



## RESEARCH ARTICLE

10.1029/2019JD030767

## Key Points:

- Continental domain fluxes and states derived by two calibrated hydrologic models (mHM and VIC) are evaluated to diagnose model functional deficiencies
- The more skillful (benchmark) model can help to identify the other model's functional deficiencies
- The new VIC parameters based on streamflow calibration exhibit better ET skill compared to two preexisting VIC configurations

## Supporting Information:

- Supporting Information S1

## Correspondence to:

O. Rakovec and N. Mizukami,  
oldrich.rakovec@ufz.de;  
mizukami@ucar.edu

## Citation:

Rakovec, O., Mizukami, N., Kumar, R., Newman, A., Thober, S., Wood, A. W., et al. (2019). Diagnostic evaluation of large-domain hydrologic models calibrated across the contiguous United States.

*Journal of Geophysical Research: Atmospheres*, 124, 13,991–14,007.  
<https://doi.org/10.1029/2019JD030767>

Received 5 APR 2019

Accepted 14 DEC 2019

Accepted article online 20 DEC 2019

Published online 28 DEC 2019

## Author Contributions:

**Conceptualization:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman, Andrew W. Wood, Martyn P. Clark, Luis Samaniego

**Data curation:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman, Stephan Thober, Andrew W. Wood, Martyn P. Clark

**Formal analysis:** Oldrich Rakovec, Naoki Mizukami

**Funding acquisition:** Martyn P. Clark, Luis Samaniego  
(continued)

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States

Oldrich Rakovec<sup>1,2</sup>, Naoki Mizukami<sup>3</sup>, Rohini Kumar<sup>1</sup>, Andrew J. Newman<sup>3</sup>, Stephan Thober<sup>1</sup>, Andrew W. Wood<sup>3</sup>, Martyn P. Clark<sup>3,4</sup>, and Luis Samaniego<sup>1</sup>

<sup>1</sup>UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany, <sup>2</sup>Faculty of Environmental Sciences, Czech University of Life Sciences, Prague, Czech Republic, <sup>3</sup>National Center for Atmospheric Research, Boulder, CO, USA, <sup>4</sup>Now at Coldwater Laboratory, University of Saskatchewan, Canmore, Alberta, Canada

**Abstract** This study presents diagnostic evaluation of two large-domain hydrologic models: the mesoscale Hydrologic Model (mHM) and the Variable Infiltration Capacity (VIC) over the contiguous United States (CONUS). These models have been calibrated using the Multiscale Parameter Regionalization scheme in a joint, multibasin approach using 492 medium-sized basins across the CONUS yielding spatially distributed model parameter sets. The mHM simulations are used as a performance benchmark to examine performance deficiencies in the VIC model. We find that after calibration to streamflow, VIC generally overestimates the magnitude and temporal variability of evapotranspiration (ET) as compared to mHM as well as the FLUXNET observation-based ET product, resulting in underestimation of the mean and variability of runoff. We perform a controlled calibration experiment to investigate the effect of varying number of transfer function parameters in mHM and to enable a fair comparison between both models (14 and 48 for mHM vs. 14 for VIC). Results of this experiment show similar behavior of mHM with 14 and 48 parameters. Furthermore, we diagnose the internal functioning of the VIC model by looking at the relationship of the evaporative fraction versus the degree of soil saturation and compare it with that of the mHM model, which has a different model structure, a prescribed nonlinear relationship between these variables and exhibits better model skill than VIC. Despite these limitations, the VIC-based CONUS-wide calibration constrained against streamflow exhibits better ET skill as compared to two preexisting independent VIC studies.

## 1. Introduction

Large-domain, spatially contiguous hydrologic modeling is used for a myriad of hydrologic applications, including forecasting of floods and seasonal runoff, drought monitoring, and multidecadal hydrologic projection assessments (e.g., Archfield et al., 2015). One important objective of large-domain hydrologic modeling is to simulate water fluxes (e.g., runoff and evapotranspiration [ET]) and states (e.g., soil moisture) across different hydroclimatic regions including both gauged and ungauged basins in a spatially consistent manner, leading to spatially consistent hydrologic assessments over the model domain (Mizukami et al., 2017; Samaniego et al., 2017).

In the contiguous United States (CONUS), one example of a large-domain modeling effort is the North America Land Data Assimilation System (NLDAS; Mitchell et al., 2004; Xia et al., 2012, 2012), which uses multiple land models to simulate continental domain water and energy fluxes. Over the past decade improvements of quality of the NLDAS simulations were attained via improvements of forcing input by ingesting additional sources of climate information (e.g., satellite-based precipitation, shortwave radiation etc.), and model physics improvements (Cai et al., 2014; Xia et al., 2016). One of the critical paths to further improve such large-domain simulation is development of effective methods to estimate spatial fields of model parameters (Xia et al., 2012). The spatial pattern of hydrologic model parameters affects not only model performance but also the spatial pattern of water and energy fluxes (Mizukami et al., 2017; Samaniego et al., 2017).

Estimating spatially distributed hydrologic parameters remains a major challenge in large-domain hydrologic modeling efforts (Archfield et al., 2015; Bierkens, 2015; Clark et al., 2016; Gupta et al., 2014; Melsen

**Investigation:** Oldrich Rakovec, Naoki Mizukami

**Methodology:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman, Stephan Thober, Andrew W. Wood, Martyn P. Clark, Luis Samaniego

**Project administration:** Martyn P. Clark, Luis Samaniego

**Software:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman

**Supervision:** Martyn P. Clark, Luis Samaniego

**Validation:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman

**Visualization:** Oldrich Rakovec

**Writing - original draft:** Oldrich Rakovec, Naoki Mizukami

**Writing - review & editing:** Oldrich Rakovec, Naoki Mizukami, Rohini Kumar, Andrew J. Newman, Stephan Thober, Andrew W. Wood, Martyn P. Clark, Luis Samaniego

et al., 2016; Paniconi & Putti, 2015; Wood et al., 2011). This is because current techniques have difficulty using available spatially distributed geophysical information (e.g., soil and land cover maps) to estimate parameter sets at ungauged basin where model calibration is not possible. Although there are several studies that performed large-domain parameter regionalization (Beck et al., 2016; Oubeidillah et al., 2014; Troy et al., 2008; Yang et al., 2019), the field of large-domain parameter estimation is still immature, and consensus regionalization methods for large domains have not been clearly established. Therefore, many large-domain hydrologic modeling efforts use default (a priori) or partially calibrated parameter sets (e.g., Xia et al., 2012; Wood & Mizukami, 2014).

One promising approach to obtain spatially continuous, calibrated parameter sets is the Multiscale Parameter Regionalization (MPR; Samaniego et al., 2010), which explicitly accounts for subgrid variability of hydrological processes. MPR has been initially implemented in the mesoscale Hydrologic Model (mHM) and several past studies have demonstrated the effectiveness of this method for large-domain parameter estimation (e.g., Kumar et al., 2013; Rakovec et al., 2016b; Samaniego et al., 2017). The first implementation of MPR into the Variable Infiltration Capacity model (VIC; Liang et al., 1994), being one of the four models used in NLDAS, produced an improved spatial pattern of the model parameters (i.e., no spatial discontinuity in parameter fields at the boundaries of calibrated watersheds), however, the evaluation of simulated streamflow revealed limited improvements in streamflow model performance over the use of a default, semi-calibrated parameter set (Mizukami et al., 2017).

The overall goal of this study is to provide a diagnostic evaluation of large-domain model functioning. Specifically, this study seeks to diagnose the internal functioning of the VIC model to reveal the underlying causes leading to limited model predictive skill. We do so by understanding the changing relationships between modeled storage and fluxes (Gharari et al., 2014; Koster & Mahanama, 2012), for example, evaporative fraction versus degree of soil saturation across different model configurations. We focus particularly on how the different model structures and associated parameter estimates affect internal model behavior leading to different water flux simulations by contrasting the VIC and mHM models over the CONUS domain.

The remainder of this paper is structured as follows. Section 2 describes the models (VIC and mHM) and the parameter estimation methods. Section 3 compares the model simulations in terms of basin streamflow and ET (section 3.1), gridded fluxes across the entire CONUS domain are contrasted in section 3.2, while the effect of vegetation and soil parameters on model behavior is analyzed in section 3.3 and 3.4, respectively. Recommendations for VIC model are listed in section 3.5 and finally, conclusions are provided in section 4.

## 2. Data and Methods

### 2.1. Study Area and Data Sets

This study employs the CONUS-wide hydrometeorological data set developed by Newman et al. (2015) and Addor et al. (2017) for model calibration and evaluation. This data set contains daily basin mean meteorological data based on various long-term data sets and observed streamflow data for unimpaired or Hydro-Climata Data Network basins across CONUS. The basin size ranges between 5 and 2,000 km<sup>2</sup> (median size of 340 km<sup>2</sup>). In this study we use 492 basins, for which the mHM basin extraction was possible. The meteorological forcing data (Maurer et al., 2002) include daily average, maximum and minimum temperature, precipitation and wind speed at 12 km resolution (hereafter, the Maurer data set). Temperature and precipitation are based on observation while wind speed is estimated based on National Center for Atmospheric Research-National Centers for Environmental Prediction reanalysis data set. The details for this data set are referred to Maurer et al. (2002).

Spatially distributed geophysical data required for parameter estimation for the hydrologic models include soil texture information (sand and clay percentage) as well as bulk density from the CONUS soil characteristic data set (Miller & White, 1998). The soil data in this data set is extracted from the State Soil Geographic database developed by the U.S. Department of Agriculture's Natural Resources Conservation Service. The soil data set includes soil polygons (310 km<sup>2</sup> of median size with 114 and 825 km<sup>2</sup> of first and third quantiles) with 11 harmonized soil layers down to 2.5 m in depth, for which the model parameters are computed. In addition to the soil texture data, land cover data (land cover type and leaf area index [LAI]) are required for the vegetation parameter estimations. This study uses climatological land cover data set (1981 and 1994) compiled by Maurer et al. (2002), which have been also used for the past hydroclimate assessment study over CONUS.



## 2.2. Hydrologic Models

### 2.2.1. VIC Model

The VIC model configuration is identical to Newman et al. (2017) and Mizukami et al. (2017), which uses VIC 4.1.2 version with water balance simulation mode at a daily time step. This mode simulates water storage and vertical transmission without explicit soil-vegetation heat transfer computation. This VIC configuration was also used for hydrologic applications such as CONUS hydroclimate projection assessments (Wood & Mizukami, 2014). The water balance mode requires only the climate variables in Maurer data set. Although there is no radiation and humidity variables required for the forcing input, VIC water balance mode (prior to VIC version 5 (Hamman et al., 2018)) performs an energy balance for snowpack accumulation and ablation using radiation and humidity estimated based on temperature and precipitation using the Mountain Microclimate Simulation Model (MT-CLIM, Hungerford et al., 1989). VIC also splits each model element into several elevation bands (five bands in this study), where meteorological variables (temperature, precipitation and pressure) are adjusted based on the lapse rates and simulated at each band separately. These features affect the simulations over mountainous regions. Full descriptions on soil water transmission and ET processes are provided by Liang et al. (1994) and Liang et al. (2003), though further details of VIC process representations (e.g., the ET process) are given as needed in later section 3.2. Lastly, to compare VIC runoff to observed streamflow for each basin, instantaneous runoff depth from VIC is routed at a basin outlet using a gamma distribution with two parameters (shape and time scale) as a lumped unit hydrograph.

The calibration soil parameters are also identical to Mizukami et al. (2017); eight VIC soil parameters related to infiltration, baseflow generation, and soil water storage and transmission. This selection of parameters is based on the past VIC modeling studies (e.g., Demaria et al., 2007; Melsen et al., 2016; Nijssen et al., 2001; Oubeidillah et al., 2014; Troy et al., 2008). See supporting information Tables S1 and S2 for more details. Mizukami et al. (2017) discuss the transfer functions implemented in MPR for each VIC soil parameter. In total, 10 transfer function (global,  $\gamma$ ) parameters related to soil parameters are calibrated. In addition to the soil parameters, a multiplier of spatially distributed LAI and stomatal resistance parameters (function of a land cover type) are also calibrated together with the soil global parameters. Streamflow routing related two gamma distribution parameters (with a shape parameter between 0.1 and 3 and a time scale between 0.5 and 7 days) are directly calibrated. In total 14 parameters are calibrated; see supporting information Table S3 for more details.

### 2.2.2. The mHM

mHM is a grid-based spatially distributed hydrologic model that is forced with daily precipitation, temperature and potential ET. It accounts for major hydrologic processes such as snow generation and snowmelt, canopy interception, soil infiltration, ET, deep percolation, baseflow generation, and surface runoff routing. Details on the model processes and parameterization can be obtained from Samaniego et al. (2010) and Kumar et al. (2013). The code is open source and has been published in online repository [git.ufz.de/mhm](https://git.ufz.de/mhm) (Samaniego et al., 2018). This study uses release version 5.7. Several previous studies successfully evaluated the skill of mHM (including streamflow, evaporation, and changes in the terrestrial water storage anomaly) over a large number of river basins world-wide (e.g., Kumar et al., 2013; Huang et al., 2017; Rakovec et al., 2016b; Samaniego et al., 2019).

The model operates at multiple spatial resolutions to account for different scales at which (1) physiological data (e.g., terrain characteristics, soil textural properties, and land cover states) are available ( $\approx 800$  m in this study); and (2) the dominant hydrologic processes can be simulated (12 km in this study). The soil storage of this study is discretized into three vertical layers having the following depth from the Earth's surface: 50, 250, and 2,500 mm. Actual ET from soil is approximated by the potential ET reduced by water stress and root fraction of each soil layer. The Muskingum-Cunge method is used for river routing. To enable fair comparison with the VIC model runs, each basin's meteorological forcing data are spatially averaged. In total, the mHM setup of the CONUS domain has 48 global parameters (provided in supporting information Table S4), which all were calibrated in this study. We carried out sensitivity analysis to demonstrate that the main differences between mHM and VIC do not originate from the different number of parameters being calibrated but rather from structural differences (see supporting information Text S3, corresponding Figures S3–S6, and Table S5).

## 2.3. MPR

The purpose of the MPR is to derive transfer function (global) parameters that are scale-independent, spatially seamless and transferable across locations. MPR offers a framework to bridge the gap between the

field scale (observations) and the catchment (model) scale. MPR was initially proposed by Samaniego et al. (2010) with an initial implementation as a subcomponent of mHM. Ongoing efforts have been undertaken to create a “model-agnostic” MPR, that is, applicable to other models (e.g., Mizukami et al., 2017). Here we provide short description of MPR and we refer the reader to Samaniego et al. (2017) for more details.

First, MPR relies on user-specific transfer functions to convert geophysical properties such as soil texture information (mass fractions of soil primary particles), organic matter content, land cover state (plant functional types, bare soil, water bodies), and topography into model parameters at the native scale of the geophysical data. Second, MPR scales the resulting (native data-scale) model parameters to the desired model resolution, using prescribed scaling operators that vary by parameter. Model optimization is performed via calibrating the transfer function coefficients rather than the model parameters directly. Following the nomenclature of Samaniego et al. (2010), the transfer function coefficients are denoted as  $\gamma$  and the model parameters as  $\beta$  in the following. Parameter values at ungauged locations (e.g., streamflow) can then be estimated with the calibrated transfer functions.

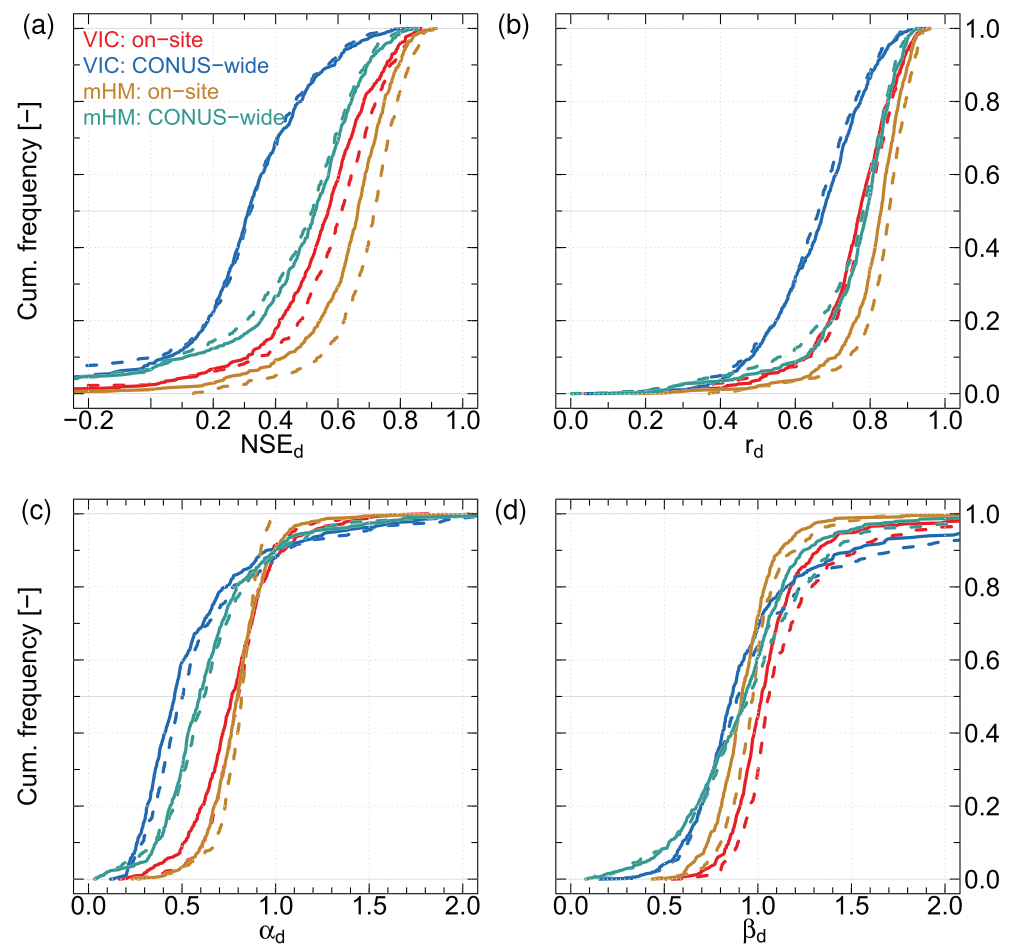
#### 2.4. Large-Sample Basin-Average Calibration and Evaluation

This study follows the approach of Mizukami et al. (2017) for the parameter estimation. Whenever possible, both models are set up in such a way that same physiographic data and model forcing data are used. First, we perform on-site (individual) basin calibrations against observed discharge using MPR yielding an optimized parameter set per basin. This on-site calibration is used solely as a “benchmark” in this paper (Newman et al., 2017). It represents an upper bound of model performance given a priori data and model structure at each basin.

Second, we carry out a joint multibasin calibration against observed discharge to obtain a single unique global parameter set across multiple basins (here 492 basins). This approach is hereafter called “CONUS-wide” calibration. The transfer functions and global parameters calibrated via the CONUS-wide calibration are applied at ungauged locations to generate spatially contiguous (i.e., seamless) parameter fields at 12 km resolution. Note that the joint multibasin calibration produces a compromise solution—a shared optimum for all participating basins. For  $N$  basins, the degree of freedom for the “benchmark” case is  $N$  times larger than the CONUS-wide case; therefore, the benchmark performance is expected to outperform the CONUS-wide case. Notably, the selected basins are biased toward humid locations (61% basins have aridity index  $< 1$ ) along the East and West Coasts, whereas fewer arid basins are used (39% basins have aridity index  $> 1$ ). This might bias the CONUS-wide parameter set toward humid locations.

Both models are calibrated against observed discharge using the Nash-Sutcliffe efficiency (NSE) at a daily time step as the objective function. The model performance is further evaluated against the three components of the Kling-Gupta efficiency (KGE, Gupta et al., 2009): the relative variability in the modeled and observed discharge ( $\alpha$ ), the ratio between the mean of the simulations and mean of the observations ( $\beta$ ), and the timing error represented by the correlation coefficient ( $r$ ). These KGE components are derived from the algebraic decomposition of NSE, therefore they provide insight into the relative contribution of bias, variability error, and correlation to NSE. Both models are optimized with two different automated calibration strategies: The VIC model is calibrated using the Shuffle Complex Evolution (Duan et al., 1993) algorithm (until convergence with 10,000 iterations at maximum), while the mHM is calibrated using the Dynamically Dimensioned Search (DDS; Tolson & Shoemaker, 2007) algorithm (1,000 iterations). These comparable settings correspond well with the recommendations of Tolson and Shoemaker (2007) who demonstrated that “DDS requires only 15–20% of the number of model evaluations used by Shuffle Complex Evolution in order to find equally good values of the objective function”.

To assess the model performance, this study includes a three-step transferability test (based on the pioneering work of Klemesš, 1986): (1) splitting the model time series into independent periods and calculating statistics for each period independently (calibration period: 1 October 1999 to 30 September 2008 and validation periods: 1 October 1989 to 30 September 1999); (2) assessing the fidelity of independent complementary variables other than those used for the parameter estimation (FLUXNET evapotranspiration—ET, Jung et al., 2011); (3) including results on a proxy basin test for model transferability of hydrologic model parameterization across different geographical regions only for mHM (parameter transferability from Europe to CONUS shown in the supporting information Figures S3–S5). The model simulations include additional 10-year period prior to both calibration and validation periods that provide “warm start” model states to initialize the study simulations.



**Figure 1.** Streamflow-based cumulative frequency distributions of (a) Nash-Sutcliffe efficiency, (b) correlation, (c) relative variability between the model and observations ( $\alpha$ ), and (d) ratio between the mean simulations and mean observations ( $\beta$ ) across 492 basins at daily time step. Dash (solid) lines indicate calibration (validation) periods. Different colors show performances for VIC and mHM models.

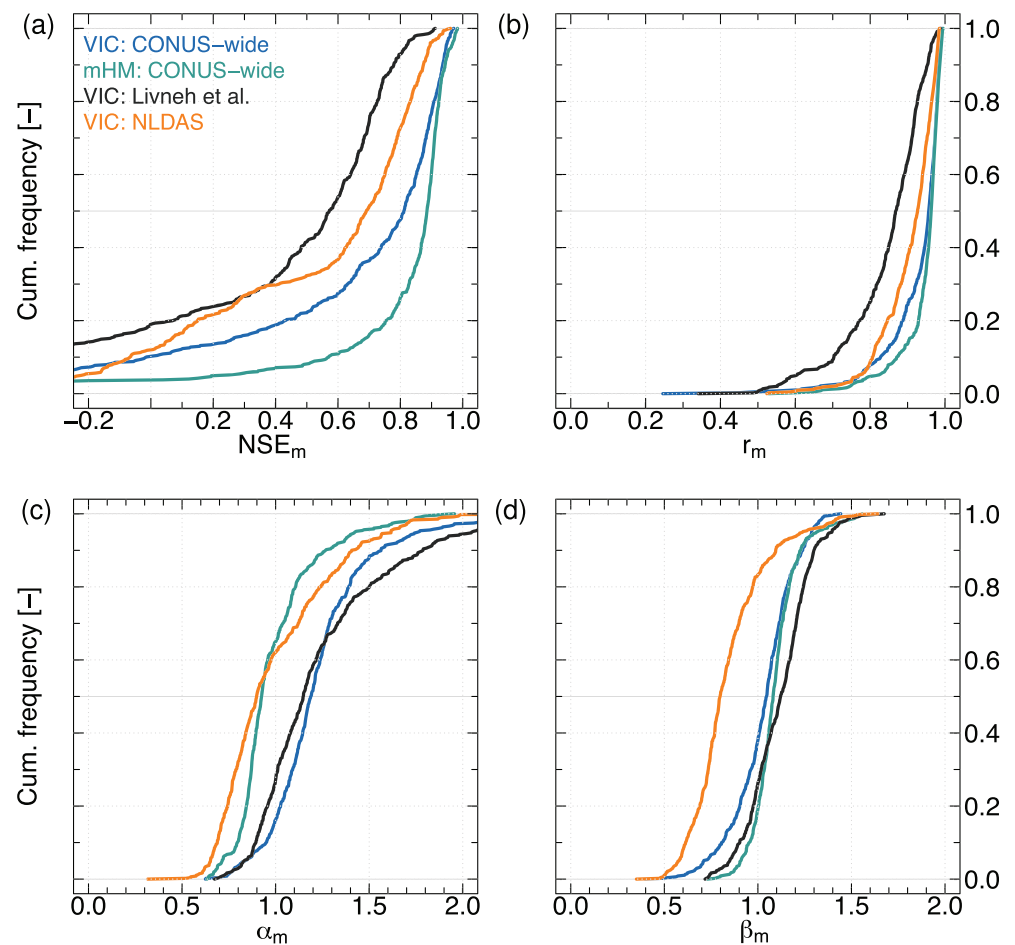
### 2.5. CONUS-Wide Seamless Simulations

To provide an informative picture of the model differences over the CONUS, we perform simulations over the CONUS with the MPR-based CONUS-wide seamless parameter fields at the 12 km spatial resolution (obtained in section 2.4). As with the calibration exercises, the simulations were performed from 1 October 1981 to 30 September 2010. Although the mHM and VIC models are forced using the same precipitation and temperature data, the models' internal methods of quantifying available energy to evaporate and transpire water differ. Additionally the ET analysis includes two additional independent CONUS domain VIC simulations: (1) Livneh 1/16th degree simulation (Livneh et al., 2013), and (2) VIC simulation within NLDAS-2 (Mocko & NASA/GSFC/HSL, 2014; Xia et al., 2012), to analyze whether the underlying differences in models result from numerical implementations. The effect of vegetation on water partitioning is also discussed. Finally, we investigate how the soil parameter fields differ between the two independent parameterizations, and how they influence the range of available water capacity of soil and how they link to evaporation and transpiration processes.

## 3. Results and Discussion

### 3.1. Large-Sample Basin-Average Calibration and Evaluation

Figure 1 shows the streamflow performance of mHM and VIC across 492 basins at a daily time step based on for the NSE, correlation ( $r$ ), relative variability ( $\alpha$ ) and bias ( $\beta$ ). The model performance using the CONUS-wide parameter set is benchmarked (contrasted) with the on-site calibration parameter sets at each basin. The results corresponding to the VIC model are identical to those presented by Mizukami et al. (2017).



**Figure 2.** Evapotranspiration-based cumulative frequency distributions of (a) Nash-Sutcliffe efficiency, (b) correlation, (c) relative variability between the hydrologic models and FLUXNET product ( $\alpha$ ), and (d) ratio between the mean hydrologic simulations and mean FLUXNET product ( $\beta$ ) across 492 basins at monthly time step. Analysis is done between 1982 and 2010. Different colors show performances for VIC and mHM. The VIC-based ET simulations of Livneh et al. (2013) and Xia, Mitchell, Ek, Sheffield, et al. (2012) are shown as an independent reference.

For the benchmark performance (VIC = red and mHM = brown), mHM produces overall higher median NSE scores of 0.72 (0.66) than VIC with 0.61 (0.57) for the calibration (validation) periods. The attribution of NSE differences mostly arises from consistently higher correlation obtained from mHM compared to VIC. The differences between the two models for variability and bias are comparatively smaller (using the on-site parameter sets). For example, mHM underestimates mean values whereas VIC slightly overestimates these (Figure 1d). Both models underestimate the observed variability to a similar extent. The difference between mHM and VIC during the calibration period is approximately of the same order of magnitude when VIC is compared to another benchmark (on-site calibration) study using the SAC-SNOW17 model (Newman et al., 2017 and their figure 2).

Comparison of model performance based on CONUS-wide calibration between VIC (blue) and mHM (green) shows a larger gap between two models in median NSE (mHM: 0.53, VIC: 0.31 for the validation period) than the benchmark case. In other words, CONUS-wide calibration deteriorates model performance compared to benchmark to a larger degree for VIC than mHM. Larger NSE deterioration of VIC performance comes from greater deterioration of the correlation and variability components ( $\alpha$ ). Furthermore, the temporal transferability of the CONUS-wide parameter sets in both models is more robust (i.e., negligible difference between calibration and validation periods) than the on-site calibrations, which are often over-fitted. We also observe that for more than 90% of the basins, both models consistently underestimate the variability, being independent from the calibration approach. This is due to NSE being used as the objective function (Gupta et al., 2008). Other objective functions, such as the KGE, tend to reduce the error in



streamflow variability (Mizukami et al., 2019), however, the choice of objective function is beyond the scope of the current study.

The hydrologic simulations based on the CONUS-wide parameter sets can also be indirectly compared to the NLDAS simulations carried out over 961 basins (size up to 10,000 km<sup>2</sup>) over CONUS (Xia et al., 2012). These NLDAS simulations are based on four models including VIC with its a priori parameter set (i.e., soil hydraulic parameters, such as saturated hydraulic conductivity are derived from soil texture data with pedotransfer functions, but conceptual parameters are regionally constant). At a daily time step, the median NSE of the NLDAS-based VIC is 0.2, while the ensemble mean over the four NLDAS models exhibits an NSE of 0.3. This is actually equal to the median skill of the MPR-based VIC simulations of this study, although the exact basin locations, underlying physiographic data and meteorological forcing data differ. The strength of the mHM-MPR is demonstrated by a median NSE of 0.53 for the unique CONUS-wide parameter set. For more information about streamflow model performance of mHM and VIC at monthly time step, including parameter transferability of mHM from Germany to CONUS see supporting information Text S1 and Figure S1. Spatial distribution of NSE score is shown in supporting information Text S2 and Figure S2.

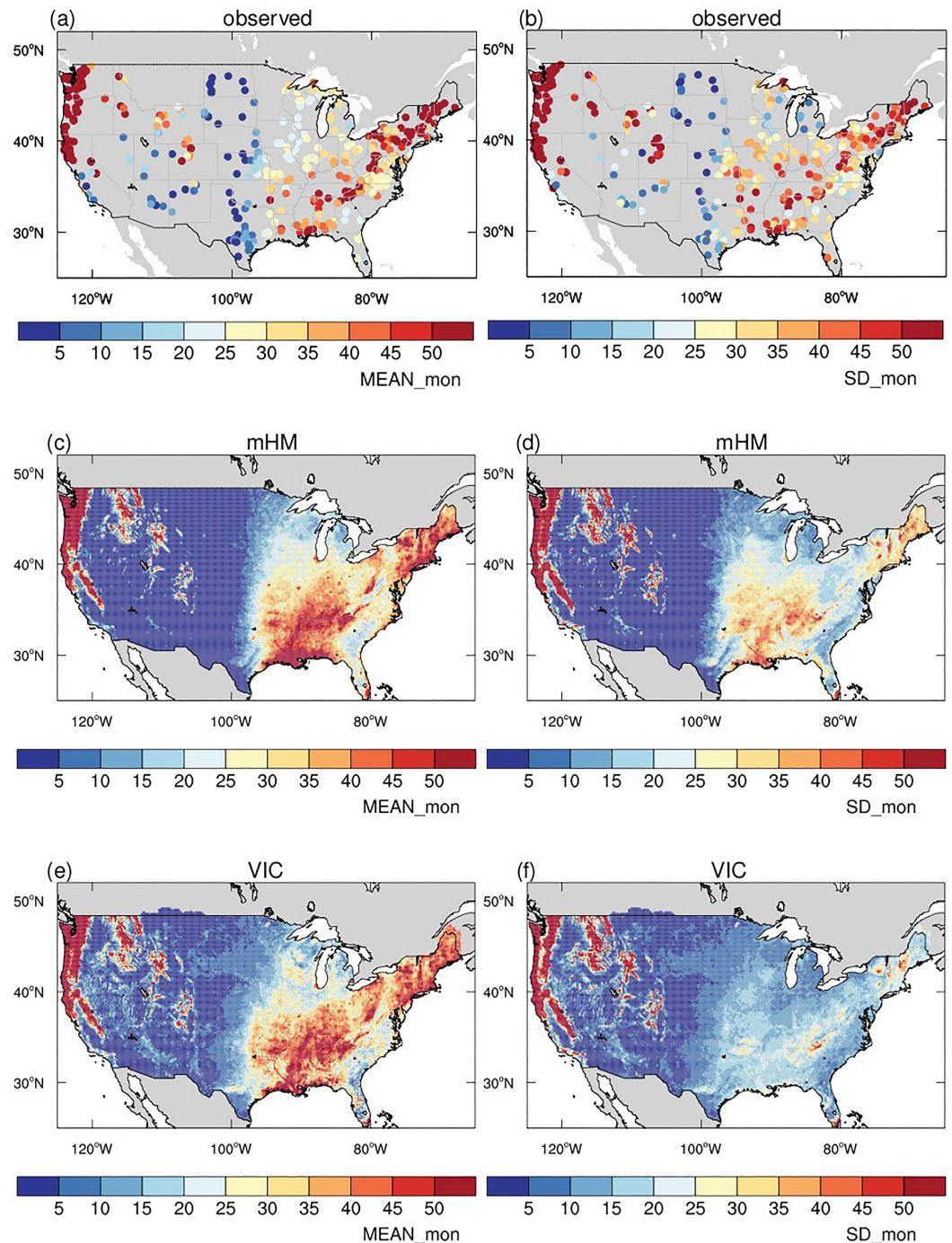
Figure 2 evaluates model ET against the long-term globally available FLUXNET data set (Jung et al., 2011), which was not used for the parameter estimation. Since FLUXNET is available at a monthly step, we computed NSE, and the KGE components similarly to the monthly streamflow analysis. Both VIC and mHM have comparable performance for ET across 492 basins in terms of the correlation ( $r$ ) and bias ( $\beta$ ), however, VIC overestimates the variability. Therefore, ET NSE difference between VIC (median of 0.81) and mHM (median of 0.89) using the CONUS-wide calibrations is primarily caused by overestimated variability in ET. Similar behavior is also observed for the VIC-based ET simulations obtained from Livneh et al. (2013) and Xia, Mitchell, Ek, Sheffield, et al. (2012), which are shown in Figure 2 as independent VIC model applications. Furthermore, note that the VIC run with Livneh et al. (2013) has lower correlation than the VIC with CONUS-wide calibration, leading to even lower NSE (median of 0.58) than for the CONUS-wide VIC implementation of Mizukami et al. (2017) (and this study). Moreover, the NLDAS-based VIC model (Xia et al., 2012) based on default parameterization (and different reanalysis meteorological data) shows low bias for ET at the monthly time step, yielding overall lower performance in NSE (median of 0.7) than the CONUS-wide constrained models.

### 3.2. CONUS-Wide Gridded Fields: Water Fluxes Comparison

The primary goal of the large-domain hydrologic modeling is to generate spatially contiguous and seamless water fluxes and states with improved accuracy. Our goal here is to evaluate the large-domain hydrologic simulations based on the MPR calibration technique applied to two hydrologic models with different structure. Therefore, we compare intermodel differences in water fluxes (total runoff and ET) using the 12 km gridded simulations (Figure 3). Both mHM and VIC produce comparable mean runoff over the eastern part of CONUS and the West Coast, while mHM tends to have less runoff in the western part of CONUS. However, the difference in month-variability of total runoff is more discernible over the entire CONUS between the models. The mHM has higher variability than VIC in the eastern part of CONUS, with less variability in the western part of CONUS at both time scales. Both models tend to underestimate the observed runoff variability in particular in the East Coast of CONUS.

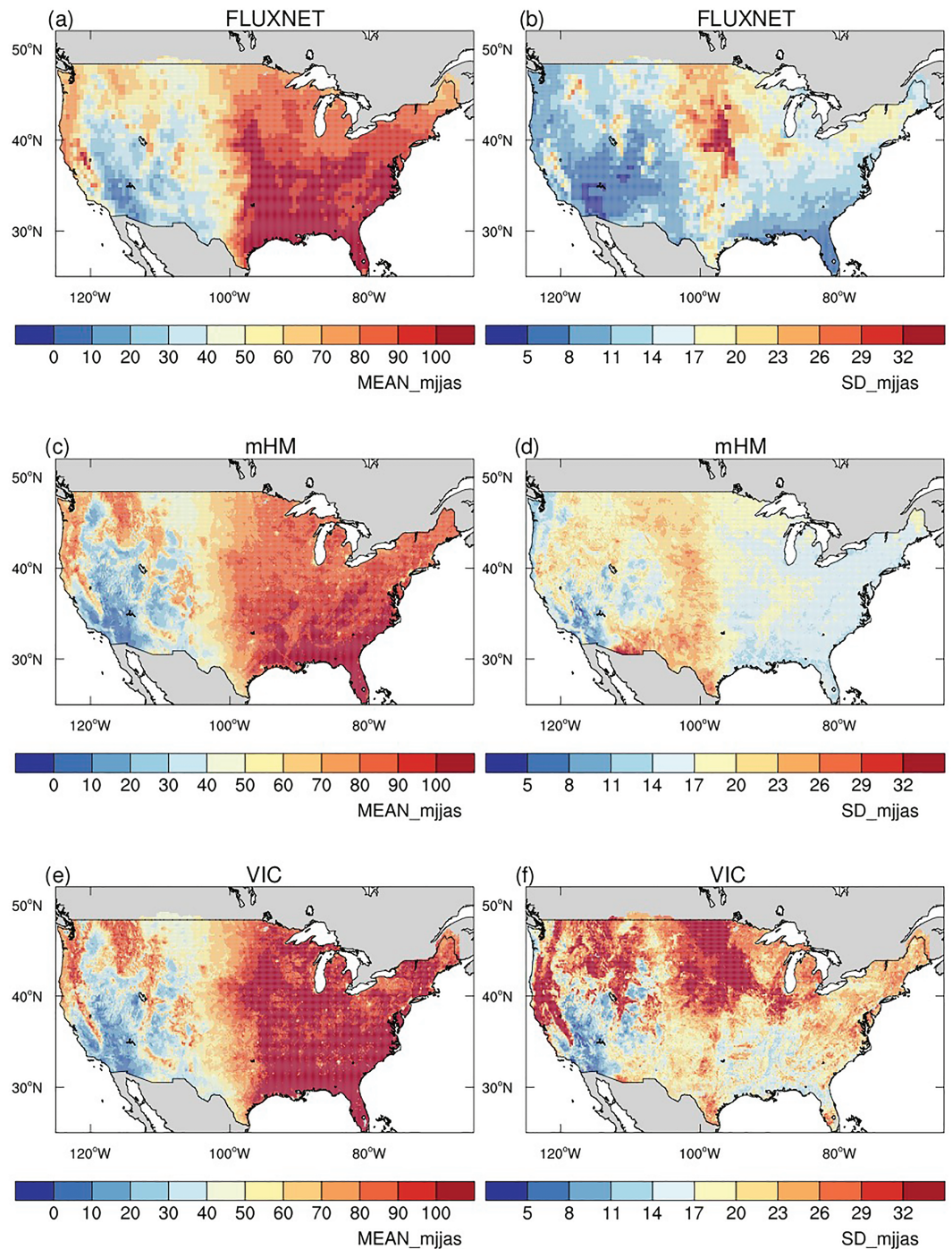
Similarly, Figure 4 compares the mean and variability of the gridded ET FLUXNET product with the ET simulations of mHM and VIC during the summer months from May to September (hereafter referred as MJJAS) when ET is the highest. VIC simulates higher monthly average ET than mHM, by 10 mm per month east of about 100° west, yet mHM and VIC produce similar mean ET over the western CONUS. Moreover, mHM ET shows a latitudinal gradient (i.e., decrease with latitude) over the eastern part of CONUS in a similar manner to FLUXNET, however, that is not present in the VIC ET. The largest variability of ET FLUXNET occurs around 100° west longitude, decreasing in both directions. While mHM resembles the spatial pattern of FLUXNET, there is no similarity in the spatial pattern of ET variability between the two models. VIC has higher summer time ET variability than mHM, in particular in the northern part of the CONUS domain.

The intermodel differences in gridded water flux simulations (i.e., runoff and ET) are consistent with the basin scale comparisons shown in the previous section 3.1. Since ET processes are closely linked to runoff behavior, we further examine the causes of ET differences. The ET difference between the two models can be due to model physics and parameters related to ET and soil water content. For ET parameterizations, mHM uses the concept of potential evapotranspiration (PET) and in this study employs the temperature-based



**Figure 3.** Basin-wide monthly mean (a) and standard deviation (b) of observed runoff. Spatial fields of monthly mean (c) and standard deviation (d) of total runoff for mHM. Spatial fields of monthly mean (e) and standard deviation (f) of total runoff for VIC. All data are shown for the period 1981–2010; units are in mm.

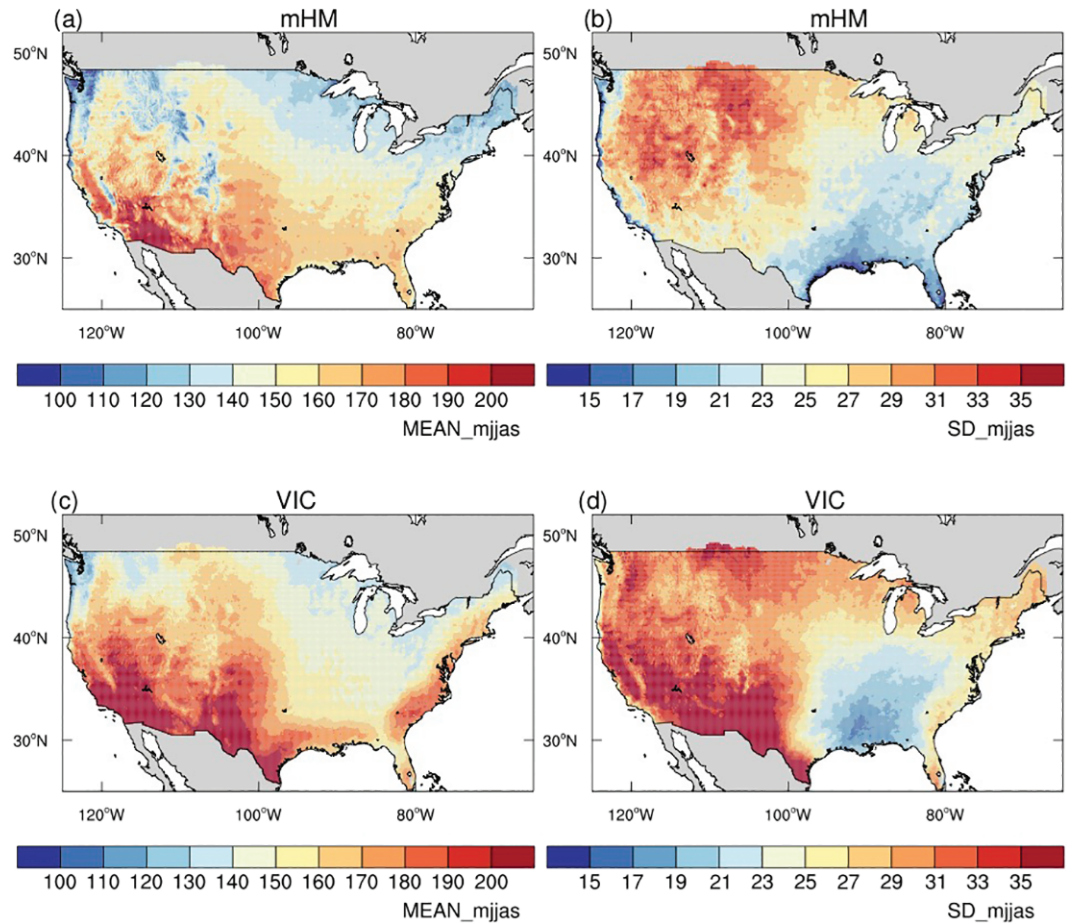
Hargreaves and Samani (1982) method. VIC uses Penman-Monteith (P-M; Allen et al., 1998) with aerodynamic and stomatal resistance unique to land cover tiles to compute actual ET (Liang et al., 1994). Therefore, we first look at energy available for ET prescribed by PET, which may be distinct for each model. Note that while mHM uses temperature for PET, VIC uses radiative energy flux and humidity computed by MT-CLIM from temperature to estimate the available energy for ET. For simplicity, P-M with open water is used to approximate available energy for evaporation for VIC, as it has the highest values out of all PET variants.



**Figure 4.** Spatial fields of actual evapotranspiration for (a, b) FLUXNET product, (c, d) mHM, and (e, f) VIC during the summer months from May to September (MJJAS) in terms of the mean (left column) and standard deviation (right column) during the period 1981–2010.

Figure 5 shows gridded PET forcings for mHM and VIC in terms of the mean and standard deviations for the five summer months (MJJAS) from 1981 to 2010. The most substantial difference in mean PET is located in Southwest through central Great plain, where VIC-based PET is overall higher than mHM-based PET. The ET differences over the same region are not noticeable (Figures 4c and 4e). On the other hand, regions where ET differences are the largest (i.e., eastern CONUS), the difference in mean PET is not substantial. PET variability differences also do not correspond with ET variability. Therefore, we conclude that PET differences are not the primary driver for ET differences between two models.





**Figure 5.** Spatial fields of potential evapotranspiration (PET) for (a, b) mHM, and (c, d) VIC during the summer months from May to September (MJJAS) in terms of the mean and standard deviation (SD) during the period 1981–2010.

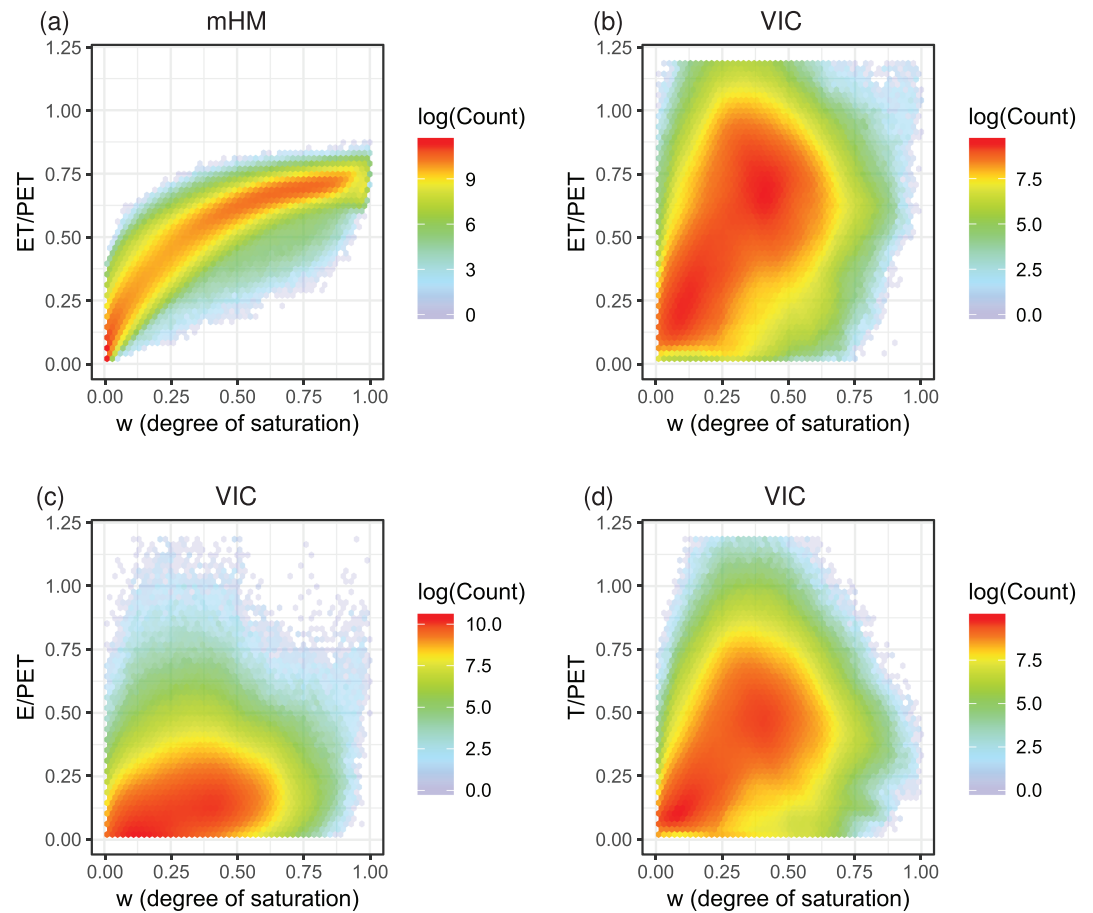
### 3.3. CONUS-Wide Gridded Fields: Effect of Vegetation on VIC-Based ET

Figure 6 shows how the evaporative fraction (ET/PET) depends on soil moisture in terms of the “wilting-point corrected” degree of saturation  $w$  (Koster & Mahanama, 2012), which is based on soil moisture available to vegetation, and is given by

$$w(t) = \frac{\theta(t) - \theta_{\text{pwp}}}{\theta_s - \theta_{\text{pwp}}}, \quad (1)$$

where  $\theta(t)$  is a fractional soil moisture [-] at month  $t$ ,  $\theta_{\text{pwp}}$  is the permanent wilting point [-],  $\theta_s$  is the porosity [-]. All the CONUS grid cells are presented for all individual MJJAS months. Overall, we observe a unique nonlinear relationship between soil moisture and ET for mHM. The internal mHM processes clearly reproduce the common functions used to mimic hydroclimatic variability at the land surface (Koster & Mahanama, 2012). However, the relationship between ET/PET and degree of saturation is linear and far weaker for VIC and also exhibits a large variability around the mean. Interestingly, the ET/PET ratio of VIC exceeds the value of one (for 2% of the grid cells-month). According to VIC-FAQ. (2018): “VIC’s canopy evaporation and sublimation are computed using a canopy resistance of 0, which allows for much higher ET rates than when architectural resistance and LAI are taken into account. Thus, when there is water stored in the canopy, actual ET can exceed PET. This generally happens during a small fraction of the time of a simulation (e.g., immediately following rain events)”. Although there is uncertainty in the PET concept in VIC, it is certain that the variability of actual ET in VIC is too high compared to both FLUXNET data set and mHM simulations accordingly.

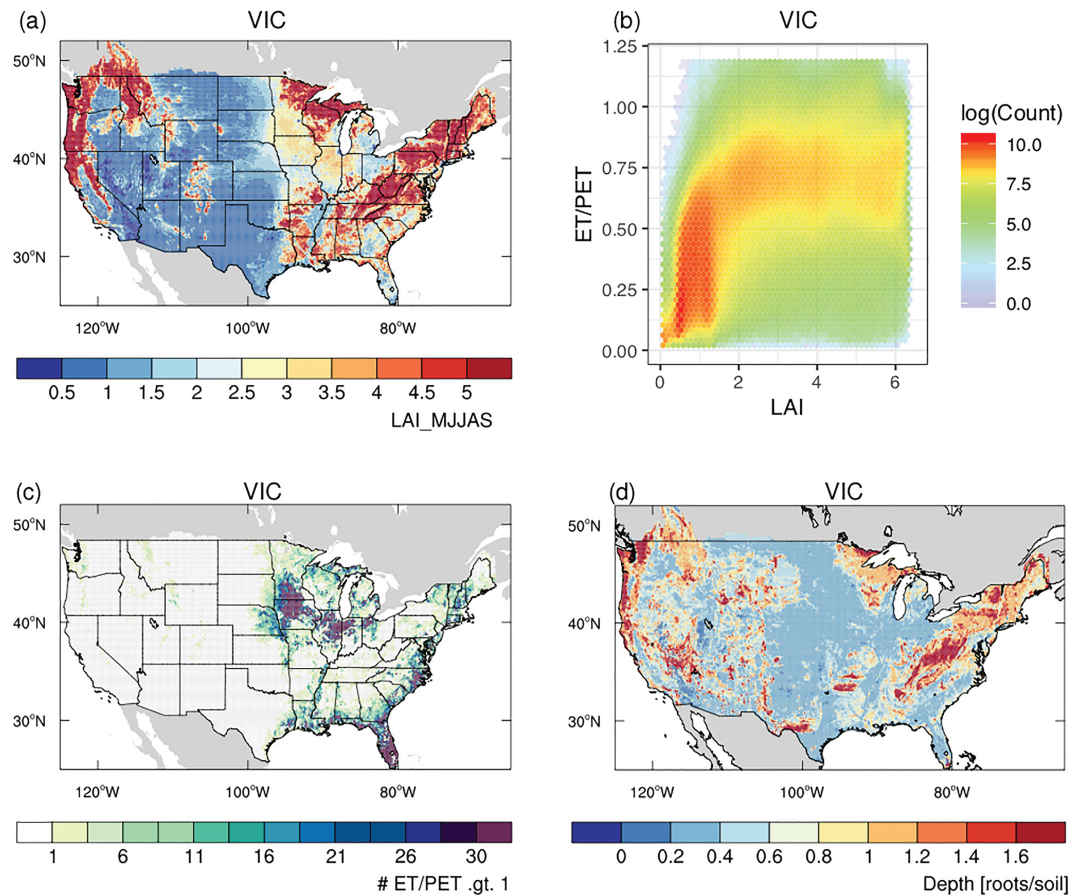




**Figure 6.** Smoothed scatterplot of monthly ET/PET with respect to the degree of soil moisture saturation for all summer months (May–September) during the period 1981–2010 for (a) mHM and (b) VIC. The top two soil layers are considered. The evaporative fraction of VIC is further decomposed into (c) evaporation (E/PET) and (d) transpiration (T/PET) components.

To better understand the causes of this ET behavior, we further decompose the ET/PET of VIC into evaporation (E/PET) and transpiration (T/PET) components. Figures 6c and 6d reveal that transpiration is the dominant component of VIC-based ET. Transpiration in VIC directly relates to the vegetation characteristics. One of the most important vegetation parameters to ET is LAI, which is defined as one half the total green leaf area per unit horizontal ground surface (e.g., Fang & Liang, 2014). The LAI values usually range between zero (bare soil) to approximately six (coniferous forest). The spatial distribution of summer LAI in VIC weighted by vegetation type fraction (Myneni et al., 1997; Maurer et al., 2002) is shown in Figure 7a. Apparent link between LAI and VIC-based ET is visible by comparing this climatological pattern of LAI, however, the heterogeneity in LAI is greater than that of the ET (Figure 4e). Overall, higher LAI and ET values are observed in the eastern part of CONUS and along the West Coast, while lower values are observed in the western part of CONUS. These patterns originate mainly due to the distribution of precipitation (not shown). Figure 7b relates the VIC-based evaporative fraction to LAI from the CONUS-wide perspective. As expected, ET/PET tends to increase along with larger LAI values until a value of three. Furthermore,  $ET/PET > 1$  occurs only for  $LAI > 1.5$ , nevertheless, this ET excess is rather independent from LAI as it is uniformly distributed across LAI values higher than three. Figure 7c depicts the regions of the ET excess across the CONUS domain. The highest abundance is observed in the Midwest, Florida, and along the East Coast, where  $ET/PET > 1$  is exceeded for more than 20% of the MJJAS months (30 out of 150 months), covering 3% of the entire CONUS domain. These locations also correspond to the largest differences between mHM and VIC in terms of the monthly total runoff variability (Figures 3d and 3f).

Other VIC vegetation parameters affecting VIC-based ET which could potentially cause the high ET are the fraction of root depth to soil depth, and the prescribed minimum stomatal resistance (DeFries & Town-



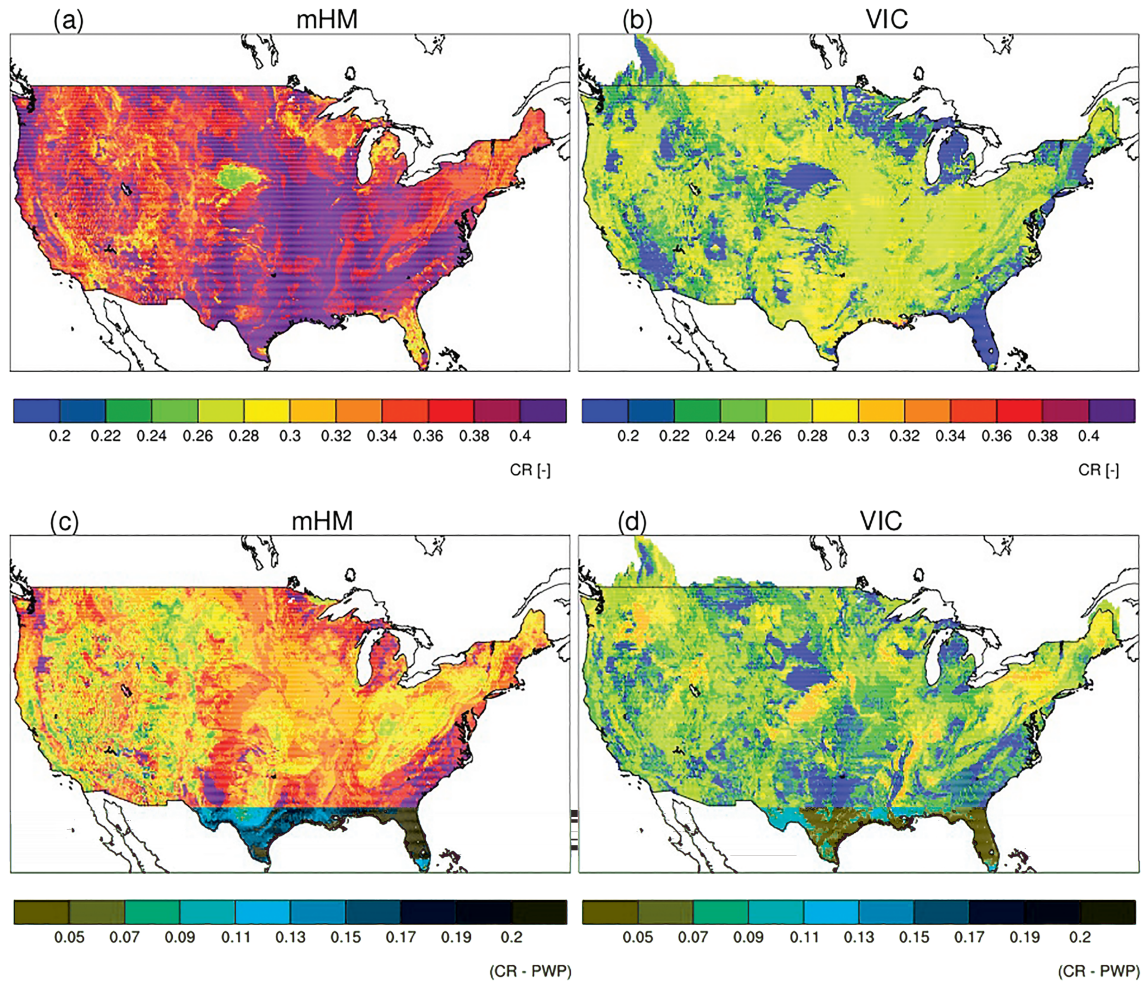
**Figure 7.** Effects of vegetation parameters on ET for VIC: (a) Spatial map of LAI (average over vegetation types) for JMMAS months (Maurer et al., 2002), (b) scatterplot of monthly ET/PET with respect to LAI using entire CONUS domain for all summer months during May–September (1981–2010), (c) abundance of monthly ET/PET > 1 for MJJAS months during 1981–2010 (i.e., ET/PET > 1 in Figure 6b). Values greater than 30 means that ET/PET was exceeded more than 20% of the simulation period. (d) Fraction of root and soil depth.

shend, 1994). However, Figure 7d shows that the spatial pattern of these model parameters does not match the spatial pattern of ET variability in VIC (e.g., compare Figures 7d with the VIC ET maps in Figure 4). Additionally, the depth of the roots in VIC is not constrained by the soil depth, which means that in this VIC parameter set, for 21% of the CONUS domain the roots are deeper than the soil (Figure 7d). Lastly, the minimum stomatal resistance is also not linked to the ET excess (not shown).

### 3.4. CONUS-Wide Gridded Fields: Effect of Soil Parameters

The next important step to better understanding ET behavior is to examine the soil parameters that define the soil water holding capacity available for ET. Such parameters are the critical point of soil ( $\theta_{cr}$ ) and permanent wilting point ( $\theta_{pwp}$ ), which are used in both models. The critical point of soil defines the transitional state below which the soil moisture state becomes unsaturated and thus the ET gets reduced from its full potential. Figures 8a and 8b show the spatial distribution of  $\theta_{cr}$  for mHM and VIC, respectively. Note that  $\theta_{cr}$  in mHM is a result of the transfer function calibration, while default  $\theta_{cr}$  of VIC is derived based on its pedotransfer function (Reynolds et al., 2000), but not constrained in the initial MPR calibration performed by Mizukami et al. (2017) and it represents the default/prior values. We observe that although the spatial pattern of  $\theta_{cr}$  is comparable between both models, the magnitude differs largely across the entire domain (0.3–0.4 for mHM and 0.2–0.3 for VIC).

Furthermore, Figures 8c and 8d depict the range from which water is available for ET as the difference between  $\theta_{cr}$  and  $\theta_{pwp}$  for mHM and VIC, respectively. It shows that VIC has a much narrower range than mHM, meaning there is less capacity of the soil to store the plant available water storage in VIC as compared



**Figure 8.** Spatial pattern of two selected soil parameters: critical point (CR) of (a) mHM and (b) VIC, and dynamic range representing the difference between CR and permanent wilting point (PWP) of (c) mHM and (d) VIC.

to mHM. This narrower range is also likely to contribute to a larger ET variability in VIC compared to mHM.

In VIC, both  $\theta_{cr}$  and  $\theta_{pwp}$  affect stomatal resistance along with LAI as given by

$$r_0(t) = \frac{r_{min}}{LAI * g^{-1}(t)} \quad (2)$$

$$g^{-1}(t) = \begin{cases} 1 & \text{if } \theta(t) \geq \theta_{cr} \\ \frac{\theta(t) - \theta_{pwp}}{\theta_{cr} - \theta_{pwp}} & \text{if } \theta_{cr} > \theta(t) > \theta_{pwp} \\ 0 & \text{if } \theta(t) \leq \theta_{pwp} \end{cases} \quad (3)$$

where  $r_{min}$  is minimum stomatal resistance, and  $\theta_{cr}$  is the fractional water content at the critical point. Where soil moisture is less than the permanent wilting point, transpiration shuts off. Given the narrow range between  $\theta_{cr}$  and  $\theta_{pwp}$  in our VIC calibration, stomatal resistance can vary drastically with a small change in soil moisture, which can contribute to high variability of transpiration. Also, with smaller  $\theta_{cr}$ , the resistance minimizes with a smaller amount of soil moisture, which contributes to high transpiration. Although there is a need for detailed sensitivity tests including these parameters, the ET behavior in VIC is consistent with characteristics of these parameter values (i.e.,  $\theta_{cr}$ ,  $\theta_{pwp}$ , and LAI) that affect stomatal resistance. Finally, we note that underlying geophysical data used for transfer function calibration might have an impact on the simulations results. We refer to earlier works of Livneh et al. (2015) and Zheng and Yang (2016) for more details.



### 3.5. Recommendations for VIC Parameter Estimation

The main motivation of this paper is to diagnose poor performance of the streamflow simulated by VIC-based MPR CONUS-wide calibration (Mizukami et al., 2017). The diagnostic analysis is performed with the aid of mHM-based MPR simulations, which has been shown to perform relatively better in previous studies (Rakovec et al., 2016b, 2016a). The result of the mHM regionalization based on MPR over the CONUS domain is promising for implementation to other models. The MPR implementation for VIC (Mizukami et al., 2017) focused on regionalization of a limited set of soil parameter (based on past VIC applications), did not assess the parameter interactions for other process parameterizations, and did not calibrate vegetation and snow-related parameters. Water balance analysis as performed in the previous sections reveals that VIC ET processes among others are not adequately constrained, which are likely determined by vegetation parameters and the soil parameters not calibrated in the VIC-based MPR. Nevertheless, despite these limitations, the VIC-based MPR CONUS-wide calibration (Mizukami et al., 2017) constrained against discharge exhibits better ET skill as compared to two independent VIC configurations (see Figure 2), which use the default parameterization.

While mHM was developed along with implementations of the transfer functions of each mHM parameter, the transfer function implementation for VIC parameters are done separately from the model development and for selection of parameters. Other soil parameters such as  $\theta_{cr}$  and  $\theta_{pwp}$  have traditionally been derived by user-specified pedotransfer functions and not adjusted further in calibration. In the MPR framework, parameter interdependency is specified with input variables of the transfer functions, which can be other model parameters. For example, parameters  $\theta_{cr}$  and  $\theta_{pwp}$  shown in Figure 8 depend on porosity and matric potential, which both need to be estimated with their own transfer functions prior to the field capacity and wilting point parameters. For VIC, baseflow parameters (D1, D2, and D3) are specified in MPR as functions of saturated hydraulic conductivity (Ks), which is computed with another transfer function. Therefore, it should be important to optimize the transfer function parameters in an integrated manner and the identification of parameter interdependency is important task. Such less comprehensive MPR calibration for VIC (than mHM) can somewhat improve the simulation at an individual basin by compensating for nonconstrained parameters. However, CONUS-wide calibration may have difficulty with finding the parameter sets that explain hydrological process across the diverse climate regions.

For the selection of calibration parameters, it is certain that more comprehensive sensitivity analysis such as Cuntz et al. (2016), who analyzed all the Noah-MP model parameters including the hard-coded parameters are directly informative. In VIC, various piecemeal sensitivity analyses (e.g., Demaria et al., 2007) and extensive applications experience have guided parameter calibration, and the sensitivities of some parameters are understood qualitatively (e.g., LAI, snow roughness, the infiltration parameter, and so on); yet the sensitivities of many soil and vegetation parameters in VIC have to be quantitatively explored. There remains a potential for essentially increasing the variability in the environmental models by adjusting a larger number of traditionally hard-coded parameters (see discussion by Houle et al., 2017; Mendoza et al., 2015, among others). Therefore, we recommend to expose all the VIC model parameters (including the hard-coded ones) to the parameter sensitivity analysis and subsequent optimization in future studies.

Lastly, Samaniego et al. (2017) presents a general protocol to describe how MPR should be implemented in a hydrologic model. One of the important aspects of MPR is the requirement of a flux matching test to determine which scaling operator should be used for each model parameter (i.e., ensuring that parameters have the same effect before and after upscaling). This is particularly important when a transfer function is calibrated in lumped modeling like our approach and applied to other spatial scales (here 12 km). Without appropriate scaling operators for each parameter, 12 km simulations are not guaranteed to preserve either fluxes or simulation skill. This is, however, out of the scope of the present study and these tests at various scales were not carried out. Similarly, multiobjective optimization (e.g., simultaneously based on Q and ET, or any other auxiliary variable), is considered beyond the scope of the present study, although it is likely to yield improved parameter regionalization and thus improved quantification of internal model processes (e.g., Rakovec et al., 2016a).

## 4. Summary and Conclusions

This study has evaluated the model performance of two “seamless” large-domain models, which were parameterized through the MPR technique. This study reveals that although the differences between



both models are marginal in terms of the major water cycle components, subannual fluxes exhibit strong differences in partitioning the precipitation into ET and runoff.

We find substantial differences in the skill for streamflow using two models of different model structures. More importantly, there is a larger gap between individual basin and CONUS-wide calibration in VIC than in mHM. Lower NSE skill in VIC with the CONUS-wide calibration (median NSE of 0.3 at daily time step and 0.4 at monthly time step) than in mHM is due to an overestimation of ET variability, leading to suppressed runoff (or streamflow) variability. The larger error in variability is an important aspect for lower NSE skill which is not attributable to different estimates of PET. Nevertheless, despite these limitations, the VIC-based MPR CONUS-wide calibration (Mizukami et al., 2017) constrained against discharge exhibits better ET skill as compared to two independent VIC setups (see Figure 2), which use the default parameterization. Moreover, VIC-based ET is not fully constrained likely because all sensitive model parameters (vegetation and remaining soil parameters) are not directly calibrated. We suggest a sensitivity analysis and implementation of transfer functions for the additional sensitive VIC parameters such that interparameter dependency can be fully covered.

### Acknowledgments

This work has received funding from the Initiative and Networking Fund of the Helmholtz Association through the project Advanced Earth System Modelling Capacity (ESM) ([www.esm-project.net](http://www.esm-project.net)) and from the U.S Army Corps of Engineers Climate Preparedness and Resilience program. This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. We thank the Editor Ruby Leung and three anonymous reviewers for their constructive comments. We appreciate the constructive discussions from Bart Nijssen (University of Washington) and Ethan Gutmann (NCAR). We thank David Schäfer (UFZ) for technical support with mHM basin boundary delineation and Johannes Brenner (UFZ) for construction of Figure 6. The mHM simulations have been computed at the High-Performance Computing (HPC) Cluster EVE, a joint effort of both the Helmholtz Centre for Environmental Research-UFZ and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. Regarding the VIC simulations, we would like to acknowledge high-performance computing support from Cheyenne (<https://doi.org/10.5065/D6RX99HX>) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Model runs can be obtained online (from <https://doi.org/10.5281/zenodo.2630558>). The content of this study is the sole responsibility of the authors; it does not represent the opinion of the Helmholtz Association, and the Helmholtz Association is not responsible for any use that might be made of the information contained.

### References

- Addor, N., Newman, A., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. Irrigation and Drainage Paper No. 56.
- Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., & Over, T. (2015). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, 51, 10,078–10,091. <https://doi.org/10.1002/2015WR017498>
- Beck, H. E., van Dijk, A. I., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52, 3599–3622. <https://doi.org/10.1002/2015WR018247>
- Bierkens, M. F. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51, 4923–4947. <https://doi.org/10.1002/2015WR017173>
- Cai, X., Yang, Z. L., Xia, Y., Huang, M., Wei, H., Leung, L. R., & Ek, M. B. (2014). Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *Journal of Geophysical Research: Atmospheres*, 119, 13,751–13,770. <https://doi.org/10.1002/2014JD022113>
- Clark, M. P., Schaeffli, B., Schymanski, S., Samaniego, L., Luce, C. H., Jackson, B. M., & Ceola, S. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, 52, 2350–2365. <https://doi.org/10.1002/2015WR017910>
- Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., & Thober, S. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, 121, 10,676–10,700. <https://doi.org/10.1002/2016JD025097>
- DeFries, R., & Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, 15(17), 3567–3586.
- Demaria, E. M., Nijssen, B., & Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research*, 112, D11113. <https://doi.org/10.1029/2006JD007534>
- Duan, Q., Gupta, V., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3), 501–521. <https://doi.org/10.1007/BF00939380>
- Fang, H., & Liang, S. (2014). Leaf area index models. Reference Module in Earth Systems and Environmental Sciences, Elsevier, <https://doi.org/10.1016/B978-0-12-409548-9.09076-X>
- Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. V., & Savenije, H. H. G. (2014). A constraint-based search algorithm for parameter identification of environmental models. *Hydrology and Earth System Sciences*, 18(12), 4861–4870. <https://doi.org/10.5194/hess-18-4861-2014>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463–477.
- Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrol. Processes*, 22(18), 3802–3813. <https://doi.org/10.1002/hyp.6989>
- Hamman, J., Nijssen, B., Bohn, T., Gergel, D. R., & Mao, Y. (2018). The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geoscientific Model Development*, 11, 3481–3496.
- Hargreaves, G., & Samani, Z. (1982). Estimating potential evapotranspiration. *Journal of the Irrigation & Drainage Division - ASCE*, 108(IR3), 225–230.
- Houle, E. S., Livneh, B., & Kasprzyk, J. R. (2017). Exploring snow model parameter sensitivity using Sobol' variance decomposition. *Environmental Modelling & Software*, 89, 144–158.
- Huang, S., Kumar, R., Flörke, M., Yang, T., Hundecha, Y., & Kraft, P. (2017). Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. *Climatic Change*, 141(3), 381–397.
- Hungerford, R. D., Nemani, R., Running, S. W., & Coughlan, J. C. (1989). MT-CLIM: A mountain microclimate simulation model. INT-414, US Forest Service.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., & Williams, C. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, 116, G00J07. <https://doi.org/10.1029/2010JG001566>
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13–24. <https://doi.org/10.1080/02626668609491024>

- Koster, R. D., & Mahanama, S. P. P. (2012). Land surface controls on hydroclimatic means and variability. *Journal of Hydrometeorology*, *13*, 1604–1620. <https://doi.org/10.1175/JHM-D-12-050.1>
- Kumar, R., Livneh, B., & Samaniego, L. (2013). Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme. *Water Resources Research*, *49*, 5700–5714. <https://doi.org/10.1002/wrcr.20431>
- Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*, 360–379. <https://doi.org/10.1029/2012WR012195>
- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14,415–14,428.
- Liang, X., Xie, Z., & Huang, M. (2003). A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model. *Journal of Geophysical Research*, *108*(16), 8613. <https://doi.org/10.1029/2002JD003090>
- Livneh, B., Kumar, R., & Samaniego, L. (2015). Influence of soil textural properties on hydrologic fluxes in the Mississippi River basin. *Hydrological Processes*, *29*, 4638–4655. <https://doi.org/10.1002/hyp.10601>
- Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., & Lettenmaier, D. P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, *26*(23), 9384–9392. <https://doi.org/10.1175/JCLI-D-12-00508.1>
- Maurer, E., Wood, A., Adam, J., Lettenmaier, D., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal Climate*, *15*(22), 3237–3251.
- Melsen, L. A., Teuling, A., Torfs, P., Uijlenhoet, R., Mizukami, N., & Clark, M. P. (2016). HESS opinions: The need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences*, *20*(3), 1069–1079.
- Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., & Uijlenhoet, R. (2016). Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. *Hydrology and Earth System Sciences*, *20*(6), 2207–2226.
- Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, *51*, 716–728. <https://doi.org/10.1002/2014WR015820>
- Miller, D. A., & White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions*, *2*(2), 1–26.
- Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., & Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, *109*, D07S90. <https://doi.org/10.1029/2003jd003823>
- Mizukami, N., Clark, M., Newman, A., Wood, A. W., Gutmann, E., Nijssen, B., & Samaniego, L. (2017). Toward seamless large domain parameter estimation for hydrologic models. *Water Resources Research*, *53*, 8020–8040. <https://doi.org/10.1002/2017WR020401>
- Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for « high-flow » estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. <https://doi.org/10.5194/hess-23-2601-2019>
- Mocko, D., & NASA/GSFC/HSL (2014). NLDAS VIC Land Surface Model L4 monthly 0.125 × 0.125 degree V002, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed: 24 August 2018, <http://doi.org/10.5067/ZO0X4QAX5WTD>
- Myneni, R. B., Ramakrishna, R., Nemani, R., & Running, S. W. (1997). Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Transactions on Geoscience and remote sensing*, *35*(6), 1380–1393.
- Newman, A., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., & Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Newman, A., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., & Nearing, G. (2017). Benchmarking of a physically based hydrologic model. *Journal of Hydrometeorology*, *18*, 2215–2225. <https://doi.org/10.1175/JHM-D-16-0284.1>
- Nijssen, B., Schnur, R., & Lettenmaier, D. P. (2001). Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93. *Journal of Climate*, *14*(8), 1790–1808.
- Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S., & Tootle, G. (2014). A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrology and Earth System Sciences*, *18*(1), 67–84. <https://doi.org/10.5194/hess-18-67-2014>
- Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*, 7090–7129. <https://doi.org/10.1002/2015WR017780>
- Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016a). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*, 7779–7792. <https://doi.org/10.1002/2016WR019430>
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., & Samaniego, L. (2016b). Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology*, *17*, 287–307. <https://doi.org/10.1175/JHM-D-15-0054.1>
- Reynolds, C. A., Jackson, T. J., & Rawls, W. J. (2000). Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, *36*(12), 3653–3662. <https://doi.org/10.1029/2000WR900130>
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale Parameter Regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. <https://doi.org/10.1029/2008WR007327>
- Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Schweppe, R., Schäfer, D., & Attinger, S. (2018). Mesoscale Hydrologic Model (version v5.9). <https://doi.org/10.5281/zenodo.1299584>
- Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., & Attinger, S. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346. <https://doi.org/10.5194/hess-21-4323-2017>
- Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., & Kumar, R. (2019). Hydrological forecasts and projections for improved decision-making in the water sector in Europe. *Bulletin of the American Meteorological Society*, *100*(12). <https://doi.org/10.1175/BAMS-D-17-0274.1>
- Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*, W01413. <https://doi.org/10.1029/2005WR004723>
- Troy, T., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, *44*, W09411. <https://doi.org/10.1029/2007WR006513>

- VIC-FAQ (2018). Variable Infiltration Capacity (VIC) macroscale hydrologic model, FAQ. Accessed: 11 October 2018, <https://vic.readthedocs.io/en/master/FAQ/FAQ/>
- Wood, A., & Mizukami, N. (2014). Project summary report: CMIP5 1/8 degree daily weather and VIC hydrology datasets for CONUS. National Center for Atmospheric Research, <http://www.corpsclimate.us/docs/cmip5.hydrology.2014.final.report.pdf>
- Wood, E. F., Roundy, J. K., Troy, T., Van Beek, L., Bierkens, M. F., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47, W05301. <https://doi.org/10.1029/2010WR010090>
- Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., & Brewer, M. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, 121, 2750–2779. <https://doi.org/10.1002/2015JD023733>
- Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., & Lohmann, D. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, 117, D03110. <https://doi.org/10.1029/2011JD016051>
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., & Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research*, 117, D03109. <https://doi.org/10.1029/2011JD016048>
- Yang, Y., Pan, M., Beck, H. E., Fisher, C. K., Beighley, R. E., Kao, S., Hong, Y., & Wood, E. F. (2019). In Quest of Calibration Density and Consistency in Hydrologic Modeling: Distributed Parameter Calibration against Streamflow Characteristics. *Water Resources Research*, 55(9), 7784–7803. <https://doi.org/10.1029/2018WR024178>
- Zheng, H., & Yang, Z. L. (2016). Effects of soil-type datasets on regional terrestrial water cycle simulations under different climatic regimes. *Journal of Geophysical Research: Atmospheres*, 121, 14,387–14,402. <https://doi.org/10.1002/2016JD025187>