

# Water Resources Research



## RESEARCH ARTICLE

10.1029/2019WR026280

## Using the Budyko Framework for Calibrating a Global Hydrological Model

P. Greve<sup>1</sup> , P. Burek<sup>1</sup> , and Y. Wada<sup>1</sup> 

<sup>1</sup>International Institute for Applied Systems Analysis, Water Program, Laxenburg, Austria

### Key Points:

- Most global hydrological models are uncalibrated using predefined sets of basic hydrological parameters
- We developed a simple calibration approach that requires no additional data and substantially improves model performance
- The new approach is based on the Budyko framework providing a conceptual representation of the long-term water and energy balance

### Supporting Information:

- Supporting Information S1

### Correspondence to:

P. Greve,  
greve@iiasa.ac.at

### Citation:

Greve, P., Burek, P., & Wada, Y. (2020). Using the Budyko framework for calibrating a global hydrological model. *Water Resources Research*, 56, e2019WR026280. <https://doi.org/10.1029/2019WR026280>

Received 3 SEP 2019

Accepted 1 MAY 2020

Accepted article online 13 MAY 2020

**Abstract** Global hydrological models (GHMs) have become an established tool to simulate water resources worldwide. Most of the GHMs are however uncalibrated and typically use a set of basic hydrological parameters, that could potentially lead to unrealistic projections of the terrestrial water cycle. The calibration of hydrological models is usually performed by using and comparing modeled to observed discharge. Accurate station data and reliable time series data of discharge are, however, often not available for many parts of the world and classic calibration approaches are therefore not feasible. In this paper, we aim to develop a new calibration approach that requires no additional data, is easy to implement, and substantially improves model performance, especially in regions where uncalibrated model performance is rather poor. This is achieved by using the Budyko framework, which provides a conceptual representation of the long-term water and energy balance. We use a state-of-the-art GHM and calibrate the model within nine river catchments of different sizes and characteristics. Since observed river discharge is available for these catchments, we are able to compare the Budyko-based calibration approach to a classic discharge-based calibration scheme and the uncalibrated model version. In all catchments, the Budyko-based calibration approach decreases biases and increases model performance compared to the uncalibrated model version although performance improvements obtained through a classic calibration approach are greater. Nonetheless, a Budyko-based calibration is a valuable, intermediate approach between use of a basic set of a priori hydrological parameters and classical calibration against discharge data.

## 1. Motivation

Global hydrological models (GHMs) have become a common tool to assess the large-scale dynamics of surface and subsurface hydrological processes, ranging from more classical hydrologic modeling approaches focusing on small catchment areas, up to continental-basin scales. At smaller scales, models are usually calibrated using a large set of parameters. The majority of GHMs are however uncalibrated, typically using reduced sets of a priori parameters. Focusing on individual or small sets of catchments, calibration is usually performed using observed discharge. Global modeling approaches suffer from multiple issues hindering a reasonable and feasible calibration procedure, most prominently the varying availability of accurate discharge observations in different world regions (Hanasaki et al., 2018; Müller-Schmid et al., 2014). While in North America, Europe, and parts of Asia dense gauging networks enable a comprehensive calibration of model parameters, this is largely impossible in Africa, South America, Australia, and other parts of Asia. In addition, GHMs usually show large uncertainties and biases and do not permit quantitative assessments of past, present, and future hydroclimatological changes, especially in these data-scarce regions. Calibrating GHMs only against observed records that are concentrated in Europe and North America does not necessarily help to provide better estimates in such data-scarce regions. The availability of station observations has also constantly been decreasing since the early 1990s (Hunger & Döll, 2008; Vörosmary et al., 2001).

The issue of heterogeneous, inaccurate observations is addressed by using recent, more comprehensive collections of discharge data (Do et al., 2018). However, these still suffer from large data gaps in many tropical to subtropical regions. In addition, observation-driven data sets of global streamflow (Barbarossa et al., 2018) have recently become available. These are spatially explicit, but still based on observed data records that are subject to data gaps. Another approach is the regionalization of model parameters. By using sets of catchments for calibration that cover a wide range of climatic and landscape characteristics, the obtained parameters can subsequently be transferred to grid cells with similar characteristics (Beck et al., 2016). Additionally, remotely sensed data sources have been shown to be useful in specific cases (Lopez et al., 2017;

©2020. The Authors.

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

Rakovec et al., 2016; Revilla-Romero et al., 2015; Wanders et al., 2014; Zink et al., 2018), especially when using combinations from multiple data sources (Nijzink et al., 2018). Even though such approaches have been shown to improve the representation of hydrologic variables within the models, they still require extensive amounts of additional data, which are themselves subject to uncertainties and limitations. More conceptual approaches that aim to overcome this issue by minimizing or even avoiding the use of additional data for calibration have been discussed in previous research (Gharari et al., 2014; Hulsman et al., 2018; Kapangaziwiri et al., 2012; Li et al., 2014). Such approaches suggest new methods to make use of insufficient data sources by, for example, using water level measurements (Hulsman et al., 2018; Jian et al., 2017) or discontinuous observational time series (Reynolds et al., 2019). More conceptual approaches use relational constraints based on expert knowledge to specify the parameter space (Gharari et al., 2014) or to constrain hydrological model output and the parameter space based on, for example, the Budyko framework (Budyko, 1974; Bouaziz et al., 2018; Kapangaziwiri et al., 2012; Li et al., 2014).

The motivation to use the Budyko framework in this context is that it provides a benchmark for the long-term coupled water and energy balance that can be used in the context of, for example, quantifying intercatchment groundwater flows (Bouaziz et al., 2018) or to evaluate hydroclimatological data set combinations (Greve et al., 2014). In the spirit of those conceptual studies exploring possibilities to make use of the Budyko framework, we aim to fully avoid the use of additional data by presenting an innovative calibration approach for large-scale and global hydrological models that is solely based on this fundamental hydrological concept. Using the Budyko framework enables us to ensure a good representation of the basic characteristics of the long-term coupled water and energy balance, and we argue that, given its simplicity and data requirements, an implementation into existing calibration tools is easy and efficient. We hypothesize that this approach will reduce runoff biases but not the subdaily to seasonal modeling of high and low flows and that it will provide a major and simple step toward more reliable large-scale hydrological modeling in data-scarce and ungauged regions.

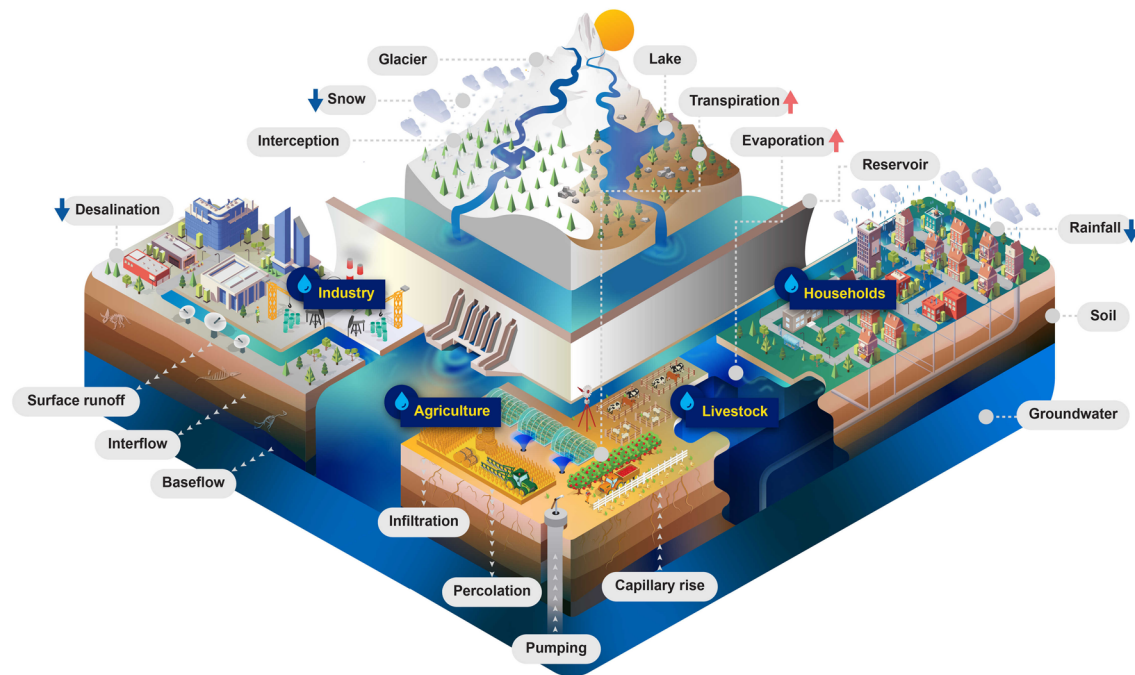
The Budyko framework conceptually represents the coupled water and energy balance. The Budyko hypothesis (Budyko, 1974) states that the evaporative index,  $E/P$ , defined as the relative fraction of mean annual (or longer) average precipitation,  $P$ , that is used for evapotranspiration,  $E$ , is a function of the aridity index  $\phi$ , defined as the ratio of potential evaporation to precipitation ( $PET/P$ ). Here we use an extended version of the Budyko hypothesis, including an additional single, free parameter (here referred to as  $\omega$ ):

$$\frac{E}{P} = f(\phi, \omega) \quad (1)$$

The Budyko hypothesis assumes no changes in water storages, therefore ideally requiring multiyear averages and catchment scales. Based on the Budyko hypothesis, it is often assumed that the first-order control of the climatological water and energy balance is represented through  $\phi$ , while the single, free parameter  $\omega$  represents the wealth of all other (second-order) controls. A widely used function fulfilling the requirements of the Budyko hypothesis is the Tixeront-Fu equation (Fu, 1981; Tixeront, 1964; Zhang et al., 2004):

$$\frac{E}{P} = (1 + \phi) - (1 + \phi^\omega)^{1/\omega} \quad (2)$$

The parameter  $\omega$  technically represents the residual influence on  $E/P$  besides  $\phi$ . If  $\omega = 2.6$ , it matches the original, empirical Budyko curve as introduced by Budyko (1974). In recent years factors such as vegetation and topography (Carmona et al., 2016; Destouni et al., 2013; Donohue et al., 2010, 2012; Jaramillo & Destouni, 2015; Jiang et al., 2015; Li et al., 2013; Ning et al., 2019; Oudin et al., 2008; Porporato et al., 2004; Shao et al., 2012; Wang & Hejazi, 2011; Williams et al., 2012; Xu et al., 2013; Yang et al., 2009; Yokoo et al., 2008; Zhou et al., 2015; Zhang et al., 2001, 2016, 2018) have been proposed to control  $\omega$ . Other studies also suggest climatic and weather influences, like phase shifts in the seasonal cycles of  $P$  and  $PET$  or average storm depth (Berghuijs et al., 2014; Carmona et al., 2016; Dean et al., 2016; Donohue et al., 2012; de Lavenne & Andréassian, 2018; Ning et al., 2019; Milly, 1994; Shao et al., 2012; Yokoo et al., 2008). However, the obtained results between the studies are generally inconclusive and sometimes contradictory (Padrón et al., 2017). Hence, instead of using deterministic  $\omega$  values within specific contexts dependent on the different factors mentioned above, a probabilistic representation of the parameter potentially provides a better, general representation of the full range of  $\omega$  values (Greve et al., 2015; Gudmundsson et al., 2016).



**Figure 1.** Schematic illustration of the processes included in the Community Water Model (CWatM).

Instead of determining a parameter set best representing observed discharge, the overarching goal of this study is to show the potential of calibrating a parameter set that best represents the Budyko curve, including variability in  $\omega$ , thereby ensuring a model setup that provides a reasonable estimate of the long-term water balance as a function of climatic aridity. Furthermore, we do not expect the Budyko-based calibration to achieve performance improvements as great as a classical discharge-based calibration.

In the first part of this study, we introduce the Community Water Model (CWatM) and specify the set of river basins that we use to evaluate the feasibility of a Budyko-based calibration approach (section 2). We then introduce the Budyko-based calibration method and outline how it is implemented (section 3). After presenting the results (section 4), we discuss the advantages and possible limitations of the Budyko-based calibration approach (section 5) before outlining the potential for future developments and further improvements (section 6).

## 2. Modeling Framework and Discharge Data

We use a state-of-the-art GHM—the CWatM—to assess differences between observed and modeled discharge in nine river basins of different sizes and characteristics. The simulated discharge is obtained from (i) the uncalibrated version of the model using a set of basic hydrological parameters, (ii) a model version calibrated against observed discharge within the nine river basins, and (iii) a model version calibrated using the newly developed Budyko-based calibration technique. It is then compared with observed discharge data.

### 2.1. CWatM

CWatM (Burek et al., 2020) is a hydrological rainfall-runoff and channel routing model developed at the International Institute for Applied Systems Analysis (IIASA). It is used to quantify water availability, human water use, and the effects of water infrastructure, including reservoirs, groundwater pumping, and irrigation. It is grid-based with recent versions for 0.5° and 5-min resolutions (with subgrid resolution taking topography and land cover into account). It operates at daily time steps (with subdaily time stepping for soil and river routing). A schematic view of all featured processes is provided in Figure 1. Please note that our main focus here will be on the simulation of runoff and discharge.

CWatM is comparable to other state-of-the-art GHMs like H08 (Hanasaki et al., 2008), WaterGAP (Alcamo et al., 2003), PCR-GLOBWB (van Beek et al., 2011; Wada et al., 2014), LPJmL (Bondeau et al., 2007) and others. It is implemented as an open-source modular structured Python program (Burek et al., 2020) and is

available at Zenodo (Burek et al., 2019) and Github. In this study, CWatM is driven by the daily meteorological forcing data set WFDEI (WATCH Forcing Data methodology applied to ERA-Interim data; Weedon et al., 2014) at a  $0.5^\circ$  spatial resolution and a daily temporal resolution. From WFDEI, the model requires daily estimates of  $P$ , as well as surface air temperature, wind speed, relative humidity, incoming longwave and shortwave radiation, and surface air pressure as inputs to calculate potential evaporation using the Penman-Monteith method based on a reference crop surface (Allen et al., 1998). A crop factor is used to account for differences in vegetation characteristics in comparison to the reference crop. Elevation data used in this study is extracted from SRTM (Jarvis et al., 2008) and the HYDRO1k Elevation Derivative Database (USGS, 2002). Soil properties are obtained from the Harmonized World Soil Database (FAO et al., 2012) and transferred to Van Genuchten model parameters (van Genuchten, 1980) using the pedotransfer model Rosetta3 by Zhang and Chaap (2017). The calculation of the fraction of direct runoff is based on the ARNO scheme (Todini, 1996). Data used for reservoir operation (water supply, flood control, hydropower generation, and others) are obtained from the HydroLakes database (Lehner et al., 2011; Messenger et al., 2016). Crop-specific calendars and growing season lengths are derived from the MIRCA2000 data set (Portmann et al., 2010). CWatM uses the kinematic wave approximation of the Saint-Venant equations (Chow et al., 1988) for river routing. CWatM uses a local drainage direction map which defines the dominant flow direction in one of the eight neighboring grid cells (D8 flow model). For the  $0.5^\circ$  version used here, the drainage direction map (DDM30) of Döll and Lehner (2002) is used. To calculate the kinematic wave in CWATM static maps of channel width, channel bankfull depth, channel gradient, Manning's coefficient, and channel length are needed. These are calculated using the river network, the flow accumulation, elevation data, and average river discharge data. In order to calculate the associated maps, the approach of Pistocchi and Pennington (2006) is used. Please see Burek et al. (2020) for a more detailed description of CWatM hydrological processes.

## 2.2. River Basins

We selected a set of nine river basins worldwide, encompassing different climate and landscape characteristics (from arid to tropical wet, from lowland to mountainous), and size (ranging from  $33,450 \text{ km}^2$  to  $4.7$  million  $\text{km}^2$  and from  $90$  to  $170,000 \text{ m}^3/\text{s}$  average annual discharge). This enables us to assess the potential of using the Budyko-based calibration approaches under varying conditions and to evaluate and analyze the potential caveats and limitations of the approach. For the basins, we used observed discharge from the Global Runoff Data Centre (GRDC), the United States Geological Survey (USGS), the Oesterreichische Wasserstrassen-Gesellschaft (viaDonau), and the Murray-Darling Basin Authority (MDBA). Table 1 provides an overview of the nine basins and their particular characteristics, sorted by their basin-wide aridity index  $\phi$  from humid to arid. For calibration we used a time period of 10 to 20 years and either daily or monthly data depending on data availability. Validation was performed based on independent periods. Please note that the time series for the Upper Nile and the Zambezi rivers were too short to enable an assessment of independent calibration and validation periods

## 3. Calibration

Calibration is performed using two approaches:

- 1: *SimBud*. The newly developed Budyko-based calibration, which determines a parameter set through calibration comprising the characteristics of the Budyko framework.
- 2: *SimDis*. Calibrating CWatM against observed discharge for each catchment, thus representing the classic, ideal approach, where measured discharge is available, assignment to a gauge is clear and values are accurate.

The parameter sets obtained are common for all grid cells within each catchment and are used to set up individual catchment-specific model runs. The model results are further evaluated against discharge estimates obtained from model runs using an a priori, uncalibrated parameter set (*Sim0*). The uncalibrated parameter set is provided in Table 2 and is based on default parameter values either set to unity or to values that are commonly associated with an average hydrological response (such as for the preferential bypass flow or infiltration capacity).

Calibration is performed using an evolutionary computation framework in Python called DEAP (Fortin et al., 2012). DEAP implemented the evolutionary algorithm NSGA-II (Deb et al., 2002) which is used here

**Table 1**  
*River Basins and Stations Used for Budyko Calibration*

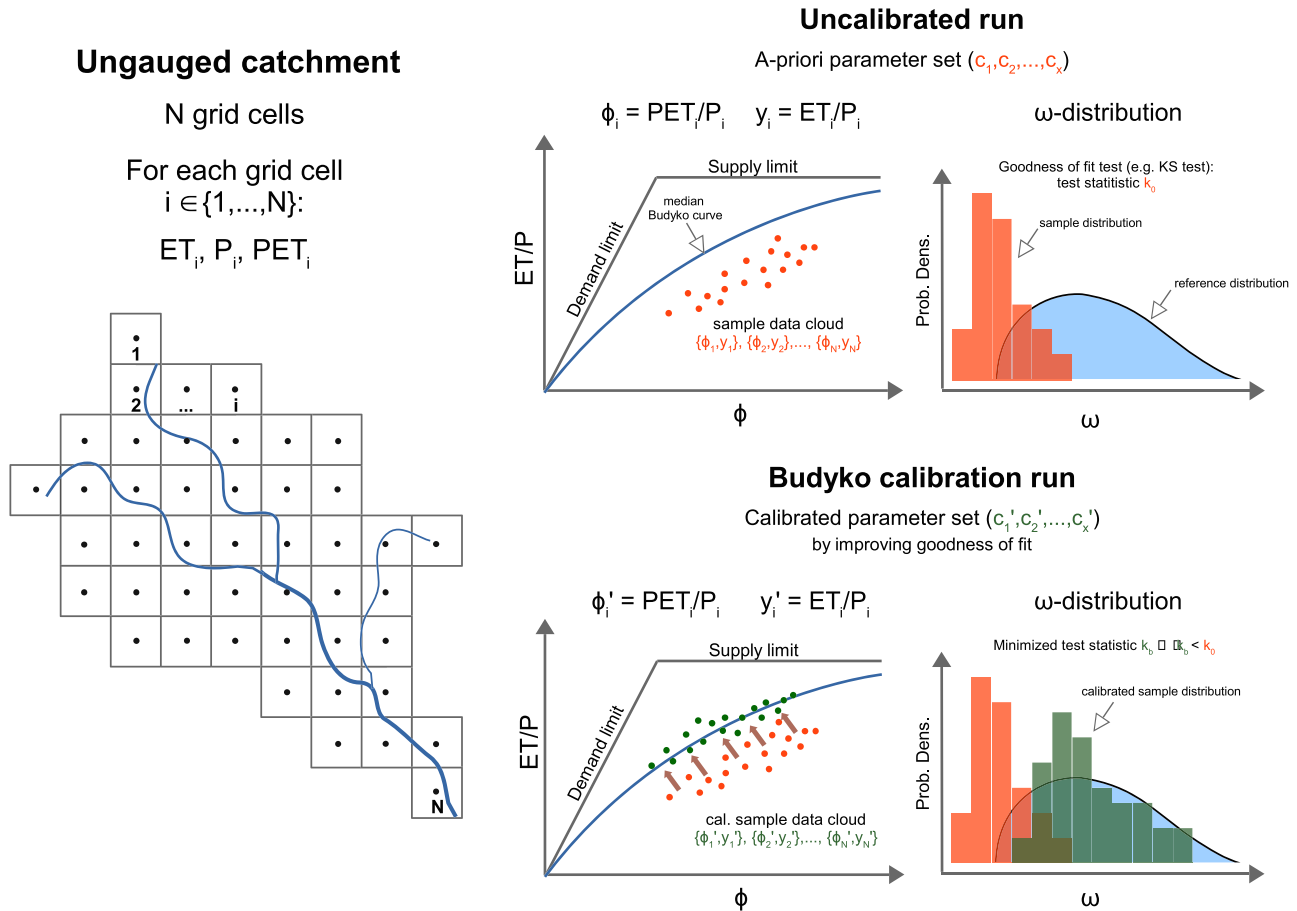
Basin	Gauges	Source	Gauge-lat	Gauge-lonSize	Basin size (km <sup>2</sup> )	Calibration period	Validation period	Average discharge (m <sup>3</sup> /s)	$\phi$
Danube	Kienstock, Austria	viadonau	43.38	15.46	95,970	1995–2010	1980–1994	1,935	0.52
Amazon	Obidos, Brazil	GRDC	−1.95	−55.51	4,680,000	1985–1998	1970–1984	170,000	0.56
Rhine	Lobith, Germany	GRDC	51.84	6.11	160,800	1995–2010	1980–1994	2,318	0.61
Danube	Zimicea, Romania	GRDC	43.63	25.35	658,400	1995–2010	1980–1994	6,119	0.77
Upper Nile	Jinja, Uganda	GRDC	0.43	33.2	262,950	1996–2005	1996–2005	992	1.21
Zambezi	Lukulu, Zambia	GRDC	−14.38	23.23	206,500	1979–1990	1979–1990	706	1.85
Indus	Beeshan, Pakistan	GRDC	34.92	72.88	164,900	1995–2010	1979–1994	2,420	1.88
Sacramento	Wilkins Slough, USA	USGS	39.01	−121.82	33,450	1991–2010	1979–1990	297	2.6
Murray-Darling	Wakool Junction, Australia	MDBA	−34.85	143.35	78,000	2000–2010	1990–1999	89	3.45

*Note.* Data are obtained from (i) GRDC (Global Runoff Data Centre; <https://www.bafg.de/GRDC>), (ii) viadonau (viadonau—Oesterreichische Wasserstrassen-Gesellschaft, <http://www.viadonau.org>), (iii) USGS (United States Geological Survey, <https://www.usgs.gov/>), and (iv) MDBA (Murray-Darling Basin Authority, <https://riverdata.mdba.gov.au>). Average discharge is based on the considered time period. Basins are ranked from humid to arid based on their basin-wide aridity index  $\phi$ . Please note that the time series for the Upper Nile and the Zambezi rivers were too short to enable an assessment of two independent periods. All available observed data were already used for calibration.

**Table 2**  
*List of Parameters Used for Calibration*

Parameter	SimBud	Unit	Process group	Description	Sim0	Min	Max
Snowmelt coefficient	x	m/K/day	Snow	coefficient used in a degree-day factor approach	0.004	0.001	0.007
Crop factor	x	—	Evaporation	adjustment to crop evapotranspiration	1	0.8	1.8
Soil depth factor	x	—	Soil	factor for the overall soil depth of soil layer 1 and 2	1	0.8	1.5
Preferential bypass flow	x	—	Soil	empirical parameter of the preferential flow relation	4	0.5	8
Infiltration capacity parameter	x	—	Soil	empirical shape parameter b of the ARNO model	0.5	0.01	1
Interflow factor	—	—	Groundwater	factor to distinguish the amount which goes to groundwater or interflow	1	0.33	3
Recession coefficient	—	—	Groundwater	factor to adjust the base flow recession constant	1	0.1	10
Runoff concentration	—	—	Routing	factor for the concentration time of run-off in each grid-cell	1	0.1	5
Channel Manning's factor	—	—	Routing	factor to adjust the roughness in channel routing	1	0.1	10
Normal storage limit	—	—	Reservoirs	fraction of total storage capacity used as normal storage limit	0.5	0.15	0.85
Lake A factor	—	—	Lakes	factor to channel width and weir coefficient	1	0.33	3
Lake and river evaporation factor	—	—	Lakes, reservoirs	factor to adjust open water evaporation	1	0.8	2

*Note.* Please note that for SimBud only parameters that are related to runoff generation are used. The other parameters are set to default (such as in Sim0). Minimum and maximum values indicate the parameter range used for calibration.



**Figure 2.** Conceptual representation of a Budyko-based calibration approach. For any (ungauged) catchment consisting of  $N$  grid cells, model estimates of  $ET$ ,  $P$ , and  $PET$  are obtained for each grid cell from the uncalibrated model. These estimates are used to assess the sample distribution of the associated  $\omega$  values and a goodness-of-fit test (e.g., Kolmogorov-Smirnov test) is performed, using a certain reference distribution. Calibration is performed by using the goodness-of-fit test metric as objective function. Calibration thus leads to model results best representing the reference distribution of  $\omega$ .

to perform a single objective optimization. The calibration generally uses a population size ( $\mu$ ) of 256 and a recombination pool size ( $\lambda$ ) of 32. We use 30 generations. For more information on the convergence, please refer to Figures S1 and S2 in the supporting information.

### 3.1. Budyko Calibration:

The Budyko-based calibration approach *SimBud* requires no additional data and employs the Budyko framework to improve the long-term water balance characteristics of the hydrological model with respect to climatic aridity. Within the calibration procedure, we compute the multiyear sums of daily  $P$ ,  $E$ , and  $PET$  at each grid cell within a specific catchment (see Figure 2). The resulting ratios  $\phi$  and  $E/P$  constitute a point cloud within the Budyko space. A straightforward approach to assess the offset from the expected median Budyko curve, would be to use the root-mean-square error (RMSE) to the original Budyko curve ( $\omega = 2.6$ ) as the objective function. However, this approach would not account for the inherent, nondeterministic nature of  $\omega$ , which led us to employ a probabilistic representation of the Budyko framework (Greve et al., 2015). This approach assumes that  $\omega$  is a (quasi-)random variable, which was found to follow a Gamma distribution based on a set of several hundred catchments following the climate gradient of the contiguous United States. The objective function is thus based on a goodness-of-fit metric (here the Kolmogorov-Smirnov distance  $D_i$ ) between the empirical distribution of the modeled  $\omega$  values ( $F(\omega_i)$ ) and a reference Gamma distribution  $F_0(x) = \Gamma(4.54, 0.37) + 1$  that is based on the analysis of observations from several hundred catchments across the contiguous United States (Greve et al., 2015):

$$F_{obj} = \min(D_i) = \min(\sup_i |F(\omega_i) - F_0(\omega_i)|), \quad (3)$$

where  $\sup$  denotes the supremum and  $D_i = \sup_i |F(\omega_i) - F_0(\omega_i)|$  defines the Kolmogorov-Smirnov distance. This means that the calibration algorithm aims to find the parameter set that minimizes  $D_i$  between the simulated and reference distribution. In summary, the calibration approach could be implemented with any appropriate goodness-of-fit metric and consists of the following general procedure:

At first, model parameters that should be calibrated and that are ideally related to the long-term water balance within a particular region or basin are selected. An appropriate single objective calibration tool (e.g., DEAP) is chosen and the respective calibration settings (number of generations, population size, etc.) that provides sufficient convergence are specified. The model is subsequently calibrated based on the following steps:

1. For each grid cell within the chosen catchment or region, compute at least annual, or better, multiyear sums of  $P$ ,  $E$ , and  $PET$  to obtain estimates of  $\phi$  and  $E/P$ .
2. Estimate  $\omega$  (numerically) for each grid cell based on the values of  $\phi$  and  $E/P$  using equation (2). All  $\omega$  values for each grid cell within the catchment are used to construct an empirical distribution. Please note here that  $\omega$  is both very sensitive to small changes close to the physical limits and not defined outside the limits. We decided to ignore data points plotting outside the physical boundaries. However, it is important to make sure that the majority of data points plot inside the limits.
3. Define a reference distribution for  $\omega$ . If no direct reference is available, we suggest using a Gamma distribution  $F_0(x) = \Gamma(4.54, 0.37) + 1$ .
4. Select a goodness-of-fit metric  $D_i$  (e.g., the Kolmogorov-Smirnov test).
5. Implement the goodness-of-fit test metric  $D_i$  (here the Kolmogorov-Smirnov distance) as objective function that should be minimized:  $F_{obj} = \min(D_i)$ .

As outlined above, this approach can be implemented using any suitable goodness-of-fit test and the choice of reference distributions is flexible. The characteristics of such distributions might, for instance, be dependent on climate type (Padrón et al., 2017) and can simply be accounted for by adjusting the reference distribution. The Budyko-based calibration can be used with any other hydrological model, as it is independent of the internal model structure and only requires standard model output (e.g., discharge, precipitation, potential, and actual evaporation).

As the Budyko framework is valid at multiannual time scales, a Budyko-based calibration approach ideally needs a long (at least several years) calibration time period. Here we use calibration periods of at least 10 years (see Table 1). We further apply the Budyko calibration at grid scale. The rather coarse spatial resolution of  $0.5^\circ$  ensures approximate water balance closure over long time scales, thus enabling the use of the Budyko framework at grid scale. Our approach further ensures that the resulting point cloud plots within the physical boundaries of the Budyko space and according to the probabilistic Budyko framework, thereby retaining the inherent variability within the Budyko space. The grid-based approach can also be used at other than basin scales.

Since the Budyko-based calibration approach focuses on the long-term water balance, being primarily related to the generation of runoff and evaporation at each grid cell, we decided to exclude parameters representing groundwater, routing, water demand, and lakes/reservoirs from the calibration procedure (see Table 2). The majority of these parameters (excluding groundwater) primarily represent short-term dynamics and calibrating them based on the long-term aridity index and precipitation partitioning is potentially spurious. Therefore, only five parameters (snowmelt coefficient, crop factor, soil depth factor, preferential bypass flow, and infiltration capacity) are finally calibrated using SimBud, whereas the other parameters are set to default values such as in Sim0. The parameter ranges are defined to cover a wide, though realistic range of the hydrological response associated with the parameters (see Table 2 for more information, units, and parameter ranges). The obtained best-fit parameter sets (based on a minimized Kolmogorov-Smirnov distance) are provided in Table 3.

### 3.2. Discharge Calibration:

We applied a modified version of the Kling-Gupta Efficiency (KGE) (Kling et al., 2012) as objective function:

$$KGE' = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (4)$$

with  $r$  denoting the correlation coefficient between simulated and observed discharge,  $\beta = \mu_s/\mu_o$  representing the bias ratio between mean, simulated ( $\mu_s$ ) and observed streamflow ( $\mu_o$ ), and  $\gamma = CV_s/CV_o$  as the



**Table 3**  
Obtained Best-Fit Parameter Sets for SimBud and SimDis

Parameter	Upper Danube		Amazon		Rhine		Danube		Upper Nile	
	SimBud	SimDis	SimBud	SimDis	SimBud	SimDis	SimBud	SimDis	SimBud	SimDis
Snowmelt coefficient	0.005	0.0026	0.0026	0.0018	0.0047	0.0032	0.0031	0.0029	0.004	0.004
Crop factor	1.24	1.16	0.9	1.2	1.13	1.58	1.1	1.57	1.72	2.14
Soil depth factor	1.4	0.9	0.97	0.94	1.07	1.08	1.46	1.2	0.8	1.24
Preferential bypass flow	4.6	7.67	7.6	7.6	6.57	7.36	5.89	5.46	5.45	5.33
Infiltration capacity	0.74	0.44	0.01	0.81	0.14	0.23	0.27	0.55	0.01	0.43
Interflow factor	1	1.1	1	0.86	1	2	1	1.85	1	0.98
Recession coefficient	1	0.05	1	0.05	1	0.21	1	0.05	1	1.65
Runoff concentration	1	0.95	1	1.57	1	2.54	1	2.37	1	2.13
Channel Manning's factor	1	0.74	1	10	1	0.5	1	1.4	1	7.63
Normal storage limit	0.5	0.56	0.5	0.71	0.5	0.52	0.5	0.85	0.5	0.63
Lake A factor	1	3.2	1	3.95	1	2.1	1	2.52	1	1.18
Lake evaporation factor	1	2	1	1.51	1	1.51	1	1.08	1	0.93
	Zambezi		Indus		Sacramento		Murray-Darling			
Parameter	SimBud	SimDis	SimBud	SimDis	SimBud	SimDis	SimBud	SimDis		
Snowmelt coefficient	0.004	0.004	0.002	0.0015	0.003	0.0022	0.004	0.004		
Crop factor	3	2.91	1.73	0.85	1.73	1.8	1.8	1.8		
Soil depth factor	1.98	1.67	1	1	1.39	1.5	1.01	1.5		
Preferential bypass flow	8	6.89	6.06	0.5	7.53	7.16	8	7.33		
Infiltration capacity	0.23	0.07	0.06	0.69	0.03	0.25	0.01	0.28		
Interflow factor	1	1.57	1	2.34	1	1.01	1	1.4		
Recession coefficient	1	2.3	1	7.1	1	5.05	1	4		
Runoff concentration	1	1.65	1	3.55	1	1.42	1	5		
Channel Manning's factor	1	1.08	1	5.91	1	4.3	1	6.61		
Normal storage limit	0.5	0.5	0.5	0.53	0.5	0.47	0.5	0.32		
Lake A factor	1	1.8	1	1.34	1	1.4	1	1.06		
Lake evaporation factor	1	0.99	1	3.36	1	1.55	1	2		

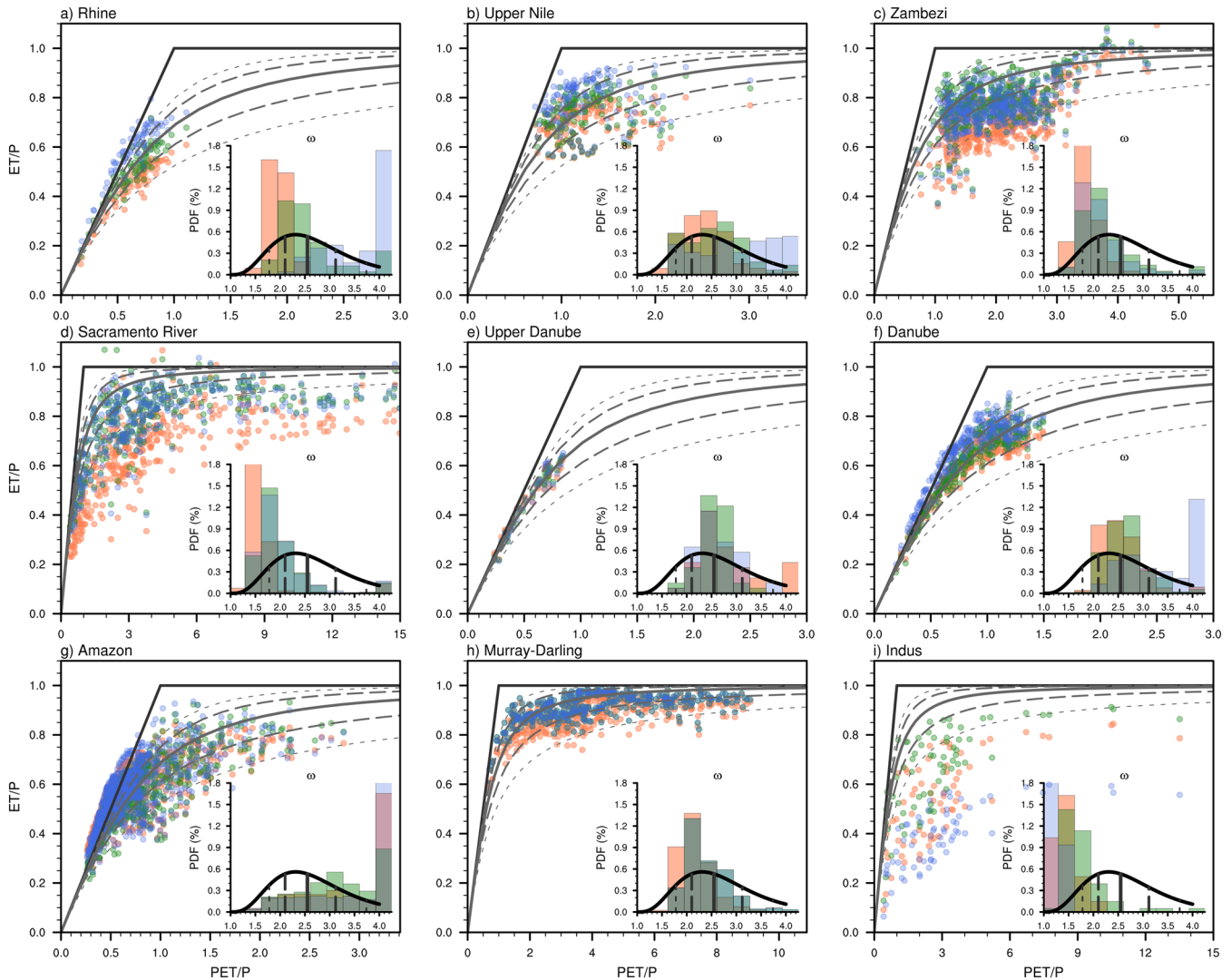
variability ratio between the simulated ( $CV_s$ ) and observed ( $CV_o$ ) coefficient of variation (unmodified KGE is computed based on the standard deviation).

Since  $r$ ,  $\beta$ , and  $\gamma$  have their optimum at unity, the KGE' could be interpreted as the Euclidean distance from the optimal value (i.e., unity) of the Pareto front and is therefore able to provide an estimate that simultaneously represents optimal correlation, mean bias, and variability. Discharge calibration is here performed by maximizing KGE against multiyear daily time series of observed discharge from the individual catchments. The calibration periods correspond to those of the Budyko-based calibration approach (see Table 1). Since validation against observed discharge can be performed within the calibration period for the Budyko calibration (as the observed discharge is not part of the calibration routine and the respective objective function), a discharge-based calibration approach ideally requires independent validation periods (see Table 1).

For discharge calibration, we use a set of 12 parameters that are optimized to maximize the KGE. These parameters represent snow, soil, groundwater, and routing characteristics, as well as evaporation processes and the treatment of lakes within the model and include the five parameters used in the Budyko-based calibration approach. Similar to the Budyko calibration approach, the parameter ranges are defined to cover a realistic range of the hydrological response associated with the respective parameters (see Table 2). The obtained best-fit parameter sets (based on a maximized KGE) are provided in Table 3.

#### 4. Results and Discussion

The nine basins chosen here represent a wide range of climate types, catchment characteristics, and catchment area. By selecting these basins, we are able to focus both on those catchments where a Budyko-based

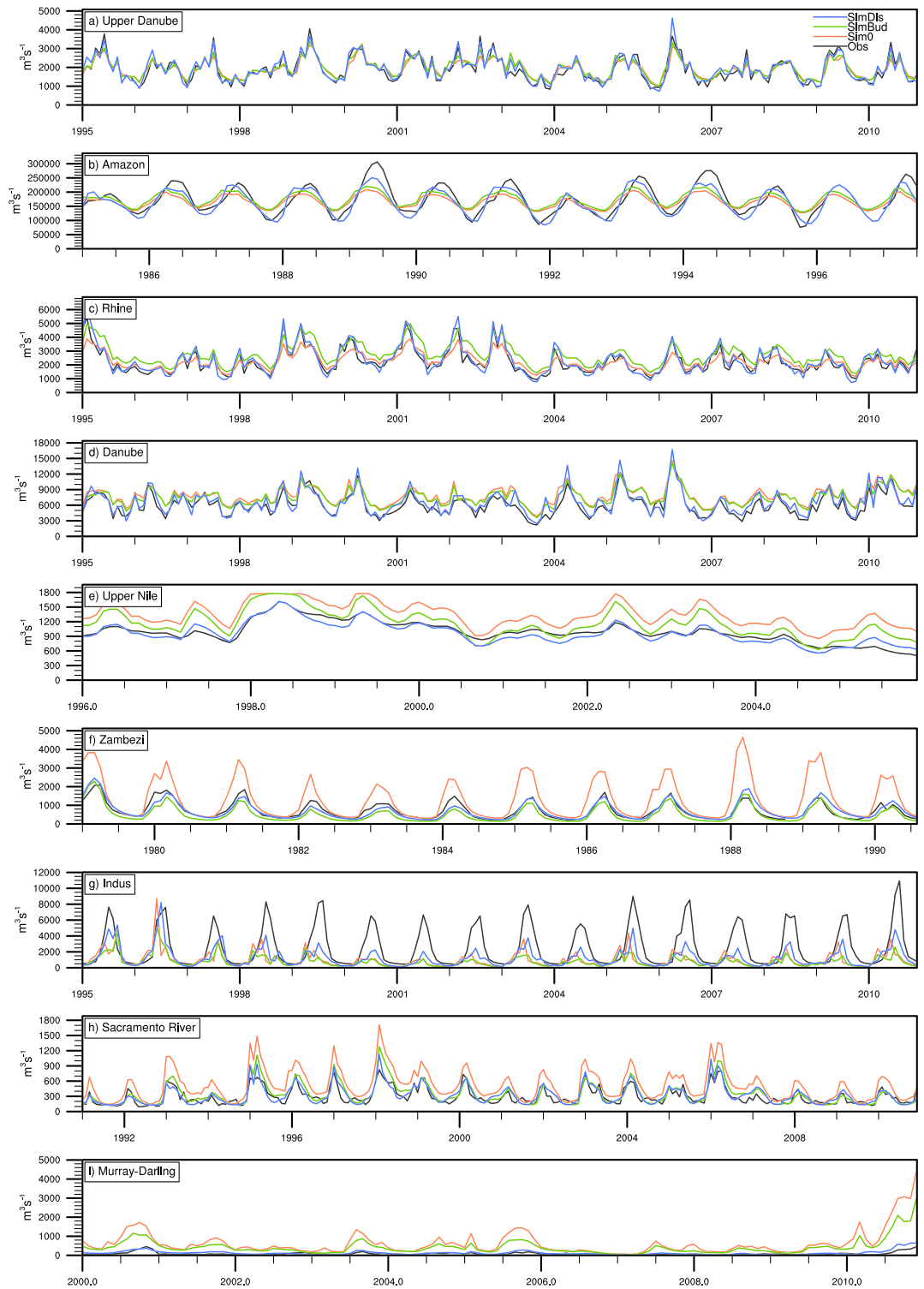


**Figure 3.** Sample data clouds of *Sim0* (red), *SimBud* (green), and *SimDis* (blue) for the nine study catchments. Each dot represents a grid cell within the catchment. The reference Budyko distribution used in this study is indicated by solid (median), dashed (25th, 75th quantile) and dotted lines (10th, 90th quantile). Subplots illustrate the associated sample distributions of  $\omega$  alongside the reference distribution of  $\omega$  as used in this study. Please note that the rightmost bar sums up all values larger than 4, including also those grid cells exceeding the theoretical limits of the Budyko framework.

calibration leads to significant increases in model performance and on those catchments illustrating the limitations and caveats of the simple Budyko-based calibration approach. We first assess the implications of the Budyko-based calibration approach (*SimBud*) within the Budyko space and compare it to the uncalibrated (*Sim0*) and discharge calibrated model runs (*SimDis*). In a following step, we discuss potential improvements regarding daily and seasonal discharge based on the obtained results. The obtained model results from the three experiments are validated against observed discharge. The model performance estimates for *Sim0*, *SimBud*, and *SimDis* presented in the following are for the validation periods indicated in Table 1. However, please note that *SimBud* does not require independent validation and calibration periods since no discharge observations are needed for calibration. In the following we present results for the respective best-fit parameter sets. Considering a larger selection of those simulations obtained during the calibration process that best minimize the objective function enables an assessment of the range of possible model results and parameter ranges and is provided in the supporting information (Figures S3, S4, and Table S1).

#### 4.1. Implications of Calibration Within the Budyko Space

By focusing on the Budyko space and the sample distributions of  $\omega$  obtained from the three considered experiments (*Sim0*, *SimBud*, and *SimDis*), we are able to assess the associated implications of choosing a



**Figure 4.** Observed (gray) and modeled time series of monthly river discharge for the (a–i) nine study catchments. Please note differences in the length of the considered time periods and the total amount of monthly river discharge.

**Table 4**  
Various Statistics of *Sim0*, *SimBud*, and *SimDis* Within the Nine Study Catchments

	KGE			NSE			R			Bias (%)			RMSE		
	Sim0	SimBud	SimDis	Sim0	SimBud	SimDis	Sim0	SimBud	SimDis	Sim0	SimBud	SimDis	Sim0	SimBud	SimDis
Upper Danube	0.53	0.64	0.81	0.53	0.62	0.621	0.74	0.79	0.81	0.31	1.39	1.19	563	505	504
Amazon	0.42	0.45	0.87	0.47	0.47	0.75	0.75	0.72	0.88	4.8	4.41	2.54	36,413	36,465	25,837
Rhine	0.49	0.81	0.9	0.48	0.59	0.81	0.85	0.84	0.91	25.4	8.92	5.53	909	810	555
Danube	0.54	0.58	0.83	0.22	0.38	0.64	0.82	0.83	0.86	27.7	22.7	6.5	2,024	1,806	1,387
Upper Nile	0.57	0.77	0.92	-1.88	0.	0.86	0.85	0.87	0.93	35.7	17.4	-2.8	377	222	87
Zambezi	0.11	0.58	0.91	-2.9	0.67	0.83	0.9	0.93	0.92	84.2	-28.4	-0.5	597	264	190
Indus	0.21	0.29	0.38	-0.1	0.05	0.2	0.76	0.67	0.77	-64.9	-55.6	-52.5	2,700	2,515	2,335
Sacramento	0.14	0.61	0.54	-1.64	0.15	0.44	0.86	0.73	0.81	82.7	26.5	27.7	276	157	128
Murray-Darling	-1.2	0.1	0.51	-2.8	0.1	0.5	0.64	0.68	0.74	214	73	-20	493	240	184

particular calibration approach (see Figure 3). For the majority of catchments (Rhine, Zambezi, Sacramento, Danube, Murray-Darling, and Indus), the uncalibrated run *Sim0* generally plots below the median Budyko curve (according to the reference Budyko distribution)—a sign that the uncalibrated model favors runoff. This is both evident within the Budyko space and from the associated sample distributions. The distributions generally peak at values below both the median and mean of the reference gamma distribution. *SimBud* generates distributions of  $\omega$  better resembling the reference distribution. For these catchments, the resulting data cloud fits the distribution of Budyko curves within the Budyko space reasonably well.

The sample distribution of  $\omega$  obtained from *SimBud* thus provides the best representation of the reference distribution in most catchments. Using discharge calibration *SimDis*, the data cloud often plots above the reference Budyko distribution (especially for the Rhine, Upper Nile, Danube, and the Amazon). In some catchments (Rhine, Danube, and the Amazon) many data points further plot above the theoretical limits of the Budyko space, therefore actually distorting the sample distributions of  $\omega$  (technically  $\omega$  is not defined for values exceeding the supply and/or demand limit). Even in catchments with a rather good representation of  $\omega$  under *Sim0* (e.g., Upper Nile and the Danube), *SimDis* usually shifts the data cloud to values above the median Budyko curve.

For the Amazon and Indus river basin, both *SimBud* and also *SimDis* do not resemble the reference distribution, either peaking above (Amazon) or below (Indus) the Budyko distribution. The largest offsets are obtained for the Indus, where especially *SimDis* plots far below the Budyko distribution. The best representations of the reference distributions are obtained in both the Upper Danube and the greater Danube, the Upper Nile, and the Murray-Darling.

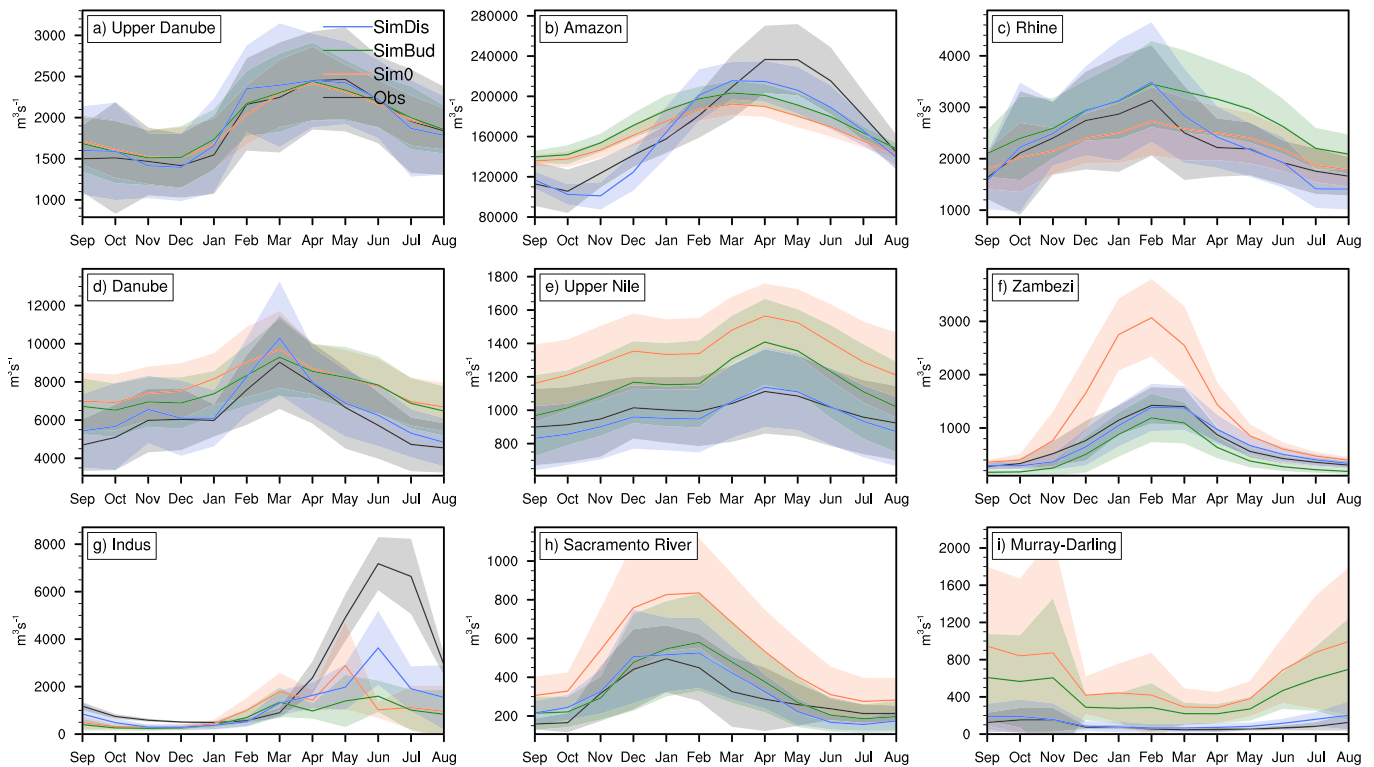
#### 4.2. Implications of Calibration Regarding Catchment Discharge

Here we analyze the implications of the calibration experiments on (i) monthly river discharge and (ii) the representation of the mean annual cycle of river discharge. We further provide an overview of the relevant statistics in comparison to the observed discharge time series (see Table 3).

##### 4.2.1. Monthly Time Series

Figure 4 provides an overview of (i) the observed time series of monthly discharge and the modeled time series obtained from (ii) *Sim0*, (iii) *SimBud*, and (iv) *SimDis*. Focusing on these time series, the most striking feature is that in those catchments with *Sim0* generally plotting below the reference Budyko distribution—an indication of the uncalibrated model to favor  $Q$  over  $E$  (see Figure 3, Zambezi, Upper Nile, Sacramento, Murray-Darling)—modeled discharge is generally overestimated in all months or specifically within the wet season (as for the Zambezi). This is largely resolved through calibration. However, while *SimBud* substantially increases the overall model performance, it is generally outperformed by *SimDis*.

By further examining the associated model statistics (Table 4), we aim to assess whether the Budyko-based calibration approach *SimBud* provides a substantial improvement in model performance against the uncalibrated model *Sim0* and how this compares to improvements achieved through the discharge calibration method *SimDis*.



**Figure 5.** Observed (gray) and modeled multiyear mean annual cycles of river discharge for the (a–i) nine study catchments. Please note differences in the total amount of monthly river discharge. Shaded areas denote the multiannual standard deviation.

For the Upper Danube and the Rhine, the uncalibrated run *Sim0* already provides a rather good representation of the observed discharge time series, showing a reasonable KGE, NSE, and correlation coefficient and small biases. The statistics are generally improved under *SimDis*, while *SimBud* slightly improves KGE and most of the other statistics. Among the humid catchments featured in this study, the Amazon shows lowest KGE in *Sim0*, while NSE, correlation, and bias are comparable to those of the Rhine and the Upper Danube. However, in the Amazon *SimBud* does not improve NSE and biases, while *SimDis* substantially improves KGE and NSE. For the greater Danube, KGE and correlation coefficient are already reasonable in *Sim0* and are slightly improved under *SimBud* compared to an overall stronger improvement under *SimDis*. However, NSE and bias are, compared to the smaller European catchments (Rhine, Upper Danube) and the Amazon, rather poor and again slightly improved under *SimBud* (while the bias still remains above 20%). *SimDis* indeed significantly reduces the bias and further improves NSE.

Moving toward more arid climates, the overall performance of *Sim0* is generally rather poor for the Upper Nile basin (including Lake Victoria), the Zambezi, the Indus, and the Sacramento rivers. Biases are especially large in *Sim0* and are largely improved in *SimBud*, even though substantial biases remain compared to improvements obtained in *SimDis* for the Upper Nile and the Zambezi. While KGE was already quite high in *Sim0* and further improved by *SimBud* in the Upper Nile, large improvements from a small KGE in *Sim0* toward a reasonably larger KGE in *SimBud* are obtained for the Zambezi and the Sacramento rivers. This also applies to NSE. Nonetheless, improvements obtained through discharge calibration *SimDis* are in any case more substantial than those obtained by *SimBud*, especially for the Upper Nile basin. Nonetheless, the Budyko-based calibration approach *SimBud* improves the large offset and largely improves the model performance for the selected arid catchments, especially for the Zambezi and Sacramento rivers. In the Zambezi basin, the uncalibrated run *Sim0* shows poor model performance, with almost no skill in simulating observed discharge, especially due to large positive biases that substantially overestimate discharge in the wet season. Further improvements are obtained using *SimDis*, almost completely resolving the bias.

From the arid catchments, the Indus and especially the Murray-Darling show poor performances in *Sim0*. For the Indus basin *SimBud* provides no potential improvement and largely underestimates observed

discharge. However, even though *SimDis* slightly improves KGE, NSE, and the correlation coefficient, the overall model performance for *SimDis* is also still rather poor, implying that CWatM may not be capable of capturing the hydrology of the Indus. Starting from a very poor overall performance of the model in *Sim0* for the Murray-Darling basin, large improvements are achieved by using *SimBud*, while, however, overall model performance remains at a poor level. *SimDis* does indeed substantially improve biases and model performance in the Murray Darling basin, providing an overall reasonable model performance. For an illustration of the percentage bias as a function of  $\phi$  please refer to Figure S5.

#### 4.2.2. Annual Cycle

Figure 5 provides representations of the observed and modeled mean annual cycles of river discharge within the study catchments. We are again able to distinguish between catchments in which *Sim0* already provides a good representation of the observed annual cycle (Rhine, Upper Danube, and also partly the Danube), and catchments in which *Sim0* generally overestimates discharge throughout all seasons (Upper Nile, Sacramento, and Murray-Darling) or specifically within the wet season (Zambezi). In the Amazon, the observed annual cycle is more pronounced compared to *Sim0*, showing more (less) discharge in the wet (dry) season, resulting in an overall rather small relative bias (see Table 3). In the Indus, modeled discharge is largely underestimated in the wet season. *SimDis* generally resembles the observed annual cycle, except for the Amazon basin, showing a slight shift in peak discharge and for the Indus, largely underestimating wet season discharge. *SimBud* also resembles observed discharge in many catchments (Rhine, Zambezi, Sacramento, and the Upper Danube), while it overestimates winter discharge in the greater Danube. In some catchments (Upper Nile, Murray-Darling, and the Danube), *SimBud* still shows substantial biases, but outperforms the uncalibrated run *Sim0* and shifts the modeled run toward the observed annual cycle. In general, for all catchments, *SimBud* leads to a more realistic representation of the observed annual cycle.

## 5. Advantages and Limitations of a Budyko-Based Model Calibration

We used a set of nine catchments of varying size, surface, and climate characteristics in order to evaluate the potential of a Budyko-based calibration approach for large-scale hydrologic models. We were able to show that a Budyko-based calibration method substantially increases model performance in the majority of catchments, but does not lead to similar improvements as those achieved through a classical discharge-based calibration approach. However, calibration based on the Budyko framework requires no additional data and is easy to implement into various modeling frameworks. As the Budyko framework provides a simple, conceptual characterization of the long-term water balance as a function of climatic aridity, it ensures its reasonable representation within a modeling context. It was especially shown to be useful in reducing biases and also in increasing model performance regarding monthly river discharge dynamics. As it is easy to implement within any calibration tool, we conclude that a Budyko-based calibration approach could be a useful alternative to uncalibrated, basic hydrological parameter sets that are widely used in large-scale hydrologic models, especially in ungauged or poorly gauged regions.

While it has advantages, a simple Budyko-based calibration is also subject to a variety of potential limitations, and caution in the interpretation and applicability of the suggested calibration approach needs to be taken. Nonetheless, the Budyko model ensures a physical consistent representation of the long-term water balance based on climatic aridity. A Budyko-based calibration approach makes use of this and enables a better representation of long-term water balance components in complex, large-scale models, thereby substantially reducing biases and improving skill in regions with poor default model performance.

We used a set of catchments that allowed us to assess the disadvantages and caveats of the Budyko calibration across different climates. Focusing, for instance, on both the greater Danube basin and an Upper Danube subbasin, could provide us with some insights related to issues arising from growing catchment size. When focusing on the large greater Danube Basin as opposed to the smaller subbasin in the Upper Danube, *SimBud* does not provide reasonable improvements, but rather reveals possible limitations of the Budyko-based calibration approach. In the larger, almost continental-sized greater Danube basin, other processes related to the routing, water extraction, and representation of lakes and reservoirs likely become more important. Such processes are not considered within the Budyko-based approach as these are assumed to be only effectively represented through discharge calibration. This becomes even more evident when considering one of the largest river basins in the world, the Amazon basin. In the Amazon, average discharge is around 170,000 m<sup>3</sup>/s and depends mostly on the timing of streamflow, which is primarily determined through river

routing and the representation of lakes and reservoirs. Hence, even though *SimBud* slightly improves model performance for the Amazon, the improvements obtained in *SimDis* are far more substantial.

The Budyko-based calibration approach further improves the long-term water balance at grid level. Certain river basins that extend into semiarid to arid regions might not be well represented within *SimBud* since most of the runoff is generated in a rather small fraction of the catchment area. We illustrate this by considering the Murray-Darling river basin. The Murray-Darling rises within the humid eastern Australian mountain ranges, while most of the basin area is semiarid and the transmission loss from the river is substantial. The transmission loss is mostly controlled through routing processes, which are not assumed to be effectively calibrated in *SimBud*. Since the Budyko approach aims to better represent the long-term water balance as a function of climatic aridity at each grid cell, a lot of emphasis is put on the large number of semiarid grid cells that do not generate substantial amounts of runoff. Hence, large improvements are expected and achieved through using *SimDis*, while *SimBud* reduces the very large bias of *Sim0*, without substantially improving most other statistics. This might also apply to other river basins with similar characteristics, such as the Niger or the Nile. Additionally, it is important to consider that many of those basins (and especially the Murray-Darling basin) are heavily managed, and water extractions for irrigation are substantial and need to be taken into account to provide realistic discharge simulations.

Generally speaking, it appears that *SimBud* shows the best performance under humid and more temperate conditions (e.g., in midlatitude regions such as Europe and North America), while already *Sim0* provides reasonable results in these regions. The highest potential for model improvements is indeed found in semiarid to arid catchments, where the overall model performance of *Sim0* is rather poor and largely improved by *SimBud*, despite still showing substantial biases (see Table 4). A reasonable representation of discharge in arid regions also requires a good setup of model parameters that are not necessarily improved through *SimBud*. Even though not explicitly assessed here, it is likely that *SimBud* will also improve model performance in high-latitude catchments, since snowmelt coefficients are directly calibrated. As mentioned above, it is also important to consider that the nine considered catchments are subject to human interventions. Previous studies suggest that population density (Jiang et al., 2015; Wang & Hejazi, 2011), irrigation (Han et al., 2011; Jaramillo & Destouni, 2015; Wang & Hejazi, 2011), and dams and reservoirs substantially impact the representation of the water balance within the Budyko framework. The results also need to be considered in this context.

Another example of basins where no significant improvement in model performance could be expected from *SimBud* is the Indus. Neither *SimBud* nor *SimDis* improve model performance in the Indus basin as a result of either poorly constrained forcing data or a poor parameterization of relevant processes within CWatM. However, even sophisticated calibration approaches are not able to overcome and address biases in the forcing data, with *SimBud* being no exception to this. Discharge in the Indus basin is thus largely underestimated in both *SimBud* and *SimDis*.

It is also important to note that the calibration method presented here could be further improved. This might, for instance, be achieved by using different reference distributions representing variations in  $\omega$ . Such distributions depend, for example, on the climate type (Padrón et al., 2017). However, this also implies that we might overestimate the variability in  $\omega$  by using a reference distribution obtained from observations covering a large climatic gradient. Another improvement could be the use of other, more comprehensive goodness-of-fit measures (such as, the Anderson-Darling test) or approaches based on maximum likelihood estimation. These could, in comparison to the Kolmogorov-Smirnov distance, provide a better representation of the characteristics of the particular reference distribution and might increase the selectivity of a parameter set in order to represent the entirety of  $\phi$  and  $E/P$  ratios. It should further be considered that, at the moment, only the water balance of a grid cell without incoming discharge and evaporation from rivers and lakes are estimated and neither the storage term, incoming discharge from upstream cells, nor evaporation from lakes and wetlands are taken into account. Taking this into consideration, possible limitations with respect to the Budyko-based calibration approach need to be evaluated in the future.

## 6. Concluding Remarks

The Budyko-based approach improved biases and model performance in comparison to an uncalibrated model simulation throughout a set of nine catchments. However, as expected, these improvements in model

performance are not as substantial in comparison to a classical approach using river discharge as the objective function for calibration. Nonetheless, our initial hypothesis, stating that a Budyko-based calibration method potentially provides better results than using a priori sets of basic hydrological parameters without calibration, was proven valid. Hence, the Budyko-based calibration method introduced here can be seen as an independent and beneficial approach over using predefined parameter sets in ungauged catchments or in regions with sparse data.

Given previous conceptual research to constrain the parameter space of hydrological models based on the Budyko framework (Kapangaziwiri et al., 2012; Li et al., 2014), we present here a novel method to directly and independently calibrate hydrological model parameters at grid scale without the use of any additional data. As calibration based on the Budyko framework ensures a good representation of the long-term water balance based on climatic aridity, improvements in annual and partly also in seasonal hydrological statistics are achieved and improve hydrological model performance most in poorly gauged basins.

The simplicity of the Budyko framework should enable an easy implementation into other existing large-scale hydrological models. Compared to calibration based on various data products, Budyko-based calibration approaches can be directly applied and require no preprocessing and data handling. However, while remote sensing products potentially provide more advanced results, especially in data scarce and ungauged regions, it is also important to note that these products obtained from satellite data are based on various (model) assumptions and are thus subject to large uncertainties and to multiple sources of uncertainty. Hence, a Budyko-based calibration approach will provide an easily accessible tool to ensure an improved representation of the long-term water balance as a function of climatic aridity within large-scale hydrological models. It could further be implemented without any data prerequisites. And, if not alone, it will at least provide a useful addition to calibration approaches using multiple information sources based on, for example, remote-sensing products (Nijzink et al., 2018).

However, it is important to note that the improvement in model performance is subject to the particular model setup and the choice of model parameters. Using other hydrological models or more sophisticated parameter sets might not permit a substantial improvement in model performance under Budyko-based calibration approaches. However, due to the general applicability of the Budyko framework, representing (and therefore ensuring) valid and consistent long-term hydrological conditions as a function of climatic aridity, it is likely that a Budyko-based calibration will improve model performance in other hydrological models as well (at least in comparison to uncalibrated setups).

In summary, the Budyko-based calibration approach presented in this study enables us to implement a calibration tool for CWatM, ideally performing a global calibration for all catchments in order to replace the basic hydrological parameter assumptions that are used in most GHMs. This will potentially overcome unrealistic representations of available water resources in river catchments.

#### Acknowledgments

Data availability: Data on river discharge are available online from GRDC, viadonau, USGS, MDBA (see Table 1), WFDEI data are available through [https://www.eu-watch.org/data\\_availability](https://www.eu-watch.org/data_availability); Model availability: CWatM is open source under GNU General Public License. More information can be found under these sites (<https://www.iiasa.ac.at/cwatm> and <https://cwatm.github.io/>).

#### References

- Alcamo, J., Doell, P., Henrichs, T., Kaspar, F., Lehner, B., Roesch, T., & Siebert, S. (2003). Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences Journal*, 48(3), 317–337. <https://doi.org/10.1623/hysj.48.3.317.45290>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56. *FAO, Rome*, 300(9), D05109.
- Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H., & Schipper, A. M. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Scientific Data*, 5, 180,052. <https://doi.org/10.1038/sdata.2018.52>
- Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles D. G., McVicar, T. R., Schellekens J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52, 3599–3622. <https://doi.org/10.1002/2015WR018247>
- Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, 4(7), 583–586. <https://doi.org/10.1038/nclimate2246>
- Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., & Smith, B. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, 13(3), 679–706. <https://doi.org/10.1111/j.1365-2486.2006.01305.x>
- Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., van Savenije, H., & Hrachowitz, M. (2018). Redressing the balance: Quantifying net intercatchment groundwater flows. *Hydrology and Earth System Sciences*, 22, 6415–6434. <https://doi.org/10.5194/hess-22-6415-2018>
- Budyko, M. I. (1974). *Climate and life*. New York: Academic Press.
- Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., et al. (2019). Community Water Model (CWatM v1.04). *Zenodo*. <https://doi.org/10.5281/zenodo.3361478>
- Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., et al. (2020). Development of the Community Water Model (CWatM v1.04): A high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development Discussions*. <https://doi.org/10.5194/gmd-2019-214>



- Burek, P., van der Knijff, J., & de Roo, A. (2013). LISFLOOD, distributed water balance and flood simulation model.
- Carmona, A. M., Poveda, G., Sivapalan, M., Vallejo-Bernal, S. M., & Bustamante, E. (2016). A scaling approach to Budyko's framework and the complementary relationship of evapotranspiration in humid environments: Case study of the Amazon River basin. *Hydrology and Earth System Sciences*, 20(2), 589–603. <https://doi.org/10.5194/hess-20-589-2016>
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). Applied hydrology.
- de Lavenne, A., & Andréassian, V. (2018). Impact of climate seasonality on catchment yield: A parameterization for commonly-used water balance formulas. *Journal of Hydrology*, 558, 266–274. <https://doi.org/10.1016/j.jhydrol.2018.01.009>
- Dean, J. F., Camporese, M., Webb, J. A., Grover, S. P., Dresel, P. E., & Daly, E. (2016). Water balance complexities in ephemeral catchments with different land uses: Insights from monitoring and distributed hydrologic modeling. *Water Resources Research*, 52(6), 4713–4729. <https://doi.org/10.1002/2016WR018663>
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. <https://doi.org/10.1109/4235.996017>
- 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>
- Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata archive (GSIM) ? Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, 10(2), 765–785. <https://doi.org/10.5194/essd-10-765-2018>
- Döll, P., & Lehner, B. (2002). Validation of a new global 30-min drainage direction map. *Journal of Hydrology*, 258, 214–231. [https://doi.org/10.1016/S0022-1694\(01\)00565-0](https://doi.org/10.1016/S0022-1694(01)00565-0)
- Donohue, R. J., Roderick, M. L., & McVicar, T. R. (2010). Can dynamic vegetation information improve the accuracy of Budyko's hydrological model? *Journal of Hydrology*, 390(1-2), 23–34. <https://doi.org/10.1016/j.jhydrol.2010.06.025>
- 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, 35–50. <https://doi.org/10.1016/j.jhydrol.2012.02.033>
- FAO, IASA, ISRIC, ISSCAS, & JRC (2012). Harmonized world soil database (version 1.2).
- Fortin, F. A., Rainville, F. M. D., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms Made Easy. *Journal of Machine Learning Research*, 13(Jul), 2171–2175.
- Fu, B. (1981). On the calculation of the evaporation from land surface (in Chinese). *Scientia Geographica Sinica*, 1(5), 23–31.
- Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & van Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, 18, 4839–4859. <https://doi.org/10.5194/hess-18-4839-2014>
- Greve, P., Gudmundsson, L., Orlovsky, B., & Seneviratne, S. I. (2015). Introducing a probabilistic Budyko framework. *Geophysical Research Letters*, 42, 2261–2269. <https://doi.org/10.1002/2015GL063449>
- 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, 716–721. <https://doi.org/10.1038/ngeo2247>
- Gudmundsson, L., Greve, P., & Seneviratne, S. I. (2016). The sensitivity of water availability to changes in the aridity index and other factors—A probabilistic analysis in the Budyko-space. *Geophysical Research Letters*, 43, 6985–6994. <https://doi.org/10.1002/2016GL069763>
- Han, S., Hu, H., Yang, D., & Liu, Q. (2011). Irrigation impact on annual water balance of the oases in Tarim Basin, Northwest China. *Hydrological Processes*, 25(2), 167–174. <https://doi.org/10.1002/hyp.7830>
- Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., & Tanaka, K. (2008). An integrated model for the assessment of global water resources ? Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, 12(4), 1007–1025. <https://doi.org/10.5194/hess-12-1007-2008>
- Hanasaki, N., Yoshikawa, S., Pokrel, Y., & Kanae, A. (2018). A global hydrological simulation to specify the sources of water used by humans. *Hydrology and Earth System Sciences*, 22, 789–817. <https://doi.org/10.5194/hess-22-789-2018>
- Hulsman, P., Bogaard, T. A., & van Savenije, H. H. G. (2018). Rainfall-runoff modelling using river-stage time series in the absence of reliable discharge information: A case study in the semi-arid Mara River basin. *Hydrology and Earth System Sciences*, 22(5081–5095). <https://doi.org/10.5194/hess-22-5081-2018>
- Hunger, M., & Döll, P. (2008). Value of river discharge data for global-scale hydrological modeling. *Hydrology and Earth System Sciences*, 12, 841–861. <https://doi.org/10.5194/hess-12-841-2008>
- 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>
- Jarvis, A., Reuter, H., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. available from the CGIAR-CSI SRTM 90 m Database <https://srtm.csi.cgiar.org>
- Jian, J., Ryu, D., Costelloe, J. F., & Su, C.-H. (2017). Towards hydrological model calibration using river level measurements. *Journal of Hydrology: Regional Studies*, 10, 95–109. <https://doi.org/10.1016/j.ejrh.2016.12.085>
- Jiang, C., Xiong, L., Wang, D., Liu, P., Guo, S., & Xu, C. Y. (2015). Separating the impacts of climate change and human activities on runoff using the Budyko-type equations with time-varying parameters. *Journal of Hydrology*, 522, 326–338. <https://doi.org/10.1016/j.jhydrol.2014.12.060>
- Kapangaziwiri, E., Hughes, D. A., & Wagener, T. (2012). Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. *Hydrological Sciences Journal*, 57, 1000–1019. <https://doi.org/10.1080/02626667.2012.690881>
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- de Lavenne, A., Andréassian, V., 2018. Impact of climate seasonality on catchment yield: A parameterization for commonly-used water balance formulas. *Journal of Hydrology* 558, 266–274. <https://doi.org/10.1016/j.jhydrol.2018.01.009>
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9), 494–502. <https://doi.org/10.1890/100125>
- 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, 969–976. <https://doi.org/10.1002/wrcr.20107>
- Li, H.-Y., Sivapalan, M., Tian, F., & Harman, C. (2014). Functional approach to exploring climatic and landscape controls of runoff generation: 1. Behavioral constraints on runoff volume. *Water Resources Research*, 50, 9300–9322. <https://doi.org/10.1002/2014WR016307>

- Lopez Lopez, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*, 21(6), 3125–3144. <https://doi.org/10.5194/hess-21-3125-2017>
- Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7, 13603. <https://doi.org/10.1038/ncomms13603>
- Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30(7), 2143–2156. <https://doi.org/10.1029/94WR00586>
- Müller-Schmid, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, 18, 3511–3538. <https://doi.org/10.5194/hess-18-3511-2014>
- Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, 54, 8332–8362. <https://doi.org/10.1029/2017WR021895>
- Ning, T., Zhou, S., Chang, F., Shen, H., Li, Z., & Liu, W. (2019). Interaction of vegetation, climate and topography on evapotranspiration modelling at different time scales within the Budyko framework. *Agricultural and Forest Meteorology*, 275, 59–68. <https://doi.org/10.1016/j.agrformet.2019.05.001>
- Oudin, L., Andréassian, V., Lerat, J., & Michel, C. (2008). Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments. *Journal of Hydrology*, 357(3–4), 303–316. <https://doi.org/10.1016/j.jhydrol.2008.05.021>
- 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, 9659–9678. <https://doi.org/10.1002/2017WR021215>
- Pistocchi, A., & Pennington, D. (2006). European hydraulic geometries for continental scale environmental modelling. *Journal of Hydrology*, 329(3–4), 553–567. <https://doi.org/10.1016/j.jhydrol.2006.03.009>
- Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change, 164(5), 25–632. <https://doi.org/10.1086/an.2004.164.issue-5>
- Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, 24, GB1011. <https://doi.org/10.1029/2008GB003435>
- Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, 52, 7779–7792. <https://doi.org/10.1002/2016WR019430>
- Revilla-Romero, B., Beck, H. E., Burek, P., Salamon, P., de Roo, A., & Thielen, J. (2015). Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent. *Remote Sensing of Environment*, 171, 118–131. <https://doi.org/10.1016/j.rse.2015.10.022>
- Reynolds, J. E., Halldin, S., Seibert, J., Xu, C. Y., & Grabs, T. (2019). Robustness of flood-model calibration using single and multiple events. *Hydrological Sciences Journal*, 1–12. <https://doi.org/10.1080/02626667.2019.1609682>
- Shao, Q., Traylen, A., & Zhang, L. (2012). Nonparametric method for estimating the effects of climatic and catchment characteristics on mean annual evapotranspiration. *Water Resources Research*, 48, W03517. <https://doi.org/10.1029/2010WR009610>
- Tixeront, J. (1964). Prediction of streamflow, *French: Prevision des apports des cours d'eau* (pp. 118–126). IAHS, Ghentbrugge: IAHS Publication 63. General Assembly of Berkeley.
- Todini, E. (1996). The ARNO rainfall-runoff model. *Journal of Hydrology*, 175(1), 339–382. [https://doi.org/10.1016/S0022-1694\(96\)80016-3](https://doi.org/10.1016/S0022-1694(96)80016-3)
- USGS (2002). Geological survey center for earth resources observation and science, Hydro1k. Land Processes Distributed Active Archive Center (LP DAAC).
- van Beek, L. P. H., Wada, Y., & Bierkens, M. F. P. (2011). Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, 47, W07517. <https://doi.org/10.1029/2010WR009791>
- van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil Science Society of America Journal*, 44(5), 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Vörosmary, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., et al. (2001). Global water data: A newly endangered species. *EOS: Transactions American Geophysical Union*, 82(5), 54–58. <https://doi.org/10.1029/01EO00031>
- Wada, Y., Wissler, D., & Bierkens, M. F. P. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics*, 5(1), 15–40. <https://doi.org/10.5194/esd-5-15-2014>
- Wanders, N., Bierkens, M. F. P., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, 50, 6874–6891. <https://doi.org/10.1002/2013WR014639>
- Wang, D., & Hejazi, M. (2011). Quantifying the relative contribution of the climate and direct human impacts on mean annual streamflow in the contiguous United States. *Water Resources Research*, 47, W00J12. <https://doi.org/10.1029/2010WR010283>
- 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, 7505–7514. <https://doi.org/10.1002/2014WR015638>
- Williams, C. A., Reichstein, M., Buchmann, N., Baldocchi, D., Beer, C., Schwalm, C., & Schaefer, K. (2012). Climate and vegetation controls on the surface water balance: Synthesis of evapotranspiration measured across a global network of flux towers. *Water Resources Research*, 48, W06523. <https://doi.org/10.1029/2011WR011586>
- 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, 6123–6129. <https://doi.org/10.1002/2013GL058324>
- Yang, D., Shao, W., Yeh, P. J. F., Yang, H., Kanae, S., & Oki, T. (2009). Impact of vegetation coverage on regional water balance in the nonhumid regions of China. *Water Resources Research*, 45, W00A14. <https://doi.org/10.1029/2008WR006948>
- Yokoo, Y., Sivapalan, M., & Oki, T. (2008). Investigating the roles of climate seasonality and landscape characteristics on mean annual and monthly water balances. *Journal of Hydrology*, 357(3–4), 255–269. <https://doi.org/10.1016/j.jhydrol.2008.05.010>
- Zhang, Y., & Chaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, 547, 39–53. <https://doi.org/10.1016/j.jhydrol.2017.01.004>
- Zhang, L., Dawes, W. R., & Walker, G. R. (2001). Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resources Research*, 37(3), 701–708. <https://doi.org/10.1029/2000WR900325>
- 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, W02502. <https://doi.org/10.1029/2003WR002710>
- Zhang, S., Yang, Y., McVicar, T. R., & Yang, D. (2018). An analytical solution for the impact of vegetation changes on hydrological partitioning within the Budyko framework. *Water Resources Research*, 54, 519–537. <https://doi.org/10.1002/2017WR022028>
- Zhang, S., Yang, H., Yang, D., & Jayawardena, A. W. (2016). Quantifying the effect of vegetation change on the regional water balance within the Budyko framework. *Geophysical Research Letters*, 43, 1140–1148. <https://doi.org/10.1002/2015GL066952>

- Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., & Su, Y. (2015). Global pattern for the effect of climate and land cover on water yield. *Nature Communications*, 6, 5918.
- Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. *Water Resources Research*, 54, 2976–998. <https://doi.org/10.1002/2017WR021346>