

# Water Resources Research



## RESEARCH ARTICLE

10.1029/2020WR028658

## Budyko-Based Long-Term Water and Energy Balance Closure in Global Watersheds From Earth Observations

Akash Koppa<sup>1,2</sup> , Sarfaraz Alam<sup>1</sup> , Diego G. Miralles<sup>2</sup> , and Mekonnen Gebremichael<sup>1</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA, <sup>2</sup>Hydro-Climate Extremes Lab (H-CEL), Ghent University, Ghent, Belgium

### Key Points:

- A Budyko-based approach to water and energy balance closure mitigates the need for runoff data
- Errors in water and energy balance closure are influenced more by uncertainties in evaporation rather than precipitation
- Inability of Earth observations to close the water and energy balance of temperate forests

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

A. Koppa,  
[akashkoppa@ucla.edu](mailto:akashkoppa@ucla.edu)

### Citation:

Koppa, A., Alam, S., Miralles, D. G., & Gebremichael, M. (2021). Budyko-based long-term water and energy balance closure in global watersheds from Earth observations. *Water Resources Research*, 57, e2020WR028658. <https://doi.org/10.1029/2020WR028658>

Received 26 AUG 2020

Accepted 24 MAR 2021

**Abstract** Earth observations offer potential pathways for accurately closing the water and energy balance of watersheds, a fundamental challenge in hydrology. However, previous attempts based on purely satellite-based estimates have focused on closing the water and energy balances separately. They are hindered by the lack of estimates of key components, such as runoff. Here, we posit a novel approach based on Budyko's water and energy balance constraints. The approach is applied to quantify the degree of long-term closure at the watershed scale, as well as its associated uncertainties, using an ensemble of global satellite data sets. We find large spatial variability across aridity, elevation, and other environmental gradients. Specifically, we find a positive correlation between elevation and closure uncertainty, as derived from the Budyko approach. In mountainous watersheds the uncertainty in closure is  $3.9 \pm 0.7$  (dimensionless). Our results show that uncertainties in terrestrial evaporation contribute twice as much as precipitation uncertainties to errors in the closure of water and energy balance. Moreover, our results highlight the need for improving satellite-based precipitation and evaporation data in humid temperate forests, where the closure error in the Budyko space is as high as  $1.1 \pm 0.3$ , compared to only  $0.2 \pm 0.03$  in tropical forests. Comparing the results with land surface model-based data sets driven by in situ precipitation, we find that Earth observation-based data sets perform better in regions where precipitation gauges are sparse. These findings have implications for improving the understanding of global hydrology and regional water management and can guide the development of satellite remote sensing-based data sets and Earth system models.

## 1. Introduction

The terrestrial water and energy cycles are tightly coupled at various spatial and temporal scales; these links can be direct, due to the use and release of energy in water phase changes, as well as indirect, due to multiple feedbacks. At the watershed scale, the transfer of heat and moisture modulates the state of the atmosphere, soil surface, subsurface, and biosphere. For example, positive and negative soil moisture-precipitation feedbacks result from the interplay between the energy available to evaporate water and soil moisture, together with the response of the atmospheric boundary layer to the resulting latent and sensible heat fluxes (Eltahir, 1998; Guillod et al., 2015; Taylor et al., 2012). Recent studies have shown that the latent heat flux (i.e., terrestrial evaporation, hereafter "ET") is strongly coupled, not just to surface soil moisture, but also vegetation physiology (Mengis et al., 2015) and groundwater storage (Fan et al., 2017; Maxwell & Condon, 2016).

An important step in improving the understanding of land-atmosphere coupling at the watershed scale is to investigate the closure of the water and energy balances. In fact, in hydrology, the closure problem using observations is well recognized as a longstanding scientific challenge (Beven, 2006). Although Earth system models close the water and energy balances by design (with the use of continuity equations), the increased complexity of these models over the recent decades has not reflected on a commensurate increase in our understanding of watershed-level processes. Beven et al. (2020) attribute this discrepancy to the significant uncertainties in hydrological observations, represented in their inability to close the water balance without considerable errors in the different fluxes. Meanwhile, several studies have highlighted the challenges of accurately closing the water and energy balance of watersheds using observations from multiple sources (Aires, 2014; Beven, 2019; Pan et al., 2012; Pellet et al., 2019; Sahoo et al., 2011). Scrutinizing this closure at regional and global scales is deemed crucial to identify gaps in our understanding of watershed-scale processes and to reveal systematic errors in the state-of-the-art observations (Hegerl et al., 2015). Thus,

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

improving the accuracy of closure from observations can improve the prediction of regional water resources, impact water management, and enable a better adaptation to hydroclimatic extremes. In addition, accurate closure can help constrain past simulations and future projections of the hydrologic cycle by land surface models (Greve, Orlowsky, et al., 2014), global climate models (GCMs) (Allen & Ingram, 2002; Prein & Pendergrass, 2019) and hydrologic models (Greve, Burek, & Wada, 2020). Due to this importance, different studies have already attempted to assess this closure, but often disregarding the strong links between water and energy cycles: the general approach so far has been to investigate water (Sheffield, Ferguson, et al., 2009) and energy (Heusinkveld et al., 2004) balances separately.

Although ground-based instruments can be used to measure all the water and energy balance components, they are inadequate for assessing the closure at watershed scales because of upscaling issues inherent to their localized and sparse nature (Blöschl & Sivapalan, 1995). As an alternative, Earth observation satellite-based remote sensing provides spatially and temporally continuous estimates of key water and energy balance components (Lettenmaier et al., 2015). The use of multiple satellites enables continuous coverage across the entire globe, including regions that are ungauged by in situ measuring instruments. The spatial resolution of existing satellite-based estimates of water and energy balance components varies from 10–100 km, which appears adequate for closure studies of large watersheds without requiring any downscaling. However, using satellite remote sensing to assess the closure of the water and energy balance through traditional budget equations is not straightforward. A primary challenge is the lack of accurate data on watershed storage changes, ground heat flux, and runoff. The former two are either not available, or unavailable at the required spatial resolution; that is the case of watershed storage changes retrieved from the Gravity Recovery and Climate Experiment (GRACE) (Rodell, Velicogna, & Famiglietti, 2009).

Here, we aim to study the capabilities of satellite-based remote sensing data in closing the water and energy balance of large watersheds ( $>10,000 \text{ km}^2$ ) in a combined manner, using a methodology that uses information from the water and energy budgets to constrain each other. Previous closure studies have focused on either small study regions, or specific watersheds where reliable streamflow data were available. For instance, Heusinkveld et al. (2004) used field measurements to evaluate energy balance closure in a small arid watershed in Israel. Likewise, Sheffield, Ferguson, et al. (2009) and Pan et al. (2012) analyzed the ability of satellite-based data sets, in combination with streamflow measurements, to close the water balance of large watersheds. To circumvent the need for watershed storage changes and ground heat flux data, we concentrate on long timescales in which heat and water storage changes are revealed to be negligible compared to the magnitude of the cumulative fluxes. Nonetheless, this long-term perspective does not negate the need for runoff estimates.

To avoid the requirement of runoff, which cannot be retrieved from current satellites nor will they be derived directly with the next generation of sensors (Biancamaria et al., 2016), we based our study on the Budyko framework (Budyko, 1974), a semi-empirical representation of the long-term water and energy balance. The approach is applied under the hypothesis that more accurate precipitation (P) and ET data sets lead to more accurate closure of water and energy balances. This particular Budyko application has proven to mimic data evaluation methodologies which use ground-based measurements (Koppa & Gebremichael, 2017). As such, our study contributes to the growing body of literature where physical consistency of P and ET data products are evaluated based on the Budyko framework (Beck, Wood, McVicar, et al., 2020; Greve, Burek, & Wada, 2020; Greve, Orlowsky, et al., 2014; Hobeichi et al., 2020; Miralles, Jiménez, et al., 2016). In contrast to previous studies which primarily used in situ or land surface model-based data sets, this study represents a novel application of the Budyko hypothesis in understanding the current state of satellite-based remote sensing in characterizing the combined closure of water and energy balance of global watersheds. In doing so, we present a first evaluation of P and ET data sets in watersheds where such an evaluation has not been done before due to the lack of in situ data. In addition, the appraisal of closure at a global scale offers insights into the different physical controls, such as topography and land cover, which hinder or aid the effective retrieval of water and energy balance components from satellite-based remote sensing.

A final hurdle rises from the large variability in sensors and retrieval algorithms which are used to estimate the water and energy balance components, and their possible inconsistencies in accuracy, resolution, and coverage. This issue is aggravated by the fact that, in most cases, estimates stemming from a single satellite may not retrieve the component of interest accurately; this holds especially for ET, but it also pertains to

surface net radiation or precipitation. To estimate these variables with higher accuracy, most retrieval methods rely on measurements of different atmospheric and terrestrial properties from a myriad of satellite sensors. The choice of source data and algorithms to combine them leads to disagreements among the different data sets of water and energy balance components, such as precipitation (Beck, Vergopolan, et al., 2017; Hong, Tang, et al., 2018; Sun et al., 2018) and ET (McCabe et al., 2016; Miralles, Jiménez, et al., 2016; Mueller et al., 2011). To account for the uncertainties and errors in closure due to variability in sensors and retrieval algorithms, we use an ensemble of satellite-based precipitation and ET data sets. Finally, we analyze the patterns in closure across gradients in topography, aridity, greenness, land use, hydrology, and climate conditions.

In the following, we present details of the remote sensing-based data sets used in our analyses (Section 2), the methodology used in the study including the selection of watersheds at a global scale, description of the Budyko framework-based metrics developed in this study to analyze the closure of water and energy balance in the Budyko space using the remote sensing-based data sets (Section 3). In Section 4 and Section 5, we present and discuss the results of our study. Finally, we summarize our main conclusions (Section 6).

## 2. Remote Sensing Data Sets

This study considers remote sensing-based P and ET for evaluating the water and energy budget closure at watershed scales. We select eight data sets for ET, seven for P, and one for net radiation—see Table 1. The selected data sets have different temporal and spatial coverage, as listed in the table. We consider the 20-year period 1998–2017, which provides a sufficiently long temporal overlap between the different products for the analysis. For each combination of P and ET, the annual averages are determined for their period of overlap (within the 1998–2017 time period). In addition, we compare our remote sensing-based results with an ensemble of P and ET data sets derived from in situ observations, reanalysis, and land surface models. We present a brief description of the P, ET, and  $R_n$  data sets below.

### 2.1. Precipitation

The remote sensing-based precipitation products are primarily based on either longwave infrared or microwave observations, or a combination of both.

- (1) Climate Hazards Group Infrared Precipitation with Stations (CHIRPSv2.0) uses cold cloud duration data from thermal infrared sensors aboard satellites of the National Oceanic and Atmospheric Administration (NOAA). The satellites include the Geostationary Operational Environmental Satellites (GOES-8,10). The retrieval algorithm is based on the cold cloud duration of a pixel, that is, the length of time a pixel is covered by high cold clouds as revealed by infrared brightness temperatures. Further, the derived precipitation is calibrated with the Tropical Rainfall Measuring Mission (TRMM) Multi Satellite Precipitation Analysis (TMPA) 3B42v7 precipitation data set (Funk et al., 2015)
- (2) Climate Prediction Center Morphing Algorithm (CMORPHv0.x.RAW) precipitation is estimated exclusively from passive microwave observations from multiple satellites, such as the Defense Meteorological Satellite Program (DMSP) 12, 14, and 15, the NOAA-15, 16, 17, and 18, the Advanced Microwave Scanning Radiometer for Earth Observation Satellite (AMSR-E), and TRMM. In addition, infrared data are used to gap-fill precipitation where microwave data are not available through a propagation algorithm. The former is sourced from GEOS-8 and GEOS-10 from NOAA, Meteosat-7 and Meteosat-10 from the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), and the Geostationary Meteorological Satellite (GMS-5) from the Japan Aerospace Exploration Agency (JAXA) (Joyce et al., 2004)
- (3) Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) uses longwave infrared from multiple satellites (GOES-8, GOES-10, GMS-5, Meteosat-6, and Meteosat-7). PERSIANN also uses cold cloud pixel information from gridded brightness temperature, together with an artificial neural network model, to estimate precipitation (Sorooshian et al., 2000)
- (4) PERSIANN Cloud Classification System (PERSIANN.CCS) incorporates a cloud classification system which categorizes clouds based on cloud height, areal extent, and texture. These classifications are used

**Table 1**  
List of Satellite-Based Precipitation, Terrestrial Evaporation, and Net Radiation Data Sets Used for Closing the Water and Energy Balance of Watersheds

Data source	Type	Spatial resolution (degree)	Temporal coverage (years)	Spatial coverage (degree)	Reference
<i>Precipitation</i>					
CHIRPSv2.0	Satellite	0.05	1981–2019	50°N–50°S	Funk et al. (2015)
CMORPHv0.x.RAW	Satellite	0.25	2002–2017	60°N–60°S	Joyce et al. (2004)
PERSIANN	Satellite	0.25	2000–2019	60°N–60°S	Sorooshian et al. (2000)
PERSIANN.CCS	Satellite	0.04	2003–2019	60°N–60°S	Hong, Gochis, et al. (2007)
PERSIANN.CDR	Satellite	0.25	1983–2016	60°N–60°S	Ashouri et al. (2015)
TRMM.3B42RT	Satellite	0.25	2000–2019	50°N–50°S	Huffman et al. (2007)
TRMM.3B43	Satellite	0.25	2000–2019	50°N–50°S	Huffman et al. (2007)
CPC.Unifiedv1.0	Observation	0.50	1979–2019	Global	(Chen et al., 2008)
ERA5.Land	Reanalysis	0.10	1981–2019	Global	(Muñoz-Sabater et al., 2018)
GPCCv7.0	Gauge	1.00	1901–2013	Global	(Becker et al., 2013)
UDELv5.0	Gauge	0.50	1900–2017	Global	(Willmott & Matsuura, 2000)
<i>Terrestrial evaporation</i>					
AVHRR.NTSG	Satellite	0.08	1983–2013	Global	K. Zhang et al. (2010)
SSEBOPv4.0	Satellite	0.01	2003–2019	Global	Senay et al. (2011)
MOD16A3	Satellite	0.05	2000–2014	Global	Mu et al. (2011)
GLEAMv3.3a	Satellite	0.25	1980–2018	Global	Martens et al. (2017)
GLEAMv3.3b	Satellite	0.25	2003–2018	Global	Martens et al. (2017)
CSIRO-PMLv2.0	Satellite	0.50	1981–2012	Global	Y. Zhang et al. (2016)
BESS	Satellite	0.50	2000–2015	Global	Jiang and Ryu (2016)
FluxCom.RS	Satellite	0.50	2001–2015	Global	Tramontana et al. (2016)
ERA5.Land	Reanalysis	0.50	1981–2019	Global	(Muñoz-Sabater et al., 2018)
FLDAS	Reanalysis	0.10	1981–2019	Global	(McNally et al., 2017)
FluxCom.RSM	Machine learning and reanalysis	0.50	1979–2019	Global	(Tramontana et al., 2016)
GLDASv2.1	Reanalysis	0.25	1948–2019	Global	(Rodell, Houser, et al., 2004)
<i>Net radiation</i>					
CERESv4.0	Satellite	1.00	2000–2018	Global	Wielicki et al. (1996)

to assign rainfall values to each pixel based on a prespecified relationship between rain rate and brightness temperature (Hong, Gochis, et al., 2007)

- (5) PERSIANN Climate Data Record (PERSIANN.CDR) uses neural networks as in PERSIANN, but trained with the National Centers for Environmental Prediction (NCEP) precipitation data. Further, the precipitation estimates are adjusted using the Global Precipitation Climatology Project (GPCP) gauge-based data (Ashouri et al., 2015)
- (6) TRMM TMPA Real Time (TRMM.3B42RT) is the real-time data product from TMPA. It uses microwave data from multiple sensors that are intercalibrated and the data gaps are filled by calibrated infrared data. TRMM uses microwave measurements from multiple satellites (TRMM, DMSP, Aqua and NOAA satellite series) (Huffman et al., 2007)

- (7) TRMM TMPA Gauge Calibrated Data set (TRMM.3B43), in which TMPA 3B42 is calibrated with the Global Precipitation Climatology Center (GPCC) gauge-based data set (Huffman et al., 2007)

The following is a brief description of the four precipitation data sets based on in situ and reanalysis data:

- (1) Climate Prediction Center Unified Gauge-Based Data set (CPC.Unified.v1.0) is constructed using in situ precipitation observations from over 30,000 stations which are collected from different agencies across the globe. The analysis of the gauges is done at 0.125° and released at a resolution of 0.5° latitude-longitude grid (Chen et al., 2008)
- (2) European Center for Medium-Range Weather Forecasts Reanalysis 5 Land (ERA5.Land) combines weather forecasts with observations from multiple sources to generate a globally consistent estimate of atmospheric and land variables. ERA5.Land is a higher resolution replay of the land component of ERA5. ERA5 assimilates a large number of observations, including ground-based radars, satellite-based observations, atmospheric sounders among others (Muñoz-Sabater et al., 2018)
- (3) Global Precipitation Climatology Center Gauge-Based Data set (GPCC.v7.0) is an in situ gauge-based analysis based on approximately 67,200 stations worldwide (Becker et al., 2013)
- (4) University of Delaware Precipitation Data set (UDEL.v5.0) is a monthly data set constructed using observations a large number of in situ stations sourced from the Global Historical Climatology Network (GHCN) and national institutions from different countries (approximately 22,000 stations) (Willmott & Matsuura, 2000)

## 2.2. Terrestrial Evaporation

ET derivation algorithms generally use combinations of satellite observations and reanalysis data of multiple meteorological variables.

- (1) Numerical Terradynamic Simulation Group (NTSG) Algorithm (AVHRR.NTSG) ET is based on data from Advanced Very High-Resolution Radiometer (AVHRR) sensors. It employs a modified Penman-Monteith (PM) method to quantify the potential canopy transpiration and soil evaporation. For open water evaporation, it uses a Priestley and Taylor (PT) approach. The PM method is modified by incorporating biome-specific canopy conductance determined using NDVI from AVHRR (K. Zhang et al., 2010). The other meteorological inputs required by the PM and PT approaches—such as maximum, minimum, and average air temperatures, water vapor pressures—are sourced from the NCEP-National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al., 1996)
- (2) Simplified Surface Energy Balance Operational Data set (SSEBOPv4.0) applies a simplified surface energy balance model that combines ET fractions from MODIS thermal imagery with PM-based reference ET (Senay et al., 2011). The meteorological variables required by the PM are derived from the Global Data Assimilation System (GDAS), an analysis system which provides initial conditions for the medium-range forecast models from NOAA
- (3) Moderate Resolution Imaging Spectrometer (MODIS) Data set (MOD16A3) uses a semi-empirical ET algorithm that relies on a PM equation to estimate potential ET (Mu et al., 2011). The input variables for the PM equation are derived from NASA's Global Modeling and Assimilation Office analysis product (GEOSDAS)
- (4) Global Land Evaporation Amsterdam Model (GLEAM) v3.3a combines multiple remote sensing measurements into a PT-based framework, where remote sensing inputs include microwave-derived soil moisture and vegetation optical depth among others (Martens et al., 2017). Interception loss is estimated through a Gash analytical model (Miralles, Gash, et al., 2010). The surface radiation and air temperature are sourced from the European Center for Medium-Range Weather Forecasts (ECMWF) reanalysis product (ERA5). P input comes from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) data set (Beck, Wood, Pan, et al., 2019)
- (5) GLEAM v3.3b is based on the same algorithm described above but, unlike other ET data sets, it only uses satellite-based inputs and no reanalysis forcing (Martens et al., 2017). For surface radiation, the Clouds and Earth's Radiant Energy System (CERES) data set is used. The air temperature is derived from the Atmospheric Infrared Sounder (AIRS) system onboard Aqua (Aumann et al., 2003)
- (6) Commonwealth Scientific and Industrial Organization (CSIRO) PM-Leuning (PM-L) (CSIRO.PMLv2.0) uses a PML-based model that utilizes reanalysis meteorological forcing and satellite-derived inputs. The



- meteorological inputs are sourced from the Princeton Global Forcing (Sheffield, Goteti, & Wood, 2006) and the Water and Global Change (WATCH) Forcing Data ERA-Interim (WFDEI) (Weedon et al., 2014). Satellite-derived inputs include leaf area index from AVHRR, and albedo from the Global Land Surface Satellite (GLASS) data set (Liang et al., 2013). Its interception loss is based on a modified Gash analytical model. At the annual timescale, the model is constrained by the Budyko framework such that annual ET is always less than annual P for grid cells not covered by crops (K. Zhang et al., 2010)
- (7) Breathing Earth System Simulator (BESS) uses a process-based model which couples atmospheric and canopy radiative transfer, and canopy photosynthesis. It uses MODIS atmosphere and land data along with other satellite, reanalysis (ERA-Interim) and ancillary data to derive ET using a PM-based potential ET estimates and stomatal conductance based on Ball et al. (1987). Unlike other models, it explicitly links canopy conductance to carbon concentrations in the atmosphere (Jiang & Ryu, 2016)
  - (8) FLUXCOM Initiative data set (FluxCom.RS) uses multiple machine-learning techniques to train algorithms with eddy covariance data from the FLUXNET global network of flux towers (Baldocchi et al., 2001), and applies MODIS observations (land surface temperature, land cover, fraction of absorbed radiation by canopy) as predictors to yield ET estimates (Tramontana et al., 2016)

ET data sets which are generated using LSMs or machine learning models driven with in situ or reanalysis forcing are described below.

- (1) European Center for Medium-Range Weather Forecasts Reanalysis 5 Land (ERA5.Land), in which, as described above, ET reanalysis estimates are produced using the land component of ERA5, albeit at a higher resolution (Muñoz-Sabater et al., 2018)
- (2) Famine Early Warning Systems Network Land Data Assimilation System (FLDAS), which is part of a global land data assimilation system created specifically to address the monitoring and forecast requirements associated with food security in data-scarce regions. We used the outputs from the Noah LSM run using NASA's Land Information System (LIS). The primary forcing is precipitation from CHIRPS (McNally et al., 2017)
- (3) FLUXCOM Initiative Remote Sensing and Model Data (FluxCom.RSM) uses a machine learning algorithm similar to FluxCOM.RS. However, the machine learning model is driven by model-based climate variables (Tramontana et al., 2016)
- (4) Global Land Data Assimilation System (GLDAS.v2.1) assimilates satellite and in situ observations into multiple LSMs to generate fields of land surface fluxes and states. In this study, we use the instance of GLDAS which uses the Noah LSM (Rodell, Houser, et al., 2004)

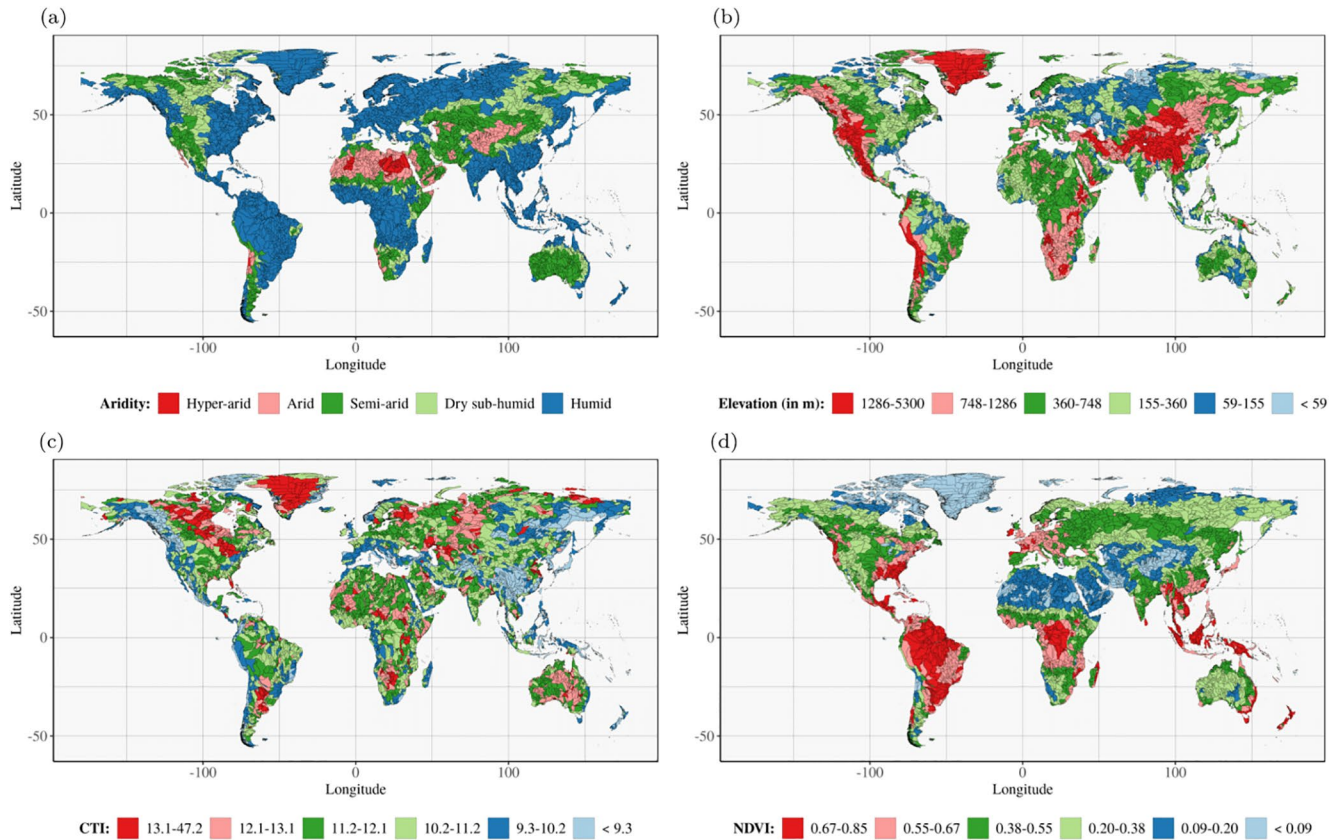
### 2.3. Net Radiation

We use net radiation ( $R_n$ ) data from the CERES instruments aboard the Terra satellites.  $R_n$  is computed in a two-step process. In the first step, uncalibrated top-of-the-atmosphere fluxes are produced using meteorological variables, cloud, aerosol, and ozone properties as input to a delta-two stream radiation transfer model. In the next step, the top-of-the-atmosphere fluxes are calibrated using observations from the CERES instruments and the input variables are modified. The calibrated input variables are then used for second run of the radiation transfer model to generate fluxes at different pressure levels (top of atmosphere to surface) (Wielicki et al., 1996).

## 3. Methodology

### 3.1. Selection of Watersheds

Spatially, most of the data sets used in this study cover the entire globe, yet some P data sets are unavailable over high latitudes. To maintain consistency, we limit our analysis to watersheds between 60°N and 60°S. We source the delineated watersheds from the HydroBASINS global watershed data set, which in turn are derived from HydroSHEDS (Lehner et al., 2008). The HydroBASINS data set consists of hierarchically nested basins at 12 levels, with level 1 representing the largest basins and level 12 the smallest. HydroBASINS classifies the basins according to the Pfafstetter coding system in which a larger basin at any level is divided into nine smaller units at the next level (Lehner & Grill, 2013). Out of the 12 levels, we choose 4,734 level-5 watersheds (Figure 1a). Choosing level 5 yields a large-enough sample; at the same time,



**Figure 1.** Study area consisting of 4,734 watersheds based on HydroBASINS (Pfafstetter level 5) and classified according to (a) aridity, (b) elevation, (c) complex topographic index (CTI), and (d) normalized difference vegetation index (NDVI).

the input satellite-based precipitation and ET data sets are available at  $<0.5^\circ$  resolution (approximately  $2,500 \text{ km}^2$ ), while the average area of the 4,734 level-5 watersheds is approximately  $28,500 \text{ km}^2$ , that is, still larger than the data set grid-cells. Figure 1a also presents a hydroclimatic classification of the 4,734 watersheds according to aridity index ( $AI = PET/P$ , where PET is potential evaporation and P is precipitation). Five aridity classes are considered, in agreement with the United Nations Environmental Program (UNEP) (Grove, 1999): (1) Humid ( $0 < AI \leq 1.5$ ), (2) Dry sub-humid ( $1.5 < AI \leq 2.0$ ), (3) Semi-arid ( $2.0 < AI \leq 5.0$ ), (4) Arid ( $5.0 < AI \leq 33.3$ ), and (5) Hyper-arid ( $AI > 33.3$ ). We also present global maps of additional watershed characteristics, including elevation (Figure 1b), compound topographic index (CTI) (Figure 1c), and Normalized Difference Vegetation Index (NDVI) (Figure 1d).

### 3.2. Long-Term Water and Energy Balance

Under the assumptions of negligible net water storage change and ground heat flux (Sposito, 2017), the long-term water (Equation 1) and energy (Equation 2) balances of watersheds can be expressed as

$$P = ET + Q \quad (1)$$

$$R_n = \lambda \cdot ET + H \quad (2)$$

where,  $Q$  is runoff,  $\lambda \cdot ET$  is surface latent heat flux ( $\lambda$  is latent heat of vaporization), and  $H$  is surface sensible heat flux. In this study, PET is calculated as  $0.8 \cdot R_n$  based on Milly and Dunne (2016).

### 3.3. Budyko Hypothesis

The Budyko hypothesis describes the long-term water and energy balance of watersheds using two indices: (1) Evaporative Index (ET/P, abbreviated as EI) and (2) Aridity Index (PET/P, abbreviated as AI). In its original formulation (Budyko, 1974), AI and EI were related by a non-parametric and nonlinear function. However, several parametric versions of the Budyko function have been formulated ever since (Choudhury, 1999; Fuh, 1981; Yang et al., 2008; L. Zhang et al., 2004). These deviations have since been attributed to different watershed characteristics such as soil (Porporato et al., 2004), vegetation (Donohue et al., 2012; Good et al., 2017), and groundwater (Istanbulluoglu et al., 2012). In this study, we utilize Fuh's equation (Fuh, 1981), a single-parameter Budyko function to close the combined water and energy balance of watersheds:

$$\frac{ET}{P} = 1 + \frac{PET}{P} - \left( 1 + \left( \frac{PET}{P} \right)^\omega \right)^{\frac{1}{\omega}} \quad (3)$$

where  $\omega$  is a parameter which has no analytical solution, but several parameterizations have been suggested relating it to watershed characteristics such as topography, soil, vegetation, and geographical location of watersheds among others (Li et al., 2013; Xu et al., 2013).

The Budyko curve (Equation 3) is constrained by the following water and energy limits:

$$\frac{ET}{P} = 1, \frac{PET}{P} > 1 \quad (\text{water limit}) \quad (4)$$

$$\frac{ET}{P} = \frac{PET}{P}, \frac{PET}{P} < 1 \quad (\text{energy limit}) \quad (5)$$

The water limit (Equation 4) implies that ET cannot exceed P in the long term. Implicit in the constraint is the assumption that the contribution of watershed storage to long-term water availability is negligible, which is not always a valid assumption (see Section 3.4). In watersheds where water availability is not a constraint, ET is limited by the available energy in the form of  $R_n$  or PET (Equation 5). Figure 2 shows a representative Budyko space (AI vs. EI) consisting of the Budyko curve (Equation 3), the water limit (Equation 4), and the energy limit (Equation 5).

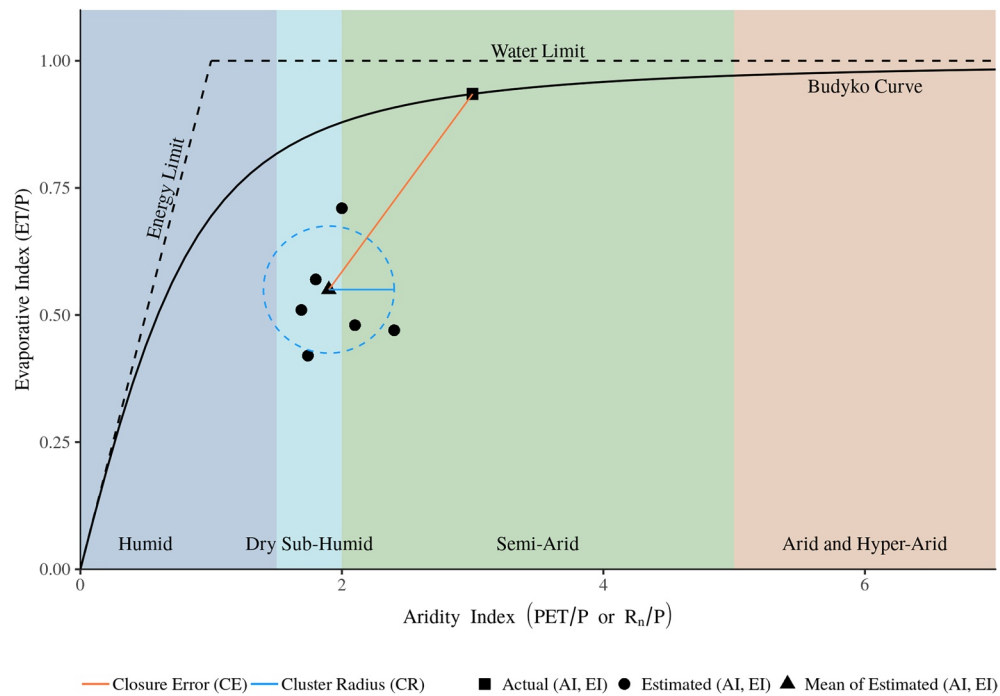
In this study, we postulate that the Budyko function (Equation 3) offers a reliable proxy for traditional water (Equation 1) and energy (Equation 2) balance equations at the watershed scale (Beck, Wood, McVicar, et al., 2020; Greve, Orlowsky, et al., 2014). We note here that Fuh's equation is semi-empirical, and while it is grounded on well-established water and energy constraints, it is not equivalent to a water and energy balance closure from first principles. However, the equation provides an effective representation of these long-term balances that has been evaluated in numerous previous studies referenced above.

### 3.4. Deviations From the Budyko Hypothesis

The efficacy of the Budyko hypothesis in representing the long-term combined water-energy balance of watersheds is well known. However, watersheds may violate the assumptions (detailed above) under which the hypothesis is applicable. A major and observed deviation happens in watersheds with continuous and prolonged groundwater extraction, which leads to the exceedance of the water limit (Equation 4) (Condon & Maxwell, 2017; Destouni et al., 2013; Jaramillo & Destouni, 2015; Moshir Panahi et al., 2020). In certain watersheds with high heat advection and interception loss, the energy limit (Equation 5) may also be exceeded (Holwerda et al., 2012).

In this study, we address the issue of the violation of the water limit arising from long-term changes in watershed storage. According to Condon and Maxwell (2017), long-term groundwater extraction (for example, irrigation) can cause significant increases in ET. Thus, the water limit (Equation 4) can be redefined to include this additional source by irrigation as  $ET/(P + \Delta TWS) = 1$ , where  $\Delta TWS$  represents the change ( $\Delta$ ) in



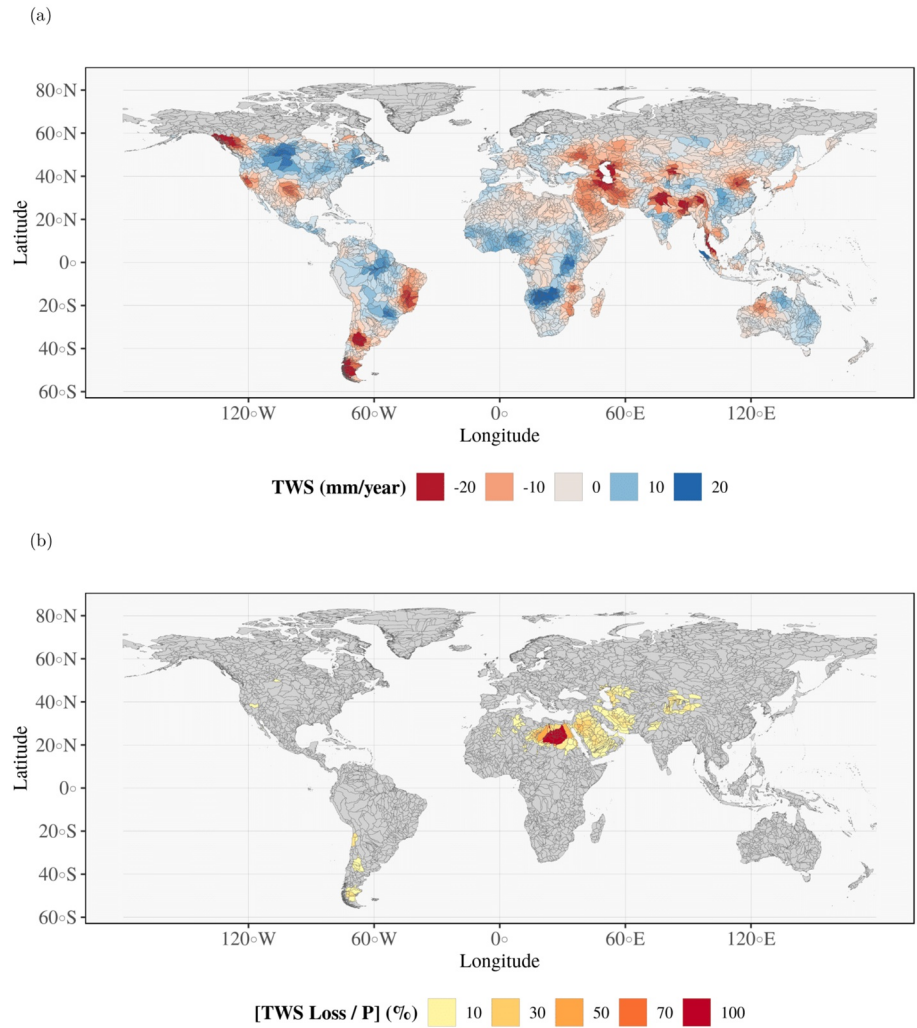


**Figure 2.** The Budyko space consisting of a representative Budyko curve constrained by the water and energy limits. We also represent the cluster radius (CR) (solid blue line) and the closure error (CE) (solid orange line) used in this study to quantify the uncertainty and error in water and energy balance closure of watersheds, respectively. Each point represents a  $[P, ET, R_n]$  combination in the Budyko space. The analyses are based on 56 such combinations per watershed. The triangle represents the centroid of the 56 combinations.

total water storage (TWS). Therefore, it can be reasonably assumed that watersheds which have substantial contribution from  $\Delta TWS$  to their water availability can deviate from the water and energy balance described by the Budyko hypothesis. In this study, we identify and exclude such watersheds using the following procedure. First, we use the global TWS anomaly data from GRACE, specifically the solution from the Jet Propulsion Laboratory (JPL) (Watkins et al., 2015), to calculate the mean (2002–2017)  $\Delta TWS$  in each watershed (Figure 3a). The procedure detailed in Rodell, Famiglietti, et al. (2018) is followed to fit a linear trend to the monthly TWS anomalies and the trend values are aggregated to mm/year for each watershed. Then, we exclude watersheds in which the mean  $\Delta TWS$  is greater than 5% of the mean  $P$ , under the assumption that if  $\Delta TWS$  is less than 5% it cannot substantially influence the appraisal of (Budyko-based) water and energy balance closure (described below). Based on this criterion, we exclude 266 watersheds from our analyses (Figure 3b). These watersheds are primarily located in intensively irrigated regions, such as North Africa, Arabian Peninsula, watersheds around the Caspian Sea, South America, and north-west India.

### 3.5. Uncertainty and Error in Water and Energy Balance Closure

We evaluate the ability of the ensemble of satellite-based  $P$ ,  $ET$ , and  $R_n$  data sets to close the water and energy balance of watersheds using two Budyko hypothesis-based metrics: (1) uncertainty in closure and (2) error in closure. Each point (AI, EI) in the Budyko space (Figure 2) represents the long-term water and energy balance of a watershed, as described by a specific combination of  $P$ ,  $ET$ , and  $R_n$  (or  $PET$ ). Therefore, the use of an ensemble of 56 possible permutations of  $P$ ,  $ET$ , and  $R_n$  data sets results in a point cloud for each of the 4,734 watersheds (Figure 1a), which represents the uncertainty in the closure of the water and energy balance due to the input data sets. First, we define cluster radius (CR), a simple metric to quantify that closure uncertainty in the Budyko space. For each watershed, CR is defined as the mean of all the Euclidean distances between the centroid of the point cloud and the individual points. In Figure 2, CR is represented by the solid blue line, while the dashed red line represents the circle corresponding to that radius.



**Figure 3.** (a) Global trends in mean annual changes in total water storage (TWS) derived from GRACE TWS anomaly data. (b) Watersheds in which TWS change is greater than 5% of the mean annual precipitation (266 watersheds), which are excluded in this study.

Likewise, the fit to the Budyko function (Equation 3) over long time scales enables us to quantify the error in the closure of water and energy balance of watersheds. Analogous to the CR as a metric of uncertainty, we also define closure error (CE) to quantify the closure error in the Budyko space. CE is defined as the Cartesian distance between a point (AI, EI) described by the satellite-based data sets and the reference (AI, EI) of the watershed described by the Budyko function (red line in Figure 2). In a previous study, we showed that CE accurately represents the error in the P and ET data sets (Koppa & Gebremichael, 2017). Specifically, CE in the Budyko space (Figure 2) is calculated as

$$CE = \sqrt{(AI_{est} - AI_{ref})^2 + (EI_{est} - EI_{mod})^2} \quad (6)$$

where,

- (1)  $EI_{est}$  and  $AI_{est}$  are the long-term average evaporative and aridity indices estimated from satellite-based estimates of P, ET and  $R_n$  considered in this study (see Section 3)
- (2)  $AI_{ref}$  is the reference long-term estimate of AI determined using observed  $R_n$  from the NASA/GEWEX Surface Radiation Budget (SRB) (Cox et al., 2006) v3.0 data and P from WorldClim v2.0 data set (Fick & Hijmans, 2017), a widely used high resolution (1 km) data set which uses between 9,000 and 6,000

weather stations for estimating long-term P climatology at a global scale. As in Koppa and Gebremichael (2017) this approach is based on the assumption that the aridity of the watershed is known and does not change significantly. The SRB data spans the years 1983–2007 with a spatial resolution of  $1^\circ \times 1^\circ$ . The temporal range of the WorldClim data set is 1970–2000 and the spatial resolution is  $1 \times 1$  km. The reasons for selecting the SRB data are threefold. First, it is the only satellite-based radiation data set which spans most of the time period WorldClim is based on. More importantly, SRB has been extensively validated (Pinker et al., 2005) and is widely used in the development of benchmark global meteorological and land surface products, such as the Princeton Global Forcing (Sheffield, Goteti, & Wood, 2006) and the North American Land Data Assimilation System (NLDAS) (Mitchell et al., 2004) data sets. Finally, none of the ET data sets use SRB  $R_n$  as an input, because SRB mission ended in 2007, and the temporal coverage of the data sets used in the study spans till the year 2019. As such, the  $AI_{ref}$  values remain independent of the data sets used for generating the evaluated ET data sets

- (3)  $EI_{mod}$  is determined from Fuh's equation (Equation 3) as  $EI_{mod} = 1 + AI_{ref} - \left(1 + (AI_{ref})^\omega\right)^{1/\omega}$  where  $\omega$  is the Budyko parameter.  $EI_{mod}$  is calculated for every year, and the long-term  $EI_{mod}$  is determined as the average of all annual values

As  $\omega$  has no analytical solution, we use the parameterization developed by Xu et al. (2013). Specifically, we make use of the following multiple linear regression (MLR) model

$$\omega = 5.05722 - 0.09322lat + 0.13085CTI + 1.31697NDVI + 0.000003A - 0.00018elev \quad (7)$$

where,

- (1) lat is the latitude of the centroid of each watershed and is derived from the HydroBASINS data set
- (2) CTI is the compound topographic index (K. J. Beven & Kirkby, 1979) defined as  $\ln(\alpha / \tan \beta)$ , where  $\alpha$  is the upstream watershed area derived from the HydroBASINS data set,  $\beta$  is the slope in degrees. We derive the average slope of each watershed from the Multi-Error-Removed Improved-Terrain (MERIT) DEM which covers the entire Earth (Yamazaki et al., 2017). Low CTI values are associated to smaller watersheds and steep slopes. On the other hand, high CTI values are associated with large watersheds and gentle slopes, and typically related to higher water accumulation and soil moisture
- (3) A is the area of watershed derived from HydroBASINS data set and elev is elevation derived from MERIT DEM
- (4) NDVI is the normalized difference vegetation index derived from MODIS, specifically MOD13A2 v6.0, available at 1 km spatial resolution (Didan et al., 2015). We determine one NDVI value per watershed by averaging across 2000–2018

The Budyko curve is highly nonlinear. As a result, in arid and hyper-arid regions, where precipitation values are very low, small changes in precipitation may lead to large changes in the distance-based metrics (CR and CE) used in this study. These issues were systematically addressed in Koppa and Gebremichael (2017). To prevent the CE metric from being biased due to the nonlinearity of the Budyko curve, Koppa and Gebremichael (2017) found that it is important to assume that the aridity index of the watershed is known, so that the Euclidean distance is calculated to the correct region in the Budyko curve. This is an extension of the approach by Greve, Orlowsky, et al. (2014), wherein the shortest distance to the Budyko curve was calculated. The approach by Greve, Orlowsky, et al. (2014) is, however, more prone to erroneous CE values. For example, if the remote sensing-based data sets are present in the humid region of the Budyko space, but the actual aridity of the watershed is arid, then the CE value must be high. However, because the changes in the humid region of the Budyko curve are small, the CE values are erroneously small. In Koppa and Gebremichael (2017), this issue is discussed in detail and corrected for with the formulation in Equation 6. To calculate the benchmark AI, we use the best available long-term climatological P data set (WorldClim) and an independent net radiation data set which matches the time period (SRB). This also enables a single benchmark EI value for each watershed against which the closure is analyzed. More importantly, to enable comparison across watersheds with different aridity indices, a normalization of the CR and CE metrics of each watershed by the corresponding AI values is required. In Koppa and Gebremichael (2017), the

methodology presented here has been comprehensively and successfully tested against in situ station-based validation methodologies.

A potential source of uncertainty in the CE metric (Equation 6) is the value of the Fuh parameter,  $\omega$ , which has no analytical solution. In this study, we use an empirically derived equation, which relates  $\omega$  to different characteristics of the watershed (Equation 7). Therefore, we quantify the uncertainty in the CE metric arising from uncertainty in the accuracy of Equation 7  $\omega$  as follows. For every watershed, we first fix the value of  $\omega$  derived from Equation 7 as the initial value ( $\omega_{ini}$ ). We then define an uncertainty range of  $\pm 2.0$  ( $[(\omega_{ini} - 2) \leq \omega_{ini} \leq (\omega_{ini} + 2)]$ ) around the initial value within which the sensitivity analysis is carried out. Specifically, we adopt the stratified single parameter sampling strategy suggested by (Saltelli et al., 2008). Accordingly, we sample the value of  $\omega$  at equal intervals of 0.1 in the range  $[\omega_{ini} - 2, \omega_{ini} + 2]$ . For each watershed, a value of  $\omega$  is sampled and used to calculate the resulting CE metric. Finally, the standard deviation of the CE metric per watershed is calculated from the set of CE values calculated using the explored range of  $\omega$ .

### 3.6. Evaluating P and ET Data Sets Based on Water and Energy Balance Closure

In this study, we validate the selected P and ET data sets based on their ability to close the water and energy balance of watersheds. The most convenient approach is to select as the best performing the P and ET data sets which, when combined, result in the lowest CE for a watershed. However, this does not completely solve the issue of compensatory biases, that is, P and ET data sets which are independently inaccurate but when combined lead to low CE values. Therefore, we adopt a more robust procedure to determine the best performing P and ET data sets. We calculate a set of CE values (Equation 6) for every P (or ET) data set by combining the P (or ET) data set with all the ET (or P) data sets. If we consider CHIRPSv2.0 as an example, we calculate a set of CE values by combining the CHIRPSv2.0 with the eight ET data sets used in this study. Then, we calculate the mean of these set of CE values for each P (or ET) data set. Finally, the P (or ET) data set with the lowest mean CE is considered as best performing for each specific watershed.

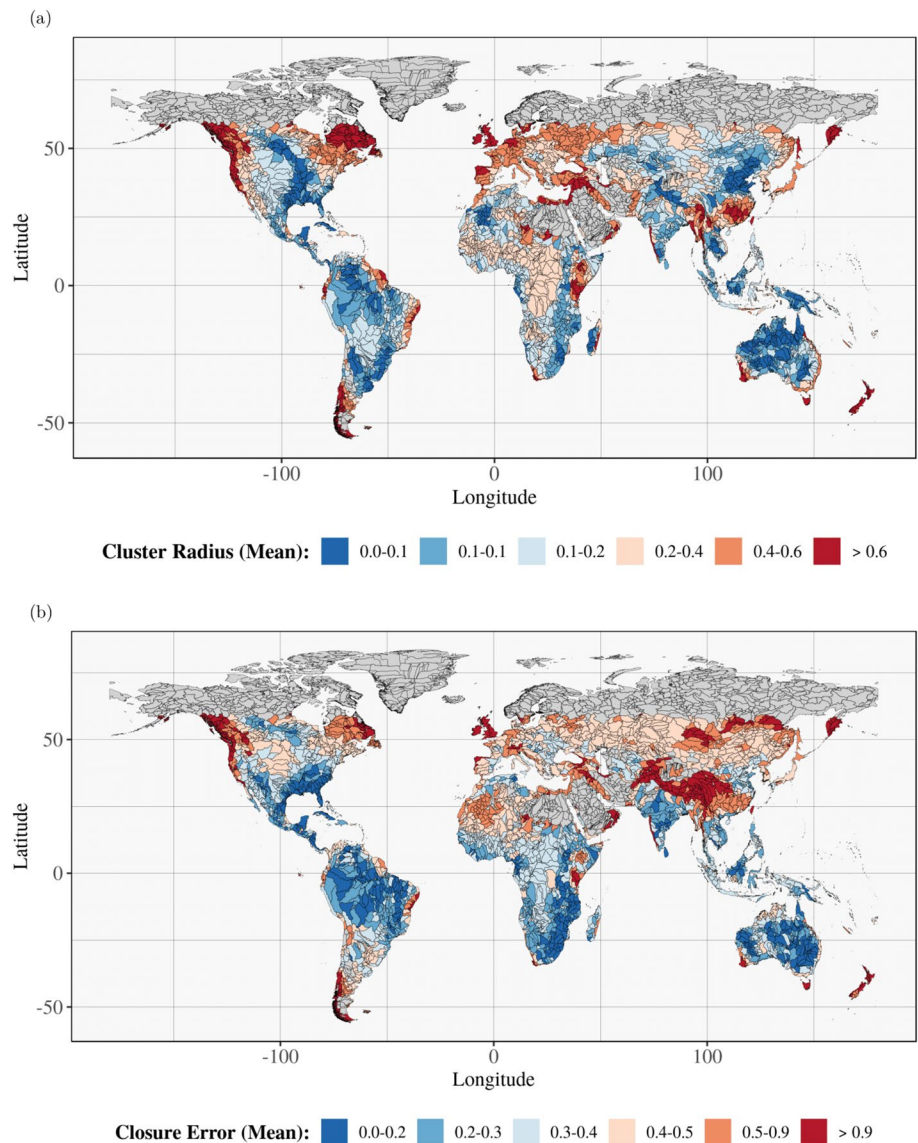
## 4. Results

First, we analyze the global distribution of uncertainty (CR) and error (CE) in satellite-based data sets with respect to the closing the water and energy balance of watersheds from (Figure 4). Next, we use spider plots to interpret the results by decomposing them along different gradients, which represent different characteristics of the watersheds such as climate (AI), topography (elevation and CTI), and greenness (NDVI) (Figure 6). In addition, we quantify the sensitivities of CR and CE to changes in P and ET (Figure 7). For every watershed, we rank different P and ET data sets based on their skill at closing the (Budyko-based) long-term water and energy balance (Figure 8). In addition, we analyze the relative performance of the P and ET data sets according to different characteristics of the watershed (Figure 9).

To aid the interpretation of results, the values presented in the CR and CE global maps are categorized into six percentile-based bins: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%. The bin limits (0%, 10%, 25%, 50%, 75%, 90%, and 100% percentiles) are calculated from the 4,753 CR and CE values (one for each watershed). Then, each CR and CE value is assigned to one of the six bins.

### 4.1. Global Patterns in the Budyko-Based Water and Energy Balance Closure

Figure 4a shows that CR values are low in humid and dry sub-humid regions of eastern United States, the rainforests of Amazonia, parts of India, and southeast Asia (see also aridity map in Figure 1a). From Figure 4b, it is evident that CE is also low in these areas. However, the CR and CE for water and energy balance closure are very high in humid regions in northern latitudes, especially in the temperate forests of North America and Europe, eastern China, and Myanmar. In semi-arid regions, we see that the watersheds in the western part of United States exhibit high CR and CE values (Figure 4). In addition, we see high CR and CE values in Patagonia and the Middle East. At the same time, several semi-arid watersheds have low CR and CE values, including most of Australia, north-eastern China, central United States, and southern

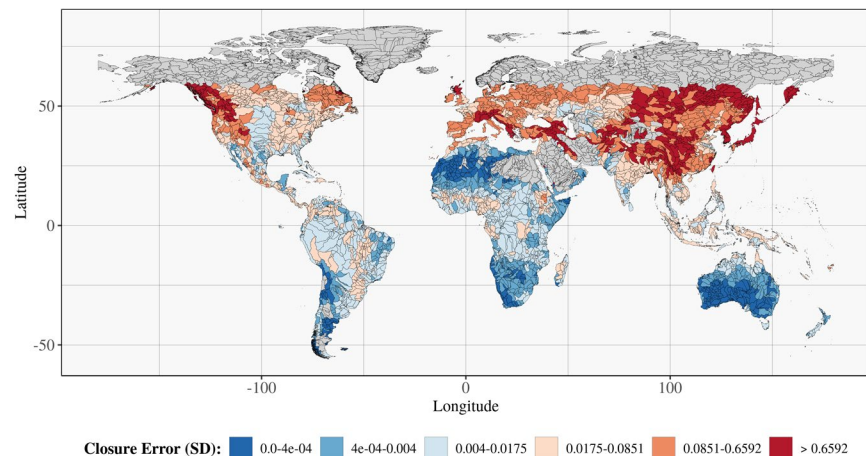


**Figure 4.** Global distribution of the mean of (a) CR and (b) CE, normalized by the aridity index of watersheds, determined using 56 combinations of P, ET, and  $R_n$  data sets. We select the following classification bins for the CR and CE metrics based on different percentile ranges: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%.

Africa. On the other hand, in arid and hyper-arid regions CR and CE are typically high, including in the East Sudanian savanna, the Greater Horn of Africa, Middle East and central Asia. While the spatial patterns of CR and CE are similar, there are certain regions where they show significant deviations. In the west Sahara and the Himalayan region, CR values are very low but the corresponding CE values are high. Likewise, the watersheds in sub-Saharan Africa and the rainforests in central Africa have high CR compared to their CE values. We interpret these similarities and differences between CR and CE in the “Discussion” section (Section 5). For reference, the standard deviation of CR and CE is found to follow the spatial patterns in mean values (Figure S2).

As detailed in the Methodology section, it is important to discuss the uncertainty in CE associated with the uncertainty in the Fuh parameter (Figure 5). In 75% of the watersheds, this uncertainty is very low (lower than 0.0813, or approximately 12% of the corresponding CE values). However, some regions—especially in eastern part of China, Russia, and the west of North America—show a high uncertainty in the CE metric propagating from the Fuh parameter. Most of these regions also show high CE values (higher latitudes in



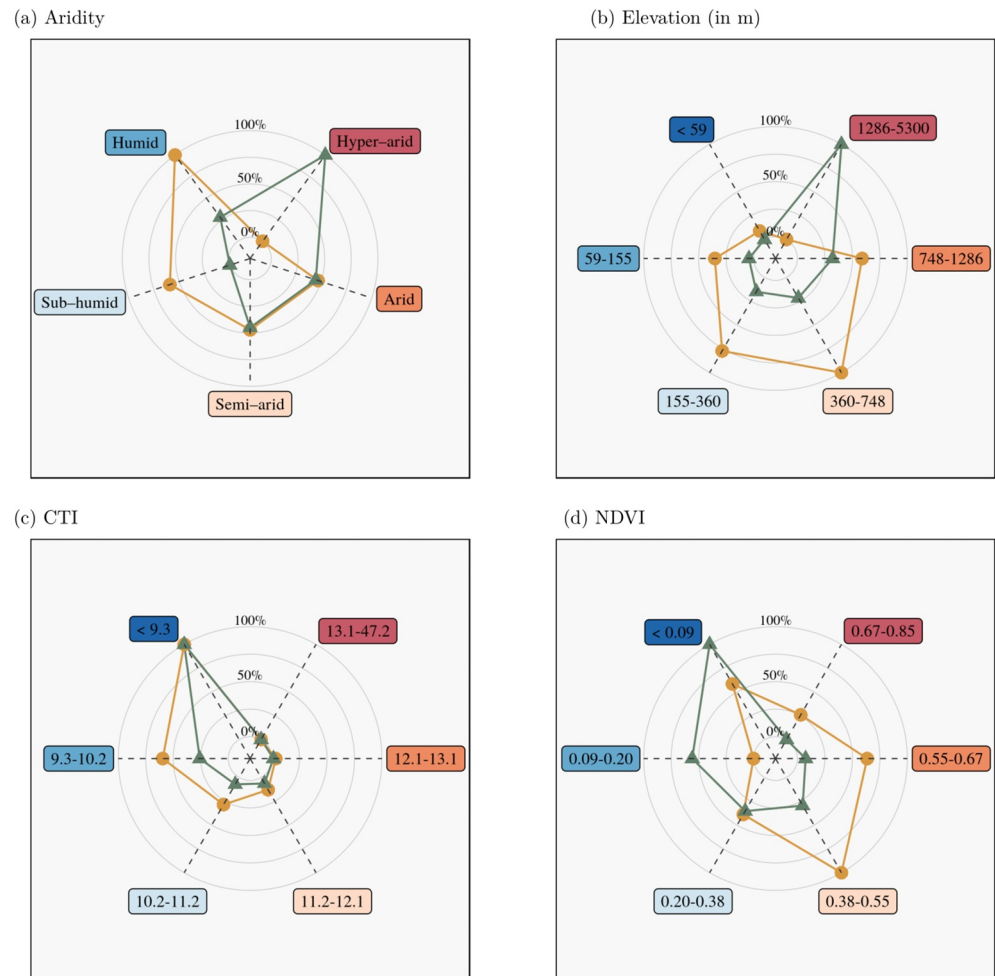


**Figure 5.** Global distribution of the uncertainty (standard deviation) in CE values due to uncertainty in the Fuh parameter (Equation 3). We select the following classification bins for the standard deviation in CE metrics based on different percentile ranges: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%.

North America, and eastern China). We note here that while performing the sensitivity analysis of the Fuh parameter, we considered a very wide range of values  $\left[\left(\omega_{ini} - 2\right) \leq \omega_{ini} \leq \left(\omega_{ini} + 2\right)\right]$ . Therefore, the uncertainty in Figure 5 represents a high ceiling of what could be expected due to the influence of uncertainties in the Fuh parameter. For the sake of clarity and considering that only in 5% of the watersheds this uncertainty is  $> 0.6$ , we do not present uncertainty ranges in the figures discussed in subsequent sections.

Making use of spider plots, Figure 6 groups the CR and CE values according to various watershed characteristics that govern the partitioning of P into ET and Q (Padrón et al., 2017). It is evident from Figure 6a that when the watersheds are aggregated according to aridity (Figure 6a), the highest CR is seen in humid regions, followed by hyper-arid regions. The highest CE values are in the arid watersheds, followed by humid regions. Figure 6a shows that the high closure uncertainty and error in humid regions is primarily due to the influence of northern latitudes. When we group the CR and CE values along an elevation gradient (Figure 6b), the highest CR is seen in relatively high percentile ranges (360–748 m and 748–1,286 m). Interestingly, we see that the in watersheds in the highest elevation percentile range (1,286–5,300 m) there is a stark difference between the closure uncertainty and error. In Figure 6c, we group the CR and CE values according to the compound topographic index (CTI). With increase in CTI, we see a decrease in both CR and CE values. Our final choice of a watershed characteristic to group the CR and CE values with is NDVI, which represents the greenness of vegetation in the watersheds (Figure 6d). The highest uncertainty and error occur in regions with very low NDVI. With increase in the greenness, the CE decreases. However, CR is relatively higher for watersheds with NDVI in the range 0.38–0.55, which corresponds to temperate forests (Figure 6d).

With a single  $R_n$  data set considered in our study, the uncertainty in the Budyko-based water and energy balance closure (CR) is computed based on the variability in the chosen P and ET data sets only. To determine the relative contribution of P and ET data to CR values for each watershed we conduct a simple experiment. First, we fix the value of ET (or P) to the mean of all the 8 ET (or 7 P) data sets for each watershed, and then determine the CR values for each of the 7 P (or 8 ET) data sets. The resulting CR values relate only to the permutation of P (or ET)—see Figures 7a and 7b. By comparing both figures, we see that ET uncertainties (Figure 7b) contribute twice as much as P uncertainties (Figure 7a) to CR (see also Figure S5). In general, the spatial patterns in both figures are similar: regions with high P and E uncertainties tend to coincide. However, several regions show deviations from this norm. In Amazonia, the CR values in the two cases are comparable, thus the uncertainty in water-energy balance closure in the Budyko space is influenced equally by both P and ET uncertainties in this region. This is also the case for southeast Asia and central India, where CR values are impacted equally by the uncertainty in P and ET.

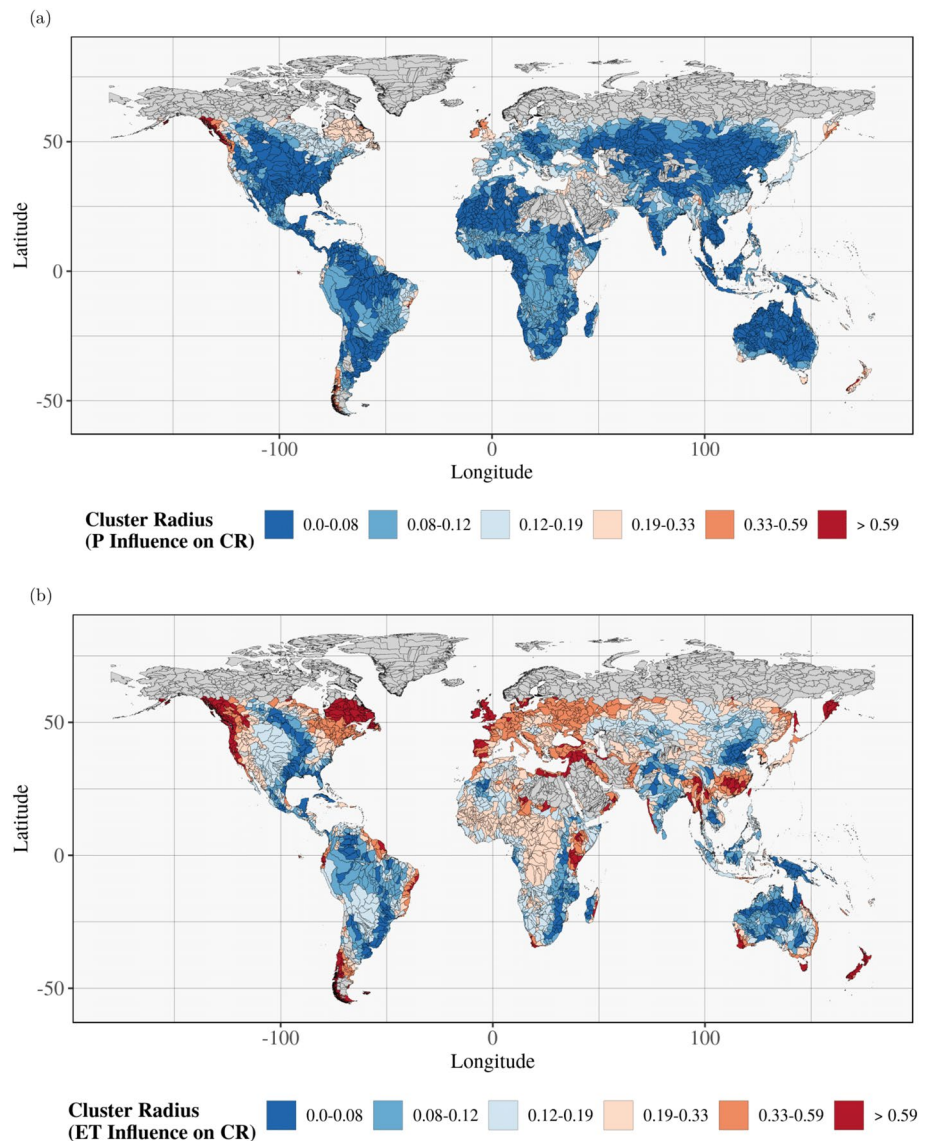


**Figure 6.** Spider plots of CR and CE, normalized by AI, for (a) aridity, (b) elevation (in m), (c) compound topographic index (CTI), and (d) NDVI. In the plot, the CR and CE values for different watershed characteristics (for example, aridity) are scaled between 0% and 100% (represented by the concentric circles), with 0% and 100% representing the lowest and highest CR (or CE) values, respectively. For example, when the watersheds are grouped according to aridity, the highest CE value is seen in hyper-arid watersheds and the lowest CE value is seen in sub-humid watersheds. The classification bins for elevation, CTI, and NDVI correspond to the following percentile ranges: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%.

#### 4.2. Best P and ET Data Sets for Water and Energy Balance Closure

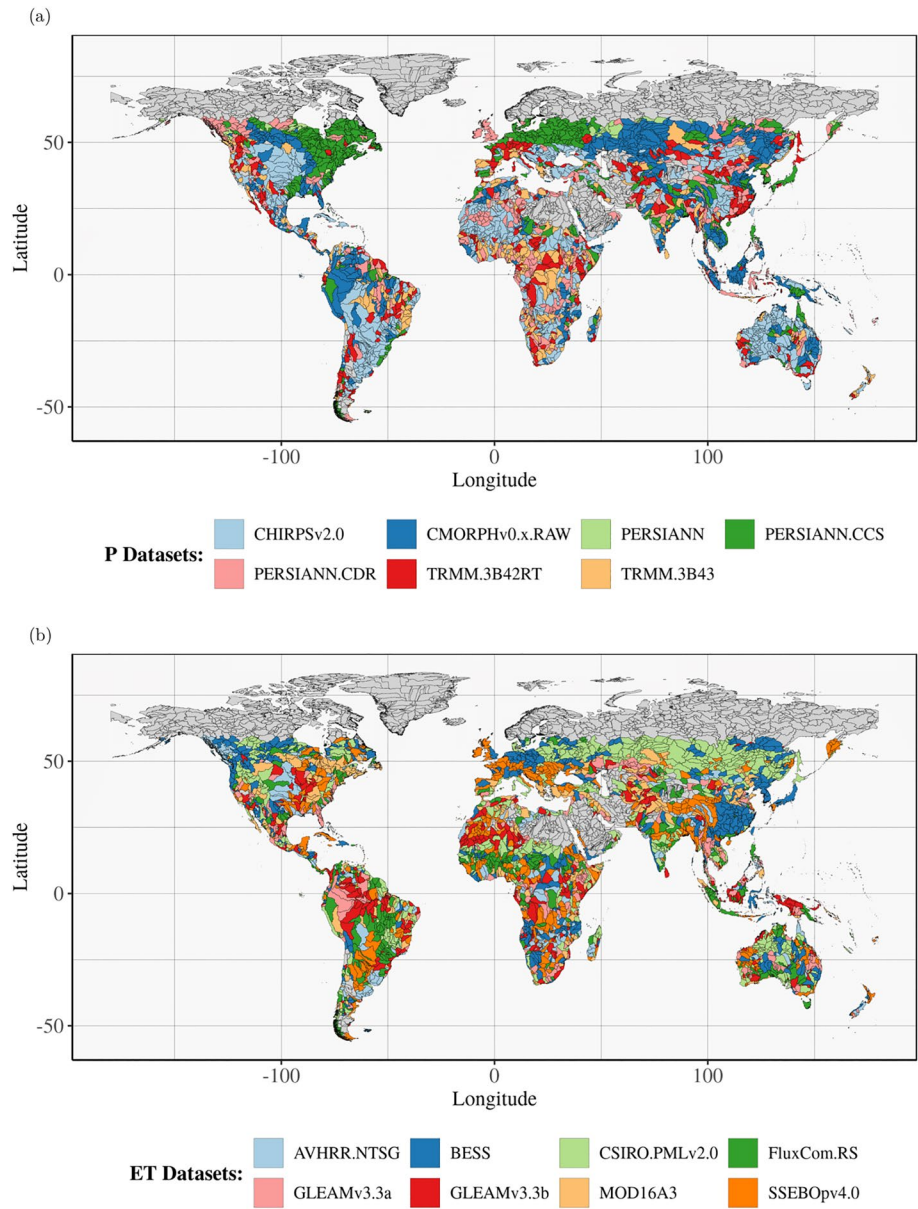
By comparing Figures 4b and S3, it is evident that the lowest errors in the Budyko-based water and energy balance closure (i.e., CE) are an order of magnitude lower than the average CE values from all the 56 combinations of P and ET data sets. This shows that a few of the existing retrieval algorithms are capable of accurately closing the water and energy balance of watersheds, or that at least they do not all suffer from common deficiencies that make them unsuitable. Therefore, analyzing the distribution of the best P and ET data sets can help us determine the most adequate P and ET retrieval algorithms for long-term closure studies in different regions.

In Figure 8, we see several consistent patterns. The PERSIANN family of P data sets along with CHIRPSv2.0 have the lowest CE in temperate forests and the rainforests of Congo and Amazonia. On the other hand, the picture is more complex for ET. In temperate forests, BESS and CSIRO.PMLv2.0 are the most performant (Figure 8b). In Amazonia, the GLEAM data sets have the least CE. In the rainforests of Congo, FluxCom, RS and SSEBopv4.0 seem to perform better than the other data sets. In the mountainous watersheds of the Himalayas and the Andes, the two TRMM-based P data sets have the least CE values. However, for ET



**Figure 7.** Global distribution of the mean CR determined with (a) mean ET (average of 8 data sets) and 7 P data sets and (b) mean P (average of 7 data sets) and 8 ET data sets, normalized by the watershed aridity index. We select the classification bins based on following percentile ranges: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%.

no single data set performs consistently better than the others. In the semi-arid Southern Hemisphere watersheds of Patagonia and Australia, we see that TRMM data sets are the most performant. As far as ET is concerned, AVHRR and CSIRO.PMLv2.0 data sets show the lowest CE. In the arid and hyper arid regions of Sahara and the Middle East, we see again that PERSIANN.CDR, PERSIANN.CSS, and CHIRPSv2.0 outperforming the other data sets.; among ET data sets, GLEAM data sets have the lowest CE along with CSIRO.PMLv2.0. Finally, we rank the P and ET data sets according to different watershed characteristics such as aridity, elevation, CTI, and NDVI in Figure 9. For both P and ET, we see that the highest CE values are in mountainous watersheds (high elevation and low CTI). Overall, we see that the TRMM.3B43 and BESS data sets have high ranks across all the different classification criteria for P and ET data, respectively. While the ET data sets seem to perform similar to each other when grouped according to different watershed characteristics, P data sets show discernible differences. We see that PERSIANN.CDR and CMORPH data sets perform consistently worse than other data sets across different grouping criteria.



**Figure 8.** Global distribution of the best (a) P and (b) ET data set per watershed with respect to the closure of the long-term water and energy balance (i.e., CE). Per watershed, we calculate the mean CE for each specific P (or ET) data set paired with all ET (or P) data sets. Then, the P (or ET) data set with the minimum value is selected as best performing.

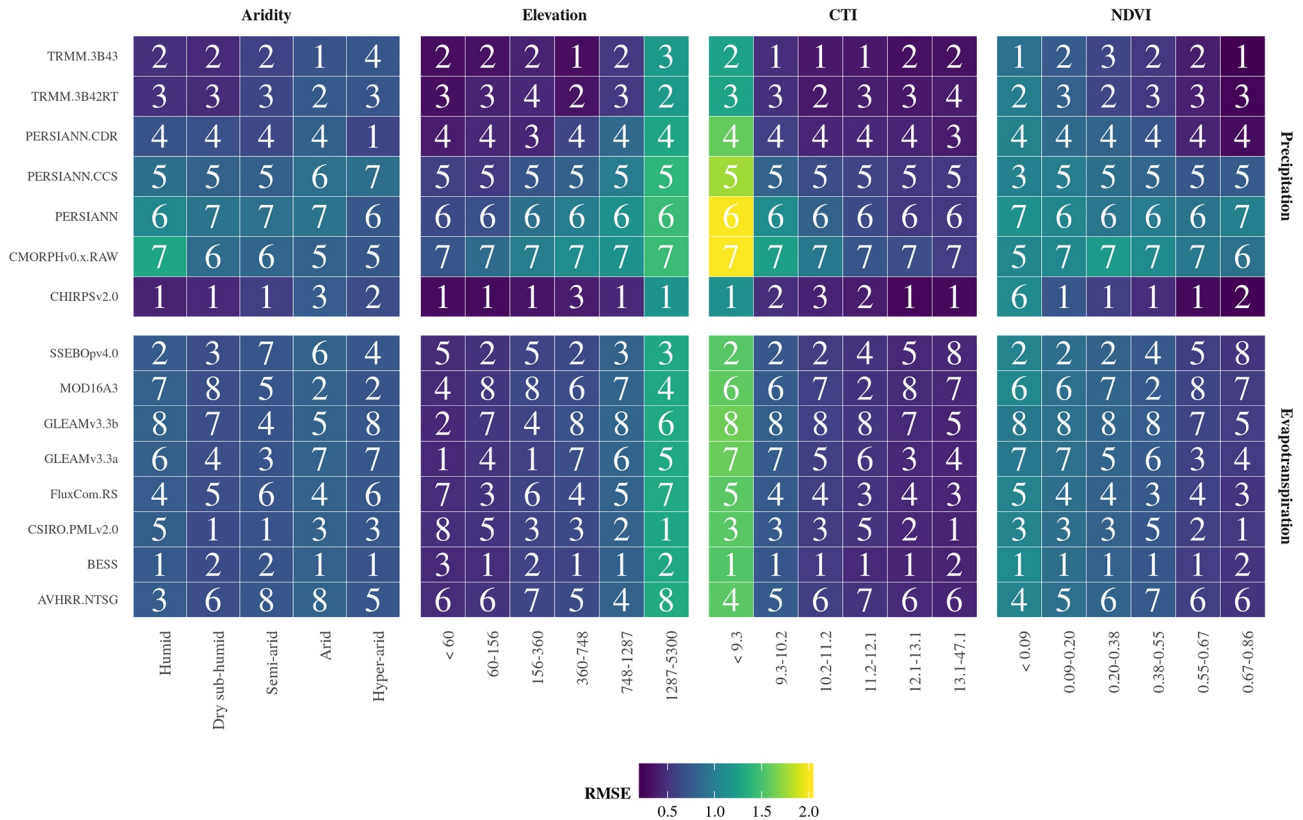
## 5. Discussion

### 5.1. Interpretation of CR and CE Patterns

CR and CE provide two distinct types of information, and differences between the two for the same watershed provide useful insights into the efficacy of existing algorithms. We analyze two contrasting scenarios: (1) low CR and high CE and (2) high CR and low CE values. For this, we use spider plots (Figure 4) and the spatial distribution of a simple ratio,  $R$ , defined as

$$R(\text{CR}, \text{CE}) = \frac{\text{CR}}{\text{CR} + \text{CE}} \quad (8)$$



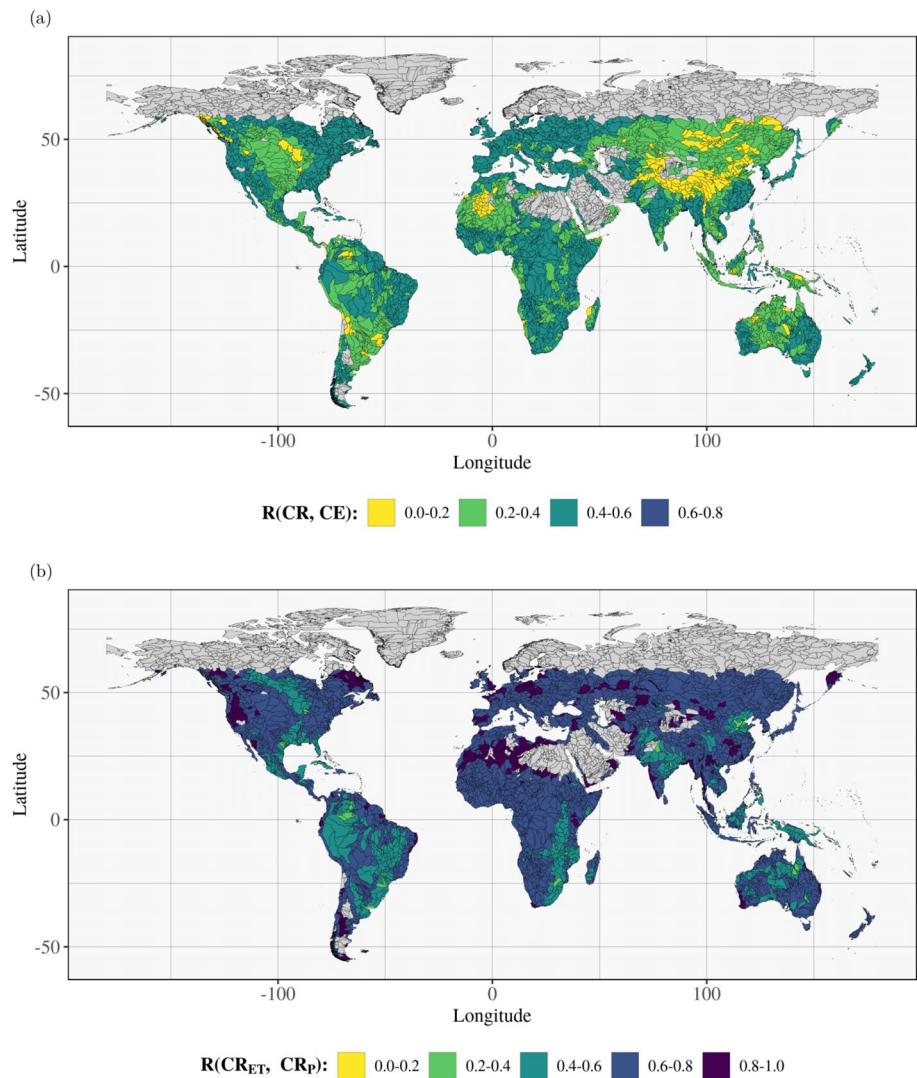


**Figure 9.** Heat maps of the root mean square errors (RMSE) of P and ET data sets for different aridity, elevation, CTI, and NDVI classes. The numbers represent the rank of the P or ET data set for each watershed characteristic. The RMSE values are calculated as follows: For each P or ET data set, the Budyko error for every watershed is determined. We then calculate the RMSE of the Budyko errors corresponding to watersheds which come under a specific class of a watershed characteristic. We select the following classification based on different percentile ranges: 0%–10%, 10%–25%, 25%–50%, 50%–75%, 75%–90%, and 90%–100%.

Figure 10a illustrates the results.  $R(\text{CR}, \text{CE})$  closer to 0 represents the first scenario (low CR and high CE), and  $R(\text{CR}, \text{CE})$  closer to 1 represents the second scenario (high CR and low CE). For the first scenario, a prime example are the arid and semi-arid regions, where the aggregated CE value is much higher than the CR value (Figure 6a). This shows that although the data sets agree very well with each other regarding the degree of closure (low uncertainty, clustered close to each other in the Budyko space), they do not agree with the Budyko function (distance to the Budyko curve is high for all combinations). Therefore, in semi-arid regions, most (if not all) existing P and ET retrieval algorithms need to be improved from the perspective of water and energy balance closure. We also see this in high elevation watersheds (1,286–5,300 m), highlighting the need for better satellite-based remote sensing retrievals in mountainous regions. It is also interesting to note that the only forested region to which this scenario is applicable is Amazonia. Although the absolute CR and CE values are low in Amazonia, the relative difference between the two suggests that P and ET estimation algorithms systematically fail to achieve a closure. While a possible explanation for such systematic deviations could be the unsuitability of the negligible watershed storage changes, we note that watersheds with significant storage losses (Figure 3b) were excluded from the analyses.

The opposite scenario (i.e.,  $R$  values close to 1) indicates regions where P and ET data sets have large uncertainty, leading to high CR values, but the corresponding CE values are relatively low. As the spider plots in Figure 6a suggest, the humid regions are an example for this scenario; more specifically the Congo rainforest, western North America, and mainland Europe (Figure 10a). In these regions, despite the uncertainties in the ensemble, a few data sets are close to the reference Budyko curve and hence have a low CE. In such cases, the combinations of P and ET data sets which are close to the Budyko curve can serve as reference in improving other algorithms. This gives rise to the possibility of creating a merged P and ET data sets using CE for determining the weights (Beck, van Dijk, et al., 2017). The data sets which are not close to the

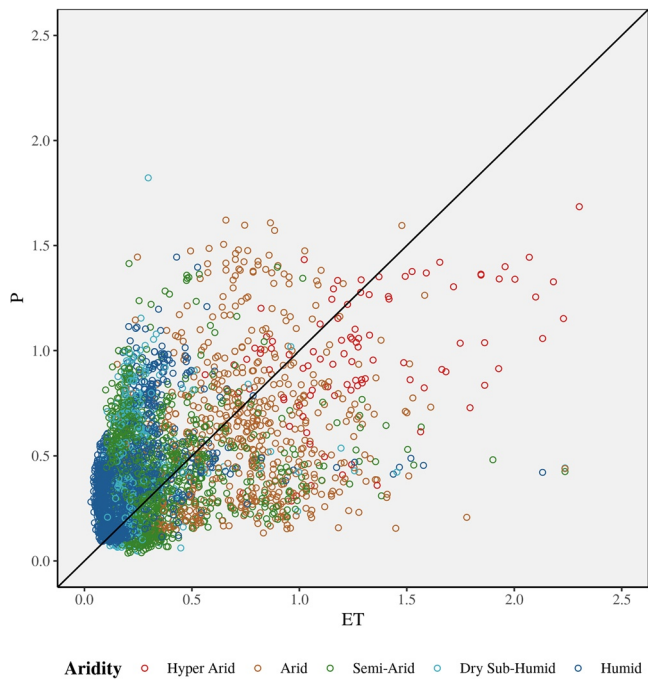




**Figure 10.** Global distribution of ratios (a)  $R(CR, CE)$  and (b)  $R(CR_{ET}, CR_P)$  using 56 combinations of P, ET, and  $R_n$  data sets.  $R(CR, CE)$  is defined as  $CR/(CR + CE)$ , where CR and CE values are taken from Figures 4a and 4b respectively.  $R(CR_{ET}, CR_P)$  is defined as  $CR_{ET}/(CR_{ET} + CR_P)$ , where  $CR_P$  and  $CR_{ET}$  values are taken from Figures 7a and 7b respectively.

Budyko curve (i.e., that show a high CE) may have common, systematic deficiencies in their input data or algorithms; this may orient product developers that seek to improve the accuracy of their data sets.

Finally, the spider plots reflect that CR and CE are highly correlated with watershed characteristics (Figure 6). Specifically, we see that CTI (Figure 6c), a topographic index, explains well the spatial patterns in CR and CE values, even better than elevation. As described in the Methodology section, lower values of CTI correspond to smaller and steeper watersheds, and higher CTI corresponds to flatter and larger watersheds. We see that the CR and CE are high for smaller and steeper watersheds, and lower for larger watersheds. This highlights the difficulties of remote sensing techniques in mountainous regions, which has been noted in previous studies (Dinku et al., 2011; Gebremichael et al., 2014). From our analysis, we see that PERSIANN.CDR and SSEBopv4.0 tend to perform the best for these regions; the higher resolution of SSEBopv4.0 (1 km) could play role in that sense. We also see that the greenness of the watershed (NDVI) explains the variability of CE, and CR, to a lesser extent. Very low NDVI values correspond to watersheds with lack of vegetation, which lie primarily in arid and hyper-arid regions, where the uncertainty is higher.



**Figure 11.** Scatter plot of the coefficient of variation of the absolute values of P (determined using 7 P data sets for each watershed) versus the coefficient of variation of the absolute values of ET (determined using 8 ET data sets for each watershed). This plot shows the absolute uncertainties (in the form of coefficient of variation) in P and ET data sets, as opposed to cluster radius (which represents the uncertainty in the closure of water and energy balance).

## 5.2. Understanding the Sources of the Budyko-Based Closure Uncertainty

The results also reveal that the CR of watersheds are primarily governed by uncertainty in ET data rather than P (Figure 7). However, there may be regions in which the uncertainty is governed by P. To better understand whether CR is affected more by P or ET uncertainties, we define, similar to  $R(\text{CR}, \text{CE})$ , another ratio:

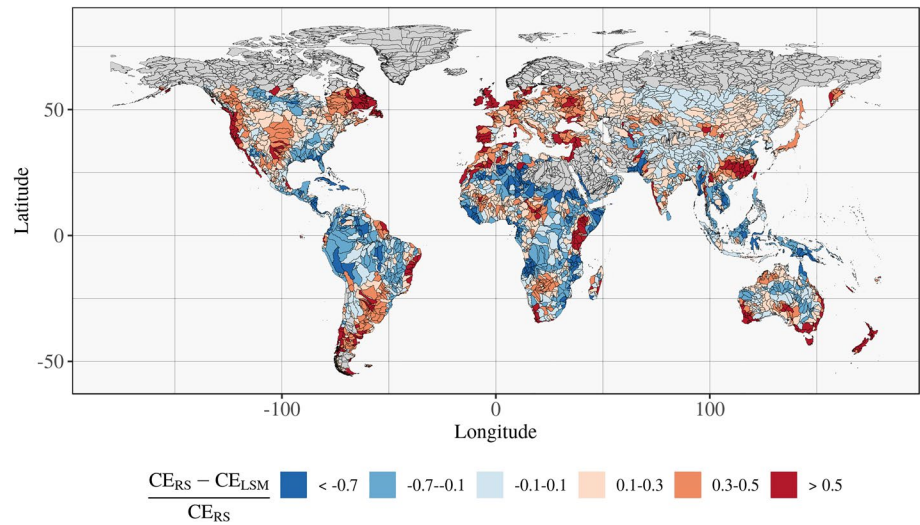
$$R(\text{CR}_{\text{ET}}, \text{CR}_{\text{P}}) = \frac{\text{CR}_{\text{ET}}}{\text{CR}_{\text{ET}} + \text{CR}_{\text{P}}} \quad (9)$$

where  $\text{CR}_{\text{P}}$  is the CR determined using mean of ET data sets (Figure 7a) and  $\text{CR}_{\text{ET}}$  is CR determined using mean of P data sets (Figure 7b). Therefore,  $\text{CR}_{\text{ET}}$  represents uncertainty in (Budyko-based) closure arising only from the ET uncertainty, and  $\text{CR}_{\text{P}}$  represents uncertainty in closure arising from P uncertainties. If  $R(\text{CR}_{\text{ET}}, \text{CR}_{\text{P}})$  tends to 0, it implies that uncertainty in P influences CR more than ET. If  $R(\text{CR}_{\text{ET}}, \text{CR}_{\text{P}})$  tends to 1, uncertainty in ET influences CR more. Figure 10b, although confirming the results from Figure 7, shows that there are several regions in which CR is influenced by both P and ET uncertainties. Examples for such a scenario are central India, parts of southern Africa and Amazonia. In Amazonia, the higher uncertainties in P are also reflected in the fact that no unique P data set outperforms all others, and that the best data sets are not gauge calibrated. This reflects that there are not enough gauges in the region, but also the general inadequacy of satellite remote sensing over tropical forests. In terms of ET, GLEAM seems to outperform the other data sets in Amazonia, with the GLEAMv3.3b outperforming GLEAMv3.3a despite using only satellite data.

It is important to note that the impact of individual uncertainties in P and ET data sets on CR and CE (Figures 10 and S5) are not linearly related to the absolute uncertainties in P and ET data sets. This is apparent in the scatter plot between the coefficient of variation (CV) in P and ET (Figure 11). From Figure 11, it is evident that the absolute uncertainty in ET is in fact marginally lower than the uncertainty in P data sets, especially in humid regions. However, from Figures 10 and S5, the uncertainties in ET propagate further into CR and CE. A mathematical reason for this is apparent from the Budyko space: the fact that P appears in the denominator in both  $x$  (AI) and  $y$  (EI) axes, but ET appears only in the numerator in the  $y$  axis (EI). A physically consistent explanation is the fact that higher precipitation implies both lower aridity and a lower evaporative index at the same time. On the other hand, higher ET affects only EI, and this effect is not compensated by changes to AI (in water-limited regions). This nonlinearity in sensitivities would not be revealed without combining the water and energy balances together.

## 5.3. Performance of P and ET Data Sets in High Latitude Watersheds

A recurring theme in the results is that the ensemble of P and ET data sets selected in this study fails to close the Budyko-based water and energy balance of watersheds in the northern latitudes, both in the temperate forests and the grasslands in mainland Europe. The decomposition of the results according to watershed characteristics (Figure 1) did not reveal any consistent reason for this high uncertainty. However, we see that the best combination of P and ET data sets for these regions results in very low CE values (0%–10% range in Figure S3). This implies that some algorithms (PERSIANN.CDR and PERSIANN.CCS for precipitation, and BESS and CSIRO.PMLv2.0 for ET) are still well suited to represent the long-term closure in these regions. The fact that a gauge-calibrated P data set (PERSIANN.CDR) performs well reveals the benefit of utilizing the dense network of rain gauges in Europe and North America. Also, the extension of the PERSIANN algorithm with a more sophisticated cloud segmentation (PERSIANN.CCS) seems to have



**Figure 12.** Global distribution of fractional difference between mean closure error (CE) of all combinations of remote sensing-based ( $CE_{RS}$ ) and in situ and LSM-based ( $CE_{LSM}$ ) P and ET data sets. Positive values imply LSM-based data sets are better at closing the Budyko-based water and energy balance. Negative values imply remote sensing-based data sets are better.

contributed to better rainfall estimation over these regions. For ET, the coupled biophysical model used in BESS may be more suitable for higher latitudes. We note that both BESS and CSIRO.PMLv2.0 are based on the PM equation, which tends to perform better in low radiation conditions than PT formulations (Miralles, Jiménez, et al., 2016). Also, the CSIRO.PMLv2.0 data set is calibrated with the Budyko curve, which could explain its better performance in these analyses. In fact, CSIRO.PMLv2.0 outperforms all the data sets across the majority of the landcover classes (Figure S4). We note that our ranking of the data sets is in the context of long-term closure, and therefore only indicative of the performance of the mean bias in the data sets and not of the temporal dynamics.

#### 5.4. Comparison With P and ET From In Situ Observations and LSMs

Finally, we contrast our results based on remote sensing data sets with those arising from the use of in situ and state-of-art LSM's. As CR is biased by the number of P and ET data sets (56 combinations for remote sensing-based data sets compared to 16 for in situ and LSM-based data sets), we focus only on comparing CE results, which provide an independent benchmark of water and energy balance closure in the Budyko space and are not biased by the number of data sets. Figure 12 presents a global map of fractional difference (Equation 10) between mean of CE values calculated using only remote sensing-based data sets ( $CE_{RS}$ ) and only in situ and LSM-based data sets ( $CE_{LSM}$ ).

$$CE_{RS-LSM} = \frac{CE_{RS} - CE_{LSM}}{CE_{RS}} \quad (10)$$

Negative  $CE_{RS-LSM}$  values point to lower remote sensing-based CE compared to in situ and LSM-based data sets, and hence a better performance in closing the water and energy balance in the Budyko space. A consistent geographical pattern emerges in Figure 12, in which negative  $CE_{RS-LSM}$  values are seen in the tropics, while positive  $CE_{RS-LSM}$  values are seen in the mid and high latitudes. A plausible explanation for this clear distinction could be that the in situ P data sets are more accurate in regions with high density of rain gauges, such as the North America and Europe. This increased accuracy translates into improved accuracy in ET, as many of the LSM and reanalysis-based ET estimates use in situ-based gridded P estimates as forcing. The opposite applies in parts of the tropics, such as the Amazon and Congo rainforests, where in situ measurements are sparse and hence satellite-based remote sensing data sets show an advantage. We also see large regions in central Asia and north-east Europe where the remote sensing data sets perform similar to in situ and LSM's as far as the Budyko-based water and energy balance closure is concerned ( $-0.1$  to  $0.1$  bin in

Figure 12). In Figure S6, we present a comprehensive ranking of remote sensing and in situ or LSM-based P and ET data set using heatmaps, similar to Figure 9. We see that GPCCv7.0, which uses the greatest number of in situ rainfall stations, consistently performs better than other in situ and remote sensing-based data sets. As far as ET is concerned, we see that the Noah LSM-based FLDAS and GLDASv2.1 perform better than the other two in situ and LSM-based data sets under different aridity, topographic, and vegetation-based gradients. Our results not only highlight the importance of satellite-based remote sensing in regions where in situ monitoring is insufficient.

## 6. Conclusion and Future Work

In this study, we quantified the potential of global satellite data to close the water and energy balance of large watersheds worldwide. To do so, we applied a Budyko-based framework that explicitly considers water and energy balance constraints. Our results helped identify regions in which current understanding of watershed-scale hydrologic processes is hindered by systematic errors in observations. Moreover, the closure analysis enabled the quantification of long-term biases in individual satellite-based data sets, thus providing a global assessment of the current skill at retrieving different fluxes.

We show that uncertainties and errors in closure are highly variable in space. Of importance is the high uncertainty in temperate forests, where both P and ET data sets are highly uncertain and exhibit significant biases. In addition, we see that the water-energy balance closure is difficult to achieve in arid and semi-arid regions, where the availability of accurate satellite data is important, as these regions are also scarcely sampled in situ. Concerted efforts are needed to improve P and ET data sets in arid and mountainous watersheds where most of the data sets are shown to agree with each other but display systematic errors. Finally, we showed that satellite-based remote sensing data sets perform better than in situ observations and model-based P and ET data sets in data-scarce regions. Overall, our findings highlight the need for improving existing algorithms in specific regions, either by revisiting their parameterizations or incorporating physical processes that may be relevant but not considered at the moment. The evaluation presented here is novel, independent of in situ measurements, and integrative in terms of considering both the water and energy balance together using the Budyko framework. Altogether, it offers a new means for the creation of merged data sets P and ET that better represent the long-term closure of water and energy cycles.

## Data Availability Statement

All data used in the study are available from the respective literature cited in the manuscript. All data processing and analysis codes are openly available at <https://doi.org/10.5281/zenodo.4002794>.

## Acknowledgments

A. Koppa acknowledges support from the NASA Applied Sciences Water Resources Application Grant NNX15AC33G. S. Alam and M. Gebremichael acknowledge support from the University of California Research Initiatives Award, grant number LFR-18-548316. D. G. Miralles acknowledges support from the European Research Council (ERC) under grant agreement no. 715254 (DRY-2-DRY) and the EU H2020 program, project no. 869550 (DOWN2EARTH).

## References

- Aires, F. (2014). Combining datasets of satellite-retrieved products. Part I: Methodology and water budget closure. *Journal of Hydrometeorology*, 15(4), 1677–1691. <https://doi.org/10.1175/jhm-d-13-0148.1>
- Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903), 228–232. <https://doi.org/10.1038/nature01092>
- Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., et al. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1), 69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., et al. (2003). AIRS/AMSU/HSB on the aqua mission: Design, science objectives, data products, and processing systems. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 253–264. <https://doi.org/10.1109/TGRS.2002.808356>
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11), 2415–2434. [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2)
- Ball, J. T., Woodrow, I. E., & Berry, J. A. (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. *Progress in photosynthesis research*, (pp. 221–224). Springer. [https://doi.org/10.1007/978-94-017-0519-6\\_48](https://doi.org/10.1007/978-94-017-0519-6_48)
- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017a). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589–615. <https://doi.org/10.5194/hess-21-589-2017>
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017b). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12), 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>



- Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garretón, C., Baez-Villanueva, O. M., et al. (2020). Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *Journal of Climate*, 33(4), 1299–1315. <https://doi.org/10.1175/jcli-d-19-0332.1>
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I. J. M., et al. (2019). MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100(3), 473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., & Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) precipitation from 1901–present. *Earth System Science Data*, 5(1), 71–99. <https://doi.org/10.5194/essd-5-71-2013>
- Beven, K. (2006). Searching for the Holy Grail of scientific hydrology:  $Q_t = (S, R, \Delta t)A$  as closure. *Hydrology and Earth System Sciences*, 10(5), 609–618. <https://doi.org/10.5194/hess-10-609-2006>
- Beven, K. (2019). Towards a methodology for testing models as hypotheses in the inexact sciences. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2224), 20180862. <https://doi.org/10.1098/rspa.2018.0862>
- Beven, K., Asadullah, A., Bates, P., Blyth, E., Chappell, N., Child, S., et al. (2020). Developing observational methods to drive future hydrological science: Can we make a start as a community? *Hydrological Processes*, 34(3), 868–873. <https://doi.org/10.1002/hyp.13622>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, 37(2), 307–337. <https://doi.org/10.1007/s10712-015-9346-y>
- Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9(3–4), 251–290. <https://doi.org/10.1002/hyp.3360090305>
- Budyko, M. I., & Mikhail, I. (1974). *Climate and life*. Academic Press. Retrieved from [https://books.google.co.in/books/about/Climate\\_and\\_Life.html?id=Ln89Y-6KwZYC&redir\\_esc=y](https://books.google.co.in/books/about/Climate_and_Life.html?id=Ln89Y-6KwZYC&redir_esc=y)
- Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, 113(D4). <https://doi.org/10.1029/2007JD009132>
- Choudhury, B. (1999). Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *Journal of Hydrology*, 216(1–2), 99–110. [https://doi.org/10.1016/S0022-1694\(98\)00293-5](https://doi.org/10.1016/S0022-1694(98)00293-5)
- Condon, L. E., & Maxwell, R. M. (2017). Systematic shifts in Budyko relationships caused by groundwater storage changes. *Hydrology and Earth System Sciences*, 21(2), 1117–1135. <https://doi.org/10.5194/hess-21-1117-2017>
- Cox, S. J., Stackhouse, P. W., Jr., Gupta, S. K., Mikovitz, J. C., Zhang, T., Hinkelman, L. M., et al. (2006). The NASA/GEWEX surface radiation budget project: Overview and analysis. In *Extended Abstracts, 12th Conference on Atmospheric Radiation*.
- Destouni, G., Jaramillo, F., & Prieto, C. (2013). Hydroclimatic shifts driven by human water use for food and energy production. *Nature climate change*, 3(3), 213–217. <https://doi.org/10.1038/nclimate1719>
- Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). *MODIS vegetation index user's guide (MOD13 series)*. University of Arizona, Vegetation Index and Phenology Lab.
- Dinku, T., Ceccato, P., & Connor, S. J. (2011). Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *International Journal of Remote Sensing*, 32(21), 5965–5979. <https://doi.org/10.1080/01431161.2010.499381>
- Donohue, R. J., Roderick, M. L., & McVicar, T. R. (2012). Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model. *Journal of Hydrology*, 436–437(437), 35–50. <https://doi.org/10.1016/J.JHYDROL.2012.02.033>
- Eltahir, E. A. B. (1998). A soil moisture-rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, 34(4), 765–776. <https://doi.org/10.1029/97WR03499>
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), 10572–10577. <https://doi.org/10.1073/pnas.1712381114>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fuh, B. (1981). On the calculation of the evaporation from land surface. *Scientia Atmospherica Sinica*, 5(1), 23–31.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, 2(1), 150066. <https://doi.org/10.1038/sdata.2015.66>
- Gebremichael, M., Bitew, M. M., Hirpa, F. A., & Tesfay, G. N. (2014). Accuracy of satellite rainfall estimates in the Blue Nile Basin: Lowland plain versus highland mountain. *Water Resources Research*, 50(11), 8775–8790. <https://doi.org/10.1002/2013WR014500>
- Good, S. P., Moore, G. W., & Miralles, D. G. (2017). A mesic maximum in biological water use demarcates biome sensitivity to aridity shifts. *Nature Ecology & Evolution*, 1(12), 1883–1888. <https://doi.org/10.1038/s41559-017-0371-8>
- Greve, P., Burek, P., & Wada, Y. (2020). Using the Budyko framework for calibrating a global hydrological model. *Water Resources Research*, 56(6). <https://doi.org/10.1029/2019wr026280>
- Greve, P., Orlovsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, 7(10), 716–721. <https://doi.org/10.1038/NGEO2247>
- Grove, A. T. (1999). World atlas of desertification. In N. J. Middleton, D. S. G. Thomas, L. Arnold (Eds.), *Earth surface processes and landforms* (2nd ed., Vol. 24). John Wiley & Sons, Ltd. [https://doi.org/10.1002/\(SICI\)1096-9837\(199903\)24:3<280::AID-ESP955>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1096-9837(199903)24:3<280::AID-ESP955>3.0.CO;2-7)
- Guillod, B. P., Orlovsky, B., Miralles, D. G., Teuling, A. J., & Seneviratne, S. I. (2015). Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature Communications*, 6(1), 6443. <https://doi.org/10.1038/ncomms7443>
- Hegerl, G. C., Black, E., Allan, R. P., Ingram, W. J., Polson, D., Trenberth, K. E., et al. (2015). Challenges in quantifying changes in the global water cycle. *Bulletin of the American Meteorological Society*, 96(7), 1097–1115. <https://doi.org/10.1175/BAMS-D-13-00212.1>
- Heusinkveld, B. G., Jacobs, A. F. G., Holtslag, A. A. M., & Berkowicz, S. M. (2004). Surface energy balance closure in an arid region: Role of soil heat flux. *Agricultural and Forest Meteorology*, 122(1–2), 21–37. <https://doi.org/10.1016/J.AGRFORMET.2003.09.005>
- Hobeichi, S., Abramowitz, G., Contractor, S., & Evans, J. (2020). Evaluating precipitation datasets using surface water and energy budget closure. *Journal of Hydrometeorology*, 21(5), 989–1009. <https://doi.org/10.1175/JHM-D-19-0255.1>
- Holwerda, F., Bruijnzeel, L. A., Scatena, F. N., Vugts, H. F., & Meesters, A. G. C. A. (2012). Wet canopy evaporation from a Puerto Rican lower montane rain forest: The importance of realistically estimated aerodynamic conductance. *Journal of Hydrology*, 414–415, 1–15. <https://doi.org/10.1016/j.jhydrol.2011.07.033>
- Hong, Y., Gochis, D., Cheng, J.-T., Hsu, K.-I., Sorooshian, S., Hong, Y., et al. (2007). Evaluation of PERSIANN-CCS rainfall measurement using the NAME event rain gauge network. *Journal of Hydrometeorology*, 8(3), 469–482. <https://doi.org/10.1175/JHM574.1>



- Hong, Y., Tang, G., Ma, Y., Huang, Q., Han, Z., Zeng, Z., et al. (2018). *Remote sensing precipitation: Sensors, retrievals, validations, and applications* (pp. 1–23). Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-662-47871-4\\_4-2](https://doi.org/10.1007/978-3-662-47871-4_4-2)
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1), 38–55. <https://doi.org/10.1175/JHM560.1>
- Istanbulluoglu, E., Wang, T., Wright, O. M., & Lenters, J. D. (2012). Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. *Water Resources Research*, 48(3). <https://doi.org/10.1029/2010WR010100>
- Jaramillo, F., & Destouni, G. (2015). Local flow regulation and irrigation raise global human water consumption and footprint. *Science*, 350(6265), 1248–1251. <https://doi.org/10.1126/science.aad1010>
- Jiang, C., & Ryu, Y. (2016). Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). *Remote Sensing of Environment*, 186, 528–547. <https://doi.org/10.1016/j.rse.2016.08.030>
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., Xie, P., Joyce, R. J., Janowiak, J. E., et al. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, 5(3), 487–503. [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2)
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO)
- Koppa, A., & Gebremichael, M. (2017). A framework for validation of remotely sensed precipitation and evapotranspiration based on the Budyko hypothesis. *Water Resources Research*, 53(10). <https://doi.org/10.1002/2017WR020593>
- Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171–2186. <https://doi.org/10.1002/hyp.9740>
- Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10), 93. <https://doi.org/10.1029/2008EO100001>
- Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51(9), 7309–7342. <https://doi.org/10.1002/2015WR017616>
- Li, D., Pan, M., Cong, Z., Zhang, L., & Wood, E. (2013). Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, 49(2), 969–976. <https://doi.org/10.1002/wrcr.20107>
- Liang, S., Zhao, X., Liu, S., Yuan, W., Cheng, X., Xiao, Z., et al. (2013). A long-term Global Land Surface Satellite (GLASS) data-set for environmental studies. *International Journal of Digital Earth*, 6(1), 5–33. <https://doi.org/10.1080/17538947.2013.805262>
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- Maxwell, R. M., & Condon, L. E. (2016). Connections between groundwater flow and transpiration partitioning. *Science*, 353(6297), 377–380. <https://doi.org/10.1126/science.aaf7891>
- McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., & Wood, E. F. (2016). The GEWEX LandFlux project: Evaluation of model evaporation using tower-based and globally gridded forcing data. *Geoscientific Model Development*, 9(1), 283–305. <https://doi.org/10.5194/gmd-9-283-2016>
- McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., et al. (2017). A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data*, 4(1), 1–19. <https://doi.org/10.1038/sdata.2017.12>
- Mengis, N., Keller, D. P., Eby, M., & Oschlies, A. (2015). Uncertainty in the response of transpiration to CO<sub>2</sub> and implications for climate change. *Environmental Research Letters*, 10(9). <https://doi.org/10.1088/1748-9326/10/9/094001>
- Milly, P. C. D., & Dunne, K. A. (2016). Potential evapotranspiration and continental drying. *Nature Climate Change*, 6(10), 946–949. <https://doi.org/10.1038/nclimate3046>
- Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., & Dolman, A. J. (2010). Global canopy interception from satellite observations. *Journal of Geophysical Research*, 115(D16), D16122. <https://doi.org/10.1029/2009JD013530>
- Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., et al. (2016). The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology and Earth System Sciences*, 20(2), 823–842. <https://doi.org/10.5194/hess-20-823-2016>
- Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCM products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, 109(D7). <https://doi.org/10.1029/2003jd003823>
- Moshir Panahi, D., Kalantari, Z., Ghajarnia, N., Seifollahi-Aghmiuni, S., & Destouni, G. (2020). Variability and change in the hydro-climate and water resources of Iran over a recent 30-year period. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-64089-y>
- Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8), 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>
- Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., et al. (2011). Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters*, 38(6), 1–7. <https://doi.org/10.1029/2010GL046230>
- Muñoz-Sabater, J., Dutra, E., Balsamo, G., Bousssetta, S., Zsoter, E., Albergel, C., & Agustí-Panareda, A. (2018). ERA5-Land: An improved version of the ERA5 reanalysis land component. *Joint ISWG and LSA-SAF Workshop IPMA*.
- Padrón, R. S., Gudmundsson, L., Greve, P., & Seneviratne, S. I. (2017). Large-scale controls of the surface water balance over land: Insights from a systematic review and meta-analysis. *Water Resources Research*, 53(11), 9659–9678. <https://doi.org/10.1002/2017WR021215>
- Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2012). Multisource estimation of long-term terrestrial water budget for major global river basins. *Journal of Climate*, 25(9), 3191–3206. <https://doi.org/10.1175/JCLI-D-11-00300.1>
- Pellet, V., Aires, F., Munier, S., Fernández Prieto, D., Jordá, G., Arnaud Dorigo, W., et al. (2019). Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle - Application to the Mediterranean region. *Hydrology and Earth System Sciences*, 23(1), 465–491. <https://doi.org/10.5194/hess-23-465-2019>
- Pinker, R. T., Zhang, B., & Dutton, E. G. (2005). Do satellites detect trends in surface solar radiation? *Science*, 308(5723), 850–854. <https://doi.org/10.1126/science.1103159>
- Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. *The American Naturalist*, 164(5), 625–632. <https://doi.org/10.1086/424970>
- Prein, A. F., & Pendergrass, A. G. (2019). Can we constrain uncertainty in hydrologic cycle projections? *Geophysical Research Letters*, 46(7), 3911–3916. <https://doi.org/10.1029/2018GL081529>
- Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoin, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, 557(7707), 651–659. <https://doi.org/10.1038/s41586-018-0123-1>

- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85(3), 381–394. <https://doi.org/10.1175/BAMS-85-3-381>
- Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, 460(7258), 999–1002. <https://doi.org/10.1038/nature08238>
- Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2011). Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sensing of Environment*, 115(8), 1850–1865. <https://doi.org/10.1016/j.rse.2011.03.009>
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis. The primer*. John Wiley & Sons. <https://doi.org/10.1002/9780470725184>
- Senay, G. B., Budde, M. E., & Verdin, J. P. (2011). Enhancing the Simplified Surface Energy Balance (SSEB) approach for estimating landscape ET: Validation with the METRIC model. *Agricultural Water Management*, 98(4), 606–618. <https://doi.org/10.1016/J.AGWAT.2010.10.014>
- Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, 36(7). <https://doi.org/10.1029/2009GL037338>
- Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19(13), 3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., Braithwaite, D., et al. (2000). Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9), 2035–2046. [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSSE>2.3.CO;2)
- Sposito, G. (2017). Understanding the Budyko equation. *Water*, 9(4), 236. <https://doi.org/10.3390/w9040236>
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1), 79–107. <https://doi.org/10.1002/2017RG000574>
- Taylor, C. M., De Jeu, R. A. M., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, 489(7416), 423–426. <https://doi.org/10.1038/nature11377>
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, 13(14), 4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>
- Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, 120(4), 2648–2671. <https://doi.org/10.1002/2014JB011547>
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50(9), 7505–7514. <https://doi.org/10.1002/2014WR015638>
- Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., Cooper, J. E., et al. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, 77(5), 853–868. [https://doi.org/10.1175/1520-0477\(1996\)077<0853:CATERE>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2)
- Willmott, C., & Matsuura, K. (2000). *Terrestrial air temperature and precipitation: Monthly and annual climatologies*. University of Delaware.
- Xu, X., Liu, W., Scanlon, B. R., Zhang, L., & Pan, M. (2013). Local and global factors controlling water-energy balances within the Budyko framework. *Geophysical Research Letters*, 40(23), 6123–6129. <https://doi.org/10.1002/2013GL058324>
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, 44(11), 5844–5853. <https://doi.org/10.1002/2017GL072874>
- Yang, H., Yang, D., Lei, Z., & Sun, F. (2008). New analytical derivation of the mean annual water-energy balance equation. *Water Resources Research*, 44(3). <https://doi.org/10.1029/2007WR006135>
- Zhang, K., Kimball, J. S., Nemani, R. R., & Running, S. W. (2010). A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research*, 46(9). <https://doi.org/10.1029/2009WR008800>
- Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., & Briggs, P. R. (2004). A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, 40(2). <https://doi.org/10.1029/2003WR002710>
- Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports*, 6(1), 19124. <https://doi.org/10.1038/srep19124>