Hydrol. Earth Syst. Sci., 25, 1529–1568, 2021 https://doi.org/10.5194/hess-25-1529-2021 © Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Implications of model selection: a comparison of publicly available, conterminous US-extent hydrologic component estimates

Samuel Saxe^{1,2}, William Farmer¹, Jessica Driscoll¹, and Terri S. Hogue^{2,3}

¹Analysis and Prediction Branch, U.S. Geological Survey, Lakewood, CO 80225, USA
 ²Hydrologic Science and Engineering, Colorado School of Mines, Golden, CO 80401, USA
 ³Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO 80401, USA

Correspondence: Samuel Saxe (ssaxe@usgs.gov)

Received: 8 May 2020 – Discussion started: 5 June 2020

Revised: 16 December 2020 - Accepted: 11 January 2021 - Published: 26 March 2021

Abstract. Spatiotemporally continuous estimates of the hydrologic cycle are often generated through hydrologic modeling, reanalysis, or remote sensing (RS) methods and are commonly applied as a supplement to, or a substitute for, in situ measurements when observational data are sparse or unavailable. This study compares estimates of precipitation (P), actual evapotranspiration (ET), runoff (R), snow water equivalent (SWE), and soil moisture (SM) from 87 unique data sets generated by 47 hydrologic models, reanalysis data sets, and remote sensing products across the conterminous United States (CONUS). Uncertainty between hydrologic component estimates was shown to be high in the western CONUS, with median uncertainty (measured as the coefficient of variation) ranging from 11% to 21% for P, 14% to 26 % for ET, 28 % to 82 % for *R*, 76 % to 84 % for SWE, and 36 % to 96 % for SM. Uncertainty between estimates was lower in the eastern CONUS, with medians ranging from 5 % to 14 % for P, 13 % to 22 % for ET, 28 % to 82 % for R, 53 % to 63 % for SWE, and 42 % to 83 % for SM. Interannual trends in estimates from 1982 to 2010 show common disagreement in R, SWE, and SM. Correlating fluxes and stores against remote-sensing-derived products show poor overall correlation in the western CONUS for ET and SM estimates. Study results show that disagreement between estimates can be substantial, sometimes exceeding the magnitude of the measurements themselves. The authors conclude that multimodel ensembles are not only useful but are in fact a necessity for accurately representing uncertainty in research results. Spatial biases of model disagreement values in the western United States show that targeted research efforts in arid and semiarid water-limited regions are warranted, with

the greatest emphasis on storage and runoff components, to better describe complexities of the terrestrial hydrologic system and reconcile model disagreement.

1 Introduction

A long-term goal of the atmospheric and hydrologic scientific communities has been to produce accurate estimates of the hydrologic cycle across continental and global domains (Archfield et al., 2015; Beven, 2006; Freeze and Harlan, 1969). Various methodologies have been applied to meet this goal, typically in the form of physically based, reanalysis, or remote-sensing-based models. Many of the resulting estimates are made publicly available by an assortment of scientific entities at both continental and global extents across a wide range of spatiotemporal resolutions. These data sets accelerate progress in the atmospheric and hydrologic sciences by filling knowledge gaps in data-sparse regions, reducing the need for time-consuming and computationally expensive modeling, and by providing numerous estimates to apply within ensemble analyses.

Publicly available modeled estimates have been applied to work on water budget analyses (Gao et al., 2010; Pan et al., 2012; Rodell et al., 2015; Smith and Kummerow, 2013; Velpuri et al., 2019; Zhang et al., 2018), the effects of climate change (La Fontaine et al., 2015; G. J. McCabe et al., 2017), and water availability and use (Landerer and Swenson, 2012; Thomas and Famiglietti, 2019; Voss et al., 2013; Zaussinger et al., 2019). Rapid increases in computational power and data accessibility following the advent of early global- and continental-extent hydrologic models in the 1980s and 1990s (Koster and Suarez, 1992; Manabe, 1969; Sellers et al., 1986; Yang and Dickinson, 1996), and increasingly higher-resolution passive and active satellite measurements of both the surface and subsurface of the Earth (Alsdorf et al., 2007; M. F. McCabe et al., 2017), have led to an explosion in the production of multidecadal, continental- to global-extent models estimating all major components of the hydrologic cycle (Peters-Lidard et al., 2018). Accordingly, this has given rise to a multitude of model comparison and evaluation projects throughout the scientific literature.

Earlier model intercomparison projects, such as the Coupled Model Intercomparison Project (CMIP; Covey et al., 2003), the Integrated Hydrologic Model Intercomparison Project (IH-MIP; Kollet et al., 2017; Maxwell et al., 2014), the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013), and the Water Model Intercomparison Project (WaterMIP; Haddeland et al., 2011), provide robust analyses aimed at attributing differences in model process representation to differences in model design, parameter selection, and meteorological forcings.

Detailed comparisons between smaller selections of models are often included in studies presenting novel products (Daly et al., 2008; Senay et al., 2011; Velpuri et al., 2013; Xia et al., 2012a, b). Furthermore, a range of studies have focused solely on the validation of third-party models, either examining multiple water budget components simultaneously (Gao et al., 2010; Sheffield et al., 2009) or focusing on specific components such as precipitation (P; Derin and Yilmaz, 2014; Donat et al., 2014; Guirguis and Avissar, 2008; Prat and Nelson, 2015; Sun et al., 2018), evapotranspiration (ET; Carter et al., 2018; McCabe et al., 2016), snow water equivalent (SWE; Broxton et al., 2016; Chen et al., 2014; Dawson et al., 2016, 2018; Essery et al., 2009; Mudryk et al., 2015; Rutter et al., 2009; Vuyovich et al., 2014), or soil moisture (SM; Brocca et al., 2011; Koster et al., 2009; Xia et al., 2014).

Most of these comparison studies have revolved around error metrics derived through validation against in situ and ex situ measurements. However, validation often yields unsatisfactory representations of model skill. Observational data sets are often spatiotemporally discontinuous, and model estimates in grid structure may not sufficiently represent the sub-grid heterogeneity that is present in point data, especially in topographically and ecologically complex regions. Furthermore, observational measurements can have significant associated uncertainty (Di Baldassarre and Montanari, 2009). Validation literature is heavily focused on the comparison of error statistics and spatial summaries of model skill and rarely discusses the effect of model disagreement on research results.

This research seeks to supplement past comparison and validation literature by evaluating how uncertainty between publicly available water budget component estimates used within a study can control outcomes and conclusions across a range of ecological regimes. We differentiate this work from previous model intercomparison and validation studies by placing results in the context of multiple water budget analyses. We seek to quantify how estimates of component magnitudes and long-term trends differ in the conterminous United States (CONUS) and within ecologically distinct regions. The primary goal of this work is to quantify model disagreement in terms of magnitude, interannual trend, and correlation against remote sensing (RS) products. The effects of model disagreement are shown in regional water budget analyses. We undertake a robust comparison of P, ET, R, SWE, and SM estimates generated through hydrologic models, reanalysis data sets, and remote sensing products. We iteratively calculate water budget imbalances in eight regions by applying a range of flux estimates to quantify how model selection may impact residuals.

2 Methods and data

2.1 Data categories

Hydrologic estimates are divided into water budget flux components of P, ET, and R, and storage components of SWE and SM, representing the primary fluxes and stores of a surface water budget. Data sets were selected by prioritizing public availability, ease of access, and relative use within the research community. These data sets are subdivided into loosely defined categories of hydrologic models, reanalysis data sets, and remote-sensing-derived products. References and spatiotemporal information for each data set are provided in Table 1, and more detailed long-form descriptions are provided in Appendix A.

2.1.1 Hydrologic models

The hydrologic model category (Table 1; Appendix A) includes any estimates generated using equations or concepts attempting to represent real-world hydrology. Model output differences are strongly controlled by forcing data sets (Elsner et al., 2014; Mizukami et al., 2014), calibration methods (Mendoza et al., 2015), applied equations (Clark et al., 2015a), model structure (Clark et al., 2015a, b), and geophysical parameter availability (Beven, 2002; Bierkens, 2015; Mizukami et al., 2017). The most common hydrologic models this study evaluates are conceptual and physically based. Conceptual models derive terrestrial hydrology estimates through empirical relationships between fluxes and stores, typically through a water balance model (WBM) in the style of Thornthwaite (1948; NHM-MWBM and TerraClimate). Physically based models utilize meteorological forcing data and solve equations describing physical conservation laws of mass, energy, and momentum. These can be further subdivided by targeted hydrologic variables as follows: land surface models (LSMs) target land-atmosphere interactions, especially ET, while catchment models (CMs) target stream**Table 1.** Summary of data sets used in this research, including assigned data category, abbreviated name, primary organization, literature reference, spatiotemporal resolution, and sourced hydrologic components. The reader is referred to Appendix A for definitions of abbreviated model names, descriptions, and further references. Components are precipitation (P), evapotranspiration (ET), runoff (R), snow water equivalent (SWE), and soil moisture in equivalent water depth and volumetric water content (SM(e) and SM(v), respectively). Hydrologic models NMH-MWBM and NMH-PRMS are based on a delineated (i.e., non-gridded), topographically derived spatial framework composed on hydrologic response units (HRUs). The reanalysis product WaterWatch is generated at hydrologic units (HUs) 2–8. The finest-resolution product, HU8, is used in this study.

Data set	Group	Reference	Spat	iotemporal	Components
Hydrologic model					
CPC	CPC	Fan and van den Dool (2004)	1/2°	1948-present	SM(e)
CSIRO-PML ^c	CSIRO	Zhang et al. (2016)	1/2°	1981-2012	ET
ERA5/H-TESSEL ^c	ECMWF	Muñoz Sabater (2019)	1/4°	1979-present	R, SWE, SM(v)
ERA5-Land/H-TESSEL	ECMFW	Muñoz Sabater (2019)	1/10°	2001-present	ET, R , SWE, SM(v)
GLDAS-CLM ^c	NASA	Rodell et al. (2007)	1°	1979-present	ET, R, SWE
GLEAM ^{a, c}	U. of BE	Martens et al. (2017)	1/4°	1980-present	ET, $SM(v)$
JRA-25/SiB	JMA	Onogi et al. (2007)	110 km	1979-present	R, SWE, SM(e)/(v)
JRA-55/SiB ^c	JMA	Kobayashi et al. (2015)	55 km	1957-present	R, SWE, SM(v)
Livneh-VIC	CIRES	Livneh et al. (2013)	1/16°	1915-2011	SWE
MERRA-Land/CLSM ^c	NASA	Reichle et al. (2011)	$1/2^{\circ}$	1980-2016	ET, R, SWE
MERRA-2/CLSM ^c	NASA	Reichle et al. (2017)	1/2°	1980-present	ET, R , SWE, SM(v)
NCEP–DOE/Eta-Noah ^c	NOAA	Kanamitsu et al. (2002)	210 km	1979-present	R, SWE
NCEP-NARR/Eta-Noah	NOAA	Mesinger et al. (2006)	32 km	1979-present	SWE
NHM-MWBM ^c	USGS	McCabe and Markstrom (2007)	HRU	1949–2010	ET, <i>R</i> , SWE, SM(<i>e</i>)
NHM-PRMS ^c	USGS	Regan et al. (2018)	HRU	1980-2016	ET, <i>R</i> , SWE, SM(<i>e</i>)
NLDAS2-Mosaic ^c	NASA	Xia et al. (2012a)	1/8°	1979-present	ET, <i>R</i> , SWE, SM(<i>e</i>)
NLDAS2-Noah ^c	NASA	Xia et al. (2012b)	1/8°	1979-present	ET, R , SWE, SM(e)
NLDAS2-VIC ^c	NASA	Xia (2012c)	1/8°	1979–present	ET, R , SWE, SM(e)
SNODAS	NWS	Barrett (2003)	1 km	2003-present	SWE
TerraClimate ^c	U. of ID	Abatzoglou et al. (2018)	1/24°	1958–present	ET, R , SWE, SM(e)
VegET ^c	USGS	Senay (2008)	1 km	2000–2014	ET, $SM(e)$
Reanalysis					
CanSISE	U. Toronto	Mudryk and Derksen (2017)	1°	1981-2010	SWE
CMAP ^{b, c}	CPC	Xie and Arkin (1997)	2 1/2°	1979-present	Р
Daymet ^c	ORNL	Thornton et al. (2018)	1 km	1980-present	P, SWE
ERA5 ^c	ECMWF	Muñoz Sabater (2019)	1/4°	1979-present	Р
ERA5-Land	ECMWF	Muñoz Sabater (2019)	1/10°	2001-present	Р
GPCC ^c	GPCC	Becker et al. (2013)	1/2°	1901-2013	Р
gridMET ^c	U. of ID	Abatzoglou (2013)	1/24°	1979-present	Р
Livneh et al. (2013) ^c	NOAA	Livneh et al. (2013)	1/16°	1915-2011	Р
Maurer et al. (2002)	U. WA	Maurer et al. (2002)	1/8°	1950–1999	Р
MERRA-Land ^c	NASA	Rienecker et al. (2011)	1/2°	1980-2016	Р
MERRA-2 ^c	NASA	Gelaro et al. (2017)	$1/2^{\circ}$	1980-present	Р
NCEP-DOE ^c	NCEP-DOE	Kanamitsu et al. (2002)	210 km	1979-present	Р
NLDAS2 ^c	NASA	Xia et al. (2009)	$1/8^{\circ}$	1979-present	Р
PRISM ^c	OSU	PRISM Climate Group (2004)	4 km	1895-present	Р
Reitz et al. (2017) ^c	USGS	Reitz et al. (2017)	800 m	2000-2013	ET
UoD-v5 ^c	U. of DE	Willmott and Matsuura (2001)	1/2°	1950-1999	Р
WaterWatch ^c	USGS	Jian et al. (2008)	HU8	1901-present	R

Table 1. Continued.

Data setGroupReferenceSpatiotemporalComponentsRemote sensingAMSR-E/AquaNSIDCTedesco et al. (2004) $25 \mathrm{km}$ $2002-2011$ SWEESA-CCIESADorigo et al. (2017) $1/4^{\circ}$ $1978-\mathrm{present}$ SM(v)GPCP-v3^cNASAHuffman et al. (2019) $1/2^{\circ}$ $1983-2016$ P MOD16-A2^cU. of MTRunning et al. (2017) $1 \mathrm{km}$ $2000-2010$ ETSMOS-L4FNCSRAl Bitar et al. (2013) $25 \mathrm{km}$ $2010-2017$ SM(v)SSEBop ^c USGSSenay et al. (2013) $1 \mathrm{km}$ $2000-2014$ ETTMPA 3PA2SNASAHuffman et al. (2010) $1/4^{\circ}$ $1008 \mathrm{mrsout}$ P						
Remote sensing AMSR-E/Aqua NSIDC Tedesco et al. (2004) 25 km $2002-2011$ SWE ESA-CCI ESA Dorigo et al. (2017) $1/4^{\circ}$ $1978-\text{present}$ SM(v) GPCP-v3^c NASA Huffman et al. (2019) $1/2^{\circ}$ $1983-2016$ P MOD16-A2^c U. of MT Running et al. (2017) 1 km $2000-2010$ ET SMOS-L4 FNCSR Al Bitar et al. (2013) 25 km $2010-2017$ SM(v) SSEBop ^c USGS Senay et al. (2013) 1 km $2000-2014$ ET TMPA 2842 ^S NASA Huffman et al. (2010) $1/4^{\circ}$ 1908 rescent P	Data set	Group	Reference	erence Spatiotemporal		Components
AMSR-E/Aqua NSIDC Tedesco et al. (2004) 25 km $2002-2011$ SWE ESA-CCI ESA Dorigo et al. (2017) $1/4^{\circ}$ 1978 -present SM(v) GPCP-v3^c NASA Huffman et al. (2019) $1/2^{\circ}$ $1983-2016$ P MOD16-A2^c U. of MT Running et al. (2017) 1 km $2000-2010$ ET SMOS-L4 FNCSR Al Bitar et al. (2013) 25 km $2010-2017$ SM(v) SSEBop ^c USGS Senay et al. (2013) 1 km $2000-2014$ ET TMPA NASA Huffman et al. (2010) $1/4^{\circ}$ 1008 , erroort B	Remote sensing					
P	AMSR-E/Aqua ESA-CCI GPCP-v3 ^c MOD16-A2 ^c SMOS-L4 SSEBop ^c TMPA-3B43 ^c	NSIDC ESA NASA U. of MT FNCSR USGS NASA	Tedesco et al. (2004) Dorigo et al. (2017) Huffman et al. (2019) Running et al. (2017) Al Bitar et al. (2013) Senay et al. (2013) Huffman et al. (2010)	25 km 1/4° 1/2° 1 km 25 km 1 km 1/4°	2002–2011 1978–present 1983–2016 2000–2010 2010–2017 2000–2014 1998–present	SWE SM (v) P ET SM (v) ET P

^a Versions 3.3a and 3.3b. ^b Standard and enhanced versions. ^c Data sets used in the water budget case study with complete data for water years 2001–2010.

flow (Archfield et al., 2015). Additionally, LSMs are often 1D models that operate at discrete spatial intervals, typically grid cells, lacking horizontal transfer of surface or subsurface water between regions. CMs, on the other hand, utilize 2D model structures by routing overland and subsurface flows between spatial domains to better realize surface and groundwater estimates using prescribed stream networks. The CLSM, described in the literature as a land surface model (Koster et al., 2000), is discussed in conjunction with the NHM-PRMS because of the sub-grid catchment network used to model horizontal runoff and streamflow fluxes. LSMs are the most common hydrologic model here (e.g., CLM, H-TESSEL, Mosaic, Noah, SiB, and VIC) because they are often coupled with global-extent atmospheric research and, thus, more commonly operated at the spatial extent of this study, while CMs are more typically operated at the catchment or basin scale. Additional products included in the hydrologic model category are those using simplified WBMs falling outside the conceptual model paradigm (CPC, CSIRO-PML, GLEAM, and VegET) and those using component-specific physically based models (SNODAS).

2.1.2 Reanalysis

Reanalysis data sets (Table 1; Appendix A) assimilate multisource in situ and ex situ observational data into spatiotemporally continuous 4D estimates of continental- or global-scale atmospheric and meteorological fluxes using numerical algorithms. This category includes reanalysis data sets derived solely from in situ and ex situ data (e.g., CMAP, Daymet, Maurer et al., 2002, and UoD-v5) and those assimilating or blending multiple reanalysis products with or without observational measurements (e.g., CanSISE, gridMET, Livneh et al., 2013, and NLDAS2). In past studies, reanalysis models were often grouped and defined separately from gridded data sets derived through statistical interpolation of in situ observations (e.g., WaterWatch, Reitz et al., 2017, and GPCC). We combine the two categories, defining the single reanalysis category as containing products not generated solely from remote sensing observations or through terrestrial hydrologic models. The reasons for this are two-fold, i.e., (1) to simplify reporting of results and discussion across water budget components, and (2) although interpolation-based products are often used as reference data sets in model validation studies, especially for precipitation, accuracy decreases in regions with sparse observations and complex topography (Hofstra et al., 2008) and, thus, should be considered modeled products.

2.1.3 Remote sensing derived

The RS-derived data sets (Table 1; Appendix A) are components of the hydrologic system derived from passive and/or active ex situ observations. We note the use of the term RSderived rather than RS because none of the data sets used here can be truly described as direct observations as they are always relying on a range of modeling techniques to create hydrologic estimates. For example, the AMSR-E/Aqua SWE product uses a sequence of steps including a snow-detection routine, followed by a physical retrieval algorithm, which is then validated against observational SWE data (Chang and Rango, 2000), and the SSEBop product uses a radiationdriven energy balance model to estimate ET (Senay et al., 2013). Remote sensing data sets include estimates modeled from single-sensor ex situ observations (MOD16-A2 and SSEBop), though most utilize an assembly of information from both passive and active sensors at various spatiotemporal resolutions (e.g., ESA-CCI and GPCP-v3).

2.2 Data processing

Data sets included in this research cover a broad spectrum of spatial resolutions, extents, coordinate reference systems, and timescales and, therefore, require a uniform spatiotemporal system to better facilitate comparison. To that end, gridded (e.g., NLDAS2-Mosaic) and polygonal (e.g., NHM-PRMS) data sets were aggregated by area-weighted mean to 10 Environmental Protection Agency Ecoregions, Level I (Omernik and Griffith, 2014), that encompass the Watershed Boundary Dataset hydrologic units (U.S. Geological **Table 2.** Sizes of study ecoregions in square kilometers (km^2) and percent (%) of study domain.

Ecoregion	Area (km ²)	Percent of CONUS (%)
	(1111)	001100 (10)
Marine West Coast Forest	91 035	1.11
Mediterranean California	168 757	2.06
Northwestern Forested Mountains*	859 801	10.52
North American Deserts*	1 533 536	18.76
Southern Semiarid Highlands	66 699	0.82
Temperate Sierras	110 038	1.35
Great Plains*	2 328 673	28.48
Eastern Temperate Forests*	2 561 292	31.33
Northern Forests	433 032	5.30
Tropical Wet Forests	22 469	0.27
Total domain area	8 175 332	100 %

* Primary ecoregions highlighted during water budget analysis.

Survey, 2016; Fig. 1; Table 2) over the CONUS. Data sets were processed in their native coordinate system to avoid interpolation of raw values, and the reference system of the ecoregions' spatial data set was transformed to match that of each target data set. Flux and storage terms were aggregated to monthly time steps. Flux values (P, ET, and R) were summed from hourly or daily when necessary, and storage values (SWE and SM) were averaged.

Units of *P*, ET, *R*, and SWE are uniformly presented as equivalent water depth in millimeters (mm). SM is provided by data sets as either equivalent water depth (mm) or volumetric soil moisture content (m³ per cubic meter) and denoted as SM(e) or SM(v), respectively. The SM(e)and SM(v) categories are merged during results and discussion when magnitude is irrelevant, such as with long-term trend and correlation against RS products. All data generated through these methods are available in an associated data release hosted on the USGS ScienceBase (Saxe et al., 2020).

2.3 Statistics

2.3.1 Annual uncertainty

Uncertainties in hydrologic component values are measured as the variability between modeled estimates using the coefficient of variation (CV) calculated as follows:

$$CV = \frac{\sigma}{\overline{x}} \times 100\%,\tag{1}$$

where σ is the standard deviation across all products of a given component in a given water year, and \overline{x} is the associated mean. The σ and CV statistics were calculated for each water year (WY; a water year is the 12 month period of 1 October through 30 September and is designated by the calendar year in which it ends) between available estimates for all components. Annual values of flux terms (*P*, ET, and

R) were derived by summing monthly values to WYs, and storage terms were derived by averaging monthly values by WY. Incomplete WYs (*n* months <12) were discarded. Because of incompatibility between data set temporal ranges, σ and CV values were divided into two 15 WY periods, namely the early period consisting of WYs 1985–1999 and the late period consisting of WYs 2000–2014.

Uncertainty metrics are impacted by the number of models available for each calculation. That is, increasing the number of models typically increases the intermodel variability. Depending on the water year, there can be 14–16 *P* models, 11–17 ET models, 12–15 *R* models, 14–20 SWE models, 7–9 SM(*e*) models, and 5–7 SM(*v*) models available. To reduce the effects of model count disparity in variability measures, CV and σ are calculated through a bootstrap approach that calculates the CV and σ for all possible combinations of five models (from the minimum SM(*v*) availability) by water year, and returns final CV and σ as the mean of the bootstrapped combinations.

2.3.2 SWE timing

Variability in seasonal timing of SWE estimates was compared for trends of both accumulation and ablation in terms of relative timing, defined here as the difference between the antecedent month of snow accumulation or ablation for each data set from the mean antecedent month of the remaining data sets, calculated for each Julian year from 1985 to 2014. The antecedent accumulation (ablation) month is defined as the first month of the June-December (January-May) period, where change in SWE is greater than (less than) 1 mm per month (-1 mm per month). Some data sets contained anomalies in January and February, showing a large negative change in SWE followed by positive rates of SWE accumulation for 1-2 more months, which introduced a negative bias into relative timing values. To eliminate these biases, ablation timing was only considered for months subsequent to the final date of the positive SWE rate of change. Resulting annual relative timing values are presented for each SWE product grouped by (a) early and late periods, (b) accumulation and ablation, and (c) the Northern Forests and Northwestern Forested Mountains ecoregions. To simplify the comparison of SWE timing between trends and ecoregions, annual directions of relative timing were summarized by percent direction, defined here as the percentage of years with a positive relative timing value, as follows:

$$p^{+} = \frac{|\mathbb{R}^{+}|}{|\mathbb{R}|} \times 100\%,$$
 (2)

where p^+ is percent direction, $|\mathbb{R}^+|$ is the cardinality (number of elements in a set) of positive, real relative timing values, and $|\mathbb{R}|$ is the cardinality of all real values. Within the figure presenting summary values (Fig. B4), percent direction values less than 50 % (i.e., those dominantly negative)



Ecoregions over CONUS Extent

Figure 1. Distribution of the 10 ecoregions covering the CONUS study domain.

are converted to a scale of 50 % to 100 % with the following:

$$pD = \begin{cases} p^+ & p^+ > 50\\ 100\% - p^+ & p^+ < 50 \end{cases},$$
(3)

where pD is the directionalized percent direction colored according to the dominant direction value. Positive pD uses a green gradient, and negative pD uses a purple gradient.

2.3.3 Interannual trend

Both the Mann–Kendall trend test (τ ; Kendall, 1938) and Sen's slope estimator (Sen, 1968) are used to identify and measure monotonic trends in annual values over WYs 1982-2010. Trend significance is evaluated using p values (p) in a binary significance test, assuming an alpha (α) of 0.05 and a null hypothesis of no monotonic trend. Under the condition of $p < \alpha$, the null hypothesis is rejected and the alternative hypothesis, i.e., that of the presence of a monotonic trend, is assumed; if $p > \alpha$, the null hypothesis is not rejected. Disagreement in the presence of a significant trend and trend direction is quantified using the unalikeability coefficient (u), which measures how often categorical variables differ on a $0 \le u \le 1$ scale, with 0 and 1 being complete agreement and disagreement, respectively (Kader and Perry, 2007). This study compares trends between 11 to 16 different products $(11 \le n \le 16)$, depending on water budget component, and uses categorical values (c) of (a) significant negative trend, (b) significant positive trend, and (c) no significant trend, derived from the direction (sign) and significance ($\alpha = 0.05$) of τ . When n > c, maximum possible *u* decreases from one exponentially to an asymptote of two-thirds as the ratio of *n* : *c* increases. With a n:c ranging from 11:3 to 16:3, maximum possible u in this trend analysis is approximately 0.71.

Spearman's rho (Spearman, 1904) is applied to quantify the correlation between RS components P, ET, SWE, and SM and the analogous components of hydrologic models and remote sensing data sets. Spearman's rho (ρ), a nonparametric rank correlation metric, is used to estimate correlation, producing values on a -1 to 1 scale, where 1 is perfect positive correlation and -1 is perfect negative correlation, as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)},\tag{4}$$

where *d* is the difference between ranks, and *n* is the sample size. A binary significance test is used to test correlation statistics, assuming $\alpha = 0.05$, and a null hypothesis that there is no relationship between data sets. If $p < \alpha$, the null hypothesis is rejected and a significant correlation between the data sets is assumed; if $p > \alpha$, the null hypothesis is not rejected. Correlation is computed and assessed along the monthly time step, requiring a minimum of 48 months of temporal overlap between the modeled estimates and remote sensing data set.

2.3.4 Water budget imbalances

Water budgets are calculated assuming a steady-state system and are solved for imbalances by the following:

$$P = AET + R + \varepsilon, \tag{5}$$

$$\varepsilon = P - \operatorname{AET} - R,\tag{6}$$

where ε is imbalances. Imbalances, ε , cannot be accurately defined as residuals. In reality, ε is the sum of excluded fluxes

in addition to model uncertainty (residuals). Excluded fluxes include both natural (e.g., groundwater recharge and changes to long-term storage) and anthropogenic (e.g., groundwater extraction) hydrologic processes. Relative imbalances ($R\varepsilon$) are calculated to better compare water budget results between regions of varying hydrologic flux by weighting water budget ε against the total input *P* as follows:

$$R\varepsilon = (\varepsilon/P) \times 100\%. \tag{7}$$

Water budget relative imbalances were calculated from summed P, ET, R, and ε over the 10 water year period of 2001–2010 for 15 P, 15 ET, and 13 R estimates (noted in Table 1) with temporally continuous monthly data for 10 ecoregions. Each ecoregion yielded 2925 water budgets by iterating through all possible combinations of models, totaling 29 250 water budgets in primary regions of the CONUS. In addition, a single water budget was calculated for each of the eight ecoregions using the ensemble means of the model estimates over the same period.

3 Results

3.1 CONUS domain average

3.1.1 Magnitude variability

Uncertainties between annual estimates of states and fluxes of the hydrologic cycle are substantial when averaged over the CONUS for each studied component (Fig. 2). The actual magnitude of model estimate differences averaged by water budget component, measured as σ , is similar for P, ET, and R (Fig. 2a) and low (6.6 mm per year) for SWE. However, comparing uncertainty relative to mean magnitude (Fig. 2b), measured as CV, shows that the vertical, atmospheric-controlled fluxes of P and ET demonstrate lower intermodel uncertainty than the horizontal flux of R or storage components SWE and SM. The CV statistic shows how, due to the low overall magnitude of SWE, small differences in estimates have a substantial effect on magnitude. SM, which is often a large overall storage term in the water budget, also shows high CV, indicating that uncertainty in estimates of this component may have the largest impact on hydrologic analyses.

Unfortunately, high-magnitude differences in SM (Fig. 3e, f) are likely strongly controlled by model-defined soil layer depth and, thus, limit the utility of disagreement statistics. For example, the TerraClimate and NLDAS2-Noah models apply spatially invariant root zone soil depth across the model domains (Abatzoglou et al., 2018; Xia et al., 2012a), whereas the NHM-PRMS assumes a variable depth according to vegetative and geophysical parameters (Regan et al., 2018). SM(e) estimates are directly controlled by soil depth, returning values of equivalent water depth. SM(v) differences are likely to be less influenced by soil depth due to inherent measurements of fractional volume rather than depth.

Mean Annual Variability of Component Estimates over the CONUS



Figure 2. Mean annual (a) standard deviation (σ) and (b) coefficient of variation (CV) calculated between CONUS-extent modeled estimates of hydrologic components precipitation (*P*), evapotranspiration (ET), runoff (*R*), snow water equivalent (SWE), soil moisture in equivalent water depth (SM(*e*)), and soil moisture in volumetric water content (SM(*v*)). Whiskers on each bar represent the standard deviation of annual values.

However, soil water profiles can change significantly with depth, so attributing model differences strictly by root zone depth definition is more difficult.

Box plots of modeled estimates by water budget component (Fig. 3) demonstrate the annual ranges of magnitude generated by various products. Precipitation models (Fig. 3a) exhibit low overall variability, with median magnitudes falling within a distinct 100 mm per year (700–800 mm per year) range. The exceptions to this are the CMAP reanalysis data sets, which have a median 647 mm per year.

Variability among ET estimates (Fig. 3b) is higher, falling within a 200 mm per year (435–650 mm per year) range, though attributing differences to model type (e.g., LSMs vs. CMs) is difficult. Generally, CMs (MERRA-2/CLSM, MERRA-Land/CLSM, and NHM-PRMS) and conceptual WBMs (NHM-MWBM, GLEAM, and TerraClimate) produce greater annual rates of ET compared to LSMs (NLDAS2-VIC, GLDAS-CLM, and NLDAS2-Noah). Exceptions to this are the LSMs NLDAS2-MOSAIC and H-TESSEL that are more similar in annual ET flux to CMs and WBMs. The RS data sets agree more at the CONUS extent with lower-magnitude estimates.

Estimates of R show three distinct clusters of CONUS magnitudes (Fig. 3c) of 130–190, 220–242, and 310–345 mm per year. The NLDAS2-Noah and NLDAS2-VIC LSMs, which produced some of the lowest ET rates, fall within the higher-magnitude R clusters, as do the JRA-driven SiB models and NHM-PRMS model. The lowest magnitude estimates are the NLDAS2-Mosaic and ERA5-driven H-



Annual Component Magnitudes by Model

Figure 3. Box plots of annual water year magnitudes of all study model estimates averaged over the CONUS extent for hydrologic components (**a**) precipitation (*P*), (**b**) evapotranspiration (ET), (**c**) runoff (*R*), (**d**) snow water equivalent (SWE), (**e**)soil moisture in units of equivalent water depth (SM(e)), and (**f**) soil moisture in units of volumetric water content (SM(v)). Data set magnitudes are subdivided into two periods, i.e., 1985–1999 (top) and 2000–2014 (bottom), for each model. Flux component estimates (*P*, ET, and *R*) are summed from monthly values to calculate annual water year rates. Storage component estimates (SWE, SM(e), and SM(v)) are averaged from monthly values to calculate annual water year average storage values. Box lower limits, midlines, and upper limits represent the 25th, 50th (median), and 75th percentiles, respectively, of the associated data. Whiskers represent 1.5 times the interquartile range. Box color denotes data set categories of hydrologic model, reanalysis, or remote sensing derived. Asterisks are used to identify the following hydrologic models types: land surface models (****), catchment models (***), water balance models (**), and miscellaneous (*).

TESSEL LSMs, which are grouped with the CM MERRA-Land/CLSM and WBM TerraClimate. LSMs are more likely to estimate greater R than WBMs or CMs.

Box plots of annually averaged SWE monthly values (Fig. 3d) show that WBMs, notably the NHM-MWBM, generate much higher SWE than most other data sets. Alternatively, there are few discernible patterns between the remaining products, with LSMs and CMs interspersed throughout the gamut of median values. To generalize, the LSMs Noah, GLDAS-CLM, and NLDAS2-Mosaic produce lower-magnitude estimates of monthly SWE. A total of three different Noah estimates, driven with different meteorological forcings and run both independently (NLDAS2-Noah)

or as part of a larger model (e.g., NCEP-NARR/Eta-Noah), all estimate lower SWE relative to other data sets. The two VIC LSMs show contrasting median monthly SWE magnitude over the CONUS, with the Livneh-VIC median exceeding the NLDAS2-VIC median by more than 200 %. The RS AMSR-E/Aqua product agrees more with lower-magnitude estimates at the CONUS extent.

Across all water budget components, most data sets demonstrate lower-magnitude values in the late period (WYs 2000–2014) compared to the early period (WYs 1985–1999). Within the P flux, almost all data sets agree on a decrease in the magnitude of hydrologic fluxes from the early to late periods. This is similarly reflected in the flux estimates of

ET and R, and, to a less uniform extent, the storage components of SWE and SM. Several data sets, however, exhibit increased estimates from the early to late periods. For example, the NCEP–DOE reanalysis precipitation, and corresponding runoff derived by forcing the Noah LSM within the Eta atmospheric model, show increased median annual flux rates from the early to late periods.

3.1.2 Trends

Only 19 out of 87 component estimates exhibit statistically significant CONUS average domain trends, evaluated with Sen's slope, from 1982 to 2010 (Fig. 4). Within most components, one or more data sets produced a significant slope that is contradicted by one or more data sets. For example, the *R* estimates from LSMs GLDAS-CLM, JRA-55/SiB, and ERA5/H-TESSEL each show a significant negative trend in annual values across the CONUS over the 1982–2010 period. Conversely, the NCEP–DOE/Eta-Noah LSM shows a significant positive trend over the same period, and the remaining data sets, a mix of hydrologic and reanalysis models, show no significant trend. While the NCEP–DOE reanalysis precipitation trend matches that of the NCEP–DOE forced Eta-Noah, model estimates of SWE indicate a significant negative trend.

3.2 Ecoregions

3.2.1 Magnitude variability

Uncertainty in modeled hydrologic component estimates in each ecoregion is presented in terms of CV (Fig. 5) and σ (Appendix A). Ecoregions are organized from west to east within each figure. Results and discussion of uncertainty are focused primarily on CV due to the disparity in annual water flux between regions. By component, *P* and ET estimates demonstrate lower uncertainty compared to other hydrologic components (Fig. 5a, b). *P* estimates typically range in median uncertainty from 5% to 14% in the eastern CONUS and 11% to 21% in the western CONUS. *P* uncertainty is highest in mountainous regions with more variable topography (e.g., Northwestern Forested Mountains and Temperate Sierras) and lowest in more humid, topographically homogeneous regions (e.g., Eastern Temperate Forests).

The median uncertainties of ET estimates typically range from 13% to 22% in the eastern CONUS ecoregions and 14% to 26% in the western CONUS. Uncertainty is greatest along the Pacific coast and northerly ecoregions and arid and semiarid ecoregions that are heavily influenced by anthropogenic water use. Of the four largest ecoregions (Table 2), Northwestern Forested Mountains and North American Deserts demonstrate mean uncertainties of 21% and 19%, respectively, and the Great Plains and Eastern Temperate Forests demonstrate mean uncertainties of 14% each. Estimates of *R* (Fig. 5c) generally exceed 29 % variability throughout the study ecoregions, with median values reaching or exceeding 72 % in arid, snow-sparse western ecoregions of North American Deserts, Southern Semiarid Highlands, and Temperate Sierras, as well as in Tropical Wet Forests influence by tropical storm systems. In the large Northwestern Forested Mountains and Great Plains ecoregions, median data set variability is 28 % to 55 % and 47 % to 81 %, respectively.

SWE estimate uncertainty is highlighted for the snowdominated Northwestern Forested Mountains and Northern Forests ecoregions (Fig. 5d), where annual peak SWE values exceed estimated snowpack in the remaining ecoregions by up to 700 % (Fig. B2). Of primary interest is the Northwestern Forested Mountains ecoregion that contains the snowmelt-dominated regimes of the Rocky, Sierra Nevada, and Cascade mountain ranges that are strong contributors to water supply for population centers such as Denver, Los Angeles, San Francisco, Portland, and Seattle. In this ecoregion, modeled estimates of monthly mean SWE storage vary by 76 % to 84 %, equating to median equivalent water depth standard deviations as high as 48 mm per month.

SM(e) estimate uncertainty (Fig. 5e) is greatest, 66 %-96%, in arid western ecoregions, though other regions fall within a 64 %–83 % uncertainty range. SM(v) shows lower overall uncertainty (Fig. 5f), with less dramatic differences between regions, ranging from 36 % to 57 %, with the greatest uncertainty (44 %–57 %) in central and eastern CONUS ecoregions. As mentioned previously, uncertainty in SM estimates, most importantly SM(e), is strongly influenced by model-defined root zone soil thickness. Spatial differences in SM(e) uncertainty measured as CV, therefore, are pronounced in regions with lower water content because CV, a measure of relative variability, is more strongly affected by magnitude differences. Because the range of values is constrained from 0 to 1 in m³ per cubic meter, uncertainty in SM(v) is less influenced by spatial variability in soil water magnitudes and is therefore a better measure for understanding regional soil water estimate disagreement. Following this logic, SM(v) uncertainty is greater in Tropical Wet Forests, Northern Forests, and Great Plains ecoregions, with an average CV of 53 % compared to an average of 42 % for the remaining ecoregions.

3.2.2 SWE accumulation and ablation

Timing of SWE estimates is presented in terms of relative timing (Figs. B3 and B4). Figures are divided by Northern Forests and Northwestern Forested Mountains ecoregions, trends of accumulation and ablation, and early and late periods. Negative (positive) values in purple (green), identify models for which accumulation or ablation begins more than 1 month earlier (later) than the mean of other data sets. Distributions of relative timing by year are presented in Fig. B3 and summarized as the percentage of years that are positive



Interannual Data set Trends from 1982-2010

Figure 4. Forest plot of interannual (water year) component estimate Sen's slope from 1982 to 2010 for components of precipitation (P), evapotranspiration (ET), snow water equivalent (SWE), soil moisture in units of equivalent water depth (SM(e)), and soil moisture in units of volumetric water content (SM(v)). Insignificant trends (p>0.05) are gray. Significant trends are colored based on direction, where negative trends are gold and positive trends are green. Data sets without complete data during the study period are represented with hollow points, and the temporal extent is noted under "Years in Period" column.



Figure 5. Box plots of uncertainty, measured as annual coefficient of variation (CV), between component estimates at 10 ecoregions. CV distributions are subdivided into late and early periods, i.e., 1985–1999 and 2000–2014. Box lower limits, midlines, and upper limits represent the 25th, 50th (median), and 75th percentiles, respectively, of the associated data. Whiskers represent 1.5 times the interquartile range. Box colors denote the ecoregion and correspond to the map colors in Fig. 1. Components displayed are precipitation (*P*), evapotranspiration (ET), runoff (*R*), snow water equivalent (SWE), soil moisture in units of equivalent water depth (SM(e)), and soil moisture in units of volumetric water content (SM(v)). Results for SWE uncertainty are omitted for ecoregions with limited annual snowpack.

or negative in Fig. B4. In both ecoregions, uncertainty among data sets is greater in ablation timing than accumulation timing, and overall uncertainty is greater in the Northwestern Forested Mountains ecoregion than in the Northern Forests.

Relative timing between models can be similar between ecoregions (Figs. B3a and B4), especially regarding accumulation. For instance, the AMSR-E/Aqua RS and NHM-MWBM CM consistently show later accumulation start dates than other data sets in both ecoregions, while the ERA5-Land/H-TESSEL, JRA-55/SiB, and SNODAS models estimate earlier accumulation dates. Conversely, the JRA-25/SiB and TerraClimate models show different trends in accumulation timing between regions with earlier SWE accumulation in Northern Forests but later accumulation in Northwestern Forested Mountains.

Regarding ablation, outliers are much more common when comparing models (Figs. B3b and B4). For example, the Daymet model consistently estimates a much later start of spring ablation while the Eta-Noah models driven with NCEP–DOE and NCEP-NARR reanalyses typically estimate much earlier ablation times, often 1–2 months after the mean antecedent month (Fig. B3). Uncertainty in ablation timing is typically much higher in the Northwestern Forested Mountains ecoregion than in the Northern Forests. The Daymet, ERA5-Land/H-TESSEL, Livneh-VIC, NLDAS2-VIC, and SNODAS models commonly estimate a later start to ablation than the remaining models.

Models that predict earlier (later) accumulation timing and later (earlier) ablation timing represent longer (shorter) periods of growing or available snowpack (Fig. B4). The ERA5-Land/H-TESSEL, Livneh-VIC, and SNODAS models estimate longer snowpack periods than other models in both ecoregions. The AMSR-E/Aqua, JRA-55/SiB, NHM-PRMS, and TerraClimate models estimate longer snowpack periods in only one ecoregion. The NCEP-NARR/Eta-Noah, NLDAS2-Mosaic, and NLDAS2-Noah models estimate shorter snowpack periods in the Northern Forests ecoregion while the GLDAS-CLM and JRA-25/SiB models estimate shorter snowpack periods in the Northwestern Forested Mountains ecoregion.

Attributing the relative timing of SWE to model type is difficult at the monthly scale. LSMs estimate both earlier and later accumulation and ablation antecedences, and the two WBMs in this study show opposing timing trends. Similarly, grouping by organization does not yield significant similarities. For example, the timing of NLDAS2-driven LSMs from the National Aeronautics and Space Administration (NASA) is dissimilar from that of the MERRA-2-driven and MERRA-Land-driven CLSM, which are also from NASA. Forcing data also do not yield useful information for identifying controlling variables. Only the NLDAS2-driven models from NASA tend to all show later (earlier) accumulation (ablation) timing. Even the Eta-Noah LSMs, driven by the NCEP-DOE reanalysis and the corresponding higherresolution version for North America NCEP-NARR show different relative timing values in both ecoregions.

3.2.3 Trend

Interannual trend analyses performed with the Mann– Kendall trend test (τ) over water years 1982–2010 show varying degrees of agreement and disagreement between ecoregions. Figures showing explicit distributions of trends with model names are supplied in the appendix (Fig. B5). Model disagreement in trend direction is quantified using the unalikeability coefficient (u; Fig. 6), with 0 and 1 being complete agreement and disagreement, respectively.

Precipitation estimates demonstrate the lowest *u*, showing no (u = 0) to low (u = 0.13) disagreement in eight ecoregions where data sets show no significant τ (Fig. B5a). Disagreement is higher in North American Deserts and Temperate Sierras ecoregions (u = 0.41-0.49) where most data sets show a negative τ , and all data sets show a negative τ in the Southern Semiarid Highlands. Coefficient *u* is higher among ET data sets than P data sets, averaging 0.21 across ecoregions, notably in the North American Deserts and Marine West Coast Forest ecoregions (u = 0.46 and 0.53), where negative and insignificant τ values are present (Fig. B5b). All data sets identify negative ET τ in the Southern Semiarid Highlands and Temperate Sierras. Data sets show positive, negative, and insignificant τ in the Eastern Temperate Forests (u = 0.31) and Northern Forests (u = 0.53). Runoff data sets show the most consistent spatial distribution of u > 0across the study ecoregions. Disagreement in the western ecoregions is caused by conflicting negative and insignificant $R\tau$ (Fig. B5c), with the greatest u in the Southern Semiarid Highlands (u = 0.50) and Temperate Sierras (u =0.49). Eastern ecoregion τ is generally caused by conflicting positive, negative, and insignificant τ values (u = 0.26-44). Ecoregions with good agreement (u = 0-0.35) are generally due to most data sets showing no significant trend.

SWE data sets, limited in this study to the Northwestern Forested Mountains and Northern Forests, show disagreement, i.e., u = 0.42 and 0.40, respectively, mostly caused by conflicting negative and insignificant τ , though most data

sets show no significant trend (Fig. B5d). Trend agreement is better among SM data sets (mean u = 0.25), though higher u is noted in the Northwestern Forested Mountains, North American Deserts, and Northern Forests and is caused by conflicting negative and insignificant τ values. Ecoregions with low u typically show no significant τ values in SM data sets, except for the Southern Semiarid Highlands where most data sets show a significant negative τ .

Generally, τ disagreement is highest in the North American Deserts and Northern Forests ecoregions and lowest in the Mediterranean California, Eastern Temperate Forests, Marine West Coast Forest, and Tropical Wet Forests ecoregions. Trend disagreement is almost always caused by conflicting negative and insignificant trends, indicating that disagreement is due to the occurrence, and not the direction, of the trend. That is to say, higher *u* values are caused by disagreement over whether there is or is not a significant trend present in the data, as opposed to higher *u* values being caused by disagreement between significant negative and significant positive trends.

3.2.4 Correlation with remote sensing

Spearman's rho (ρ) correlation was calculated for all hydrologic models and reanalysis data sets against remote sensing products with 48 or more months of intersecting temporal extents (Fig. 7; Table 3). Precipitation data sets (Fig. 7a, b), correlated against the GPCP-v3 and TMPA-3B43 products, show high correlation compared against 13 data sets $(\rho = 0.93 - 0.99)$ with no statistically insignificant values. Variability in *P* estimate correlations are low $(\pm 0.01-0.09)$. ET data sets correlated against the MOD16-A2 and SSE-Bop products (Fig. 7c, d) show poorer correlation in western ecoregions, with mean ρ ranging from 0.59 to 0.91. Extreme cases are found when correlating against MOD16-A2 in North American Deserts ($\rho = -0.28$) and SSEBop in Mediterranean California ($\rho = 0.25$), where four data sets show insignificant ρ . The exception in the western CONUS is in Northwestern Forested Mountains, where models correlate against both RS data sets well (MOD16-A2 $\rho = 0.91$ and SSEBop $\rho = 0.89$). Correlation against both products is high in eastern ecoregions, ranging from $\rho = 0.88$ to 0.96, with standard deviation decreasing as ρ increases.

SWE data sets correlated against the AMSR-E/Aqua product are highlighted for the Northwestern Forested Mountains and Northern Forests ecoregions (Fig. 7e). Modeled estimates correlate better in the Northwestern Forested Mountains ($\rho = 0.91$) than Northern Forests ($\rho = 0.80$). SM data sets show worse and more variable ρ against remote sensing products ESA-CCI and SMOS-L4 (Fig. 7f, g) than other components. At the CONUS scale, estimates correlate very poorly against the ESA-CCI product ($\rho = 0.17$) and only moderately well against SMOS-L4 ($\rho = 0.68$). Generally, correlation is best in the Mediterranean California and Eastern Temperate Forests ecoregions ($\rho = 0.74$ –0.92).

data sets correlated against remote sensing products is shown as (n = X) for each column. Mean and standard deviation (σ) of ρ are provided for each ecoregion and remote sensing products as $\rho \pm \sigma$. Counts of statistically insignificant correlations (Y) are denoted when n > 0 as (Y). SM combines data sets produced in units of equivalent water depth (SM(e) in (able 3. Spearman's tho (ρ) correlation of monthly hydrologic model and reanalysis data set estimates against remote sensing products quantifying hydrologic components precipitation (P), evapotranspiration (ET), snow water equivalent (SWE), and soil moisture (SM). Correlation was calculated at 10 ecoregions and at the CONUS extent. The number of modeled text) and volumetric soil moisture water content (SM(v) in text). Correlation of SWE products is included only for regions of high annual mean SWE storage. Hydrologic model and reanalysis data sets (Table 1) used in each column are listed in the footnotes.

		Ρ	AE	T	SWE	R	ZSM
	GPCP-v3 ^a	TMPA-3B43 ^b	MOD16-A2 ^c	SSEBop ^c	AMSR-E/Aqua ^d	ESA-CCI ^e	SMOS-L4 ^f
Ecoregion	(n = 15)	(n = 14)	(n = 15)	(n = 15)	(n = 18)	(n = 15)	(n = 13)
Marine West Coast Forest	0.99 ± 0.01	0.99 ± 0.01	0.82 ± 0.18	0.80 ± 0.19	α Ι	0.72 ± 0.09	0.78 ± 0.09
Mediterranean California	0.98 ± 0.01	0.97 ± 0.02	0.71 ± 0.20	0.25 ± 0.42 (4)	ад І	0.79 ± 0.13	0.92 ± 0.04
Northwestern Forested Mountains	0.97 ± 0.05	0.96 ± 0.05	0.91 ± 0.06	0.89 ± 0.08	0.91 ± 0.07	0.58 ± 0.10	0.27 ± 0.12 (2)
North American deserts	0.95 ± 0.05	0.94 ± 0.05	-0.28 ± 0.17 (4)	0.81 ± 0.13	ad 1	0.61 ± 0.18	0.28 ± 0.11 (2)
Southern Semiarid Highlands	0.97 ± 0.04	0.96 ± 0.05	0.67 ± 0.17	0.64 ± 0.17	α Ι	0.63 ± 0.20	0.79 ± 0.16
Temperate Sierras	0.97 ± 0.04	0.96 ± 0.04	0.59 ± 0.15	0.68 ± 0.18	ад І	0.51 ± 0.19	0.43 ± 0.21 (1)
Great Plains	0.99 ± 0.02	0.98 ± 0.01	0.92 ± 0.02	0.93 ± 0.03	80 	0.62 ± 0.15	0.69 ± 0.06
Eastern Temperate Forests	0.95 ± 0.09	0.94 ± 0.07	0.96 ± 0.03	0.94 ± 0.03	ac I	0.74 ± 0.10	0.83 ± 0.08
Northern Forests	0.96 ± 0.03	0.93 ± 0.02	0.94 ± 0.03	0.94 ± 0.02	0.80 ± 0.03	0.44 ± 0.08	$0.12 \pm 0.21 (10)$
Tropical Wet Forests	0.97 ± 0.03	0.96 ± 0.03	0.88 ± 0.05	0.88 ± 0.09	00 	0.40 ± 0.19 (1)	0.57 ± 0.16
CONUSh	0.94 ± 0.07	0.91 ± 0.06	0.95 ± 0.02	0.95 ± 0.03	0.91 ± 0.03	0.17 ± 0.05 (4)	0.68 ± 0.09
^a CMAP enhanced and standard, Daymet, ERA ^b CMAP enhanced and standard, Daymet, ERA ^c CSIRO-PML, ERA5-Land/H-TESSEL, GLD, ^{Davie-st el} (2017), Tarreolinious and VanFT	5, ERA5-Land, GP 5, ERA5-Land, GP 4S-CLM, GLEAM-	CC, gridMET, Livneh CC, gridMET, Livneh ·v3.3a and v3.3b, MER	et al. (2013), Maurer et al. et al. (2013), MERRA-2, N RA-2 and MERRA-Land	(2002), MERRA-2, MF MERRA-Land, NCEP-J CLSM, NHM-MWBM	ERRA-Land, NCEP-DOE OE, NLDAS2, PRISM, i and NHM-PRMS, NLDA	, NLDAS2, PRISM, and and UoD-v5. S2-Mosaic, NLDAS2-N	l UoD-v5. Ioah, NLDAS2-VIC,

¹ CanSISE, Daymet, ERA5-Land/H-TESSEL, GLDAS-CLM, JRA-55/SiB, Livneh-VIC, MERRA-2 and MERRA-Land/CLSM, NCEP-DOE and NCEP-NARR/Ea-Noah, NHM-MWBM and

NHM-PRMS, NLDÁS2-Mosaic, NLDÁS2-Noah, and NLDÁS2-VIC, SNODÁS, and TerraClimate. ² CPC, ERA5-Land and ERA5/H-TESSEL, GLEAM-v3.3a and v3.3b, JRA-25/SIB, MERRA-2/CLSM, NHM-MWBM and NHM-PRMS, NLDÁS2-Mosaic, NLDÁS2-Noah and NLDÁS2-VIC, FerraClimate, and VegET.

CPC, ERA5-Land and ERA5/H-TESSEL, GLEAM-v3.3a and v3.3b, JRA-55/SiB, MERRA-2/CLSM, NHM-PRMS, NLDAS2-Mosaic, NLDAS2-Noah, and NLDAS2-VIC, TerraClimate, and VegET. Regions omitted

Calculated from data sets spatially aggregated by weighted area mean at the CONUS extent (Fig. 1) and not as a mean of individual ecoregion ρ estimates above.

Unalikeability of Component Estimate Trend Directions from 1982–2010

Figure 6. Bar plots summarizing the disagreement in Mann–Kendall trend (τ) direction using the unalikeability coefficient (*u*) after categorizing τ as significant negative trend, significant positive trend, or no significant trend for 10 ecoregions from 1982 to 2010. Trends were assumed to be significant if p < 0.05. *u* can range from complete agreement (u = 0) to complete disagreement (u = 1) when sample size (*n*) is less than the number of categorical values (*c*). Because here n > c, maximum possible *u* is approximately 0.71 (dotted red line). Text to the right of each bar shows, in order, the *u* value and the number of data sets showing the negative, insignificant, or positive trend. Bars are color matched to ecoregion colors from Fig. 1 and are ordered from west (top) to east (bottom). Results are provided for water balance components of precipitation (*P*), evapotranspiration (ET), runoff (*R*), snow water equivalent (SWE), and soil moisture (SM) in units of either equivalent water depth or volumetric water content.

Correlation is worst in Northwestern Forested Mountains, North American Deserts, and Northern Forests (ESA-CCI $\rho = 0.44-0.61$; SMOS-L4 $\rho = 0.12-0.28$). Explicit distributions of correlation for each data set by ecoregion are provided in Fig. B6.

3.2.5 Case study – impact of model selection on water budget imbalances

A case study calculating 2925 10 year water budgets (WY 2001–2010) using 15 *P*, 15 ET, and 13 *R* estimates for each of the 10 ecoregions was performed to demonstrate quantitatively how model selection can affect research results. Each water budget was solved for a relative imbalance ($R\varepsilon$) with Eq. (7), which estimated the model error as a fraction of total water flux in *P*.

Histograms of the results, overlain with box plots (Fig. 8), demonstrate the range and distribution of potential water budget $R\varepsilon$ values. In most ecoregions, $R\varepsilon$ values exhibit an approximately normal distribution. The following four ecoregions dominating the area of the CONUS (Fig. 8a–d), constituting 90% of the CONUS area (Table 2), are the focus of these results: the western ecoregions of the Northwestern Forested Mountains and North American deserts and the eastern ecoregions of the Great Plains and Eastern Temperate Forests (11 %, 19 %, 29 %, and 31 % of the CONUS area, respectively).

Of these larger domains, water budget $R\varepsilon$ values in the eastern ecoregions show much lower variability (Table 4), with the Great Plains and Eastern Temperate Forests yielding σ values of 18.2 percentage points and 14.3 percentage points, respectively, and medians of 3.9% and -0.4%. respectively. Similarly, the 10th and 90th percentiles (P10 and P90) are -23.2% and 23.6% for the Great Plains and -19.9% and 17.2% for the Eastern Temperate Forests. In contrast, the major western ecoregions exhibit much higher variability, with Northwestern Forested Mountains and North American Desert σ values of 45.6% and 27.0%, respectively, and medians of -6.7% and 4.7%, respectively. P10 and P90 are higher than eastern regions as well, yielding -83.3 % and 23.8 % for the Northwestern Forested Mountains and -33.2% and 36.6% for the North American Deserts. Smaller ecoregions (Fig. 8e-j) yield similar spatial trends in variability, with the eastern Northern Forests ecoregion showing lower σ and P10 and P90 than the western Marine West Coast Forest and Mediterranean California ecoregions.

Correlation of Component Estimates Against Remote-Sensing-Derived Products

Figure 7. Spearman's rho (ρ) correlation values of hydrologic model and reanalysis data set component estimates against remote-sensingderived products for components of precipitation (*P*), evapotranspiration (ET), soil moisture in units of both equivalent water depth (SM(*e*)) and volumetric water content (SM(*v*)), and snow water equivalent (SWE). The title of each sub-plot (e.g., Fig. 7a) provides the hydrologic component (e.g., *P*) and the remote sensing product against which the data set was correlated (e.g., GPCP-v3). Statistically significant ρ values (p<0.05) are shown as colored circles. Insignificant ρ values (p>0.05) are shown as black squares. Horizontal bars within each ecoregion denote mean ρ . Point and bar colors correspond to ecoregions, as shown in Fig. 1. Each point or square corresponds to a single modeled data set, with points sorted by descending ρ value. More detailed information can be found in Fig. B6.

The Northwestern Forested Mountains region, accounting for 11% of the CONUS area, is unique among ecoregions in this study. It exhibits the greatest magnitude of variability in iterative water budget $R\varepsilon$, P10–P90 $R\varepsilon$ ranges, skewness (-1.6), and kurtosis (3.7). Furthermore, the difference between the ensemble water budget $R\varepsilon$ and median iterative water budget $R\varepsilon$ is greater than any other ecoregion (-6.3 points).

4 Discussion

Results of model comparisons demonstrate the effect that model disagreement can have on both regional- and continental-scale research. Comparisons of *P* estimates agree with previous findings (Derin and Yilmaz, 2014; Guirguis and Avissar, 2008; Sun et al., 2018), noting increased uncertainty between products in regions of complex topography. ET comparisons showed that LSMs are more likely to produce lower annual ET than CMs and WBMs, similar to results found by Mueller et al. (2011), and that model disagreement is higher in Pacific regions, similar to the comparison and validation of MOD16-A2, NLDAS2-Noah, and ALEXI by Carter et al. (2018). However, this paper finds some differences from the work of Carter et al. (2018) and notes a higher

uncertainty of ET estimates in arid, semiarid, and mountainous regions, as well as lower uncertainty in eastern regions. Haddeland et al. (2011) found that LSMs underestimate Rpartitioning relative to global hydrologic models (not used in this paper). Here, we find that LSMs are more likely to estimate greater R than either WBMs or CMs. Xia et al. (2012a, b) noted greater model disagreement in R estimates between several LSMs in the northeastern and western mountainous regions of the CONUS. This paper, using CV as a measure of relative uncertainty, alternatively shows that modeled runoff uncertainty is higher in the arid and semiarid regions of the western CONUS where annual R rates are lower.

Previous analyses of modeled SWE estimates focused largely on the validation of winter snowpack magnitudes against the SNODAS model or on the timing of accumulation and ablation periods. Broxton et al. (2016) found that LSMs estimated earlier ablation timing than observational measurements, in contrast with the results of Essery et al. (2009) and Rutter et al. (2009), who noted later timing in LSMs. Our results indicate the alternative argument that neither snow accumulation nor ablation timing can be accurately attributed solely to model design and note differences in relative timing of as much as 2 months between LSMs. In contrast to the arguments by Murdryk et al. (2015) and Broxton et al. (2016),

Table 4. Summary statistics of ecoregion water budget calculations summed for complete water years 2001–2010. Median (P_{50}) and standard deviation (σ) summarize the iterative water budgets' (n = 2925) relative imbalances ($R\varepsilon$) in units of percent. The 10th, 25th, 75th, and 90th percentiles (P_i) quantify the distribution ranges, along with skew (γ) and kurtosis (K). The ensemble column μ provides water budget $R\varepsilon$ produced from the ensemble mean of all data sets available for water years 2001–2010.

Ecoregion	σ	<i>P</i> ₁₀	P ₂₅	P50	P ₇₅	P90	μ	γ	K
Marine West Coast Forest	±29.5	-46.0	-21.3	-2.3	12.2	26.8	-3.8	-0.9	1.1
Mediterranean California	± 27.0	-45.4	-18.6	-3.6	9.8	23.0	-4.7	-1.0	1.3
Northwestern Forested Mountains*	± 45.6	-83.3	-35.0	-6.7	10.5	23.8	-13.0	-1.7	4.3
North American Deserts*	± 27.0	-33.2	-13.7	4.7	21.1	36.6	3.8	-0.5	0.3
Southern Semiarid Highlands	± 30.0	-40.5	-7.2	4.9	19.0	45.2	5.8	-0.7	1.5
Temperate Sierras	± 37.2	-64.6	-15.5	3.0	14.9	31.4	0.4	-1.4	2.2
Great Plains*	± 18.2	-23.2	-9.1	3.9	14.8	23.6	2.5	-0.6	0.5
Eastern Temperate Forests*	± 14.3	-19.9	-10.4	-0.4	8.8	17.2	-0.5	-0.2	-0.1
Northern Forests	± 18.9	-24.3	-13.0	-0.2	12.6	23.9	0.4	-0.1	0.0
Tropical Wet Forest	± 24.6	-35.5	-16.7	0.5	15.2	27.4	-0.7	-0.6	0.4

* Primary focus ecoregions.

who concluded that uncertainty in estimates may be controlled by differences in model structure, the results here show that attributing model differences to structure is likely impossible without controlling for forcing data, model parameters, and calibration methods. Properly identifying controls on model differences requires robust MIPs wherein hydrologic models are operated within the confines of strict input data and calibration schemes, such as those performed by Haddeland et al. (2011), Kollet et al. (2017), or Rosenzweig et al. (2013).

Perhaps the most important variable controlling uncertainties in SM(e) is the discrepancies in model-defined root zone depth. Typically, the largest surface storage component in the hydrologic cycle, i.e., differences in SM(e) estimates, will more greatly affect water availability calculations than any other component. Koster et al. (2009) argued that modeled soil moisture, related in units of equivalent water depth (i.e., SM(e)), should not be considered direct measures of actual soil water content but should instead be used as relative values to identify seasonal to annual trends and responses to changing climatology. However, LSM estimates of soil moisture depth are commonly applied in research applications as direct measures of soil water content. For example, groundwater storage trends are calculated from Gravity Recovery and Climate Experiment (GRACE) solutions of terrestrial water storage anomalies (Rodell et al., 2007; Scanlon et al., 2012; Thomas and Famiglietti, 2019; Voss et al., 2013). Groundwater storage is calculated as the remainder after removing surface storage components, such as modeled surface water storage, SWE, and SM(e) storage values, from GRACE terrestrial water storage. Because estimates of SM(e) can vary by 64 % to 96 %, groundwater storage values derived from GRACE solutions will be significantly affected. Groundwater storage values are most commonly extracted using the NLDAS2- or GLDAS-driven LSMs of Mosaic, Noah, VIC, or CLM that agree more in monthly SM(e) magnitude than with other estimates from WBMs or CMs. However, even those products show differences in SM(e) estimates of up to 250 mm per year. Therefore, quantifying differences in SM(e) is useful in understanding the uncertainty propagation in research results.

In terms of evaluating surface water availability, snowmelt from the Northwestern Forested Mountains ecoregion is a primary supply of water for many western population centers. However, disagreement between models of annual snowpack in this study, measured as SWE, averages 97 %, and relative variability and timing on accumulation or ablation can vary by up to 2 months. These levels of disagreement strongly affect the accuracy of predictive and retrospective snow water analyses of both present snowpack and long-term trends.

LSMs, typically targeted to estimate the vertical ET flux, generally overestimate R compared to CMs. Conversely, CMs, targeted to estimate the horizontal R flux, often overestimate ET compared to LSMs. This is likely an example of equifinality (Beven, 2006), wherein target variable accuracy may be reached without accurately representing the complete hydrologic system. In comparing model output by organization, estimates of P, ET, and SWE produced by NASA are often lower in magnitude than those produced by the USGS and European Centre for Medium-Range Weather Forecasts (ECMWF), indicating organizational differences in model operation methods, such as forcing data selection or calibration methodology.

Results of the water budget case studies demonstrate that water budget $R\varepsilon$ values are more variable in the western CONUS, with σ of $\pm 27.0\%$ to 45.6% by ecoregion, although even eastern CONUS $R\varepsilon$ values range in variability from $\pm 14.3\%$ to 18.9% by ecoregion. In all regions, various combinations of products can result in both positive and negative imbalances (ε), yielding alarmingly different research implications. Positive ε (P > ET + R) means that more water

Figure 8. Histograms of 10 ecoregion water budget relative imbalances calculated from combinations of all annual precipitation (15), evapotranspiration (15), and runoff (13) data sets with complete data from water years 2001–2010 (Table 1). Each region yielded 2925 water budgets. Relative imbalances are calculated as a percentage of annual P. The 10th and 90th percentiles of each ecoregion distribution are provided as vertical dotted lines. Dashed vertical lines represent the relative imbalances of 10 year (2001–2010) water budgets from the ensemble mean of all modeled products.

is entering a system than leaving. Assuming a limited influence of model uncertainty, this would imply that excluded natural or anthropogenic fluxes are removing water from the P = ET+R system. This could be interpreted as the presence of long-term natural contributions to storage components, such as soil moisture or groundwater, or as the presence of anthropogenic extractions. The same region, depending on the models selected, could yield negative ε (P < ET+R), implying the presence of additive fluxes such as releases from surface storage or increased irrigation from imported water.

Results in this paper supplement the work of Haddeland et al. (2011), who argued that climate change effect modeling on large-scale hydrologic processes should utilize a range of model estimates rather than rely on a single model realization. Indeed, we show that model variability in storage components often exceeds the magnitude of the measurements themselves, calling into question the functional application of terrestrial storage estimates. Our work shows that ensembles of model estimates are not only useful but are in fact a necessity for better understanding and quantifying current knowledge of terrestrial hydrology. Improved ensembling methods, as was done by Zaherpour et al. (2019), will likely be invaluable resources in the fields of terrestrial hydrology, climatology, and meteorology modeling. Work by Newman et al. (2015) and Addor et al. (2017) organized numerous modeled data sets across hundreds of watersheds in the USA through the Catchment Attributes and MEteorology for Large-sample Studies (CAMELs) data set, compiling meteorological, geophysical, and hydrologic response variables. Data hosted within the CAMELs data set, and the data sets prepared for this study (Saxe et al., 2020), reduce the workload for users for a better understanding of the inherent uncertainty in modeled products. This comprehensive and systematic comparison helps elucidate where we, as a scientific community, could better target research efforts to reconcile large-extent estimates of the hydrologic cycle. Most importantly, differences in modeled SM(e), both in terms of magnitude depth and correlation against RS products, would need to be addressed at the continental scale to better understand the largest surface store. Furthermore, estimates in SWE, ET, and R need to be reconciled in much of the western CONUS to provide a more accurate retrospective and operational overview of the hydrologic cycle. Finally, differences in hydrologic estimates warrant incorporation into future analysis that use these products to provide the scientific community with a more robust understanding of results confidence and areas of uncertainty. All data sets explored in this paper are easily obtained from publicly available sources, and modern data-processing workflows are applied to multiple data sets without difficulty.

5 Conclusions

Model selection can significantly alter results and findings when used as a substitute for observational data. Publicly available modeled estimates are commonly used within the scientific community, though the literature that effectively evaluates differences in magnitude and trend in the context of application is rare. MIPs abound within the scientific literature, often focusing on the causes of model disagreement (e.g., forcing data, model structure, and calibration) and validating against in situ observational data despite limited availability. In contrast, this study investigated how publicly available modeled estimates of hydrologic flux and storage components can affect scientific analyses by quantifying the disagreement between numerous hydrologic models, reanalysis data sets, and remote sensing products.

Results show that flux and storage magnitudes disagree most greatly in the western CONUS, with uncertainty (measured as a coefficient of variation) ranging from 11 % to 21 % for precipitation (*P*), 14 % to 26 % for evapotranspiration (ET), 28 % to 82 % for runoff (*R*), 76 % to 84 % for snow

water equivalent (SWE), and 36 % to 96 % for soil moisture (SM). In the eastern CONUS, uncertainty is somewhat lower, ranging from 5 % to 14 % for P, 13 % to 22 % for ET, 28 % to 82 % for R, 53 % to 63 % for SWE, and 42 % to 83 % for SM. Interannual trends in estimates from 1982 to 2010 show more comprehensive agreement for P and ET fluxes, but common disagreement for R, SWE, and SM. Disagreement in trends (i.e., positive versus negative versus insignificant) between models is typically a result of conflicting negative and insignificant trends rather than between negative and positive trends, indicating that the disagreement is due to the occurrence and not the direction of trend. Correlating fluxes and stores against remote-sensing-derived products show poor overall correlation in the western CONUS for ET and SM. P correlates well in all regions, and SWE correlates well in the primary regions of Northwestern Forested Mountains and Northern Forests.

A water budget analysis, performed by iterating through all combinations of publicly available modeled products highlighted in this research, shows that in large eastern ecoregions the model selection can result in relative imbalances ranging from -50% to 50%. In larger western ecoregions, relative imbalances can range from -150% to 60%.

Publicly available modeled estimates of the hydrologic system accelerate the development of scientific research by reducing the necessary workload, processing time, and technical expertise for researchers. In addition, estimates fill knowledge gaps in fluxes and stores where observational data are incomplete. Metrics of disagreement in component estimates presented here help to describe the uncertainty of the scientific community's current state of knowledge. The uncertainty inherent in modern data sets can affect the results of studies as diverse as satellite-derived groundwater estimates and predictive snowmelt analyses. Our results highlight problem areas within CONUS-extent hydrologic estimation efforts that warrant better understanding and addressing in future endeavors. Most importantly, our results show that the most important issues to reconcile are disagreement, in terms of both magnitude and long-term trend, and of (a) SM storage across the CONUS, (b) SWE storage in western mountains, and (c) hydrologic fluxes in western arid and semiarid regions. Additionally, our results support the findings of previous studies and agree that future work applying modeled data would prove to be more accurate, more informative, and more robust by incorporating model ensembles to provide confidence intervals, better quantify result uncertainty, and improve overall accuracy.

Appendix A

Acronym Product description

Hydrologic models

CPC	Climate Prediction Center Root zone soil moisture (equivalent water depth) estimated through a one-layer water balance model forced by CPC reanalysis precipitation and temperature. Resolutions are monthly at 1/2°, from 1948 to the present for the globe (Fan and van den Dool, 2004). (Author's acknowledgement – CPC soil moisture data are provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www. esrl.noaa.gov/psd/, last access: 1 September 2019.)
CSIRO-PML	Commonwealth Scientific and Industrial Research Organisation Estimates of evapotranspiration generated by a spatially explicit Penman–Monteith–Leuning (PML) model constrained annually by the Fu hydroclimatic model. The model is driven by precipitation, tem- perature, vapor pressure, wind speed, and radiation estimates. Data sets used include reanalysis clima- tology and meteorology products and ex situ surface and vegetation data. Resolutions are monthly at $1/2^{\circ}$ spatial resolution, from 1981 to 2012, for the globe (Zhang et al., 2016).
ERA5/	Hydrology revised-Tiled ECMWF Scheme for Surface Exchanges over Land forced with the ERA5
H-TESSEL	reanalysis From the European Centre for Medium-Range Weather Forecasts (ECMWF), a land surface model, improved from the original TESSEL model, estimates land surface and subsurface fluxes and stores. The model is structured with four soil layers (0–7, 7–28, 28–100, and 100–289 cm), a single snow layer, two sub-grid vegetation types, and a spatially variant soil type. This paper uses model output forced with ERA5 reanalysis data (see below), provided by Copernicus Climate Change Service (C3S), in conjunction with the reanalysis product. Resolutions are hourly at 0.25°, from 1979 to the present, for the globe (Balsamo et al., 2009; Muñoz Sabater, 2019). (Author's acknowledgement – contains modified Copernicus Climate Change Service Information, 2019.)
ERA5-Land/	Hydrology revised-Tiled ECMWF Scheme for Surface Exchanges over Land forced with the ERA5-
H-TESSEL	Land reanalysis From the ECMWF, a land surface model, improved from the original TESSEL model, estimates land surface and subsurface fluxes and stores. See above for a description of the model structure. This paper uses model output forced with ERA5-Land reanalysis data (see below), provided by C3S, in conjunc- tion with the reanalysis product. Resolutions are hourly at 0.1° from 2001 to the present for the globe (Balsamo et al., 2009; C3S, 2019). (Author's acknowledgement – contains modified Copernicus Cli- mate Change Service Information, 2019.)
GLDAS-CLM	Community Land Model, V2, driven by the NASA GLDAS From the National Aeronautics and Space Administration (NASA) Global Land Data Assimilation System (GLDAS), a single-column land surface model that estimates surface and subsurface fluxes and stores within independent spatial domains. The model is structured with variable layer spacing for soil temperature and moisture, multilayer snow process parameterization, TOPMODEL-concept runoff parameterization, a canopy photosynthesis–conductance model, and tiling treatments for sub- grid energy and water balances. Model input requirements are land surface types, soil and vegetation parameters, model initialization states, and climatological forcing data. This study uses the operational version driven by the NASA GLDAS. Resolutions are sub-daily at 1° from 1979 to the present for the globe (Dai et al., 2003; Rodell et al., 2004, 2007).

Acronym	Product description
Hydrologic models	S
GLEAM	Global Land Evaporation Amsterdam Model Estimates land evaporation, surface soil moisture, root zone soil moisture, potential evaporation, and evaporative stress conditions. Potential evapotranspiration is calculated using the Priestly–Taylor equation on reanalysis, in situ, and ex situ meteorological measurements. Actual evapotranspiration is calculated using an evaporative stress factor based on root zone soil moisture estimates and mi- crowave ex situ observations. Root zone soil moisture is estimated through a water balance model. Model generation 3.3 replaces reanalysis measurements from ERA-Interim with ERA-5 (see below). Model version 3.3b differs from 3.3a by utilizing primarily ex situ data and excluding reanalysis products. Resolutions are daily at 1/4° from 1980 to the present for the globe (Martens et al., 2017; Miralles et al., 2011).
JRA-25/SiB	Simple Biosphere model forced with Japanese 25-year Reanalysis Land surface flux and store estimates generated as output from the Japan Meteorological Agency (JMA)-operated Simple Biosphere model (SiB) forced with Japanese 25-year Reanalysis (JRA-25) reanalysis. The predecessor to modern land surface models such as Mosaic, the model is biophysi- cally based and structured with two vegetation layers (canopy and ground cover) and three soil lay- ers. Resolutions are sub-daily to daily at T106 (\sim 110 km) from 1979 to 2004 for the globe (NCAR Research Staff, 2016; Onogi et al., 2007; Sellers et al., 1986).
JRA-55/SiB	Simple Biosphere model forced with Japanese 55-year Reanalysis Land surface flux and store estimates generated as output from the JMA-operated SiB model forced with Japanese 55-year Reanalysis (JRA-55). See above for the model description. Resolutions are sub-daily to daily at T319 (~ 55 km) from 1957 to the present for the globe (Kobayashi et al., 2015; Kobayashi and NCAR Research Staff, 2019; Sellers et al., 1986).
Livneh-VIC	Variable Infiltration Capacity (VIC) model forced and calibrated by Livneh et al. (2013) Modeled estimates of soil moisture, snow water equivalent, discharge, and surface heat fluxes gen- erated by forcing the VIC model with the Livneh et al. 2013 reanalysis meteorological data set (see below). See below (NLDAS2-VIC) for a description of the model structure. Resolutions are sub- daily to daily at 1/16° from 1915 to 2011 for the conterminous United States (Liang et al., 1994; Livneh et al., 2013).
MERRA- Land/CLSM	Catchment land surface model (CLSM) forced with MERRA-Land reanalysis A land surface model developed to improve the treatment of the horizontal hydrologic process in response to previous land surface model over-attention to vertical processes. The model is structured around tiled hydrologic catchments defined by topography, two soil layers, and three snow layers. This paper uses surface estimates generated from the MERRA-Land forced version from GEOS-5, released in conjunction with the MERRA-Land reanalysis product (see below). Resolutions are sub- daily at 1/2° from 1980 to 2016 for the globe (Koster et al., 2000; Reichle et al., 2011; Rienecker et al., 2011).
MERRA-2/ CLSM	Catchment land surface model forced with MERRA-2 reanalysis A land surface model estimating surface and subsurface hydrologic fluxes and stores. See above for

CLSM A land surface model estimating surface and subsurface hydrologic fluxes and stores. See above for the model description. This paper uses estimates generated from the MERRA-2-forced version from GEOS-5, released in conjunction with the MERRA-2 reanalysis product (see below). Resolutions are sub-daily at 1/2° from 1980 to the present (Gelaro et al., 2017; Reichle et al., 2017).

1548

Acronym Product description

Hydrologic models

- NCEP-DOE/ Noah land surface model forced with National Centers for Environmental Prediction-Department of Eta-Noah Energy, R-2 reanalysis Noah land surface model estimates surface and subsurface hydrologic fluxes. See below (NLDAS2-Noah) for the model description. Model estimates here are generated through the Noah land surface model component of the National Centers for Environmental Prediction (NCEP) Eta atmospheric model forced with NCEP-Department of Energy (DOE) reanalysis (see below), released in conjunction with the NCEP–DOE product. Resolutions are sub-daily at T62 (\sim 210 km) gridding from 1979 to the present for the globe (Kalnay et al., 1996; Kanamitsu et al., 2002). (Author acknowledgment - NCEP_Reanalysis 2 data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.) NCEP-NARR/ Noah land surface model forced with National Centers for Environmental Prediction North American Eta-Noah **Regional Reanalysis** Noah land surface model estimates produced as a component of the larger Eta atmospheric model. See below (NLDAS2-Noah) for the model description. Model estimates here are generated through the Noah surface model component of the NCEP Eta atmospheric model forced with NCEP North American Regional Reanalysis (NCEP-NARR), released in conjunction with the NCEP-NARR product. Resolutions are sub-daily at 32 km from 1979 to the present for North America (Mesinger et al., 2006). (Author's acknowledgement- NCEP Reanalysis data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.) NHM-MWBM National Hydrologic Model framework Monthly Water Balance Model A water balance model applied within the U.S. Geological Survey's National Hydrologic Model framework that utilizes a monthly accounting procedure to estimate evapotranspiration, runoff, soil moisture, and snow water equivalent based on methodology from Thornthwaite (1948). Resolutions are monthly at Hydrologic Response Units from 1949 to 2010 for the conterminous United States (McCabe and Markstrom, 2007).
- NHM-PRMS National Hydrologic Model framework Precipitation Runoff Modeling System A process-based, deterministic hydrologic model applied within the framework of the U.S. Geological Survey's National Hydrologic Model that estimates various surface and subsurface fluxes and stores using a conceptualized watershed composed of a series of reservoirs, stream segments, and lakes maintained with a balanced water budget. The model utilizes reanalysis and ex situ physical characteristic data of topography, soils, vegetation, geology, and land use to derive required parameters and is driven by precipitation and temperature reanalysis products. Resolutions are daily at Hydrologic Response Units from 1980 to 2016 for the conterminous United States (Markstrom et al., 2015; Regan et al., 2018).
- NLDAS2-Mosaic Mosaic model driven by the NASA NLDAS, Phase 2
 A land surface model, directly descended from the SiB land surface model, that estimates surface and subsurface fluxes and stores, originally targeted to be coupled with climate and weather models. The model is structured to allow vegetation control over surface energy and water balances, a canopy interception reservoir, three soil reservoirs (thin surface, middle root zone, and lower recharge), and a complete snow budget. The model name is derived from the mosaic approach of tiling subgrid cells into homogeneous vegetation types with independent energy balances. This study uses the operational version driven by the NASA NLDAS Phase 2 (see below), wherein the model was configured to support up to 10 tiles per grid cell, each with specified predominant soil types and three spatially invariant soil layers. Resolutions are hourly at 1/8° from 1979 to the present for North America (Koster and Suarez, 1996; Xia et al., 2012a, b).

1550	S. Saxe et al.: A comparison of publicly available, CONUS-extent hydrologic component estimates
Acronym	Product description
Hydrologic mode	els
NLDAS2-Noah	Noah model driven by the NASA NLDAS, Phase 2 A grid-based land surface model, simpler than the Noah-MP version, originally developed as a com- ponent of the NOAA-NCEP Eta model and updated for use in the NASA NLDAS, Phase 2 (see below) and is used within the Weather Research and Forecasting (WRF) atmospheric model, the National Oceanic and Atmospheric Administration-NCEP Climate Forecast System (NOAA-NCEP CFS), and the NOAA Global Forecast System. The model is structured with four spatially invariant thickness soil layers (three layers forming the root zone system in non-forested regions and four in forested regions) and can simulate soil freeze–thaw processes. This study uses the operational ver- sion (Noah-2.8) driven by the NASA NLDAS, Phase 2 (see below). Resolutions are hourly at 1/8° from 1979 to the present for North America (Ek et al., 2003; Xia et al., 2012b).
NLDAS2-VIC	Variable Infiltration Capacity (VIC) model driven by NLDAS, Phase 2 A semi-distributed, grid-based hydrologic model estimating surface and subsurface fluxes and stores. Model structure is composed of three soil layers (top spatially invariant 10 cm thickness; others spa- tially variant) with root zone depth controlled by vegetation. The model utilizes sub-grid vegetation tiling similar to the Mosaic land surface model (see above) and a two-layer energy balance snow model. This study uses the operational version driven by the NASA NLDAS, Phase 2 (see below). Resolutions are hourly at 1/8° from 1979 to the present for North America (Liang et al., 1994; Xia et al., 2012b, c).
SNODAS	SNOw Data Assimilation System Estimates snow cover and snow water equivalent by integrating modeled snow estimates with ob- servational and reanalysis data from in situ and ex situ sources. The primary estimation method is a physically based, spatially distributed energy and mass balance snow model forced by downscaled output from the National Weather Service Rapid Refresh weather forecast model. In situ and ex situ data sources are applied, depending on difference fields between modeled and observed values, and used to perform immediate model calibrations. Resolutions are daily at 1 km from 2003 to the present for North America (Barrett, 2003; National Operational Hydrologic Remote Sensing Center, 2004).
TerraClimate	TerraClimate A data set utilizing both reanalysis and water budget modeling to generate estimates of climate, mete- orological, and hydrologic variables. Climate and meteorological estimates are calculated by apply- ing interpolated time-varying anomalies from both Climate Research Unit (CRU) Ts4.0 and JRA-55 to the higher-resolution WorldClim climatology. Hydrologic estimates of evapotranspiration, precip- itation, temperature, and soil water capacity are generated through a modified Thornthwaite–Mather climatic water balance model. Water-balance-derived resolutions are monthly at 1/2° from 1958 to the present for the globe (Abatzoglou et al., 2018).
VegET	Vegetation ET Estimates root zone soil moisture and evapotranspiration through a precipitation-driven 1D root zone water balance model. Because the model is parameterized to operate on a control volume us- ing water-holding capacity, it captures only evapotranspiration from vegetation sources (i.e., natural conditions) and does not consider anthropogenic (i.e., non-natural) water use. Data were provided via personal correspondence with the authors. Resolutions are daily at 1 km from 2000 to 2014 for the conterminous United States (Senay, 2008; Velpuri and Senay, 2017).

Acronym	Product description
Reanalysis	
CanSISE	Canadian Sea Ice and Snow Evolution Network, V2 Estimates snow water equivalent (SWE) by merging five observation-based estimates through an adapted ensemble mean methodology. Merged products are (1) GlobSnow-combined SWE (merges ex situ passive microwave and in situ weather station observations), (2) ERA-Interim reanalysis, (3) MERRA reanalysis, (4) SWE from the Interactions Soil–Biosphere–Atmosphere (ISBA) land surface model forced with ERA-Interim reanalysis (see below), and (5) NASA Global Land Data Assimilation System reanalysis. Resolutions are daily at 1° from 1981 to 2010 for the Northern Hemisphere (Mudryk et al., 2015; Mudryk and Derksen, 2017).
СМАР	Climate Prediction Center (CPC) Merged Analysis of Precipitation From the CPC, this model estimates precipitation by blending various in situ, ex situ, and re- analysis data sets. The standard product estimates precipitation by blending ex situ Global Pre- cipitation Index (GPI), Outgoing Longwave Radiation (OLR) Precipitation Index (OPI), Special Sensor Microwave/Imager (SSM/I) emission, and Microwave Sounding Unit (MSU) data with in situ precipitation gauge measurement data. The enhanced product additionally includes blended NCEP-NCAR reanalysis (see below) precipitation estimates. Resolutions are monthly at 2 1/2° from 1979 to the present for the globe (Xie and Arkin, 1997). (Author's acknowledgement – CMAP precipitation data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.)
Daymet	Daily Surface Weather Data on a 1 km Grid for North America, V3 Version 3 of the Daymet model generates estimates of temperature, precipitation, radiation, vapor pressure, snow water equivalent, and day length from in situ observations of temperature and pre- cipitation and digital elevation data. Interpolation is performed using the spatial convolution of a truncated Gaussian weighting filter with an iterative station density algorithm applied to heteroge- neous observation distribution in complex terrain. Snow water equivalent is estimated using a simple temperature-based snowmelt model from Running and Coughlan (1988). This paper assigns Daymet snow water equivalent estimates to the reanalysis category because of the snow model's simplicity and lack of a physical or conceptual hydrologic framework. The best descriptions of the current model framework are found on the ORNL DAAC website (https://daac.ornl.gov/DAYMET/guides/ Daymet_V3_CFMosaics.html, last access: 1 October 2019). Resolutions are daily at 1 km from 1980 to the present for North America (Thornton et al., 1997, 2000, 2017).
ERA5	ECMWF reanalysis, fifth product Generates numerous estimates of climate and meteorological variables by assimilating observational data from 55 ex situ and 19 in situ sources using the 4D-Var variational method. Previous generations of the ERA product were FGGE, ERA-15, ERA-40, and ERA-Interim, which are all now out of ser- vice. Resolutions are hourly at 0.25° from 1979 to the present for the globe (Muñoz Sabater, 2019). (Author's acknowledgement – contains modified Copernicus Climate Change Service Information, 2019.)
ERA5-Land	ECMWF reanalysis, fifth product Estimates of climate, meteorological, and surface hydrology fluxes by replaying the ERA5 reanalysis (see above) land component at a higher spatial resolution. Resolutions are hourly at 0.1° from 2001 to the present for the globe (Balsamo et al., 2009; Muñoz Sabater, 2019). (Author's acknowledgement – contains modified Copernicus Climate Change Service Information, 2019.)

1552	S. Saxe et al.: A comparison of publicly available, CONUS-extent hydrologic component estimates
Acronym	Product description
Reanalysis	
GPCC	Global Precipitation Climatology Centre Monthly accumulated precipitation estimates based on 67 200 in situ observational stations with decadal or longer temporal spans. The data set used in this paper is V7, one of two GPCC full data products, which the product's authors suggest has the highest accuracy of their data sets. Res- olutions are monthly at 1/2° from 1901 to 2013 for the globe (Becker et al., 2013; Schneider et al., 2011). (Author's acknowledgement – GPCC precipitation data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.)
gridMET	gridMET (or METDATA) Estimates temperature, precipitation, downward shortwave radiation, wind velocity, humidity, rela- tive humidity, and specific humidity by blending PRISM reanalysis (see below) climate data with NLDAS2 reanalysis (see below) data. Resolutions are daily at 1/24° from 1979 to the present for the conterminous United States (Abatzoglou, 2013).
Livneh et al. (2013)	Livneh daily CONUS near-surface gridded observed meteorological data Near-surface meteorological estimates generated from approximately 20 000 NOAA Cooperative Observer station daily data sets. Precipitation and temperature in situ observations are converted to grids using the synergraphic mapping system, which is wind data linearly interpolated from the lower resolution NCEP-NCAR reanalysis data set. Other variables were derived using methods from the mountain microclimate simulator. Daily temperature data is converted to 3 h using spline interpola- tion. Resolutions are daily at 1/16° from 1915 to 2011 for the conterminous United States (Livneh et al., 2013). (Author's acknowledgement – Livneh data provided by NOAA/OAR/ESRL PSD, Boul- der, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.)
Maurer et al. (2002)	Maurer et al. (2002) Meteorological and surface hydrologic estimates generated through both reanalysis and hydrologic modeling. Precipitation gridded from in situ daily measurement stations using the synergraphic map- ping system algorithm scaled to match PRISM (see below) long-term averages. Surface estimates derived through the VIC (see above) hydrologic model. Resolutions are daily at 1/8° from 1950 to 1999 for North America (Maurer et al., 2002).
MERRA-Land	Modern-Era Retrospective analysis for Research and Applications with improved land surface vari- ables The predecessor to MERRA-2 (see below) estimates meteorological and surface hydrologic com- ponents. Meteorological estimates are derived through the GEOS-5 atmospheric general circulation model by assimilating in situ (3DVAR analysis algorithm) and ex situ (Community Radiative Trans- fer Model) observational data. This paper uses the land surface diagnostics precipitation. Resolutions are sub-daily at 1/2° from 1980 to 2016 for the globe (Rienecker et al., 2011).
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, V2 The most recent version of the MERRA product differs from MERRA-Land by including updates to models, algorithms, observing systems, and ex situ processing methods, as well as 14 additional ex situ data sources. Precipitation is derived from global precipitation products disaggregated to hourly time steps using MERRA-Land precipitation. Precipitation estimates used in this paper are from the land surface diagnostics category. Resolutions are sub-daily at 1/2° from 1980 to the present (Gelaro et al., 2017).

1	EE 2
т	333

Acronym	Product description
Reanalysis	
NCEP-DOE	National Centers for Environmental Prediction–Department of Energy, R-2 The updated (R-2) version of the NCEP-NCAR reanalysis estimates climate, meteorological, and surface hydrologic components. Reanalysis estimates of atmospheric variables are performed through the assimilation and modeling of in situ and ex situ observational data. Resolutions are sub- daily at T62 (\sim 210 km) gridding from 1979 to the present for the globe (Kalnay et al., 1996; Kana- mitsu et al., 2002). (Author's acknowledgment – NCEP–DOE data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https://www.esrl.noaa.gov/psd/, last access: 1 October 2019.)
NLDAS2	North American Land Data Assimilation System, Phase 2 Compiled data used to drive land surface models (see above), including (1) land surface parameters of vegetation, soil, topography, and temperature and (2) surface-forcing fields of precipitation, radi- ation, temperature, humidity, wind, and pressure derived by blending various in situ and ex situ data sets. Precipitation, specifically, is derived by blending temporally disaggregated reanalysis (CPC and NARR; see above) and ex situ (Doppler Stage II and CPC morphing technique, CMORPH) esti- mates. Resolutions are hourly at 1/8° from 1979 to the present for North America (Xia et al., 2009, 2012b).
PRISM	Parameter-elevation Regressions on Independent Slopes Model Reanalysis estimates of precipitation and temperature calculated using a climate-elevation regres- sion model, utilizing information of the location, elevation, coastal proximity, and other geophysical parameters. Resolutions are daily at 800 m (paid) and 4 km (free) from 1895 to the present for the United States (PRISM Climate Group, OSU, 2004).
Reitz et al. (2017)	Reitz et al. (2017) Reanalysis estimates of groundwater recharge, quick-flow runoff, and evapotranspiration. Runoff and evapotranspiration are calculated using regression equations derived from observational water balance data using land cover, temperature, and precipitation information. Groundwater recharge is calculated as the remainder of a water balance using runoff, evapotranspiration, and precipitation. Resolutions are annual (monthly provided via personal correspondence with authors) at 800 m from 2000 to 2013 for the conterminous United States (Reitz et al., 2017).
UoD-v5	University of Delaware (UoD) Air Temperature and Precipitation, V5 Reanalysis precipitation and temperature estimates calculated from in situ observational data with an interpolation algorithm based on the spherical Shepard's distance-weighting method with digi- tal elevation model information. This paper uses the fifth product version. Resolutions are monthly at 1/2° from 1950 to 1999 (Willmott and Matsuura, 2001). (Author's acknowledgement – UoD precipitation data provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from https: //www.esrl.noaa.gov/psd/, last access: 1 October 2019.)
WaterWatch	WaterWatch Estimates of runoff derived from the U.S. Geological Survey's in situ stream gauge network. Resolutions are monthly at watershed, with Hydrologic Unit levels 2–8, and state polygons from 1901 to the present (Jian et al., 2008).

1554	S. Saxe et al.: A comparison of publicly available, CONUS-extent hydrologic component estimates							
Acronym	Product description							
Remote sensing								
AMSR-E/Aqua	AMSR-E/Aqua Monthly L3 Global Snow Water Equivalent EASE-Grids, V2 Scientifically identical to version 1 (there are updates to the product maturity code in V2), this model estimates SWE using passive microwave data collected from the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) instrument hosted on NASA's Aqua satellite. Mi- crowave measurements are converted to SWE estimates using the AMSR-E snow water equivalent algorithm that utilizes brightness temperature differences calculated using the dense media radiative transfer equation and preselected snowpack profiles to develop an artificial neural network. Probable SWE estimate ranges are then restricted using surface temperature and land cover attributes derived from the ex situ MODIS sensor hosted on the NASA Terra satellite. Resolutions are daily at 25 km from 2002 to 2011 for the globe (Chang et al., 2003; Chang and Rango, 2000; Tedesco et al., 2004).							
ESA-CCI	European Space Agency-Climate Change Initiative Volumetric soil moisture estimates generated by merging both active and passive ex situ microwave data-derived soil moisture products into three products, namely active, passive, and combined. Pas- sive microwave products are from the Scanning Multichannel Microwave Radiometer (SMMR), SSM/I, Tropical Microwave Imager (TMI), Advanced Microwave Scanning Radiometer-Earth Ob- serving System (AMSR-E), WindSat, Soil Moisture and Ocean Salinity (SMOS), and AMSR2 sen- sors. Active microwave products are from the Active Microwave Instrument Wind Scatterometer (AMI-WS) and Advanced SCATterometer, with satellite A or satellites A/B (ASCAT-A or ASCAT- A/B) sensors. Merging is performed by resampling target products to a uniform spatiotemporal struc- ture, scaling to match ranges, and weighting using triple collocation. This research utilizes the com- bined product. Resolutions are daily at 1/4° from 1978 to the present for the globe (Dorigo et al., 2017).							
GPCP-v3	Global Precipitation Climatology Project, V3 – precipitation data (beta) Estimates of precipitation generated by assimilating radiometer data from Special Sensor Mi- crowave/Imager (SSM/I) and SSM/Is, infrared data from the Precipitation Estimation from Re- motely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN- CDR), TIROS Operational Vertical Sounder (TIROS-TOVS) sensor and Tropical Rainfall Measur- ing Missing (TRMM) Combined Climatology precipitation estimates with the GPCC global gauge analysis (see above in the reanalysis section). This paper uses the combined satellite gauge product rather than the satellite-only product. Resolutions are monthly at 1/2° from 1983 to 2016 for the globe (Huffman et al., 2019).							
MOD16-A2	MODIS Global Evapotranspiration Project The Level 4 Moderate Resolution Imaging Spectroradiometer (MODIS) land data product estimates evapotranspiration from ex situ observational MODIS vegetation data and reanalysis meteorological data using an improved evapotranspiration model (Mu et al., 2011) based on the Penman–Monteith equation. Resolutions are 8 d at 1 km from 2000 to 2010 for the globe (Running et al., 2017).							
SMOS L4	Soil Moisture and Ocean Salinity, level 4 Volumetric root zone soil moisture estimates derived using Centre Aval de Traitement des Données SMOS (CATDS)-generated ascending and descending orbit SMOS L3 surface volumetric soil mois- ture estimates in conjunction with MODIS sensor-derived vegetation information, reanalysis NCEP climate data, FAO soil textures, and ECOCLIMAP surface cover. A double bucket model, composed of a simple water budget model (5–40 cm depth) and a budget model based on a linearized Richard's equation formulation (40–200 cm depth), is used to extrapolate surface SMOS L3 soil moisture (0– 5 cm) to the root zone domain. This study merged the ascending and descending orbit products by mean. Resolutions are daily at 25 km EASE Grids from 2010 to 2017 for the globe (Al Bitar et al., 2013).							

Acronym	Product description						
Remote sensing							
SSEBop	Operational Simplified Surface Energy Balance model A parameterization of the Simplified Surface Energy Balance approach that estimates evapotranspi- ration from ex situ MODIS vegetation, ex situ Shuttle Radar Topography Mission (SRTM) eleva- tion data, reanalysis meteorological data, and other modeled data using an energy balance approach where actual evapotranspiration is calculated as the difference between net surface radiation, sen- sible heat flux, and ground heat flux. Resolutions are monthly at 1 km from 2000 to 2014 for the conterminous United States and portions of surrounding countries (Senay et al., 2011, 2013).						
TMPA-3B43	TRMM Multisatellite Precipitation Analysis, V7 From the larger Tropical Rainfall Measuring Mission (TRMM), this model generates spatiotempo- rally continuous precipitation estimates by merging precipitation estimates from numerous ex situ data sources. Primary data sources include (1) passive microwave data from various low Earth or- bit satellites (e.g., TRMM and AMSR-E) that are converted to precipitation estimates using source- specific algorithms, (2) infrared data collected by geosynchronous Earth orbit satellites, (3) a TRMM Combined Instrument (TCI) estimate, which is a merged passive microwave and active radar prod- uct, and (4) in situ GPCP and CAMS monthly precipitation measurements. Resolutions are sub-daily at 1/4° from 1998 to the present for the globe (Huffman et al., 2007, 2010).						

Appendix B

Figure B1. Box plots of annual standard deviation (σ) between component estimates at 10 ecoregions. σ distributions are subdivided into two periods, namely 1985–1999 and 2000–2014. Box lower limits, midlines, and upper limits represent the 25th, 50th (median), and 75th percentiles, respectively, of the associated data. Whiskers represent 1.5 times the interquartile range. Box colors denote ecoregions and correspond to map colors in Fig. 1. Components displayed are precipitation (*P*), evapotranspiration (ET), runoff (*R*), snow water equivalent (SWE), soil moisture in units of equivalent water depth (SM(*e*)), and soil moisture in units of volumetric water content (SM(*v*)).

Figure B2. Box plot summarizing annual peak snow water equivalent (SWE) depths by ecoregion derived from the distribution of modeled estimates used in this paper. Box lower limits, midlines, and upper limits represent the 25th, 50th (median), and 75th percentiles, respectively, of the associated data. Whiskers represent 1.5 times the interquartile range.

Comparing Annual SWE Accumulation and Ablation Timing

Figure B3. Heat maps showing the relative timing of the beginning of snow water equivalent (SWE) accumulation (a) and ablation (b) periods for two ecoregions from 1985 to 2014. Relative timing is the difference between the month of the beginning of a model's accumulation or ablation period and the mean of all other model's timing. The beginning of accumulation and ablation is defined as when the rate of change of SWE between months is greater than 1 mm per month or less than -1 mm per month, respectively.

Percent Annual Positive vs. Negative Timing

Figure B4. Heat maps summarizing the annual relative timing values of SWE accumulation (accum.) and ablation from Fig. B3 for two ecoregions. Grid cells are colored by the percentage of years that have a positive relative timing (green) or by the percentage of years that have a negative relative timing (purple), whichever is greater.

Comparison of Model Trends by Ecoregion (a) Precipitation $\tau \ge 0.50$ 0 < τ < 0.50 R NCEP-DOE R NCEP-DOE UoD-v5 NCEP-DOE UoD-v5 PRISM NLDAS2 NCEP-DOE MERRA-Land UoD-v5 D R R R R R S R R UoD-v5 UoD-v5 UoD-v5 UoD-v5 PRISM UoD-v5 PRISM RRRRR RRRRR R R R PRISM NLDAS2 PRISM NLDAS2 NCEP-DOE MERRA-Land PRISM PRISM PRISM FRA5 NLDAS2 NCEP-DOE MERRA-Land NCEP-DOE neh et al. 2013 GPCP-v3 NLDAS2 NCEP-DOE MERRA-Land GPCP-v3 CMAP-Standard CMAP-Enhanced NLDAS2 NLDAS2 NCEP-DOE MERRA-Land DayMET MERRA-L and R MERRA-L and R R R Insignificant MERRA-2 MERRA-2 > 0.05) Livneh et al. 2013 ĥ ERA5 MERRA-2 MERRA-2 R R MERRA-2 MERRA-2 MERRA-2 DavMET ivneh et al. 2013 neh et al. 2013 gridMET UoD-v5 Livneh et al. 2013 Livneh et al. 2013 eh et al. 2013 CMAP-Standard neh et al. 2013 Livneh et al. 201 aridMET gridMET gridMET GPCP-v3 gridMET GPCP-v3 gridMET GPCP-v3 ĥ CMAP-Enhanced gridMET gridMET GPCP-v3 GPCP-v3 GPCC ERA5 ERA5 RRRRRRR PRISM GPCP-v3 NLDAS2 DayMET NLDAS2 GPCC ERA5 ٩ CMAP-Enhanced GPCC ERA5 GPCC ERA5 GPCC ERA5 GPCC ERA5 DayMET CMAP-Standard CMAP-Standard GPCC gridMET A MERRA-Land R PRISM DayMET R DavMET DavMET DayME DayMET CMAP-Standard R R R R R R R MERRA-2 DayMET CMAP-Standard R R RRRR CMAP-Standard CMAP-Enhanced CMAP-Standard CMAP-Enhanced GPCC CMAP-Standard RRRR R R CMAP-Standard MERRA-2 MERRA-2 gridMET CMAP-Enhanced R CMAP-Enhanced CMAP-Enhanced CMAP-Enhanced MERRA-Land MERRA-Land GPCC NLDAS2 NCEP-DOE neh et al. 2013 -0.50 < τ < 0 **τ** ≤ -0.50 Trop. Wet Forests Marine W. Coast Forest Medit. CA NW Forested Mtns N. Amer. Deserts So. Semiarid Highlands Temp. Sierras Great Plains E. Temp. Forests North. Forests (b) Evapotranspiration τ ≥ 0.50 $0 < \tau < 0.50$ H GLEAM-v3.3a CSIRO-PMI H H GLEAM-v3.3a TerraClimate NLDAS2-VIC NLDAS2-Noah NLDAS2-Mosair NHM-PRMS NHM-MWBM TerraClimate NLDAS2-VIC NLDAS2-Noah NLDAS2-Mosaid NHM-PRMS TerraClimate TerraClimate TerraClimate H H TerraClimate HHHH NLDAS2-Noah NLDAS2-Mosaic nsignificant NLDAS2-VIC H MERRA-Land/CLSM . 0.05) # # # Н TerraClimate NLDAS2-Noah NLDAS2-VIC NI DAS2-Noah NLDAS2-VIC NLDAS2-Noah NLDAS2-Mosai NHM-PRMS NHM-MWBM H NEDAS2-VIC NLDAS2-Noah NLDAS2-Mosai NHM-PRMS NHM-MWBM H H MERRA-2/CLSM NLDAS2-Woald NLDAS2-Wosaid NLDAS2-VIC ERRA-Land/CLSM H H H NLDAS2-Noah Н NLDAS2-VIC NLDAS2-Mosai NHM-PRMS GLEAM-v3.3a NLDAS2-Noah H H NHM-PRMS < d) // // NHM-MWBM NHM-PRMS NLDAS2-Mosaic NHM-MWBM NHM-PRMS NLDAS2-Mosaic MERRA-2/CLSM H NLDAS2-Noah H H H H GLEAM-v3.3a VERRA-Land/CLSN TerraClimate NHM-MWBM NHM-PRMS ERRA-Land/CLSM H MERRA-Land/CLSM TerraClimate H MERRA-Land/CLSM MERRA-2/CLSM GLEAM-v3.3a GLDAS-CLM CSIRO-PML CSIRO-PMI MERRA /IERRA-2/CLSM GLEAM-v3.3a CSIRO-PMI h CSIRO-PML IERRA-Land/CLSN MERRA-2/CLSM NHM-MWBM H H H H H H MERRA-2/CLSM H H H MERRA-2/CLSM NLDAS2-Mosaic MERRA-2/CLSM TerraClimate Н MERRA-2/CLSM GLEAM-v3.3a CSIRO-PML GLEAM-v3.3 H MERRA-Land/CLSM H MERRA-2/CLSM H NLDAS2-VIC GLEAM-v3.3a MERRA-Land/CLSM NHM-MWBM GLEAM-v3.3a NHM-PRMS NHM-MWBM NLDAS2-VIC Н CSIRO-PML CSIRO-PML CSIRO-PML NHM-MWBM CSIRO-PML H H -0.50 < τ < 0 GLDAS-CLM Н Н GLDAS-CLM Н GLDAS-CLM H GLDAS-CLM H MERRA-Land/CLSM τ ≤ -0.50 Н GLDAS-CLM GLDAS-CLM H GLDAS-CLM Н GLDAS-CLM Marine W. Coast Forest Medit CA N. Amer. Deserts So. Semiarid Highlands NW Forested Temp. Sierras Great Plains E. Temp. North. Trop. Wet Forests Mtns Forests Forests (c) Runoff **τ** ≥ 0.50 HNCEP-DOE/Eta-Noah HNCEP-DOE/Eta-No $0 < \tau < 0.50$ HNCEP-DOE/Eta-Noah HNCEP-DOE/Eta-Noah GLDAS-CLM н JRA-55/SiB н JRA-25/SiB H JRA-55/SiB WaterWatch WaterWatch WaterWatch WaterWatch R H H H WaterWatch WaterWatch WaterWatch WaterWatch TerraClimate NLDAS2-Mosaic Н TerraClimate TerraClimate TerraClimate H NLDAS2-Mosaic H H H H H H H TerraClimate H H TerraClimate WaterWatch R WaterWatch NLDAS2-VIC NLDAS2-VIC NLDAS2-Noah NLDAS2-VIC NHM-PRMS H H H Н NLDAS2-VIC NLDAS2-VIC NLDAS2-VIC H TerraClimate NLDAS2-Noah NLDAS2-Noah NHM-PRMS Н NLDAS2-Noah NHM-MWBM NLDAS2-Noah H H NLDAS2-Noah NLDAS2-VIC Insignificant NLDAS2-Mosaic NHM-MWBM Н NLDAS2-VIC (p > 0.05) H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H NCEP-D0E/Eta-Noah H MERRA-Land/CLSM H MERRA-2/CLSM H JRA-55/SIB H JRA-55/SIB NLDAS2-Mosai NLDAS2-Mosai NLDAS2-Mosaid HNCEP-DOE/Eta-Noa NLDAS2-Mosaid NLDAS2-Mosaic H NLDAS2-Noah NCEP-DOE/Eta-Noah NHM-PRMS NHM-PRMS H MERRA-Land/CLSM Н NLDAS2-Noah NHM-PRMS NHM-PRMS NHM-PRMS NLDAS2-Mos NHM-MWBN NHM-MWBM NLDAS2-Mosaic NHM-MWBM Н MERRA-2/CLSM NHM-MWBM NHM-MWBM NHM-PRMS н FRA5/H-TESSEL IERRA-Land/CLSM DOE/Eta-No NHM-PRMS JRA-55/SiB JRA-25/SiB GLDAS-CLM HNCEP-DOE/Eta-No H MERRA-Land/CLSM MERRA-Land/CLSN NHM-MWRM н Н MERRA-2/CLSM H MERRA-Land/CLSM NHM-MWBM H MERRA-Land/CLSM H MERRA-2/CLSM Н GLDAS-CLM MERRA-2/CLSM JRA-25/SiB H MERRA-2/CLSM ERRA-Land/CLSM JRA-55/SiB JRA-25/SiB н Н MERRA-2/CLSM NLDAS2-VIC NLDAS2-Noah H MERRA-Land/CLSM IRA-25/SiP JRA-55/SiB MERRA-2/CLSM Н JRA-25/SiB NLDAS2-Noah MERRA-Land/CLSM MERRA-2/CLSM H JRA-25/SiB ERA5/H-TESSEL IRA-25/SiB Н MERRA-2/CLSM IRA-25/SiB GLDAS-CLM GLDAS-CLN GLDAS-CLM JRA-55/SiB Н ERA5/H-TESSEL ERA5/H-TESSEL ERA5/H-TESSEL Н ERA5/H-TESSEL H H H ERA5/H-TESSEL ERA5/H-TESSEL ERA5/H-TESSEL JRA-55/SiB ERA5/H-TESSEL GLDAS-CLM H H Н GLDAS-CLM H TerraClimate NLDAS2-VIC JRA-55/SiB н GLDAS-CLM -0.50 < T < 0 GLDAS-CLM τ ≤ -0.50 HNCEP-DOE/Eta-Noah JRA-55/SiB JRA-25/SiB Н Marine W. Medit. NW Forested N. Amer. Deserts So. Semiarid Temp. Sierras Great E. Temp. North. Trop. Wet Coast Forest CA. Mtns Highlands Plains Forests Forests Forests

Figure B5.

		ator Equit	aloni											
τ≥0.50	H NHM-MWBM													
0 < т < 0.50	H GLDAS-CLM		HNCEP-NARR/Eta-Noah	HNCEP-NARR/Eta-Noah						H JRA-55/SiB				
0 < 1 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0 > 0 < 0	H GLDAS-CLM H CEP-NARR/Eta-Noaf H TerraClimate H NLDAS2-ViC H NLDAS2-VoiA H NLDAS2-Noaic H MERRA-2/CLSM H MERRA-2/CLSM H Livnet-ViC H JRA-55/SiB R DayMET R CanSISE	H TerraClimate H NLDAS2-VIC H NLDAS2-VIOA H NLDAS2-Noah H NLDAS2-Noah H NLDAS2-Noah H NLM-RWosh H MERA-Land/CLSM H MERA-2/CLSM H GLDAS-CLM H GLDAS-CLM H ERAS/H-TESSEL R DayMET C GaRISE	H TerraClimate H TerraClimate H NLDAS2-VIC H NLDAS2-VIC H NLDAS2-VIC H NLDAS2-VIC H NLDAS2-VIC H NLM-MWBM H MERR-4_20/CLSM H MERR-2/CLSM H JRA-55/SIB H GLDAS-CLM H ERAS/H-TESSEL R CanSISE	H CLEP-NARKVETa-Noar H TerraClimate H NLDAS2-VIC H NLDAS2-Voah H NLDAS2-Voah H NLDAS2-VOAH H NHM-MWBM H MERA-Land/CLSM H MERA-2/CLSM H GLDAS-CLM H ERAS/H-TESSEL R CanSISE	H TerraClimate H NLDAS2-VIC H NLDAS2-VIC H NHM-PRMS H NHM-PRMS H NHM-RHVETA-NoaH H MCEP-NARFIETA-NoaH H MCEP-NARFIETA-Noa H JRA-55/SIB H JLAS-5CLM H GLDAS-CLM H ERA5/H-TESSEL R DayMET C CARISF	H TerraClimate H NLDAS2-ViC H NLDAS2-ViOa H NLDAS2-Noah H NHM-PRMS H NHM-PRMS H MERA-1and/CLSM H MERA-2/CLSM H MERA-2/CLSM H JRA-25/SIB H JRA-25/SIB H JRA-25/SIB H GLDAS-CLM H ERA5/H-TESSEL	H TerraClimate H NLDAS2-VIC H NLDAS2-Voah H NLDAS2-Noah H NHM-PRMS H NHM-RVETA- MCEP-1ARR/ETa-Noah H MERRA-Land/CLSM H MERRA-2/CLSM H JRA-55/3B JRA-55/3B H ERA5/H-TESSEL R DayNET	H TerraClimate H NLDAS2-VIC H NLDAS2-Voah H NLDAS2-Mosaic H NHM-PRMS H NHM-RWBM H NHM-RNMBM H MERRA-Land/CLSM H MERRA-2/CLSM H JRA-25/SIB H JRA-25/SIB H GLDAS-CLM H GLDAS-CLM H ERA5/H-TESSEL R DavVET	H TerraClimate H NLDAS2-ViC H NLDAS2-Vioah H NLDAS2-Noah H NLDAS2-Mosaic H NLHA-Wosaic H NHM-PRMS H NHM-WWBM H HERA-7/CLSM H JRA-25/SiB H JRA-25/SiB H JRA-25/SiB H JRA-25/SiB H ERAS/H-TESSEL R CanSISE HN/CF2-D0F/Fra-Moah	 JIAK-33/SIB TerraClimate H NLDAS2-VIC H NLDAS2-VIC H NLDAS2-VIC H NLDAS2-Noah H RERA-2/CLSM H RERA-2/CLSM H RERAS/H-TESSEL R DarWET 				
-0.50 < τ < 0	H ERA5/H-TESSEL H JRA-25/SiB	<u> </u>	H NHM-PRMS	H JRA-25/SiB	H NLDAS2-Noah	HNCEP-DOE/Eta-Noah	R CanSISE	R CanSISE	HNCEP-NARR/Eta-Noah	R CanSISE				
0 50	H NHM-PRMS	HNCEP-DOE/Eta-Noah	HNCEP-DOE/Eta-Noah	HNCEP-DOE/Eta-Noah	H MERRA-2/CLSM	R CanSISE	H GLDAS-CLM		H GLDAS-CLM					
l ≤ -0.50	HNCEP-DOE/Eta-Noah	H JRA-25/SiB	R DayMET	R DayMET			HNCEP-DOE/Eta-Noah	HNCEP-DOE/Eta-Noah						
	Marine W. Coast Forest	Medit. CA	NW Forested Mtns	N. Amer. Deserts	So. Semiarid Highlands	Temp. Sierras	Great Plains	E. Temp. Forests	North. Forests	Trop. Wet Forests				
(τ≥0.50	e) Soil Moi	sture (equi	valent wate	er depth)										
0 < т < 0.50									H IRA-25/SiB					
nsignificant (p > 0.05)	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H JRA-25/SIB	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H JRA-25/SIB	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H CPC	H TerraClimate H NHM-PRMS H NHM-MWBM H JRA-25/SiB H CPC H NLDAS2-Mosaic	H NHM-MWBM H NLDAS2-Noah H NLDAS2-VIC H NLDAS2-Mosaic H NHM-PRMS H Torre Climate	H TerraClimate H NHM-PRMS H NHM-MWBM H JRA-25/SiB H NLDAS2-Noah H CPC	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H JRA-25/SiB	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM H JRA-25/SiB	H TerraClimate H NLDAS2-VIC H NLDAS2-Voah H NLDAS2-Mosaic H NHM-PRMS H NHM-MWBM	H TerraClimate H NLDAS2-VIC H NLDAS2-Noah H NLDAS2-Nosaic H NHM-PRMS H NHM-MWBM H JRA-25/SIB				
-0 50 < T < 0	H CPC	n CFC	H IRA-25/SiB	H NLDAS2-VIC H NLDAS2-Noah	H CPC	H NLDAS2-VIC H NLDAS2-Mosaic	n Gro	n Crc	H CPC	n crc				
-0.30 < (< 0	<i>n</i> 010		11 SIXA 23/315		H IRA-25/SiB									
1 2 0.00	Marine W. Coast Forest	Medit. CA	NW Forested Mtns	N. Amer. Deserts	So. Semiarid Highlands	Temp. Sierras	Great Plains	E. Temp. Forests	North. Forests	Trop. Wet Forests				
(f) Soil Moisture (volumetric content)														
τ≥0.50				,										
0 < T < 0.50									H IRA-25/SiB	H IRA-55/SiB				
ŧ						C 504.001			H JRA-55/SiB					
5) 5	H JRA-55/SiB	H JRA-55/SiB	H MERRA-2/CLSM	H GLEAM-v3.3a	S ESA-CCI	3 ESA-GUI	H MERRA-2/CLSM	H JRA-55/SiB	// NEDDA 0/01/01/	H MERRA-2/CLSM				
0.0	H JRA-25/SiB	H JRA-25/SiB	H GLEAM-v3.3a	S ESA-CCI	H FRA5/H-TESSEI	H JRA-25/SiB	H GLEAM-v3.3a	H JRA-25/SiB	H CLEAM-2/CLSM	H GLEAM-V3.3a				
^ du	H GLEAM-v3.3a	H GLEAM-v3.3a	S ESA-CCI	H ERA5/H-TESSEL	H JRA-55/SiB	H JRA-55/SiB	S ESA-CCI	H GLEAM-v3.3a	S ESA-CCI	S ESA-CCI				
nsi (p	H ERA5/H-TESSEL	H ERA5/H-TESSEL	H ERA5/H-TESSEL	H JRA-25/SiB	H GLEAM-v3.3a	H MERRA-2/CLSM	H ERA5/H-TESSEL	H ERA5/H-TESSEL	H ERA5/H-TESSEL	H ERA5/H-TESSEL				
-0 50 < T < 0			H JRA-25/SiB	H MERRA-2/CLSM	H MERRA-2/CLSM	H GLEAM-v3.3a	H JRA-25/SiB							
T < -0.50			H JRA-55/SiB	H JRA-55/SiB	H JRA-25/SiB		H JRA-55/SiB							
1 2 0.00	Marine W.	Medit.	NW Forested	N. Amer.	So. Semiarid	Temp.	Great	E. Temp.	North.	Trop. Wet				
	Coast Forest	CA	Mtns	Deserts	Highlands	Sierras	Plains	Forests	Forests	Forests				

Figure B5. Binned distributions of interannual model trends from 1982 to 2010, measured with Mann–Kendall's tau (τ) for 10 ecoregions, for water balance components of precipitation, evapotranspiration, runoff, snow water equivalent, and soil moisture. Model names are binned into very positive ($\tau \ge 0.50$; dark green), positive ($0 < \tau < 0.50$; light green), negative ($-0.50 < \tau < 0$; light purple), and very negative ($\tau \le -0.50$; dark purple) significant trends, as well as an insignificant trend box (p > 0.05; gray). Trend significance is assumed when the p value does not exceed 0.05 (p < 0.05). Model categories are denoted with a single letter identifying either a hydrologic (H), reanalysis (R), or remote sensing (S) model.

(d) Snow Water Equivalent

Model Correlation against Remote Sensing Products

Figure B6. Distribution of hydrologic model and reanalysis data set correlation against remote sensing products using Spearman's rho (ρ), provided by water budget component. Values of ρ are binned by >0.90, 0.50–0.90, <0.50, and statistically insignificant (p>0.05). Rectangles are color paired with their associated ecoregions (Fig. 1). Water budget components shown are precipitation (precip), evapotranspiration (ET), snow water equivalent (SWE), and soil moisture (SM).

Data availability. All data used in this study were obtained from public domains and are freely available. Primary data sources for hydrologic, reanalysis, and remote sensing models are provided in Appendix A. Data used for this study were converted from their primary structures (rasters and shapefiles) to EPA Ecoregions Level 1 and are available through the U.S. Geological Survey ScienceBase in Saxe et al. (2020; https://doi.org/10.5066/P9588YM2).

Author contributions. SaS, WF, JD, and TSH conceptualized and developed the methodology for this paper. Funding acquisition was performed by WF, JD, and TSH. Data curation, formal analysis, visualization, and writing of original draft was performed by SaS. WF, JD, and TSH provided project supervision and reviewed the prepublished work.

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

Disclaimer. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Acknowledgements. The authors are grateful for the early assistance of Flannery Dolan (Tufts University), Shirley Leung (University of Washington), and Adam Price (University of California Santa Cruz), as well as the invaluable input from journal reviewers.

Financial support. This work was performed under the U.S. Geological Survey's Water Availability and Use Science Program and the Hydrologic Sciences and Engineering fellowship from the Colorado School of Mines.

Review statement. This paper was edited by Ryan Teuling and reviewed by Nans Addor and Janneke Remmers.

References

- Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological applications and modelling, Int. J. Climatol., 33, 121–131, https://doi.org/10.1002/joc.3413, 2013.
- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Scientific Data, 5, 170191, 2018.
- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.
- Al Bitar, A., Kerr, Y. H., Merlin, O., Cabot, F., and Wigneron, J.-P.: Global drought index from SMOS soil moisture, Geoscience

and Remote Sensing Symposium (IGARSS), 21 February 2013, Melbourne, Australia, 2013.

- Alsdorf, D. E., Rodrigues, E., and Lettenmaier, D. P.: Measuring surface water from space, Rev. Geophys., 45, 2, https://doi.org/10.1029/2006RG000197, 2007.
- Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., and Over, T.: Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51, 10078–10091, https://doi.org/10.1002/2015WR017498, 2015.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system, J. Hydrometeorol., 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.
- Barrett, A.: National Operational Hydrologic Remote Sensing Center, Snow Data Assimilation System (SNODAS), products at NSIDC, Special Report, National Snow and Ice Data Center, Boulder, Colorado, USA, available at: https://nsidc.org/ sites/nsidc.org/files/files/nsidc_special_report_11.pdf (last access: 1 November 2019), 2003.
- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, Earth Syst. Sci. Data, 5, 71–99, https://doi.org/10.5194/essd-5-71-2013, 2013.
- Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process., 16, 189–206, https://doi.org/10.1002/hyp.343, 2002.
- Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.
- Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51, 4923–4947, https://doi.org/10.1002/2015WR017173, 2015.
- Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., and Bittelli, M.: Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, Remote Sens. Environ., 115, 3390–3408, https://doi.org/10.1016/j.rse.2011.08.003, 2011.
- Broxton, P. D., Zeng, X., and Dawson, N.: Why do global reanalyses and land data assimilation products underestimate snow water equivalent?, J. Hydrometeorol., 17, 2743–2761, https://doi.org/10.1175/JHM-D-16-0056.1, 2016.
- Carter, E., Hain, C., Anderson, M., and Steinschneider, S.: A water balance–based, spatiotemporal evaluation of terrestrial evapotranspiration products across the contiguous United States, J. Hydrometeorol., 19, 891–905, https://doi.org/10.1175/JHM-D-17-0186.1, 2018.
- Chang, A. T. C. and Rango, A.: Algorithm theoretical basis document (ATBD) for the AMSR-E Snow Water Equivalent Algoritm, NASA/GSFC, available at: https://nsidc.org/sites/nsidc.

org/files/technical-references/amsr_atbd_snow.pdf (1 November 2019), 2000.

- Chang, A. T. C., Kelly, R. E. J., Foster, J. L., and Hall, D. K.: Global SWE monitoring using AMSR-E data, in: Proceedings IGARSS 2003, IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003, 680–682, https://doi.org/10.1109/IGARSS.2003.1293880, 2003.
- Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G.-Y., Wen, L., and Yang, Z.-L.: Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study, J. Geophys. Res.-Atmos., 119, 13795–13819, https://doi.org/10.1002/2014JD022167, 2014.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498–2514, https://doi.org/10.1002/2015WR017198, 2015a.
- Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models, Water Resour. Res., 51, 5929–5956, https://doi.org/10.1002/2015WR017096, 2015b.
- Covey, C., Achuta Rao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J., and Taylor, K. E.: An overview of results from the Coupled Model Intercomparison Project, Global Planet. Change, 37, 103–133, https://doi.org/10.1016/S0921-8181(02)00193-5, 2003.
- Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.-L.: The Common Land Model, B. Am. Meteorol. Soc., 84, 1013– 1024, https://doi.org/10.1175/BAMS-84-8-1013, 2003.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, Int. J. Climatol., 28, 2031–2064, https://doi.org/10.1002/joc.1688, 2008.
- Dawson, N., Broxton, P., Zeng, X., Leuthold, M., Barlage, M., and Holbrook, P.: An evaluation of snow initializations in NCEP global and regional forecasting models, J. Hydrometeorol., 17, 1885–1901, https://doi.org/10.1175/JHM-D-15-0227.1, 2016.
- Dawson, N., Broxton, P., and Zeng, X.: Evaluation of remotely sensed snow water equivalent and snow cover extent over the contiguous United States, J. Hydrometeorol., 19, 1777–1791, https://doi.org/10.1175/JHM-D-18-0007.1, 2018.
- Derin, Y. and Yilmaz, K. K.: Evaluation of multiple satellite-based precipitation products over complex topography, J. Hydrometeorol., 15, 1498–1516, https://doi.org/10.1175/JHM-D-13-0191.1, 2014.
- Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, https://doi.org/10.5194/hess-13-913-2009, 2009.
- Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel analysis and implications for our per-

ception of the land surface, B. Am. Meteorol. Soc., 87, 1381–1398, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

- Donat, M. G., Sillmann, J., Wild, S., Alexander, L. V., Lippmann, T., and Zwiers, F. W.: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets, J. Climate, 27, 5019–5035, https://doi.org/10.1175/JCLI-D-13-00405.1, 2014.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sens. Environ., 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.
- Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res.-Atmos., 108, D22, https://doi.org/10.1029/2002JD003296, 2003.
- Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., and Clark, M. P.: How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?, J. Hydrometeorol., 15, 1384–1403, https://doi.org/10.1175/JHM-D-13-083.1, 2014.
- Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., Barr, A., Bartlett, P., and Elder, K.: SNOWMIP2: An evaluation of forest snow process simulations, B. Am. Meteorol. Soc., 90, 1120–1136, https://doi.org/10.1175/2009BAMS2629.1, 2009.
- Fan, Y. and van den Dool, H.: Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, J. Geophys. Res., 109, D10102, https://doi.org/10.1029/2003JD004345, 2004.
- Freeze, R. A. and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, J. Hydrol., 9, 237–258, https://doi.org/10.1016/0022-1694(69)90020-1, 1969.
- Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, Int. J. Remote Sens., 31, 3955–3978, https://doi.org/10.1080/01431161.2010.483488, 2010.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.
- Guirguis, K. J. and Avissar, R.: A precipitation climatology and dataset intercomparison for the western United States, J. Hydrometeorol., 9, 825–841, https://doi.org/10.1175/2008JHM832.1, 2008.
- Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten,

D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeorol., 12, 869–884, https://doi.org/10.1175/2011JHM1324.1, 2011.

- Hofstra, N., Haylock, M., New, N., Jones, P., and Frei, C.: Comparison of six methods for the interpolation of daily, European climate data, J. Geophys. Res., 113, D21, https://doi.org/10.1029/2008JD010100, 2008.
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol., 8, 38–55, https://doi.org/10.1175/JHM560.1, 2007.
- Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J.: The TRMM Multi-Satellite Precipitation Analysis (TMPA), in: Satellite Rainfall Applications for Surface Hydrology, edited by: Gebremichael, M. and Hossain, F., Springer, Dor-drecht, The Netherlands, 3–22, https://doi.org/10.1007/978-90-481-2915-7_1, 2010.
- Huffman, G. J., Adler, R. F., Bolvin, D. T., Hsu, K., Kidd, C., Nelkin, E. J., Tan, J., and Xie, P.: Algorithm Theoretical Basis Document (ATBD) for Global Precipitation Climatology Project Version 3.0, Precipitation Data, MEaSUREs project, Greenbelt, Maryland, USA, 26 pp., 2019.
- Jian, X., Wolock, D., and Lins, H. F.: WaterWatch Maps, graphs, and tables of current, recent, and past streamflow conditions, Report, U.S. Geological Survey, Reston, Virginia, USA, 2008.
- Kader, G. D. and Perry, M.: Variability for categorical variables, Journal of Statistics Education, 15, 2, https://doi.org/10.1080/10691898.2007.11889465, 2007.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, B. Am. Meteorol. Soc., 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP–DOE AMIP-II Reanalysis (R-2), B. Am. Meteorol. Soc., 83, 1631–1644, https://doi.org/10.1175/BAMS-83-11-1631, 2002.
- Kendall, M. G.: A new measure of rank correlation, Biometrika, 30, 81–93, https://doi.org/10.1093/biomet/30.1-2.81, 1938.
- Kobayashi, S. and NCAR Research Staff: The Climate Data Guide: JRA-55, available at: https://climatedataguide.ucar.edu/ climate-data/jra-55, last access: 1 January 2019.
- Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 reanalysis: General specifications and basic characteristics, J. Meteorol. Soc. Jpn., 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.
- Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E. T., Cordano, E., Endrizzi, S., Kikinzon, E., Mouche, E., Mügler, C., Park, Y.-J., Refsgaard, J. C., Stisen, S., and Sudicky, E.: The integrated hydrologic model intercompari-

son project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour. Res., 53, 867–890, https://doi.org/10.1002/2016WR019191, 2017.

- Koster, R. D. and Suarez, M. J.: Modeling the land surface boundary in climate models as a composite of independent vegetation stands, J. Geophys. Res.-Atmos., 97, 2697–2715, https://doi.org/10.1029/91JD01696, 1992.
- Koster, R. D. and Suarez, M. J.: Energy and water balance calculations in the Mosaic LSM, Goddard Space Flight Center, Greenbelt, Maryland, USA, available at: https://gmao.gsfc.nasa.gov/ pubs/docs/Koster130.pdf (last access: 1 October 2019), 1996.
- Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, J. Geophys. Res.-Atmos., 105, 24809–24822, https://doi.org/10.1029/2000JD900327, 2000.
- Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., and Puma, M. J.: On the nature of soil moisture in land surface models, J. Climate, 22, 4322–4335, https://doi.org/10.1175/2009JCLI2832.1, 2009.
- La Fontaine, J. H., Hay, L. E., Viger, R. J., Regan, R. S., and Markstrom, S. L.: Effects of climate and land cover on hydrology in the southeastern US: Potential impacts on watershed planning, J. Am. Water Resour. As., 51, 1235–1261, https://doi.org/10.1111/1752-1688.12304, 2015.
- Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, 4, https://doi.org/10.1029/2011WR011453, 2012.
- Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation model, J. Geophys. Res., 99, 415– 428, 1994.
- Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Adreadis, K. M., Maurer, E. P., and Lettenmaier, D. P.: A longterm hydrologically based dataset of land surface fluxes and states for the conterminous United States, J. Climate, 26, 9384– 9392, https://doi.org/10.1175/JCLI-D-12-00508.1, 2013.
- Manabe, S.: Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface, Mon. Weather Rev., 97, 739–774, https://doi.org/10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2, 1969.
- Markstrom, S. L., Regan, S. R., Hay, L. E., Viger, R. J., Webb, R. M. T., Payn, R. A., and La Fontaine, J. H.: PRMS-IV, the Precipitation-Runoff Modeling System, Version 4, US Geological Survey, Reston, Virginia, USA, available at: https://pubs. usgs.gov/tm/6b7/ (1 June 2019), 2015.
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903– 1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.
- Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, J. Climate, 15, 3237–3251, 2002.
- Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S. J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y.-J., Phanikumar, M. S., Shen, C., Sudicky, E. A., and Sulis, M.: Surface-subsurface

model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour. Res., 50, 1531–1549, https://doi.org/10.1002/2013WR013725, 2014.

- McCabe, G. J. and Markstrom, S. L.: A monthly water-balance model driven by a graphical user interface, USGS Numbered Series, US Geological Survey, Reston, Virginia, USA, 2007.
- McCabe, G. J., Wolock, D. M., and Austin, S. H.: Variability of runoff-based drought conditions in the conterminous United States, Int. J. Climatol., 37, 1014–1021, https://doi.org/10.1002/joc.4756, 2017.
- McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data, Geosci. Model Dev., 9, 283–305, https://doi.org/10.5194/gmd-9-283-2016, 2016.
- McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The future of Earth observation in hydrology, Hydrol. Earth Syst. Sci., 21, 3879–3914, https://doi.org/10.5194/hess-21-3879-2017, 2017.
- Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., and Arnold, J. R.: Effects of hydrologic model choice and calibration on the portrayal of climate change impacts, J. Hydrometeorol., 16, 762–780, https://doi.org/10.1175/JHM-D-14-0104.1, 2015.
- Mesinger, F., Di Mego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, B. Am. Meteorol. Soc., 87, 343–360, https://doi.org/10.1175/BAMS-87-3-343, 2006.
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.
- Mizukami, N., Clark, M. P., Slater, A. G., Brekke, L. D., Elsner, M. M., Arnold, J. R., and Gangopadhyay, S.: Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions, J. Hydrometeorol., 15, 474– 488, https://doi.org/10.1175/JHM-D-13-036.1, 2014.
- Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, Water Resour. Res., 53, 8020–8040, https://doi.org/10.1002/2017WR020401, 2017.
- Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, https://doi.org/10.1016/j.rse.2007.04.015, 2007.
- Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.
- Mudryk, L. R. and Derksen, C.: CanSISE Observation-Based Ensemble of Northern Hemisphere Terrestrial Snow Water Equivalent, Version 2, NSIDC: National Snow and Ice Data Center,

Boulder, Colorado, USA, https://doi.org/10.5067/96ltniikJ7vd, 2017.

- Mudryk, L. R., Derksen, C., Kushner, P. J., and Brown, R.: Characterization of northern hemisphere snow water equivalent datasets, 1981–2010, J. Climate, 28, 8037–8051, https://doi.org/10.1175/JCLI-D-15-0229.1, 2015.
- Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, 6, https://doi.org/10.1029/2010GL046230, 2011.
- Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.68d2bb3, 2019.
- National Operational Hydrologic Remote Sensing Center: Snow Data Assimilation System (SNODAS) data products at NSIDC, Version 1, NSIDC: National Snow and Ice Data Center, Boulder, Colorado, USA, 2004.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: 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, Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.
- NCAR Research Staff: The Climate Data Guide: JRA-25, available at: https://climatedataguide.ucar.edu/climate-data/jra-25 (last access: 1 December 2019), 2016.
- Omernik, J. M. and Griffith, G. E.: Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework, Environ. Manage., 54, 1249–1266, https://doi.org/10.1007/s00267-014-0364-1, 2014.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 Reanalysis, J. Meteorol. Soc. Jpn., 85, 369–432, https://doi.org/10.2151/jmsj.85.369, 2007.
- Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R., K., Sheffield, J., and Wood, E. F.: Multisource estimation of long-term terrestrial water budget for major global river basins, J. Climate, 25, 3191– 3206, https://doi.org/10.1175/JCLI-D-11-00300.1, 2012.
- Peters-Lidard, C. D., Hossain, F., Leung, L. R., McDowell, N., Rodell, M., Tapiador, F. J., Turk, F. J., and Wood, A.: 100 years of progress in hydrology, Meteor. Mon., 59, 25.1–25.51, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0019.1, 2018.
- Prat, O. P. and Nelson, B. R.: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012), Hydrol. Earth Syst. Sci., 19, 2037–2056, https://doi.org/10.5194/hess-19-2037-2015, 2015.
- PRISM Climate Group, OSU: PRISM, available at: http://prism. oregonstate.edu (last access: 1 June 2019), 2004.
- Regan, R. S., Markstrom, S. L., Hay, L. E., Viger, R. J., Norton, P. A., Driscoll, J. M., and La Fontaine, J. H.: Description of the National Hydrologic Model for use with the Precipitation-Runoff

Modeling System (PRMS), Report, Reston, Virginia, USA, U.S. Geological Survey, 2018.

- Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, J. Climate, 24, 6322–6338, https://doi.org/10.1175/JCLI-D-10-05033.1, 2011.
- Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment of MERRA-2 land surface hydrology estimates, J. Climate, 30, 2937–2960, https://doi.org/10.1175/JCLI-D-16-0720.1, 2017.
- Reitz, M., Sanford, W. E., Senay, G. B., and Cazenas, J.: Annual estimates of recharge, quick-flow runoff, and evapotranspiration for the contiguous US using empirical regression equations, J. Am. Water Resour. As., 53, 961–983, https://doi.org/10.1111/1752-1688.12546, 2017.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.
- Rodell, M., Beaudoing, H. K., and NASA/GSFC/HSL: GLDAS CLM Land Surface Model L4 3 hourly $1.0 \times 1.0^{\circ}$ Subsetted V001, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA, https://doi.org/10.5067/83NO2QDLG6M0, 2007.
- Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The observed state of the water cycle in the early twenty-first century, J. Climate, 28, 8289–8318, 2015.
- Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J. M.: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies, Agr. Forest Meteorol., 170, 166–182, https://doi.org/10.1016/j.agrformet.2012.09.011, 2013.
- Running, S., Mu, Q., and Zhao, M.: MOD16A2 MODIS/Terra Net Evapotranspiration 8-day L4 Global 500 m SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, Sioux Falls, South Dakota, USA, https://doi.org/10.5067/MODIS/MOD16A2.006, 2017.
- Running, S. W. and Coughlan, J. C.: A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production pro-

cesses, Ecol. Model., 42, 125–154, https://doi.org/10.1016/0304-3800(88)90112-3, 1988.

- Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellström, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W.-P., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T. G., Stähli, M., Stöckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y., and Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), J. Geophys. Res.-Atmos., 114, D6, https://doi.org/10.1029/2008JD011063, 2009.
- Saxe, S., Farmer, W. H., Driscoll, J. M., and Hogue, T. S.: Collection of hydrologic models, reanalysis datasets, and remote sensing products aggregated by ecoregion over the CONUS from 1900–2018, US Geological Survey data release, https://doi.org/10.5066/P9588YM2, 2020.
- Scanlon, B. R., Longuevergne, L., and Long, D.: Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA, Water Resour. Res., 48, 4, https://doi.org/10.1029/2011WR011312, 2012.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly land-surface precipitation from rain-sauges built on GTS-based and historic data, https://doi.org/10.5676/DWD_GPCC/FD_M_V7_050, 2011.
- Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A.: A Simple Biosphere Model (SIB) for use within general circulation models, J. Atmos. Sci., 43, 505–531, https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2, 1986.
- Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc., 63, 1379–1389, https://doi.org/10.2307/2285891, 1968.
- Senay, G. B.: Modeling landscape evapotranspiration by integrating land surface phenology and a water balance algorithm, Algorithms, 1, 52–68, https://doi.org/10.3390/a1020052, 2008.
- Senay, G. B., Budde, M. E., and Verdin, J. P.: Enhancing the Simplified Surface Energy Balance (SSEB) approach for estimating landscape ET: Validation with the METRIC model, Agr. Water Manage., 98, 606–618, https://doi.org/10.1016/j.agwat.2010.10.014, 2011.
- 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, J. Am. Water Resour. As., 49, 577–591, https://doi.org/10.1111/jawr.12057, 2013.
- Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., and McCabe, M. F.: Closing the terrestrial water budget from satellite remote sensing, Geophys. Res. Lett., 36, 7, https://doi.org/10.1029/2009GL037338, 2009.
- Smith, R. A. and Kummerow, C. D.: A comparison of in situ, reanalysis, and satellite water budgets over the Upper Colorado river basin, J. Hydrometeorol., 14, 888–905, https://doi.org/10.1175/JHM-D-12-0119.1, 2013.
- Spearman, C.: The proof and measurement of association between two things, Am. J. Psychol., 15, 72–101, 1904.

- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Rev. Geophys., 56, 79–107, https://doi.org/10.1002/2017RG000574, 2018.
- Tedesco, M., Kelly, R., Foster, J. L., and Chang, A. T.: AMSR-E/Aqua Monthly L3 Global Snow Water Equivalent EASE-Grids, Version 2, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/AMSR-E/AE_MOSNO.002, 2004.
- Thomas, B. F. and Famiglietti, J. S.: Identifying climate-induced groundwater depletion in GRACE observations, Sci. Rep., 9, 4124, https://doi.org/10.1038/s41598-019-40155-y, 2019.
- Thornthwaite, C. W.: An approach toward a rational classification of climate, Geogr. Rev., 38, 55–94, 1948.
- Thornton, M., Thornton, P. E., Wei, Y., Mayer, B. W., Cook, R. B., and Vose, R. S.: Daymet: Monthly climate summaries on a 1 km grid for North America, Version 3, ORNL DAAC, Oak Ridge, Tennessee, USA, 2018.
- Thornton, P. E., Running, S. W., and White, M. A.: Generating surfaces of daily meteorological variables over large regions of complex terrain, J. Hydrol., 190, 214–251, https://doi.org/10.1016/S0022-1694(96)03128-9, 1997.
- Thornton, P. E., Hasenauer, H., and White, M. A.: Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria, Agr. Forest Meteorol., 104, 255–271, https://doi.org/10.1016/S0168-1923(00)00170-2, 2000.
- Thornton, P. E., Thornton, M. M., Mayer, B. W., Wei, Y., Devarakonda, R., Vose, R. S., and Cook, R. B.: Daymet: Daily surface weather data on a 1 km grid for North America, Version 3, ORNL DAAC, Oak Ridge, Tennessee, USA, 2017.
- U.S. Geological Survey: Watershed Boundary Dataset (WBD) data model, available at: https://nhd.usgs.gov/wbd.html (last access: 1 January 2018), 2016.
- Velpuri, N. M. and Senay, G. B.: Partitioning evapotranspiration into green and blue water sources in the conterminous United States, Sci. Rep., 7, 6191, https://doi.org/10.1038/s41598-017-06359-w, 2017.
- 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 Sens. Environ., 139, 35–49, https://doi.org/10.1016/j.rse.2013.07.013, 2013.
- Velpuri, N. M., Senay, G. B., Driscoll, J. M., Saxe, S., Hay, L., Farmer, W., and Kiang, J.: Gravity Recovery and Climate Experiment (GRACE) storage change characteristics (2003– 2016) over major surface basins and principal aquifers in the conterminous United States, Remote Sens.-Basel, 11, 35–49, https://doi.org/10.3390/rs11080936, 2019.
- Voss, K. A., Famiglietti, J. S., Lo, M.-H., de Linage, C., Rodell, M., and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, Water Resour. Res., 49, 904–914, https://doi.org/10.1002/wrcr.20078, 2013.
- Vuyovich, C. M., Jacobs, J. M., and Daly, S. F.: Comparison of passive microwave and modeled estimates of total watershed SWE

in the continental United States, Water Resour. Res., 50, 9088–9102, https://doi.org/10.1002/2013WR014734, 2014.

- Willmott, C. J. and Matsuura, K.: Terrestrial air temperature and precipitation: Monthly and annual time series (1950–1999), available at: http://climate.geog.udel.edu/~climate/html_pages/ README.ghcn_ts2.html (1 June 2019), 2001.
- Xia, Y., NCEP/EMC: NLDAS Primary Forcing Data L4 Hourly 0.125 × 0.125° V002, edited by: Mocko, D., Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA, https://doi.org/10.5067/6J5LHHOHZHN4, 2009.
- Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: 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, J. Geophys. Res.-Atmos., 117, D3, https://doi.org/10.1029/2011JD016051, 2012a.
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: 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, J. Geophys. Res.-Atmos., 117, D3, https://doi.org/10.1029/2011JD016048, 2012b.
- Xia, Y.: NLDAS Mosaic Land Surface Model L4 Hourly $0.125 \times 0.125^{\circ}$ V002, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA, https://doi.org/10.5067/EN4MBWTCENE5, 2012c.
- Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., Meng, J., and Wood, E. F.: Evaluation of multi-model simulated soil moisture in NLDAS-2, J. Hydrol., 512, 107–125, https://doi.org/10.1016/j.jhydrol.2014.02.027, 2014.
- Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, B. Am. Meteorol. Soc., 78, 2539–2558, 1997.
- Yang, Z.-L. and Dickinson, R. E.: Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the Soil Moisture Workshop and evaluation of its performance, Global Planet. Change, 13, 117–134, https://doi.org/10.1016/0921-8181(95)00041-0, 1996.
- Zaherpour, J., Mount, N., Gosling, S. N., Dankers, R., Eisner, S., Gerten, D., Liu, X., Masaki, Y., Schmied, H. M., Tang, Q., and Wada, Y.: Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models, Environ. Modell. Softw., 114, 112–128, https://doi.org/10.1016/j.envsoft.2019.01.003, 2019.
- Zaussinger, F., Dorigo, W., Gruber, A., Tarpanelli, A., Filippucci, P., and Brocca, L.: Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data, Hydrol. Earth Syst. Sci., 23, 897–923, https://doi.org/10.5194/hess-23-897-2019, 2019.
- Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y. Y., Miralles, D. G., and Pan, M.: Multi-decadal trends in global terrestrial evapotranspiration and its components, Sci. Rep., 6, 19124, https://doi.org/10.1038/srep19124, 2016.

Zhang, Y., Pan, M., Sheffield, J., Siemann, A. L., Fisher, C. K., Liang, M., Beck, H. E., Wanders, N., MacCracken, R. F., Houser, P. R., Zhou, T., Lettenmaier, D. P., Pinker, R. T., Bytheway, J., Kummerow, C. D., and Wood, E. F.: A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010, Hydrol. Earth Syst. Sci., 22, 241–263, https://doi.org/10.5194/hess-22-241-2018, 2018.