the calibration phase, but their performance declined in the validation period. To address this performance inconsistency, we made the decision to refrain from using calibration parameters for these basins and instead rely on the regionalization procedure to assign parameters for these basins. In total, 64 basins were dropped, as their calibrated parameters were deemed unreliable due to poor model performance during the validation period.

Attributes of all donor basins and all uncalibrated HUC10 areas are characterized using two different basin attributes (namely HLR and CAMEL), to establish two sets of similarity measures and parameter mappings to the WRF-Hydro grid cells within each HUC10 area. Once this is complete, WRF-Hydro is run over the full CONUS domain model simulations for both regionalization scenarios, with the goal of choosing the best mapping on a per-HUC8 basis for the final model application configuration. Water year 2018 serves as the spin-up period for the model simulations, while water years 2019 to 2021 are utilized for evaluating the regionalization performance. A water year is defined as the 12-month period October 1, for any given year through September 30, of the following year and is designated by the calendar year in which it ends. Streamflow error metrics show comparable performance for both regionalization scenarios, with CAMEL exhibiting a slight superiority over HLR. To optimize performance across the CONUS, we employ a strategic mix-and-match approach. For each HUC8 basin, we select the regionalization scenario that yields the best KGE for daily streamflow.

Following the implementation of the mix-and-match approach and the establishment of the final configuration of the IWAAs WRF-Hydro CONUS model application, we conduct model simulations spanning the period from October 2009 to October 2021, encompassing the entire 10-year timeframe of the IWAAs program. In Figure 9, we present daily streamflow error metrics, including KGE, percent bias, and correlation coefficient, derived from the conclusive WRF-Hydro IWAAs model run across all USGS gages nationwide.

It is crucial to acknowledge that the model application does not account for human interventions to unimpeded channel flow. Consequently, suboptimal model performance is anticipated in regions with extensive stream regulation, such as large rivers where flows are heavily managed for water supply or hydropower. Despite this limitation, we opted to include regulated gages for comprehensive reporting and analysis.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

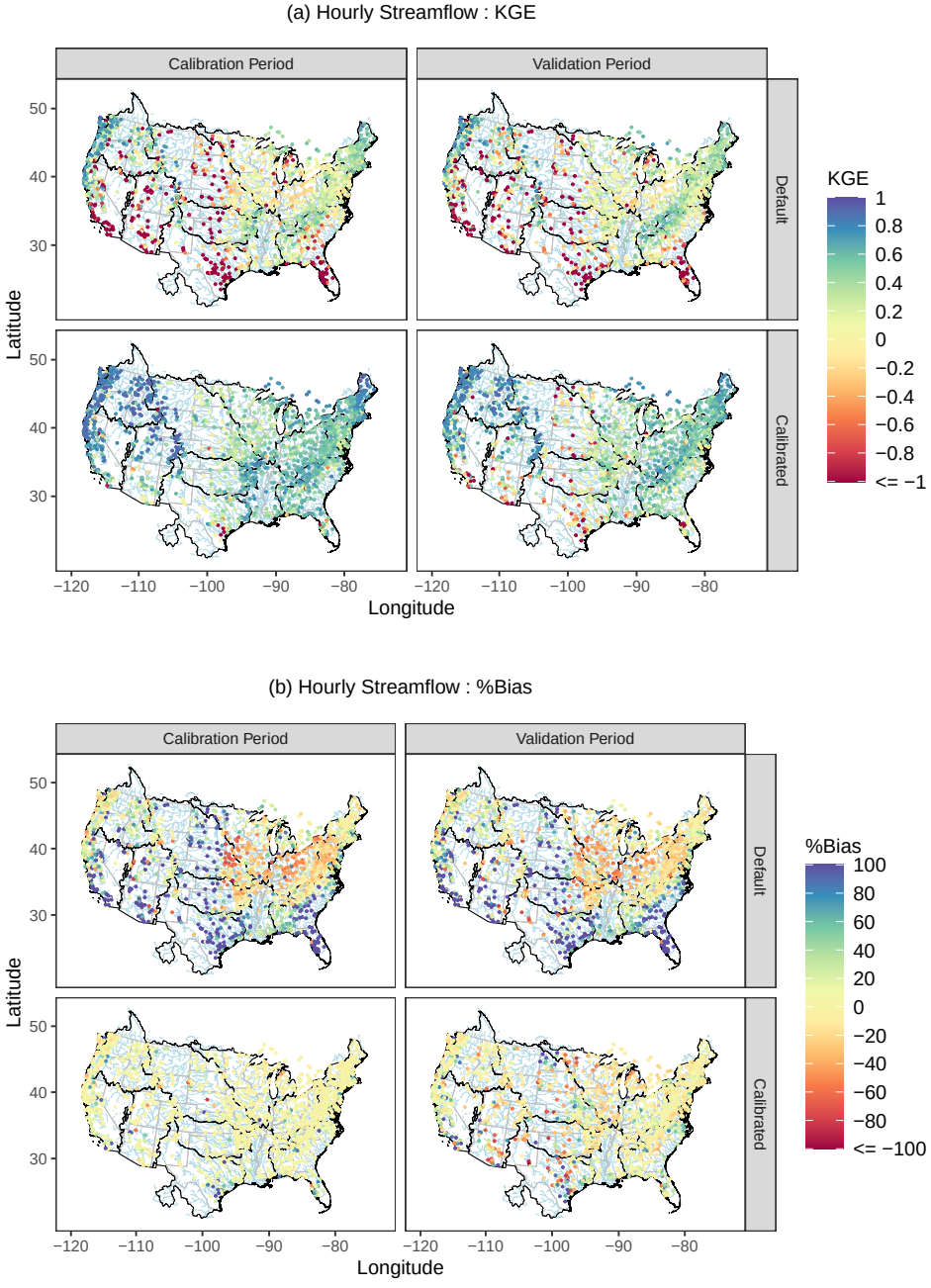
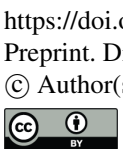


Figure 7. Spatial maps of (a) hourly streamflow KGE and (b) percent bias (%Bias) at 1,522 calibration gages over the calibration (left panels) and validation (right panels) periods for default (upper panels) and calibrated (lower panels) parameters.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

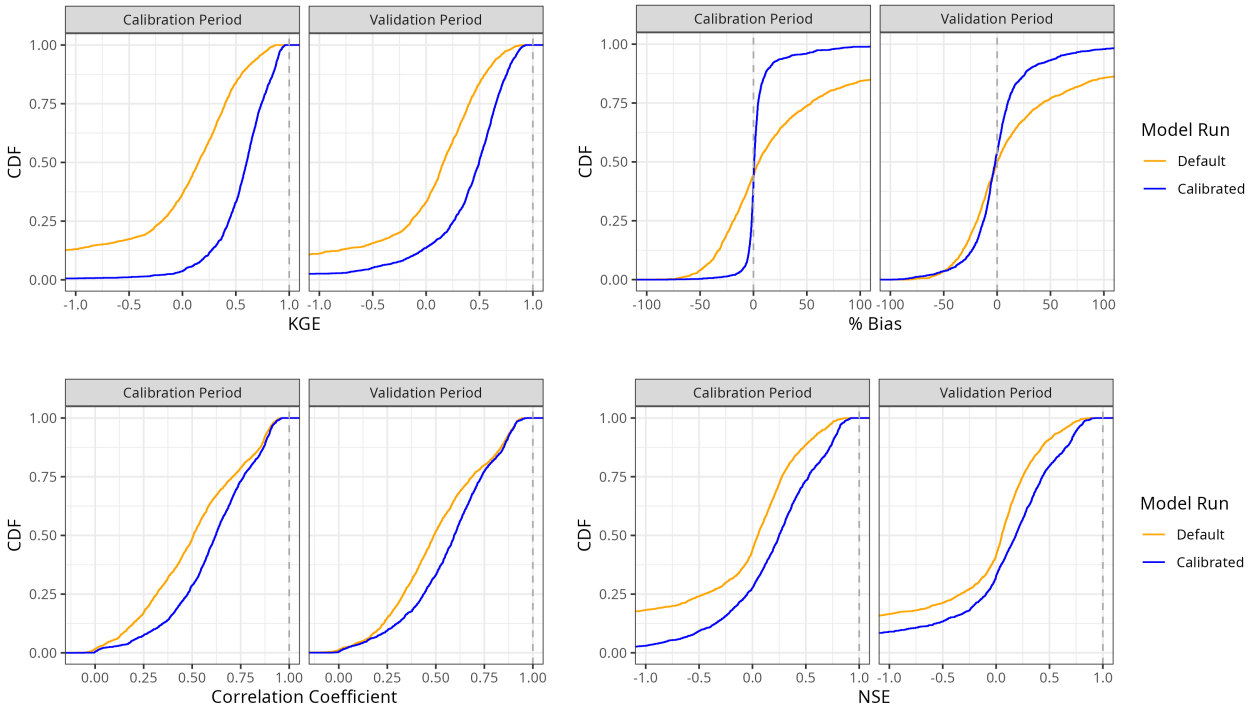


Figure 8. Cumulative density function of hourly streamflow metrics (KGE, %Bias, correlation coefficient, NSE) across the 1,522 calibration gages; dashed line shows the optimal value of each error metric.

As illustrated in Figure 9, streamflow exhibits generally low biases across the majority of gages, particularly in the eastern US. Florida emerges as a region with conspicuous high biases (indicated by the bluish color) in numerous basins. While calibration substantially reduced high biases in this region, this improvement has not been uniformly propagated across the entire area during the regionalization process. Consequently, streamflow is overestimated in numerous basins that were not subject to calibration.

The Central Plains region (defined here as the area of CONUS between -105 and -95 degrees roughly) is characterized by diverse model behavior. Initial high biases in the model were mitigated through the calibration process over the designated calibration period. However, during the subsequent validation period, model response was mixed, with some gages reporting positive biases and others reporting negative biases, as depicted in Figure 7 (b). This mixed behavior extends into the regionalized basins, where a spectrum of gages displays either high biases (indicated by the blue color) or low biases (indicated by the red color). Figure 9 underscores this region's challenges further, revealing the lowest correlation coefficients in simulated streamflow in Missouri Basin (MBRFC), Arkansas-Red Basin (ABRFC), and West Gulf (WGRFC) River Forecast Centers (RFC). This, coupled with the observed poor model biases, contributes to lower KGE values. Complexities in these areas, including regional groundwater systems, surface-water and groundwater disconnections, intensive management and diversions, and potential atmospheric forcing errors, likely contribute to the observed challenges in model performance.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

The relatively low correlation coefficients of the WRF-Hydro IWAAs simulation compared to the NWM (Cosgrove et al., 2024) is likely due to the use of CONUS404 which is a model-based dataset. NWM retrospective analysis uses Analysis of Record for Calibration (AORC) dataset (Fall et al., 2023), an observation-based precipitation dataset. Employing CONUS404, however, offers additional advantages. Notably, modeled precipitation often outperforms observation-based products in mountainous regions (Lundquist et al., 2019), thereby enhancing hydrological model accuracy in these areas. Additionally, CONUS404 boasts capability to generate future climate scenarios, a feat not attainable with observation-based atmospheric forcing. Hence, while acknowledging the trade-offs, the utilization of CONUS404 offers several advantages over other available products.

Figure 10 encapsulates comparable information through the use of boxplots, delineating key performance metrics for each RFC. The metrics, including KGE, percent bias, correlation coefficient, and NSE, are presented, specifically categorized into GAGES-II reference basins (with minimal human impacts) and non-reference basins (with more significant human impacts). Notably, across all metrics and for every RFC, the model consistently demonstrates superior performance in reference basins, aligning with anticipated outcomes. In the majority of the non-reference basins, the model is likely missing a critical process, such as water diversion for irrigation or hydropower regulation. This deficiency highlights the challenges associated with accurately representing complex hydrological processes in non-reference basins, impacting overall model performance in these areas.

The Northwest (NWRFC) and California-Nevada (CNRFC) RFCs emerge as the top performers across virtually all metrics for the GAGES-II reference basin. The correlation coefficient in these regions (> 0.8) stands out as notably better than in all other areas, contributing to the higher KGE and NSE metrics. The median of NSE values across basins is 0.73 and 0.57 over the GAGESII reference basins for NWRFC, and CNRFC respectively, which is above what Moriasi et al. (2007) defines as satisfactory performance. The Northeast (NERFC), Middle Atlantic (MARFC), Ohio (OHRFC), Southeast (SERFC) and Lower Mississippi (LMRFC) and Colorado Basin (CBRFC) RFCs demonstrate reasonable performance across all metrics for GAGES-II reference basins, with particularly notable strength in model biases. These RFCs all have a positive median NSE values indicating the model is performing superior to using mean observations, and median KGE higher than 0.5 and median correlation coefficient greater than 0.6. Conversely, the Missouri Basin (MBRFC), North Central (NCRFC), West Gulf (WGRFC) and Arkansas-Red Basin (ABRFC) RFCs represent a subset of regions where the model exhibits poor performance with median KGE values below 0.5 and median NSE values close to 0. These regions experience challenges that warrant attention for further improvement in the model's predictive capabilities.

Figure 11 provides a spatial map of mean annual WRF-Hydro IWAAs runoff ratio (in percent) across CONUS from October 2009 to October 2021. Here, runoff ratio is defined as the summation of mean surface flow and mean baseflow divided by mean precipitation for each HUC12 basin, defined in the National Watershed Boundary Dataset as ranging in size from 40–160 km^2 (Blodgett, 2023). It is multiplied by 100 to be in percent. Importantly, the HUC12 scale adopted for this visualization aligns with the spatial units employed for the USGS IWAAs analysis.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

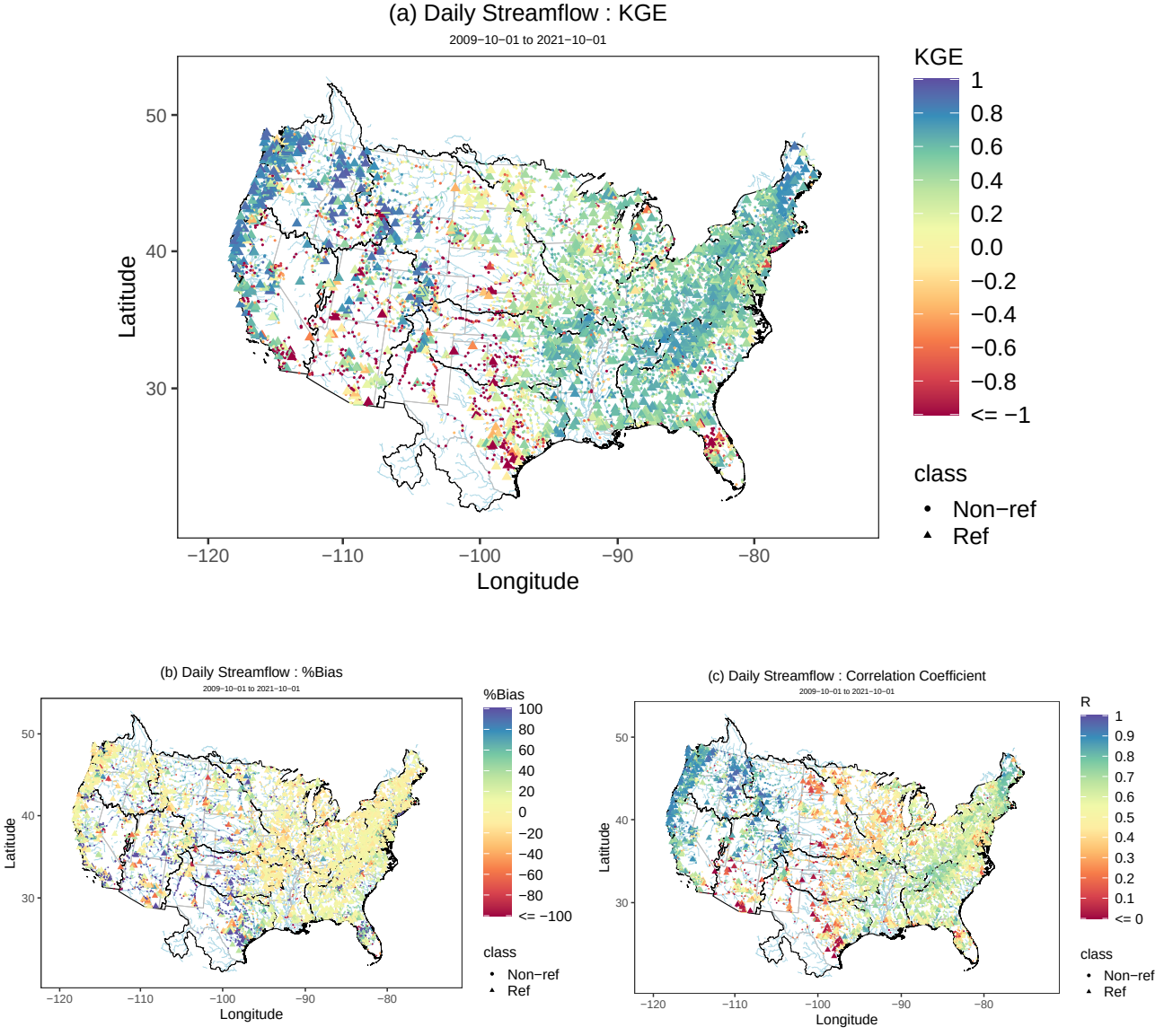


Figure 9. Spatial maps of (a) KGE, (b) percent bias (%Bias), and (c) correlation coefficient of daily streamflow over water year (WY) 2010 through WY2021 at USGS reference (Ref) and non-reference gages (Non-ref).

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Figure 10. Boxplot of KGE, NSE, percent bias and correlation coefficient of daily streamflow at USGS reference (Ref) and non-reference gages (Non-ref) across each river forecast center. The median of the error metrics are given in each figure above the boxplots and the number of gauges in each category is indicated below the boxplot in the lower panel.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

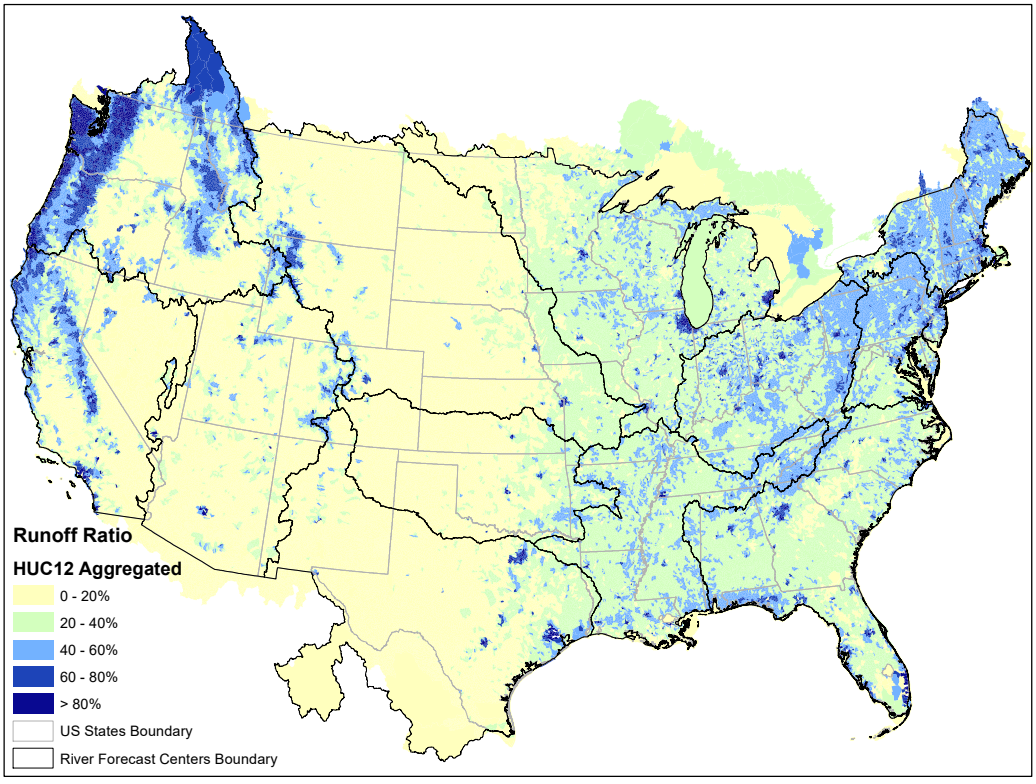


Figure 11. Cumulative density function of hourly streamflow metrics (KGE, %Bias, correlation coefficient, NSE) across the 1,522 calibration gages; dashed line shows the optimal value of each error metric.

5.3 Snow Analysis

Snow performance in the WRF-Hydro IWAAs model is compared with the NOAA National Weather Service Snow Data Assimilation System (SNODAS) snow water equivalent (SWE). SNODAS is a modeling and data assimilation system which ingests a number of in situ and remotely sensed observations to produce high-resolution gridded estimates of snow depth and SWE (Carroll et al., 2001). SNODAS SWE and snow depth are available over CONUS at 1-km spatial resolution and daily temporal resolution from water year 2004 to present (NOHRSC, 2004). SNODAS SWE is an estimate of true SWE rather than a direct observation. We thus treat SNODAS as a convenient and widely used (e.g. Hedrick et al., 2015; Lv et al., 2019; Yang et al., 2023) estimate of snow performance against which to evaluate WRF-Hydro IWAAs SWE, rather than an objective truth. For ease of comparison, SNODAS SWE is interpolated onto the WRF-Hydro IWAAs model grid using a bilinear interpolation. Comparisons are made in terms of both grid-scale seasonal mean SWE and regional monthly mean SWE within RFCs that receive large seasonal snow accumulation (>5 mm peak annual SWE).

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

Broadly, relative to SNODAS, WRF-Hydro IWAAs SWE exhibits low biases which develop over the course of the snow accumulation season (December-February), peak near the time of peak SWE (February-April, depending on the region), and persist through the snow ablation season (March-May) (Figures 12 and 13). In the snow accumulation season, mixed performance is evident in western US mountain ranges, with slight low biases across the Sierra Nevada (CNRFC), areas of both high and low biases in the middle Rockies and Cascades (NWRFC), and minimal biases in the Colorado Rockies. A low bias is evident across the upper Midwest (NCRFC), which is small in absolute magnitude but large relative to the amount of regional SWE (Figure 12 a and b). In the snow ablation season, low biases exist throughout the western US mountains and the upper Midwest and northeastern US (Figure 12 c and d). The timing of peak SWE in WRF-Hydro IWAAs coincides with that of SNODAS in most RFCs, excluding the California-Nevada, Colorado Basin, and Northeast RFCs, in which WRF-Hydro IWAAs peak SWE occurs approximately one month early (Figure 13). The magnitude of WRF-Hydro IWAAs peak SWE is low relative to SNODAS in all RFC regions. This low bias is most pronounced in the North Central (NCRFC), where WRF-Hydro IWAAs peak SWE is only 50% of SNODAS peak SWE. WRF-Hydro IWAAs SWE exhibits slightly slower melt rates than SNODAS, particularly during the late ablation season, resulting in similar dates of snow disappearance despite the sometimes substantial low biases in WRF-Hydro IWAAs.

Differences in seasonal SWE between WRF-Hydro IWAAs and SNODAS have numerous potential causes; a full investigation of these causes is beyond the scope of this overview. SWE errors during the snow accumulation season may be attributable to differences in rain-snow partitioning between the respective models, or due to errors in winter precipitation inputs. Errors in both the CONUS404 forcing data and the Daymet data used for precipitation and air temperature bias correction are possible. Meanwhile, disagreement in snow ablation rates between WRF-Hydro IWAAs and SNODAS may indicate differing melt parameterizations between the two models. WRF-Hydro IWAAs snowpack parameters are calibrated to streamflow and not to snow, so snow model performance may be degraded to optimize streamflow. Lastly, as a data assimilation system, SNODAS ingests in situ and remotely sensed snow observations, which is likely to introduce additional disagreement between WRF-Hydro IWAAs and SNODAS SWE.

5.4 Evapotranspiration Analysis

For ET evaluation, there are multiple observation-based datasets that have been used as a reference in the literature such as MODIS16, Operational Simplified Surface Energy Balance (SSEBop), and Global Land Evaporation Amsterdam Model (GLEAM) (Saxe et al., 2020; Mazrooei et al., 2020, 2021; Mai et al., 2022). Although there are major similarities and agreements between these datasets, each of them has its own uncertainty and there are discrepancies between them. The uncertainties associated with ET estimates pose a great challenge to the scientific research community and operational water supply managers. For instance, SSEBop ET estimates are found to be universally higher than the other products across the country (Mazrooei et al., 2024; Senay et al., 2013; Velpuri et al., 2013). Thus, none of the currently available model- and satellite-based ET products can be determined as an absolute reference for ET evaluation efforts.

We evaluate ET outputs from the WRF-Hydro IWAAs model against the GLEAM dataset. GLEAM is an observation-based reanalysis benchmark commonly used in ET literature (Martens et al., 2017; Zhan et al., 2019). GLEAM is a global evaporation

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

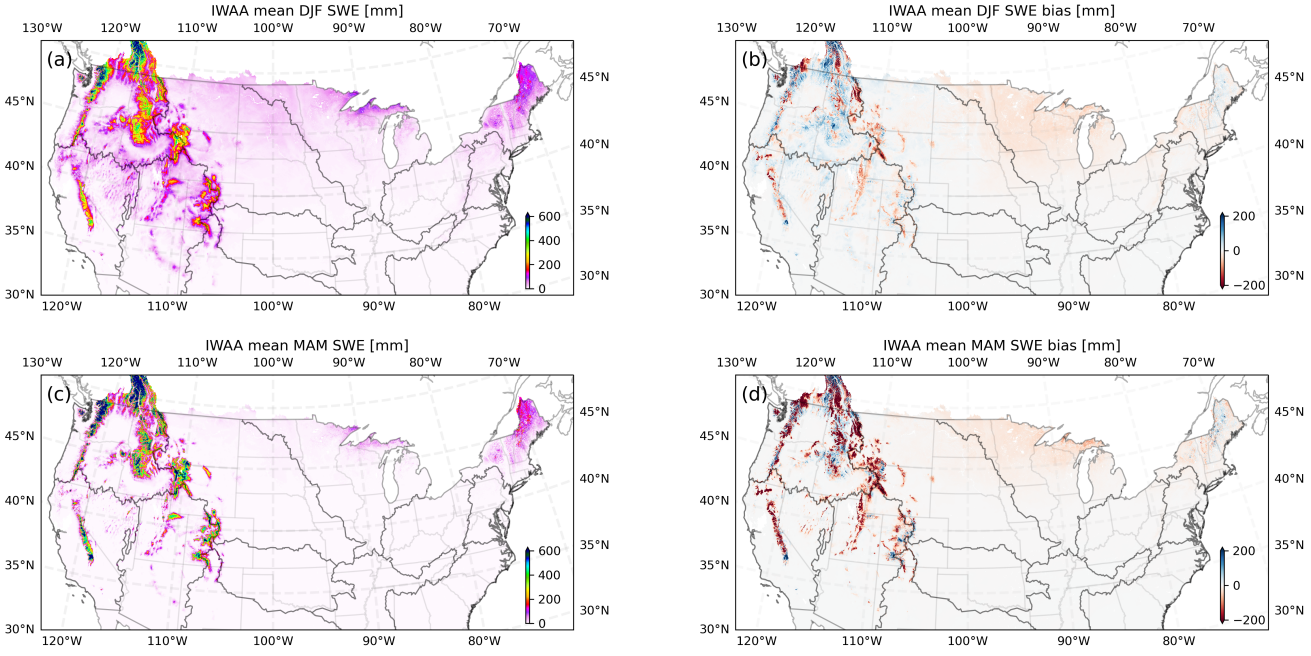


Figure 12. (a) WRF-Hydro IWAAs mean December-January-February (DJF) SWE. (b) WRF-Hydro IWAAs mean DJF SWE bias relative to SNODAS. (c) WRF-Hydro IWAAs mean March-April-May (MAM) SWE. (d) WRF-Hydro IWAAs mean MAM SWE bias relative to SNODAS. RFC (dark grey) and state (light grey) boundaries are shown.

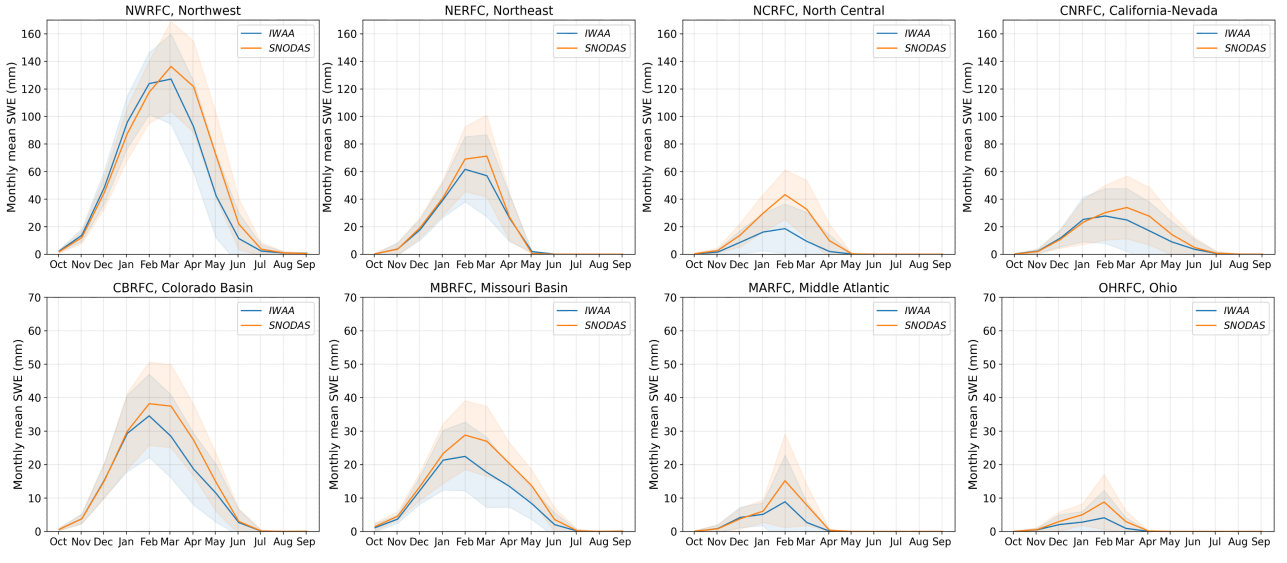


Figure 13. Time series of monthly mean SWE over the WRF-Hydro IWAAs simulation period (water years 2010-21), separated by RFC regions that receive large seasonal snow accumulation (> 5mm peak annual SWE).

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

model that is driven largely by observed forcings from satellite remote sensing and provides estimates of ET, surface and root-zone soil moisture (SM), etc. The GLEAM dataset is available at 0.25° spatial resolution with a daily temporal resolution from 2003 until present (Miralles et al., 2011). GLEAM gridded data are regridded from their native spatial resolution to a 1-km grid matching the WRF-Hydro IWAAs grid cells using bilinear interpolation. We also conduct regional analyses of modeled ET based on the 12 RFCs across CONUS.

Figure 14 a and b shows the CONUS maps of mean annual ET estimates from the WRF-Hydro IWAAs and GLEAM models. Both datasets show a very similar cumulative distribution of ET estimates from zero to 120 mm/month with the exception of the low tail of the distribution. Looking at the lower tail of the CDF plots (Figure 14 c), we find that WRF-Hydro IWAAs generates a substantial number of near-zero ET values, mainly over urban areas. This is due to the WRF-Hydro land surface model (Noah-MP) treatment of urban land cover types, which severely limits ET. In terms of spatial distribution, both models show the lowest ET estimates over arid regions (e.g., Arizona, southern California, New Mexico) and the highest values over southeastern states (e.g., Florida, Alabama, Georgia). The difference map (Figure14 d) shows that WRF-Hydro IWAAs has slightly higher estimates of ET estimates over the Northeast (NERFC), Southeast (SERFC), and Pacific Northwest regions (NWRFC).

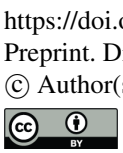
The correlation map in Figure 14 e highlights that WRF-Hydro IWAAs and GLEAM monthly ET time series are well correlated across the entire study region, except urban areas. More than 90% of 1-km WRF-Hydro cell values show correlation coefficients higher than 0.7. The median of correlation coefficients across CONUS is about 0.9. Generally, the highest correlation coefficients occur over the northeast and Midwest regions (NERFC, MARFC, OHRFC, NCRFC), and lower coefficients occur over the southwest and west arid regions (CNRFC, CBRFC, WGRFC). Strong correlation between the two products is found over regions with lots of cropland, such as the Missouri Basin and Lower Mississippi RFCs, despite the fact that irrigation is not represented in the WRF-Hydro IWAAs model. GLEAM also does not account for irrigation explicitly. However, GLEAM benefits from data assimilation using European Space Agency Climate Change Initiative (ESA_CCI) soil moisture observations. Therefore, it is likely that GLEAM is capturing some of the irrigation signals. The near-zero correlation coefficients over urban areas are due to poor performance of the WRF-Hydro IWAAs model in estimating latent heat fluxes over urban land covers. As a result, urban ET estimates from WRF-Hydro IWAAs are very small when compared to GLEAM.

Next, ET over the 1-km grid cells within these maps are grouped by the RFC regions and summarized in terms of long-term monthly mean and standard deviation bounds (Figure 15). The seasonality in ET fluxes is captured well by the WRF-Hydro IWAAs model. Overall, WRF-Hydro IWAAs ET estimates are higher than GLEAM for the central RFC regions during warm months. Over the east and west coast RFC regions, WRF-Hydro IWAAs ET values are usually lower than GLEAM, except during warm months when both models estimate ET fluxes at a very similar level.

5.5 Soil Moisture Analysis

GLEAM is a series of algorithms tuned to estimate terrestrial ET, however, it also estimates intermediate variables including potential evaporation, surface soil moisture (SM), and root-zone SM. We conduct comparison analyses between WRF-Hydro IWAAs and GLEAM in terms of SM estimates for the ground surface (i.e., top 10-cm soil layer) and root-zone. Figure 16

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024



Hydrology and Earth System Sciences
Discussions

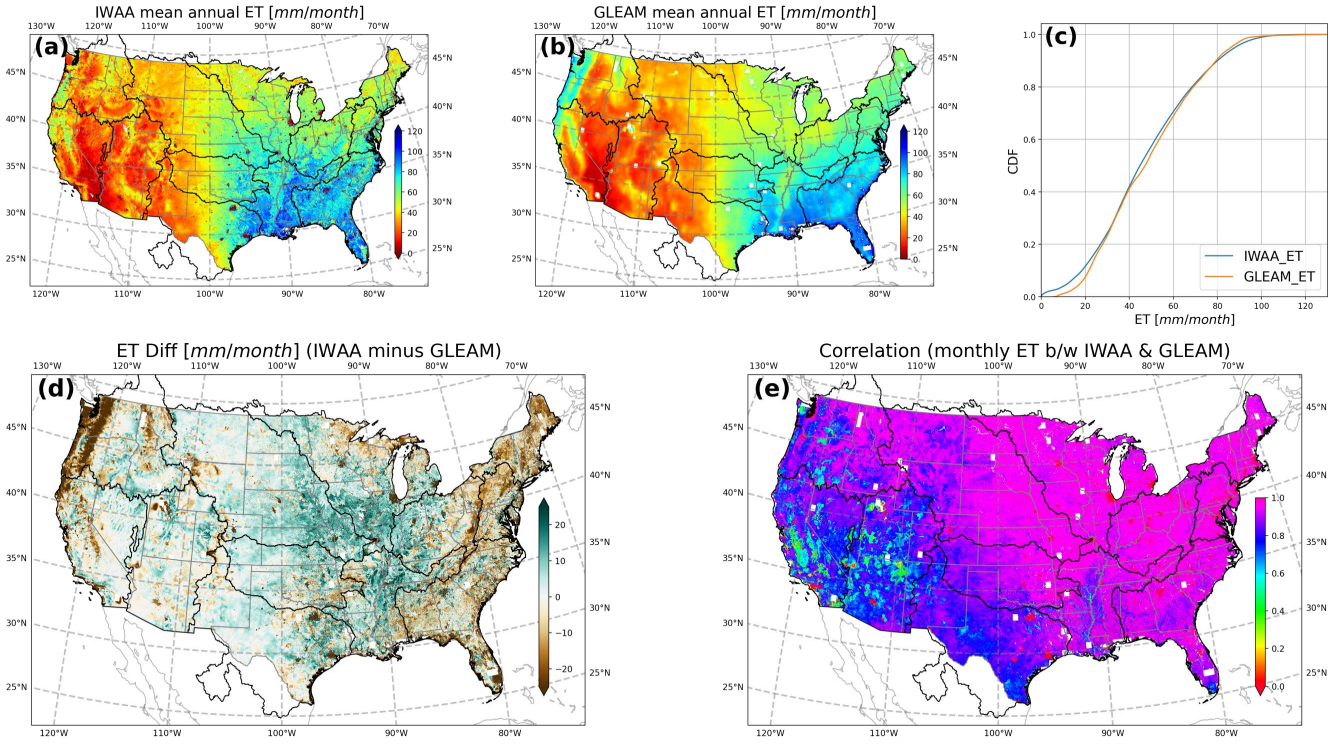


Figure 14. (a) WRF-Hydro IWAAs and (b) GLEAM mean annual ET estimates. (c) CDF of mean annual values for both WRF-Hydro IWAAs and GLEAM. (d) Long-term difference (WRF-Hydro IWAAs minus GLEAM) of the two product and (e) correlation coefficients between WRF-Hydro IWAA and GLEAM.

a and b illustrates the distributions of surface (top 10-sm) SM estimates from both datasets. In terms of spatial distribution, there is generally a strong agreement between the products across CONUS. However, WRF-Hydro IWAAs estimates surface SM lower than GLEAM, particularly over the eastern US. This low (negative) bias is even more evident over the Southeast (SERFC) coastline and North Central RFC region. The differences in surface SM between the two products are minimal over the Western RFC regions such as NWRFC, MBRFC, CBRFC, ABRFC, CNRFC, and WGRFC (Figure 17 d). Overall, the median of surface SM differences between the products is equal to -0.023 [m^3/m^3] across the country. Figure 17 e shows the correlation coefficient between the WRF-Hydro IWAAs and GLEAM surface SM fluxes. Again, we see that the two products are well correlated, with the median value across the country being equal to 0.79 and with more than 90% of grid cells having correlation coefficient greater than 0.6.

In GLEAM, three soil layers are defined (0–10 cm, 10–100 cm, and 100–250 cm) and the root zone is a function of the land cover and vegetation type (Martens et al., 2017). In WRF-Hydro IWAAs, there are four soil layers defined in the Noah-MP LSM (0–10 cm, 10–40 cm, 40–100 cm, and 100–200 cm) and root zone depth is also a function of land cover type. From the Noah-MP model parameter table, we use the parameter NROOT (i.e., number of soil layers from top down reached by

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

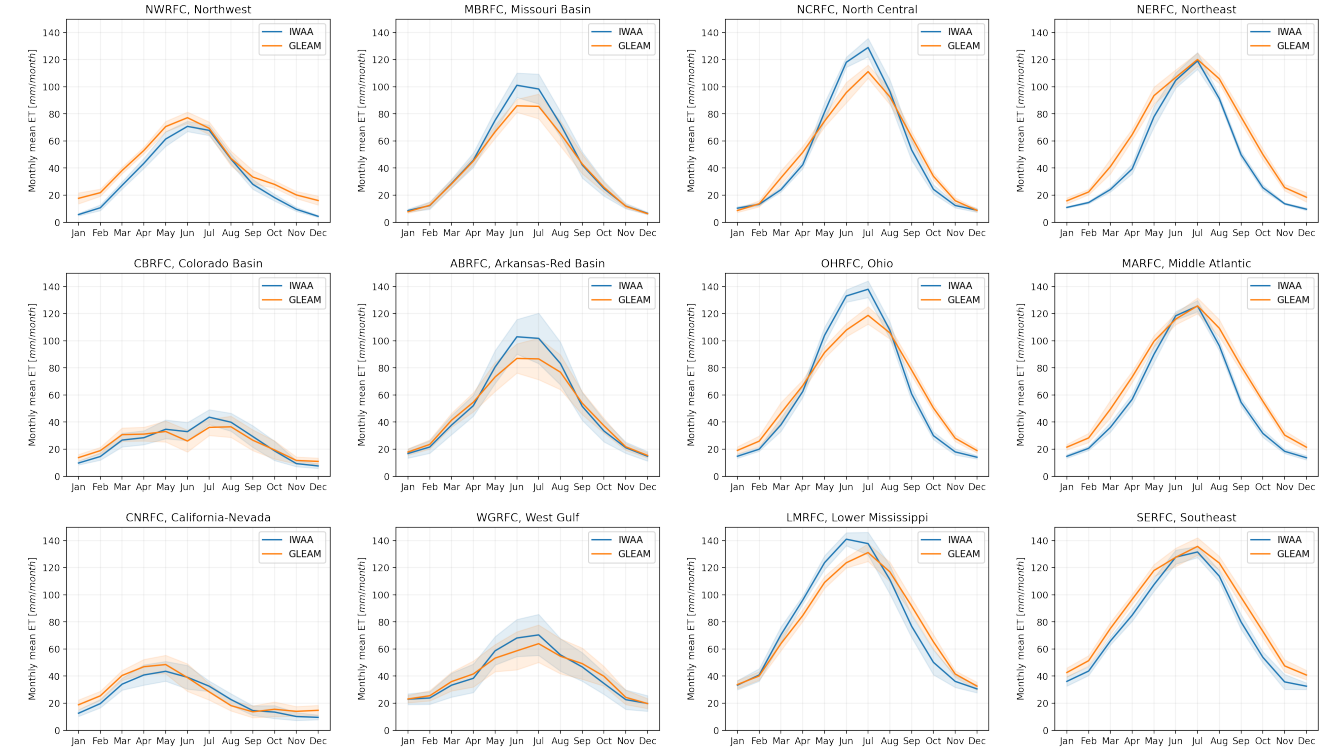


Figure 15. Monthly mean ET estimates from WRF-Hydro IWAAs and GLEAM grouped by 12 RFC regions. The shaded area reflects the monthly standard deviation of ET estimates computed over the study period.

roots) to calculate the depth-mean root zone SM and compare against the GLEAM root zone SM product. While root zone SM calculations are internally consistent for each model, WRF-Hydro IWAAs and GLEAM use different land cover datasets in addition to the different soil layer structures and root depth parameters, so the root zone depths do not match at all locations.

Figure 18 a, b and c shows the spatial and cumulative distributions of mean annual root-zone SM estimates from the WRF-Hydro IWAAs model and GLEAM product. Both products show approximately a similar magnitude of root-zone SM across the country. WRF-Hydro IWAAs exhibits a smoother CDF plot with a slightly narrower range of root-zone SM estimates with the 10th–90th percentile range of 0.20–0.36 as opposed to GLEAM with the 10th–90th percentile range of 0.16–0.38. Also in terms of extremes, GLEAM shows a maximum root-zone SM of 0.77, while the maximum from IWAAs is 0.57. Over the southwest desert region, GLEAM shows that the root-zone SM is the driest, in the same level as its surface SM (Figure 16). However, WRF-Hydro IWAAs suggests that root-zone SM over this region is substantially wetter than its surface SM. Figure 19 shows the RFC regional analyses of root-zone SM by grouping 1-km grid cells within RFC boundaries and summarizing them in terms of long-term monthly mean values wrapped by standard deviation bands. Northeast, Missouri Basin, and Arkansas-Red Basin RFC regions yield a minimal difference between the two products. Over the western RFC regions, such as Northwest, California-Nevada, Colorado Basin, and West Gulf, WRF-Hydro IWAAs estimates root-zone SM higher than GLEAM. This

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

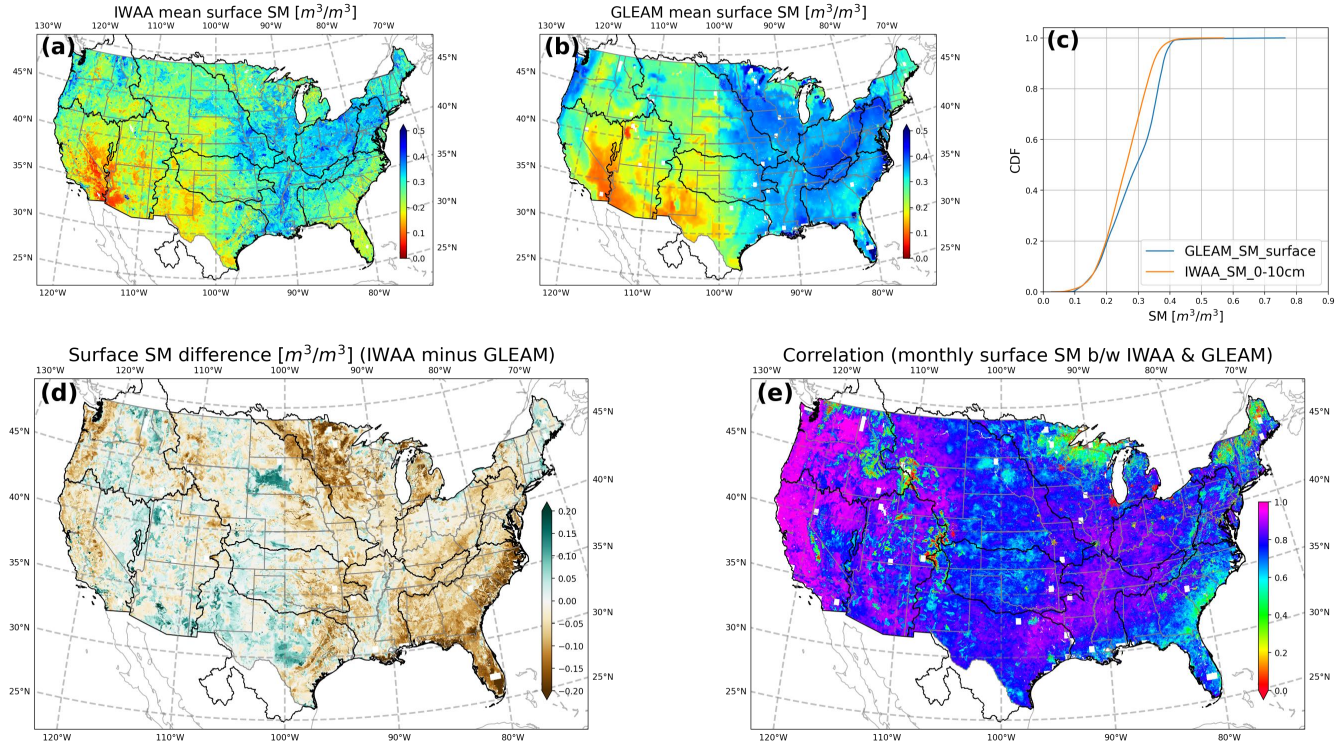


Figure 16. (a) WRF-Hydro IWAAs and (b) GLEAM mean annual surface soil moisture estimates. (c) CDF of mean annual surface soil moisture estimates for both WRF-Hydro IWAAs and GLEAM. (d) Long-term difference (WRF-Hydro IWAAs minus GLEAM) and (e) correlation coefficients between WRF-Hydro IWAA and GLEAM.

difference between WRF-Hydro IWAAs and GLEAM is even more evident during the summer season. Over the eastern regions, particularly the southeast, IWAAs estimates are drier than GLEAM. Despite these differences, both products are generally in harmony regarding the seasonality of root-zone SM. Although the differences in root-zone SM between the two products are greater when compared to the surface SM fluxes, the median difference value in the Figure 18 d is still very small (0.006 m^3/m^3). The positive difference values over the Western US and negative difference values over the Eastern US generally cancel each other at a CONUS scale. The correlation coefficient map in Figure 18 e shows the two products are strongly correlated (median across the country is 0.75). Near-perfect correlation coefficients are found over the West Coast, while near-zero correlations occur over urban areas where there is no temporal variation in SM estimates within the WRF-Hydro IWAAs model.

6 Conclusions and Potential Model Enhancement

In this paper, we describe the Weather Research and Forecasting model hydrological modeling extension package (WRF-Hydro) modeling effort under the USGS Integrated Water Availability Assessments (IWAAs), a nationwide water supply study

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

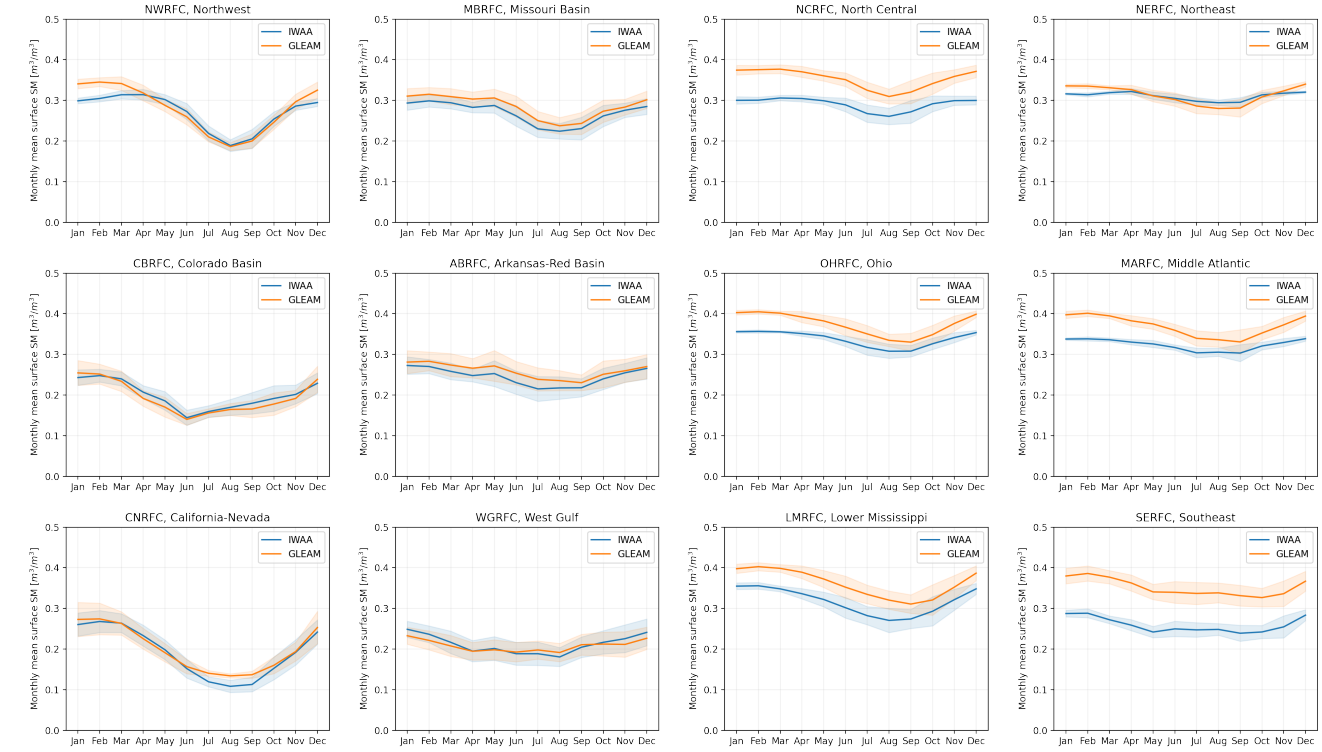


Figure 17. Monthly mean surface soil moisture (SM) estimates from WRF-Hydro IWAAs and GLEAM grouped by 12 RFC regions. The shaded area reflects the monthly standard deviation of SM estimates computed over the study period.

across the conterminous United States (CONUS). The atmospheric forcing used is the publicly available CONUS404 dataset, a mesoscale hydroclimate dataset available over the CONUS for the most recent 43 years. The CONUS404 precipitation and temperature were bias-adjusted relative to the Daymet data. WRF-Hydro IWAAs calibration is performed across 1,522 basins in the US resulting in substantial improvements in streamflow model simulations in the majority of basins. The model parameters are then extrapolated from high quality calibration basins to all other uncalibrated locations based on similarity between the calibration basins and the regionalization units (HUC10 scale). Then, we conduct model simulations spanning the period from October 2009 to October 2021 and evaluate model performance.

WRF-Hydro IWAAs streamflow performance is superior at Northwest (NWRFC) and California-Nevada (CNRFC) RFCs with median NSE values of 0.73 and 0.57 and median KGE values of 0.81 and 0.67 at the GAGESII reference basins, respectively. The Northeast (NERFC), Middle Atlantic (MARFC), Ohio (OHRFC), Southeast (SERFC) and Lower Mississippi (LMRFC) and Colorado Basin (CBRFC) RFCs also demonstrate reasonable performance with median NSE values greater than 0, median KGE higher than 0.5 and median correlation coefficient greater than 0.6 at all these RFCs. On the other hand, the model has a poor performance in the Missouri Basin (MBRFC), North Central (NCRFC), West Gulf (WGRFC) and Arkansas-Red Basin (ABRFC) RFCs where the median KGE values are below 0.5 and median NSE values are close to 0. These suboptimal

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions

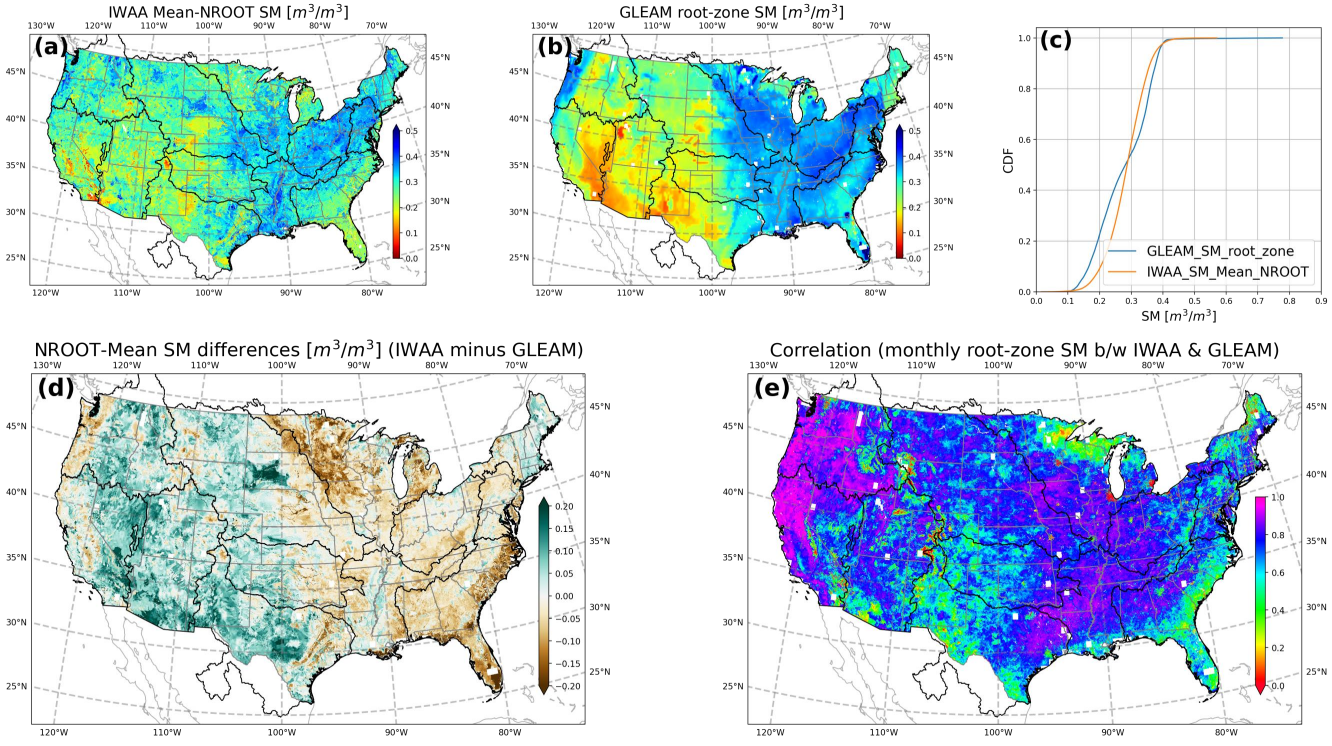


Figure 18. (a) WRF-Hydro IWAAs and (b) GLEAM mean annual root-zone soil moisture estimates. (c) CDF of mean annual root-zone soil moisture estimates for both WRF-Hydro IWAAs and GLEAM. (d) Long-term difference (WRF-Hydro IWAAs minus GLEAM) and (e) correlation coefficients between WRF-Hydro IWAA and GLEAM.

model behaviors could be rooted in deficiencies in model process representation or atmospheric forcing errors. In particular, WRF-Hydro IWAAs streamflow simulations show relatively low correlation coefficients at a daily timescale that could be attributed to the use of the model-based CONUS404 as atmospheric forcing in place of an observation-based dataset. However, CONUS404 offers additional advantages, including consistent coverage in areas with sparse observations and consistency in forcings over long time periods (versus observation networks which change over time). In addition, there is a planned future scenario of CONUS404 that could provide an opportunity for studying climate change impacts on water budget components.

Apart from streamflow, model performance is evaluated using other hydrologic components, such as snow water equivalent, soil moisture, and evapotranspiration, to paint a more complete picture of the model behavior. Snow performance is evaluated using Snow Data Assimilation System (SNODAS) snow water equivalent (SWE), regridded to the WRF-Hydro IWAAs domain, over water years 2010 through 2021. Results show a reasonable agreement between SNODAS and WRF-Hydro IWAAs SWE across CONUS during the snow accumulation season; however, a broad low bias develops during the ablation season. Aggregated SWE values across different River Forecast Center (RFC) regions confirms these findings, showing low biases in the WRF-Hydro IWAAs simulations during the ablation season and also lower peak SWE values relative to SNODAS. The

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

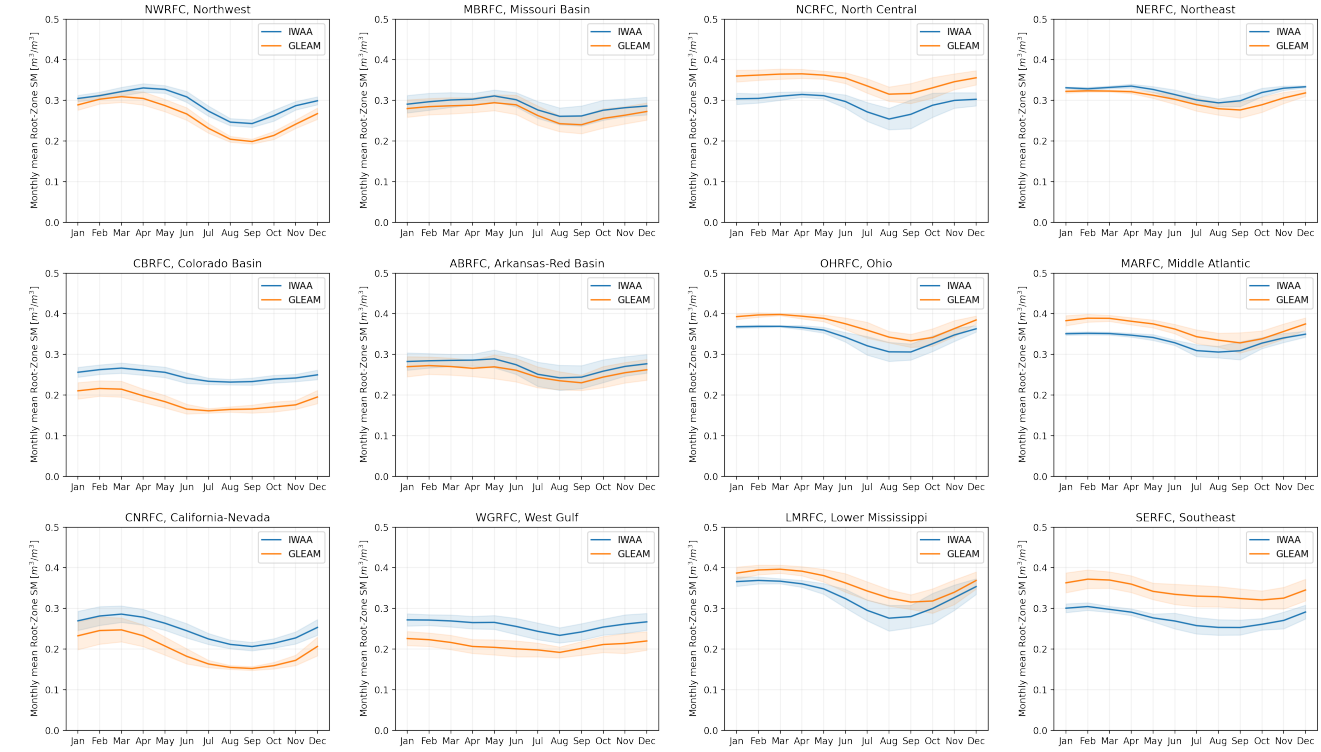


Figure 19. Monthly mean root-zone soil moisture estimates from WRF-Hydro IWAAs and GLEAM grouped by 12 RFC regions. The shaded area reflects the monthly standard deviation of SM estimates computed over the study period.

timing of peak SWE in WRF-Hydro IWAAs coincides with that of SNODAS in most RFCs. However, WRF-Hydro IWAAs SWE peaks early in certain RFCs, notably the California-Nevada, Colorado Basin, and Northeast regions. We reiterate that, as a model-based product, SNODAS SWE is considered a benchmark rather than ground truth. Evaluating IWAAs snow state variables against in situ and remotely sensed snow observations could support the full characterization of IWAAs snow performance.

WRF-Hydro IWAAs ET and soil moisture simulations are evaluated against the Global Land Evaporation Amsterdam Model (GLEAM) dataset. As with SNODAS, the model-based GLEAM serves as a benchmark rather than ground truth. ET comparison reveals a reasonable agreement between the two models when comparing the cumulative distribution functions across CONUS, except over urban areas where the WRF-Hydro IWAAs implementation underestimates ET severely. In addition, WRF-Hydro IWAAs has slightly higher estimates of ET over the central US and lower ET estimates over the Northeast (NERFC), Southeast (SERFC), and Pacific Northwest regions (NWRFC). Surface and root-zone soil moisture analyses suggest a strong agreement between WRF-Hydro IWAAs simulations and GLEAM. Generally, over eastern RFCs, IWAAs SM estimates are slightly lower than GLEAM, and over western RFCs IWAAs SM estimates are slightly higher than GLEAM. All the analyses performed indicate value in using the WRF-Hydro model simulations.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

NCAR is currently extending model simulations from the existing 12-year span to cover 43 years, leveraging the comprehensive CONUS404 dataset for long-term analysis. This expansion facilitates various studies, including trend analysis. Additionally, a forthcoming CONUS404 future scenario is planned, which can be used as atmospheric forcing for the WRF-Hydro model to support investigation of climate-change effects on water budget components.

Code availability. The WRF-Hydro model is based on the WRF-Hydro community code tagged as *applications/IWAA* available at https://github.com/NCAR/wrf_hydro_nwm_public/releases/tag/applications%2FIWAA

Data availability. The WRF-Hydro IWAAs final domains (after calibration) along with the model namelists used for the retrospective model run can be accessed under the following data release:

Rafieeinasab, A., Gochis, D., Srivastava I., Dugger, A., Sampson, K., Omani, N., Mazrooei, A., Zhang, Y., Casalli, M., 2024, Application of the WRF-Hydro Modeling System for the Conterminous United States, CONUS404BA Atmospheric Forcings, 2010-2021, U.S. Geological Survey data release, https://doi.org/10.5066/P1KZGLU2

The CONUS404 bias-adjusted dataset can be accessed under the following data release:

Zhang, Y., Grimm, J., Cabell, R., Srivastava, I. Gochis, D., Prein, A., Rasmussen, R., Ikeda, K., and Schneider, T., 2024, CONUS404 climate forcing variable subset for hydrologic models, 1979-2022: downscaled to 1 km and bias-adjusted for precipitation and temperature: U.S. Geological Survey data release, https://doi.org/10.5066/P9JE61P7.

Author contributions. Rafieeinasab has performed the calibration, and model simulations, we well as writing more than 50% of the paper. Srivastava has contributed to the work with execution of calibration for a subset of calibration basins and providing software engineer support along the project. Mazrooei has performed the ET and soil moisture analysis and wrote sections 5.4 (Evaportranspiration Analysis) and 5.5 (Soil Moisture Analysis). Enzminger has performed the snow analysis and wrote section 5.3 (Snow Analysis). Grim and Zhang have developed the bias corrected CONUS404 forcing dataset used to force the WRF-Hydro model simulations, as well as Grim writing section 3 (Atmosperfic Forcing). Liu and Dugger has developed the regionalization scripts, and Omani has performed the regionalization for this study. Liu has contribute to the paper writing in subsection "regionalization" in Section 4 (Model Calibration and Regionalization). Omani has contributed to the streamflow verfication performed for study (Section 5.2). Lafontaine, Viger, Dugger, Gochis and Schneider have been involved in planning of the project, execution of different phases, performing review of the content and manuscript, and public data release.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Acknowledgements. This research was supported by the U.S. Geological Survey (USGS) Water Mission Area's Water Availability and Use Science Program and Integrated Water Prediction Program. National Center for Atmospheric Research (NCAR) is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement 1852977. We would like to acknowledge high-performance computing support from Cheyenne doi:10.5065/D6RX99HX provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024


Hydrology and Earth System Sciences
Discussions

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, 2017.

Arnault, J., Wagner, S., Rummler, T., Fersch, B., Bliefernicht, J., Andresen, S., and Kunstmann, H.: Role of runoff–infiltration partitioning and resolved overland flow on land–atmosphere feedbacks: A case study with the WRF-Hydro coupled modeling system for West Africa, Journal of Hydrometeorology, 17, 1489–1516, 2016.

ASOS: Automated surface observing system: Asos user's guide.[Washington, DC]: US Dept. of Commerce, National Oceanic and Atmospheric Administration: Federal Aviation Administration: US Navy: US Dept. of the Air Force, URL https://www. weather. gov/media/asos/aum-toc. pdf, 1998.

Bao, D., Xue, Z. G., Warner, J. C., Moulton, M., Yin, D., Hegermiller, C. A., Zambon, J. B., and He, R.: A numerical investigation of Hurricane Florence-induced compound flooding in the Cape Fear Estuary using a dynamically coupled hydrological-ocean model, Journal of Advances in Modeling Earth Systems, 14, e2022MS003 131, 2022.

Blodgett, D. L.: Mainstem Rivers of the Conterminous United States (ver. 2.0, February 2023), U.S. Geological Survey data release, https://doi.org/10.5066/P92U7ZUT, 2023.

Carroll, T., Cline, D., Fall, G., Nilsson, A., Li, L., and Rost, A.: NOHRSC operations and the simulation of snow cover properties for the coterminous US, in: Proc. 69th Annual Meeting of the Western Snow Conf, pp. 1–14, 2001.

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., Cabell, R., Casiday, N., Cui, Z., et al.: NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling, JAWRA Journal of the American Water Resources Association, 2024.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, Journal of Applied Meteorology and Climatology, 33, 140–158, 1994.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water resources research, 28, 1015–1031, 1992.

Duan, Q., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, Journal of optimization theory and applications, 76, 501–521, 1993.

Falcone, J.: US Geological Survey GAGES-II time series data from consistent sources of land use, water use, agriculture, timber activities, dam removals, and other historical anthropogenic influences, US Geological Survey [data set], https://doi. org/10.5066/F7HQ3XS4, 2017.

Fall, G., Kitzmiller, D., Pavlovic, S., Zhang, Z., Patrick, N., St. Laurent, M., Trypaluk, C., Wu, W., and Miller, D.: The Office of Water Prediction's Analysis of Record for Calibration, version 1.1: Dataset description and precipitation evaluation, JAWRA Journal of the American Water Resources Association, 59, 1246–1272, 2023.

Fredj, E., Silver, M., and Givati, A.: An integrated simulation and distribution system for early flood warning, International Journal of Computer and Information Technology, 4, 517–526, 2015.

Givati, A., Gochis, D., Rummler, T., and Kunstmann, H.: Comparing one-way and two-way coupled hydrometeorological forecasting systems for flood forecasting in the Mediterranean region, Hydrology, 3, 19, 2016.

Gochis, D., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., and Zhang, Y.: The WRF-Hydro® modeling system technical description, (Version 5.2.0), NCAR Technical Note, 2020.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Gower, J. C.: A general coefficient of similarity and some of its properties, Biometrics, pp. 857–871, 1971.

Grim, J. A., Zhang, Y., and Gochis, D. J.: Impact of the Alamosa gap-filling radar on streamflow in the National Water Model, Frontiers in Earth Science, 10, 995 424, 2023.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80–91, 2009.

He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., Schneider, T., Rasmussen, R., Niu, G., Yang, Z., et al.: The community Noah-MP land surface modeling system technical description version 5.0, Tech. rep., NCAR Technical Note NCAR/TN-575+ STR, doi: 10.5065/ew8g-yr95, 2023.

Hedrick, A., Marshall, H.-P., Winstral, A., Elder, K., Yueh, S., and Cline, D.: Independent evaluation of the SNODAS snow depth product using regional-scale lidar-derived measurements, The Cryosphere, 9, 13–23, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., and Muñoz-Sabater, J.: The ERA5 global reanalysis, quarterly journal of the royal meteorological society, 2020.

Jones, K. A., Niknami, L. S., Buto, S. G., and Decker, D.: Federal standards and procedures for the national watershed boundary dataset (wbd): chapter 3 of section a, federal standards, book 11, collection and delineation of spatial data, Tech. rep., US Geological Survey, 2022.

Kerandi, N., Arnault, J., Laux, P., Wagner, S., Kitheka, J., and Kunstmann, H.: Joint atmospheric-terrestrial water balances for East Africa: a WRF-Hydro case study for the upper Tana River basin, Theoretical and Applied Climatology, 131, 1337–1355, 2018.

Knoben, W. J., Freer, J. E., and Woods, R. A.: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, 2019.

Lee, J., Kim, Y., and Wang, D.: Assessing the characteristics of recent drought events in South Korea using WRF-Hydro, Journal of Hydrology, 607, 127 459, 2022.

Liu, Y., Durcik, M., Gupta, H. V., and Wagener, T.: Developing distributed conceptual hydrological models from geospatial databases., edited by Abesser et al., IAHS Publ. 320, 2008.

Liu, Y., Rafieeinasab, A., Dugger, A., Feng, X., Cosgrove, B., Wu, W., and Gochis, D.: Enhancing NWM Parameter Regionalization to Improve Physical Similarity Representation While Accounting for Uncertainties, in: AGU Fall Meeting Abstracts, vol. 2021, pp. H45N–1333, 2021.

Lundquist, J., Hughes, M., Gutmann, E., and Kapnick, S.: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks, Bulletin of the American Meteorological Society, 100, 2473–2490, 2019.

Lv, Z., Pomeroy, J. W., and Fang, X.: Evaluation of SNODAS snow water equivalent in western Canada and assimilation into a Cold Region Hydrological Model, Water Resources Research, 55, 11 166–11 187, 2019.

Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., Fortin, V., Fry, L. M., Gauch, M., Klotz, D., et al.: The great lakes runoff intercomparison project phase 4: the great lakes (GRIP-GL), Hydrology and Earth System Sciences, 26, 3537–3572, 2022.

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903–1925, 2017.

Mason, L. A., Gronewold, A. D., Laitta, M., Gochis, D., Sampson, K., Read, L., Klyszejko, E., Kwan, J., Fry, L., Jones, K., et al.: New transboundary hydrographic data set for advancing regional hydrological modeling and water resources management, Journal of Water Resources Planning and Management, 145, 06019 004, 2019.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Mauricio, Z.-B.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-10, 2017.

Mazrooei, A., Sankarasubramanian, A., and Wood, A. W.: Potential in improving monthly streamflow forecasting through variational assimilation of observed streamflow, Journal of Hydrology, 600, 126 559, 2021.

Mazrooei, A., Dugger, A., Rafieeinasab, A., Towler, E., Khazai, B., Zhang, Y., and etc: Performance of U.S. National Water Model in Simulating Evapotranspiration Fluxes, Journal of Hydrological Processes, p. under review, 2024.

Mazrooei, A. H., Rafieeinasab, A., Dugger, A. L., Valayamkunnath, P., Gochis, D., Feng, X., and Cosgrove, B. A.: Evaluation of National Water Model in simulating Evapotranspiration fluxes across multiple spatio-temporal scales, in: AGU Fall Meeting Abstracts, vol. 2020, pp. H111–0001, 2020.

McKay, L., Bondelid, T., Rea, A., Johnston, C., Moore, R., Dewald, T., McKay, L., Bondelid, T., Rea, A., Johnston, C., et al.: User Guide (Data Model Version 2.1), 2012.

Mehboob, M. S., Kim, Y., Lee, J., and Eidhammer, T.: Quantifying the sources of uncertainty for hydrological predictions with WRF-Hydro over the snow-covered region in the Upper Indus Basin, Pakistan, Journal of Hydrology, 614, 128 500, 2022.

Miller, M. P., Clark, B. R., Eberts, S. M., Lambert, P. M., and Toccalino, P.: Water priorities for the nation—US Geological Survey integrated water availability assessments, Tech. rep., US Geological Survey, 2020.

Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, 2011.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885–900, 2007.

Naabil, E., Lamptey, B., Arnault, J., Olufayo, A., and Kunstmann, H.: Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa, Journal of Hydrology: Regional Studies, 12, 196–209, 2017.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I?A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., et al.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116, 2011.

Oyler, J. W., Dobrowski, S. Z., Ballantyne, A. P., Klene, A. E., and Running, S. W.: Artificial amplification of warming trends across the mountains of the western United States, Geophysical research letters, 42, 153–161, 2015.

R Core Team, R.: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna accessed on July 3rd, 2023, at https://www.R-project.org, (No Title), 2023.

Rafieeinasab, A., Norouzi, A., Seo, D.-J., and Nelson, B.: Improving high-resolution quantitative precipitation estimation via fusion of multiple radar-based precipitation products, Journal of Hydrology, 531, 320–336, 2015.

Rafieeinasab, A., Gochis, D., Srivastava, I., Dugger, A., Sampson, K., Omani, N., Mazrooei, A., Zhang, Y., Matthew, C., and Jacob, L.: Application of the WRF-Hydro Modeling System for the Conterminous United States at the NHDPlus version 2 Spatial Resolution Using the Bias Adjusted Version of the CONUS404 Atmospheric Forcings (CONUS404BA), Water Years 2010-2021: U.S. Geological Survey data release, https://doi.org/10.5066/P1KZGLU2, 2024.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Rasmussen, R., Chen, F., Liu, C., Ikeda, K., Prein, A., Kim, J., Schneider, T., Dai, A., Gochis, D., Dugger, A., et al.: CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS, Bulletin of the American Meteorological Society, 104, E1382–E1408, 2023.

Read, L., Yates, D., McCreight, J., Rafieeinasab, A., Sampson, K., and Gochis, D.: Development and evaluation of the channel routing model and parameters within the National Water Model, JAWRA Journal of the American Water Resources Association, 59, 1051–1066, 2023.

Robertson, D. E., Chiew, F. H., and Potter, N.: Adapting rainfall bias-corrections to improve hydrological simulations generated from climate model forcings, Journal of Hydrology, 619, 129 322, 2023.

Saxe, S., Farmer, W., Driscoll, J., and Hogue, T. S.: Implications of model selection: A comparison of publicly available, CONUS-extent hydrologic component estimates, Hydrology and Earth System Sciences Discussions, 2020, 1–70, 2020.

Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., and Kunstmann, H.: Fully coupled atmosphere-hydrology simulations for the central M editerranean: Impact of enhanced hydrological parameterization for short and long time scales, Journal of Advances in Modeling Earth Systems, 7, 1693–1715, 2015.

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., and Verdin, J. P.: Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach, JAWRA Journal of the American Water Resources Association, 49, 577–591, 2013.

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., and Pulwarty, R. S.: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data, Water Resources Research, 35, 2145–2160, 1999.

Thornton, P., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R., Vose, R., and Cook, R.: Daymet: daily surface weather data on a 1-km grid for North America, version 3. ORNL DAAC, Oak Ridge, Tennessee, USA, USDA-NASS, 2019. 2017 Census of Agriculture, Summary and State Data, Geographic Area Series, Part 51, AC-17-A, 51, 2016.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resources Research, 43, 2007.

Varlas, G., Katsafados, P., Papadopoulos, A., and Korres, G.: Implementation of a two-way coupled atmosphere-ocean wave modeling system for assessing air-sea interaction over the Mediterranean Sea, Atmospheric Research, 208, 201–217, 2018.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., and Verdin, J. P.: A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET, Remote Sensing of Environment, 139, 35–49, 2013.

Verri, G., Pinardi, N., Gochis, D., Tribbia, J., Navarra, A., Coppini, G., and Vukicevic, T.: A meteo-hydrological modelling system for the reconstruction of river runoff: the case of the Ofanto river catchment, Natural Hazards and Earth System Sciences, 17, 1741–1761, 2017.

Vrugt, J. A. and Ter Braak, C. J.: DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems, Hydrology and Earth System Sciences, 15, 3701–3713, 2011.

Winter, T. C.: The concept of hydrologic landscapes 1, JAWRA Journal of the American Water Resources Association, 37, 335–349, 2001.

Wolock, D. M., Winter, T. C., and McMahon, G.: Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses, Environmental management, 34, S71–S88, 2004.

Xiang, T., Vivoni, E. R., Gochis, D. J., and Mascaro, G.: On the diurnal cycle of surface energy fluxes in the North American monsoon region using the WRF-Hydro modeling system, Journal of Geophysical Research: Atmospheres, 122, 9024–9049, 2017.

https://doi.org/10.5194/hess-2024-262
Preprint. Discussion started: 18 November 2024

Yang, K., Rittger, K., Musselman, K. N., Bair, E. H., Dozier, J., Margulis, S. A., Painter, T. H., and Molotch, N. P.: Intercomparison of snow water equivalent products in the Sierra Nevada California using airborne snow observatory data and ground observations, Frontiers in Earth Science, 11, 1106 621, 2023.

Yin, D., Xue, Z. G., Bao, D., RafieeiNasab, A., Huang, Y., and Morales, M.: A Coupled Numerical Investigation of the Cape Fear River Basin during Hurricane Florence (2018), Authorea Preprints, 2021.

Yin, D., Xue, Z. G., Bao, D., RafieeiNasab, A., Huang, Y., Morales, M., and Warner, J. C.: Understanding the role of initial soil moisture and precipitation magnitude in flood forecast using a hydrometeorological modelling system, Hydrological Processes, 36, e14 710, 2022.

Yucel, I., Onen, A., Yilmaz, K., and Gochis, D.: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall, Journal of Hydrology, 523, 49–66, 2015.

Zhan, S., Song, C., Wang, J., Sheng, Y., and Quan, J.: A global assessment of terrestrial evapotranspiration increase due to surface water area change, Earth's future, 7, 266–282, 2019.

Zhang, Y., Grim, J., Cabell, R., Srivastava, I., Gochis, D., Prein, A., Rasmussen, R., Ikeda, K., and Schneider, T.: CONUS404 climate forcing variable subset for hydrologic models, 1979-2022: downscaled to 1 km and bias-adjusted for precipitation and temperature: U.S. Geological Survey data release, https://doi.org/10.5066/P9JE61P7, 2024.

Zheng, F., Maier, H. R., Wu, W., Dandy, G. C., Gupta, H. V., and Zhang, T.: On lack of robustness in hydrological model development due to absence of guidelines for selecting calibration and evaluation data: Demonstration for data-driven models, Water Resources Research, 54, 1013–1030, 2018.