JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD031485

Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments

Hylke E. Beck[1], Ming Pan[1], Peirong Lin[1], Jan Seibert[2,3,4], Albert I. J. M. van Dijk[5], and Eric F. Wood[1]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [2]Department of Geography, University of Zurich, Zurich, Switzerland, [3]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden, [4]Department of Physical Geography, Stockholm University, Stockholm, Sweden, [5]Fenner School for Environment and Society, Australian National University, Canberra, ACT, Australia

Key Points:
- We produced seamless parameter maps for the HBV hydrological model covering the entire land surface including ungauged regions
- Improvements in daily streamflow simulation performance were obtained for 88% of the 4,229 independent validation catchments
- Improvements were obtained even for highly isolated validation catchments, demonstrating the value of the approach for poorly gauged regions

Correspondence to:
H. E. Beck,
hylke.beck@gmail.com

Citation:
Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., & Wood, E. F. (2020). Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031485. https://doi.org/10.1029/2019JD031485

Received 7 AUG 2019
Accepted 3 JUL 2020
Accepted article online 30 JUL 2020

Abstract All hydrological models need to be calibrated to obtain satisfactory streamflow simulations. Here we present a novel parameter regionalization approach that involves the optimization of transfer equations linking model parameters to climate and landscape characteristics. The optimization was performed in a fully spatially distributed fashion at high resolution (0.05°), instead of at lumped catchment scale, using an unprecedented database of daily observed streamflow from 4,229 headwater catchments (<5,000 km^2) worldwide. The optimized equations were subsequently applied globally to produce parameter maps for the entire land surface including ungauged regions. The approach was evaluated using the Kling-Gupta efficiency (KGE) and a gridded version of the hydrological model HBV. Tenfold cross validation was used to evaluate the generalizability of the approach and to obtain an ensemble of parameter maps. For the 4,229 independent validation catchments, the regionalized parameters yielded a median KGE of 0.46. The median KGE improvement (relative to uncalibrated parameters) was 0.29, and improvements were obtained for 88% of the independent validation catchments. These scores compare favorably to those from previous large catchment sample studies. The degree of performance improvement due to the regionalized parameters did not depend on climate or topography. Substantial improvements were obtained even for independent validation catchments located far from the catchments used for optimization, underscoring the value of the derived parameters for poorly gauged regions. The regionalized parameters—available via www.gloh2o.org/hbv—should be useful for hydrological applications requiring accurate streamflow simulations.

1. Introduction

All hydrological models, whether physical or conceptual, need to be calibrated to obtain satisfactory streamflow simulations, due to (i) the impossibility of measuring all required model parameters at the model application scale, and (ii) the simplification and spatiotemporal discretization of complex, highly heterogeneous rainfall-runoff processes (Beven, 1989; Blöschl & Sivapalan, 1995; Duan et al., 2001, 2006; McDonnell et al., 2007; Refsgaard, 1997; Vereecken et al., 2019). Conventional calibration approaches aim to improve the correspondence between observed and simulated streamflow by adjusting the model parameters either manually (e.g., Bajracharya et al., 2017; Mishra et al., 2018) or automatically (e.g., Hirpa et al., 2018; Nijssen et al., 2001; for overviews, see Gupta et al., 2013; Moradkhani & Sorooshian, 2009; Refsgaard, 1997; Yilmaz et al., 2010). These approaches typically result in uniform parameter values for each catchment and spatial discontinuities between adjacent catchments. Additionally, they are not applicable in ungauged regions, which comprise the large majority of the global land surface (Fekete & Vörösmarty, 2007; Hannah et al., 2011; Sivapalan, 2003).

Regionalization approaches are designed to estimate model parameters in ungauged regions through the transfer of knowledge from gauged to ungauged catchments (see reviews by Beck et al., 2016; Blöschl et al., 2013; Hrachowitz et al., 2013; Parajka et al., 2013; Razavi & Coulibaly, 2013; Samaniego et al., 2010). The most widely used regionalization approaches involve catchment-by-catchment calibration followed by (i) regression with landscape and climate predictors (e.g., Abdulla & Lettenmaier, 1997; Döll et al., 2003;

Livneh & Lettenmaier, 2013), (ii) transfer of parameter sets based on geographic proximity (e.g., Merz & Blöschl, 2004; Widén-Nilsson et al., 2007), or (iii) transfer of parameter sets based on physical or climatic similarity (e.g., Beck et al., 2016; Nijssen et al., 2001). However, these approaches assume lumped (i.e., spatially uniform) parameter values for each catchment and thus neglect the often pronounced within-catchment heterogeneity in landscape and climate (Kling & Gupta, 2009; Rouholahnejad-Freund et al., 2019). Additionally, they ignore the discrepancy in scale and thus rainfall-runoff behavior between catchments and grid cells (Becker & Braun, 1999; Blöschl & Sivapalan, 1995). Furthermore, the regression method is confounded by parameter equifinality (i.e., different parameter sets yielding the same results Beven, 2006; Kokkonen et al., 2003), while the geographic-proximity method should only be used in data-rich regions (Oudin et al., 2008; Reichl et al., 2009; Vandewiele & Elias, 1995). Another type of regionalization approach employs observation-based maps of streamflow signatures (such as baseflow index) to constrain model parameters (e.g., Boughton & Chiew, 2007; Troy et al., 2008; Yadav et al., 2007). However, this approach is confounded by the limited predictability and information content of streamflow signatures (Addor et al., 2018; Beck et al., 2015). Yet another type involves the simultaneous optimization of model parameters for catchments grouped based on landscape and climate (e.g., Arheimer et al., 2019; Huang et al., 2019), but this approach fails to account for the within-catchment heterogeneity in landscape and climate and yields uniform parameters for each group.

A different regionalization approach that overcomes most of the aforementioned limitations involves the optimization of coefficients of transfer equations linking model parameters to landscape and climate predictors. Notable examples include Hundecha and Bárdossy (2004), who calibrated the HBV model for 95 catchments in the European Rhine basin; Bastola et al. (2008), who calibrated TOPMODEL for 26 catchments in the UK; Rakovec et al. (2016), who calibrated the mHM model for 400 catchments across Europe; and Mizukami et al. (2017), who calibrated the VIC model for 531 catchments in the conterminous United States. A limitation of these studies, however, is their regional focus and therefore limited generalizability. Additionally, three of these studies (Hundecha & Bárdossy, 2004; Mizukami et al., 2017; Rakovec et al., 2016) did not incorporate climate-related predictors, despite several large-scale regionalization studies highlighting their value (e.g., Beck et al., 2016; Nijssen et al., 2001; Singh et al., 2014; Young, 2006). Furthermore, Hundecha and Bárdossy (2004) and Bastola et al. (2008) used lumped predictor and parameter values for the catchments and thus neglected the within-catchment heterogeneity in hydrological processes (Kling & Gupta, 2009; Rouholahnejad-Freund et al., 2019; Samaniego et al., 2017). Rakovec et al. (2016) and Mizukami et al. (2017) ran their hydrological models in a spatially distributed fashion and thus did account for within-catchment heterogeneity, although they used fairly coarse spatial grids (0.25° and 0.125°, respectively).

Here we present a novel parameter regionalization approach using transfer equations to derive an ensemble of parameter maps for the global land surface including ungauged regions. The transfer equations were optimized in a spatially distributed fashion at high resolution (0.05°) based on an unprecedented database of daily observed streamflow from 4,229 headwater catchments (<5,000 km^2) worldwide. The approach was implemented using a gridded version of the HBV hydrological model forced with state-of-the-art downscaled meteorological data. Tenfold cross validation was used to evaluate the generalizability of the approach and to quantify uncertainty globally. We address the following aspects of the approach: (i) the model performance, (ii) the factors determining the performance, and (iii) the spatial patterns of the regionalized parameters.

2. Data and Methods

2.1. Regionalization Approach

Our regionalization approach involves the optimization of coefficients in transfer equations linking model parameters to predictors related to climate, land cover, topography, and soils (Figure 1). For the optimization, the hydrological model was run at a daily temporal and a 0.05° spatial resolution, and the runoff outputs were spatially aggregated for each catchment. The aggregated runoff was subsequently compared to the observed streamflow of each catchment through the computation of a performance score, after which a mean performance score was calculated over all catchments. We only used observed streamflow from catchments up to 5,000 km^2, due to the dominance of channel routing effects in larger catchments at the daily

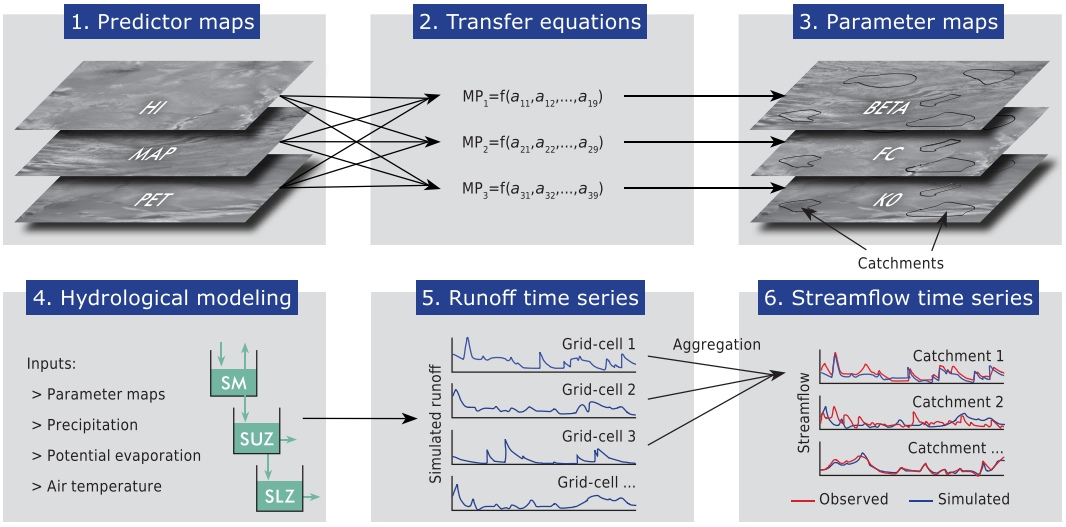

Figure 1. Schematic diagram illustrating the main steps of our model parameter regionalization approach. These steps are executed for each climate group, each cross-validation iteration, and each optimization algorithm evaluation.

time scale (Gericke & Smithers, 2014). The model was only run for the 171,128 grid cells which were included in any of the catchments (comprising 2.2% of the entire land surface), to use computational resources efficiently.

The transfer equations established for the model parameters have the following form:

$$\mathrm{MP}_i = a_{i1}\,\mathrm{HI} + a_{i2}\,\mathrm{MAP} + a_{i3}\,\mathrm{PET} + a_{i4}\,\mathrm{NDVI} + a_{i5}\,\mathrm{OW} + a_{i6}\,\mathrm{Slope} + a_{i7}\,\mathrm{Sand} + a_{i8}\,\mathrm{Clay} + a_{i9}, \quad (1)$$

where MP_i is the ith model parameter; a_{i1_9} are the coefficients that need to be optimized; and HI, MAP, PET, NDVI, OW, Slope, Sand, and Clay are the clipped and standardized predictors (defined in Table 1). We only considered linear relationships with the predictors (except for MAP), to avoid highly skewed predictor distributions which might lead to overfitting. Although this might not be ideal for every parameter-predictor combination (Heuvelmans et al., 2006; Samaniego et al., 2010), determining the most appropriate shape for each parameter-predictor relationship is beyond the scope of the current study.

Table 1
Predictors Used in the Transfer Equations for the Model Parameter Regionalization

Predictor	Description	Resolution	Data source
HI	Humidity index (−), ratio of long-term precipitation to potential evaporation	1 km	WorldClim V2 (Fick & Hijmans, 2017) (www.worldclim.org)
MAP	Mean annual precipitation (mm yr^{-1}), square root transformed	1 km	See HI
PET	Mean annual potential evaporation (mm yr^{-1}) calculated using Hargreaves (1994) from minimum and maximum daily temperature	1 km	See HI
NDVI	Mean Normalized Difference Vegetation Index (NDVI; Tucker, 1979)	1 km	SPOT-VEGETATION and PROBA-V (Maisongrande et al., 2004) (www.vito-eodata.be)
OW	Fraction of open water (lakes and reservoirs)	1 km	GLWD Level-3 (Lehner & Döll, 2004) (www.worldwildlife.org/pages/global-lakes-and-wetlands-database)
Slope	Topographic slope (°)	90 m	MERIT (Yamazaki et al., 2017) (https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/)
Sand	Soil sand content (%), average over all layers	250 m	SoilGrids250m (Hengl et al., 2017) (https://soilgrids.org/)
Clay	Soil clay content (%), average over all layers	250 m	See Sand

Among the eight predictors, three were related to climate (AI, MAP, and PET), two to land cover (NDVI and OW), one to topography (Slope), and two to soils (Sand and Clay). Climate-related predictors were used because climate is known to influence vegetation, soils, and geomorphology and thus exerts a major indirect influence on the rainfall-runoff response (Gentine et al., 2012; Troch et al., 2013). Additionally, several continental- and global-scale regionalization studies have highlighted the value of incorporating climate-related predictors (e.g., Beck et al., 2016; Nijssen et al., 2001; Singh et al., 2014; Young, 2006). The MAP predictor was square root transformed to render the data more normally distributed. The NDVI predictor was added as vegetation influences both evaporation and runoff (Zhang et al., 2001; Donohue et al., 2007; Peel, 2009). The OW predictor was added because of its predictive power in previous regionalization studies (Beck, van Dijk, et al., 2013; Van der Velde et al., 2013). The Sand and Clay predictors were used because soil texture has a strong influence on rainfall-runoff processes (Hewlett, 1961; Price, 2011; Zecharias & Brutsaert, 1988) and is correlated with several streamflow signatures (Beck, van Dijk, et al., 2013; Santhi et al., 2008). The Slope predictor was included because surface slope is correlated with soil depth (Tesfa et al., 2009) and because more steeply sloping aquifers are expected to drain faster (Brutsaert & Nieber, 1977; Vogel & Kroll, 1996; Zecharias & Brutsaert, 1988).

Each predictor was preprocessed as follows. First, we upscaled the data to 0.05° using bilinear averaging. Second, we filled any gaps using nearest neighbor. Third, we clipped the values using the 1st and 99th percentiles of the area covered by the catchments, to avoid application of the transfer equations outside the range of the catchment distribution of predictor values. Fourth and last, we standardized the values by subtracting the mean and dividing by the standard deviation of the area covered by the catchments, to make all predictors intercomparable. Since all predictor fields have a native resolution $\leq$1 km (Table 1), the optimized transfer equations can be used to derive parameter maps at up to 1-km resolution globally.

2.2. Observed Streamflow and Catchment Selection

We used an initial database of daily observed streamflow and catchment boundaries for 21,955 stations worldwide (Beck et al., 2020). The database was compiled from seven different national and international sources (listed in descending order of the number of catchments): (i) the United States Geological Survey (USGS) National Water Information System (NWIS; https://waterdata.usgs.gov/nwis) and GAGES-II database (Falcone et al., 2010; 9,180 catchments); (ii) the Global Runoff Data Centre (GRDC; https://grdc.bafg.de; Lehner, 2012; 4,628 catchments); (iii) the HidroWeb portal of the Brazilian Agência Nacional de Águas (https://www.snirh.gov.br/hidroweb; 3,029 catchments); (iv) the European Water Archive (EWA) of EURO-FRIEND-Water (https://ne-friend.bafg.de) and the CCM2-JRC CCM River and Catchment Database (https://inspire-geoportal.ec.europa.eu/demos/ccm; Vogt et al., 2007; 2,260 catchments); (v) Water Survey of Canada (WSC) National Water Data Archive (HYDAT; https://www.canada.ca/en/environment-climate-change; 1,479 catchments); (vi) the Australian Bureau of Meteorology (BoM; https://www.bom.gov.au/waterdata; Zhang et al., 2013; 776 catchments) and (vii) the Chilean Center for Climate and Resilience Research (CR2) website (https://www.cr2.cl/recursos-y-publicaciones/bases-de-datos/datos-de-caudales) and the CAMELS-CL data set (Alvarez-Garreton et al., 2018; 531 catchments).

Unsuitable catchments were identified using the following criteria. First, GRDC catchments without daily streamflow data were discarded. Second, "nonreference" GAGES-II catchments were discarded. Third, catchments smaller than 50 km^2 were discarded, to ensure that each catchment covers at least two 0.05° grid cells. Fourth, catchments larger than 5,000 km^2 were discarded, as channel routing effects become apparent at the daily time scale in larger catchments (Gericke & Smithers, 2014). Fifth, catchments with less than 5 yr of streamflow data during 2000–2016 (the calibration period) were discarded. Sixth, we discarded catchments with potentially erroneous streamflow data (identified through visual screening). The final database used for the regionalization comprised 4,229 catchments (median size 695 km^2; Figure 2).

2.3. Climate Groups and Cross Validation

To reduce the computational load, minimize the possibility of underfitting, and reduce the heterogeneity among catchments, we subdivided the land surface and the catchment set into three groups based on the Köppen-Geiger (KG) classification (Beck et al., 2018): (i) tropical (Class A; 566 catchments); (ii) arid and temperate (Classes B and C, respectively; 1,802 catchments); and (iii) cold and polar (Classes D and E,

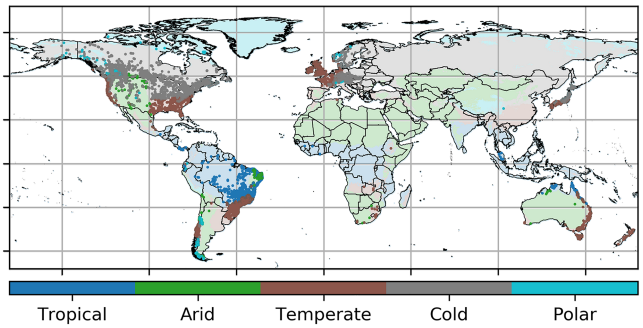

Figure 2. The dominant Köppen-Geiger climate classes of the 4,229 catchments.

respectively; 1,861 catchments; Figure 2). The arid and polar classes were represented by a small number of catchments (122 and 150, respectively) and were therefore merged with the temperate and cold classes, respectively. The dominant class of each catchment was used to subdivide the catchment set.

The transfer equation coefficients (Equation 1) were optimized for each of the three climate groups separately. For the optimization, we used tenfold cross validation, allowing us to (i) estimate the generalizability of the derived parameters, and (ii) obtain an ensemble of parameter maps, the spread of which provides an indication of the uncertainty. For each cross-validation iteration, the catchment set was partitioned into subsets of 90% for calibration and 10% for validation. The partitioning was random with each catchment being used only once for validation.

2.4. Hydrological Model and Meteorological Forcing

While the parameter regionalization approach can be applied to any hydrological model, we tested it using a gridded implementation of the HBV model (Bergström, 1976, 1992; Seibert & Vis, 2012). HBV was used because of its low complexity, high agility, and computational efficiency. Additionally, the model has been successfully used in numerous studies spanning a wide range of climate and physiographic conditions (e.g., Beck, Bruijnzeel, et al. 2013; Breuer et al., 2009; Deelstra et al., 2010; Demirel et al., 2015; Driessen et al., 2010; Jódar et al., 2018; Plesca et al., 2012; Steele-Dunne et al., 2008; Te Linde et al., 2008; Vetter et al., 2015), including several parameter regionalization studies (e.g., Bárdossy, 2007; Beck et al., 2016; Booij, 2005; Jin et al., 2009; Hundecha & Bárdossy, 2004; Masih et al., 2010; Merz & Blöschl, 2004; Parajka et al., 2005, 2007; Seibert, 1999). The model runs at a daily time step, has one unsaturated zone store, two groundwater stores, and 12 free parameters (Table 2).

Transfer equations were established for all 12 model parameters, resulting in $12 \times 9 = 108\,a$ coefficients that required optimization (Equation 1). We thus calibrated all model parameters and linked each to all eight predictors, to maximize the model agility (Mendoza et al., 2015). This was computationally feasible and did not lead to overfitting issues (i.e., low validation scores compared to calibration) because of the large sample of catchments used. We do not expect that all coefficients are well constrained after the optimization, as some model parameters might be "insensitive" (i.e., have little influence on the simulated runoff). See Abebe et al. (2010) and Zelelew and Alfredsen (2012) for sensitivity analyses of HBV.

The model requires daily time series of precipitation, potential evaporation, and air temperature as input. For precipitation, we used the gauge-, satellite-, and reanalysis-based MSWEP data set (V2.2; 0.1° resolution; 1979–present; www.gloh2o.org/mswep; Beck, van Dijk, Levizzani, et al., 2017; Beck, Wood, et al., 2019). MSWEP was chosen for its superior performance in numerous precipitation data set evaluation studies (e.g., Alijanian et al., 2017; Bai & Liu, 2018; Beck, Pan, et al., 2019; Beck, Vergopolan, et al., 2017; Casson et al., 2018; Sahlu et al., 2017; Satgé et al., 2019; Zhang et al., 2019). The precipitation data were corrected for gauge

Table 2
Parameters of the HBV Model and Their Permissible Ranges

Parameter	Description	Range
BETA	Shape coefficient of recharge function	1 to 6
FC	Maximum soil moisture storage (mm)	50 to 1,000
K0	Recession coefficient of upper zone (day^{-1})	0.05 to 0.9
K1	Recession coefficient of upper zone (day^{-1})	0.01 to 0.5
K2	Recession coefficient of lower zone (day^{-1})	0.001 to 0.2
LP	Soil moisture value above which actual evaporation reaches potential evaporation	0.2 to 1
PERC	Maximum percolation to lower zone ($mm\ day^{-1}$)	0 to 10
UZL	Threshold parameter for extra outflow from upper zone (mm)	0 to 100
TT	Threshold temperature (°C)	−2.5 to 2.5
CFMAX	Degree-day factor ($mm\ °C^{-1}\ day^{-1}$)	0.5 to 10
CFR	Refreezing coefficient	0 to 0.1
CWH	Water holding capacity	0 to 0.2

undercatch and orographic effects using the PBCOR WorldClim V2 data set (Beck et al., 2020; www.gloh2o.org/pbcor) and downscaled to 0.05° using nearest neighbor resampling. Potential evaporation was estimated using the Hargreaves (1994) equation from daily minimum and maximum air temperature. Temperature estimates were obtained by averaging two reanalysis data sets, ERA-Interim (~0.75° resolution; 1979–present Dee et al., 2011; https://cds.climate.copernicus.eu) and JRA-55 (~0.56° resolution; 1959–present Kobayashi et al., 2015; https://jra.kishou.go.jp). Prior to the averaging, both data sets were downscaled to 0.05° and bias corrected on a monthly basis through an additive approach using the comprehensive station-based WorldClim V2 climatology (1-km resolution Fick & Hijmans, 2017; www.worldclim.org).

We only ran the model for the period 1990–2016 rather than for the entire period of forcing data availability (1979–2016), to reduce the computational demand and to take advantage of the better quality of the precipitation data after 2000 (Beck, Wood, et al., 2019); the first 10 yr of the record (1990–1999) were used only to initialize the model stores.

2.5. Performance Metric

For each catchment, we calculated scores of a transformed version of the Kling-Gupta efficiency (KGE_B) from daily time series of observed streamflow and simulated streamflow (obtained by aggregating the simulated runoff; Figure 1) for the period 2000–2016. The original (untransformed) KGE is an objective performance metric combining correlation, bias, and variability introduced by Gupta et al. (2009) and modified by Kling et al. (2012). It is defined according to

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \tag{2}$$

where the correlation component r is represented by Pearson's correlation coefficient, the bias component β by the ratio of estimated and observed means, and the variability component γ by the ratio of the estimated and observed coefficients of variation:

$$\beta = \frac{\mu_s}{\mu_o} \quad \text{and} \quad \gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}, \tag{3}$$

where μ and σ represent the distribution mean and standard deviation, respectively, and the subscripts s and o indicate estimate and reference, respectively. A drawback of the KGE is that it does not have a lower limit. Accordingly, the mean over a large sample of catchments can be dominated by a single catchment with an extremely low KGE value. To avoid this, we applied the following transformation (Mathevet et al., 2006):

$$\mathrm{KGE}_B = \frac{\mathrm{KGE}}{2 - \mathrm{KGE}}. \tag{4}$$

The resulting KGE_B values range from −1 to 1, with higher values indicating better performance.

2.6. Optimization Algorithm

For all three climate group and all 10 cross-validation iterations, we optimized the coefficients of the transfer equations using the $(\mu + \lambda)$ evolutionary algorithm implemented using the Distributed Evolutionary Algorithms in Python (DEAP) toolkit (Fortin et al., 2012). The population size (μ) was set at 16 and the recombination pool size (λ) at 32. Each generation produced λ offspring from the population. Offspring were evaluated after which the population of the next generation was selected from both offspring and population. Crossover and mutation probabilities were set at 0.9 and 0.1, respectively. The number of generations was set at 25, as this was found to be sufficient for achieving convergence. This resulted in $3 \times 10 \times 25 \times 32 = 24{,}000$ gridded model evaluations.

2.7. Local Calibration

To obtain an upper limit of what is feasible using the combination of meteorological forcing, observed streamflow, and catchment boundary data (Seibert et al., 2018), we calibrated HBV on a catchment-by-catchment basis against locally observed streamflow data. For this purpose, we ran the model in a lumped (i.e., nondistributed) fashion with catchment-mean meteorological forcing data for each of the 4,229 catchments. The first half of the record of simultaneous forcing and observed streamflow data (1979–2016) was used for validation and the second half for calibration. The stores were initialized by running the model

for the first 10 yr of the record if the record length was 10 yr or by running the model twice for the entire record if the record length was <10 yr. The KGE was used as objective function, and the $(\mu + \lambda)$ evolutionary algorithm was used for the optimization, with μ set to 20, λ set to 40, and the number of generations set to 12 (resulting in $4{,}229 \times 20 \times 40 \times 12 = 40.6$ million lumped model evaluations).

3. Results and Discussion

3.1. Model Performance

Figure 3 presents the streamflow simulation performance obtained using uncalibrated parameters (randomly generated in the first generation of the optimization process), regionalized parameters for the independent evaluation catchments, and local calibration for the validation period. The median daily KGE using uncalibrated parameters over all catchments was 0.17 (Figure 3c), whereas the median daily KGE obtained using regionalized parameters for the calibration catchments was 0.46. The median daily KGE obtained using regionalized parameters for the independent validation catchments was also 0.46 (Figure 3b), suggesting that the approach generalizes well to catchments that were not used for the optimization and hence to ungauged regions. The difference in median daily KGE between uncalibrated parameters (0.17) and regionalized parameters (0.46) was thus 0.29, and improvements were obtained for 88% of the validation catchments (Figure 3a). These results confirm the efficacy of our regionalization approach in improving the streamflow simulation performance. Our median daily KGE using uncalibrated parameters of 0.17 (Figure 3c) exceeds the range of median daily KGE values of −0.25 to 0.13 obtained by Beck et al. (2016) for a diverse set of 10 uncalibrated models in catchments comparable to ours, suggesting that we are not overestimating the benefit of the regionalization.

The median daily KGE obtained using local calibration was 0.77 for the calibration period and 0.69 for the validation period (Figure 3d). The difference in median KGE between regionalized parameters (0.46) and locally calibrated parameters (0.69) was thus 0.23. Such a difference is to be expected, not least because scores obtained using local calibration compensate for local errors in meteorological forcing (Beck, Pan, et al., 2019; Beck, Vergopolan, et al., 2017), observed streamflow (Di Baldassarre & Montanari, 2009; McMillan et al., 2010), and catchment boundary data (Kauffeldt et al., 2013; Lehner, 2012).

Our local calibration scores (Figure 3d) were similar to or higher than those from previous studies, suggesting that our meteorological forcing data are of good quality (in agreement with several precipitation data set evaluations; e.g., Beck, Pan, et al., 2019; Beck, Vergopolan, et al., 2017) and that our model setup is sufficiently agile (i.e., capable of simulating a wide range of catchment behavior; Mendoza et al., 2015). Alfieri et al. (2020), for example, calibrated the LISFLOOD model for 1,226 catchments globally (mean size 42,000 km^2) and obtained median daily KGE values of 0.67 and 0.61 for the calibration and validation periods, respectively, while we obtained higher median scores of 0.77 and 0.69, respectively (Figure 3d), despite the much smaller size of our catchments (mean size 1,165 km^2). Filipova and Leedal (2018) calibrated the IHACRES model for ~3,000 catchments globally similar in size to ours and obtained a median daily calibration KGE of 0.60, considerably lower than our median score of 0.77. For the conterminous United States, we obtained a median daily validation KGE of 0.73, whereas Mizukami et al. (2019) obtained median scores of 0.63 and 0.74 for the VIC and mHM models, respectively, and Gao et al. (2019) obtained median scores of 0.62 and 0.61 for HBV and TOPMODEL, respectively.

To our knowledge, only seven previous regionalization studies had a global scope (Arheimer et al., 2019; Beck et al., 2016; Döll et al., 2003; Filipova & Leedal, 2018; Nijssen et al., 2001; Van Dijk et al., 2013; Widén-Nilsson et al., 2007). However, none of these studies accounted for the within-catchment heterogeneity in landscape and climate (Kling & Gupta, 2009; Rouholahnejad-Freund et al., 2019) or the scale discrepancy between catchments and grid cells (Becker & Braun, 1999; Blöschl & Sivapalan, 1995). Additionally, three studies (Döll et al., 2003; Nijssen et al., 2001; Widén-Nilsson et al., 2007) only used observed streamflow from large catchments (≫10,000 km^2) in which routing effects tend to dominate at the daily time scale (Gericke & Smithers, 2014), while one study (Widén-Nilsson et al., 2007) used a regionalization approach based on spatial proximity despite a lack of gauges in many regions across the globe (Fekete & Vörösmarty, 2007; Hannah et al., 2011; Sivapalan, 2003). Furthermore, with the exception of Beck et al. (2016), Filipova and Leedal (2018), and Arheimer et al. (2019), these studies used observed

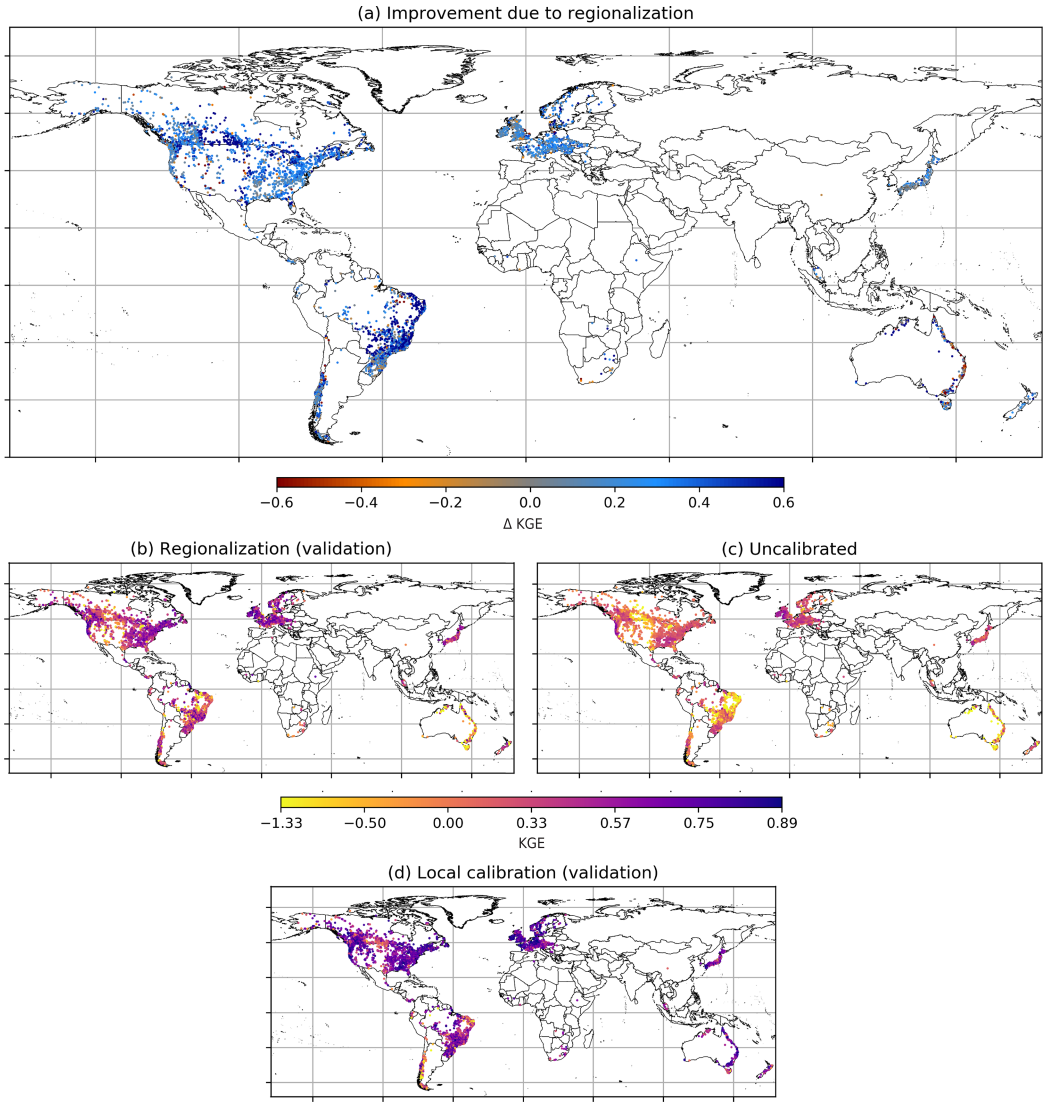

Figure 3. (a) The improvement in KGE after regionalization calculated as the difference between (b) KGE obtained for the independent validation catchments using regionalized parameters and (c) KGE obtained for the first generation of the optimization process (i.e., using uncalibrated parameters). (d) KGE obtained for the validation period using parameters calibrated against locally observed streamflow. Each data point represents a catchment centroid ($N = 4{,}229$).

streamflow from a relatively small number of catchments (9 to 311) and did not evaluate the performance of the regionalized parameters in independent validation catchments.

In probably the most similar previous global regionalization study to date, Beck et al. (2016) produced parameter maps (0.5° resolution) for HBV using a regionalization approach based on climatic and physiographic similarity. Although they used another performance metric to optimize their parameters, they reported a median daily KGE improvement using regionalized parameters of 0.08, considerably lower than our median improvement of 0.29 (Figure 3a). Furthermore, they reported performance improvements for 79% of the validation catchments using the performance metric that they used for the optimization, substantially less than the 88% obtained by us (Figure 3a). This suggests that our new regionalization approach provides a better generalization capability.

Filipova and Leedal (2018) used an approach based on climatic and physiographic similarity to regionalize parameters of the IHACRES model for ~3,000 catchments globally similar in size to ours. They obtained a median daily KGE of 0.40 using leave-one-out cross validation, which is lower than our median score of 0.46 (Figure 3b).

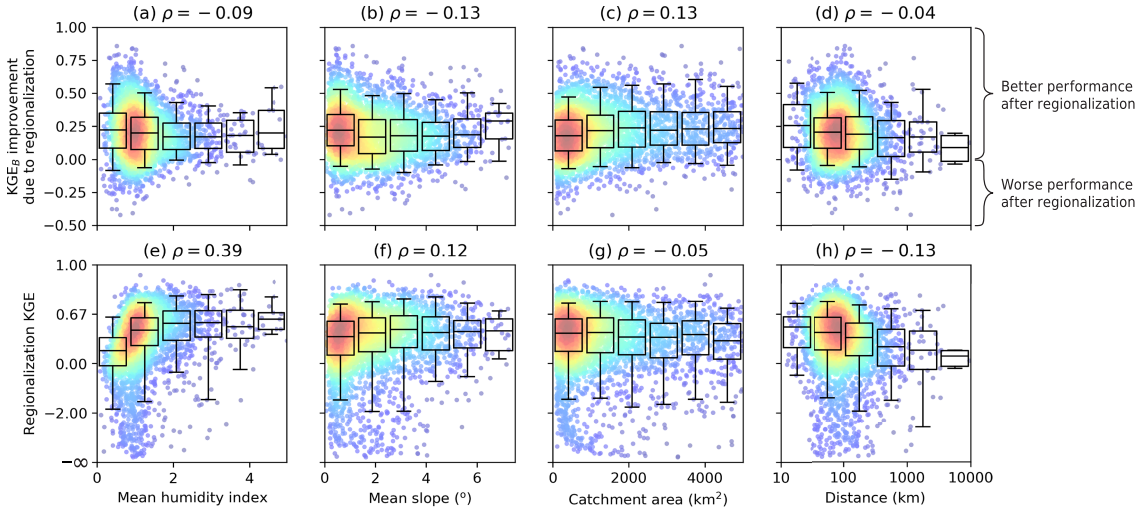

Figure 4. Relationships between catchment characteristics and (a–d) KGE_B improvement values (scores using regionalized parameters minus scores using uncalibrated parameters) and (e–h) regionalization KGE values (obtained for the independent validation catchments). The uncalibrated scores were calculated by averaging the scores obtained in the first generation of the optimization process. "Distance" is the mean distance to the 10 closest catchments used for the optimization. ρ denotes Spearman's rank correlation coefficient. Each data point represents a catchment ($N = 4{,}229$). The boxes indicate the 25th and 75th percentiles, the line across the box indicates the median, and the whiskers indicate the 5th and 95th percentiles.

Arheimer et al. (2019) produced global parameter maps for the HYPE hydrological model by jointly optimizing the parameters of groups of catchments. They obtained a median monthly KGE of 0.40 for 2,863 large catchments ($\gg$1,000 km^2), whereas we obtained a median daily KGE of 0.46 for 4,229 small- to medium-sized catchments (median area 695 km^2; Figure 3b). Thus, their median score is lower than ours, despite the fact that simulation performance might be expected to be better at monthly resolution (Beck et al., 2016; Xia et al., 2012; Zink et al., 2017) and in large catchments, where (i) local forcing errors and landscape heterogeneity may be averaged out, (ii) the hydrograph is smoother due to channel routing, and (iii) the catchment delineation is likely more accurate (Merz et al., 2009; Parajka et al., 2013; Rakovec et al., 2016).

3.2. Factors Determining the Performance

The improvement in KGE using regionalized parameters relative to uncalibrated parameters showed no clear relationship with catchment-mean humidity index (Figure 4a) or topographic slope (Figure 4b), suggesting that the benefit of the regionalized parameters does not depend on climate or topography. Conversely, there was a weak positive relationship between KGE improvement values and catchment area (Figure 4c), suggesting that larger catchments benefit more from the regionalization approach. This could be because larger catchments aggregate the runoff over multiple grid cells, potentially canceling out the parameter and forcing errors present in individual grid cells. In a similar vein, several previous studies obtained better performance by averaging the outputs from multiple parameter sets than by taking the output of a single parameter set (e.g., Bao et al., 2012; Beck et al., 2016; Garambois et al., 2015; McIntyre et al., 2004, 2005; Oudin et al., 2008; Reichl et al., 2009; Viney et al., 2009; Zhang & Chiew, 2009). A weak negative relationship was found between KGE improvement values and "distance" (defined as the mean distance to the 10 closest catchments used for the optimization; Figure 4d), suggesting that the regionalization approach provides slightly less (but still substantial) benefit in poorly gauged regions.

We found a strong positive correlation (Spearman's rank correlation coefficient $\rho = 0.39$) between regionalization KGE values and catchment-mean humidity index (Figure 4e), indicating that model performance tends to be better in humid regions, confirming numerous studies (e.g., Arheimer et al., 2019; Beck et al., 2016; Beck, van Dijk, et al., 2017; Beck, Vergopolan, et al., 2017; Essou et al., 2016; Newman et al., 2015; Parajka et al., 2013) and previously attributed to the localized short-lived convective rainfall, the high evaporative losses, and the nonlinear rainfall-runoff relationship (Pilgrim et al., 1988; Ye et al., 1997). Conversely, regionalization KGE values were not clearly related to catchment-mean topographic slope (Figure 4f). The slightly negative relationship between regionalization KGE values and catchment area

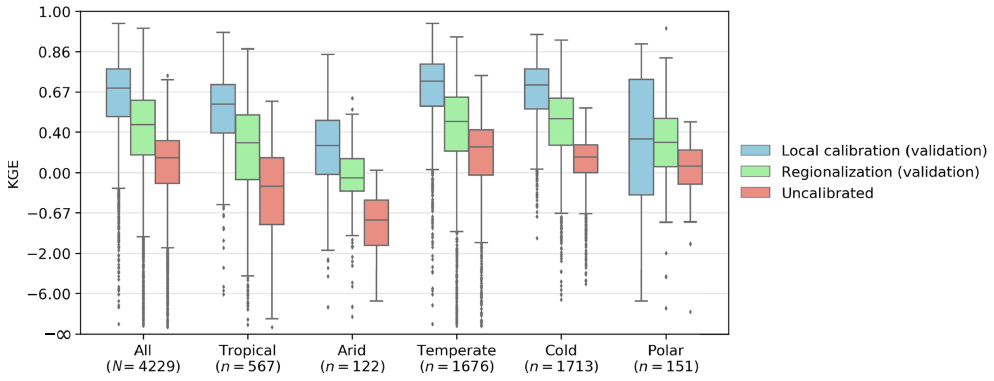

Figure 5. Box-and-whisker plots of KGE values obtained using local calibration, regionalization, and without calibration for all catchments and for the five major Köppen-Geiger climate classes. The local calibration scores represent the validation period, whereas the regionalization scores represent the validation catchments. The uncalibrated scores were calculated by averaging the scores obtained in the first generation of the optimization process. The lines in each box represent the median value, the bottom and top edges of the boxes represent the 25th and 75th percentile values, respectively, while the "whiskers" represent the 5th and 95th percentile values. The catchments were grouped based on the dominant Köppen-Geiger climate class. *n* denotes the number of catchments in each group.

(Figure 4g) is contrary to previous studies (e.g., Gericke & Smithers, 2014; Parajka et al., 2013) and is probably partly due to the more arid nature of the larger catchments in our data set ($\rho = -0.19$ between catchment area and catchment-mean humidity index), and partly because we did not use a routing model to account for streamflow delays in larger catchments. Similarly, the negative relationship between regionalization KGE values and distance (Figure 4h) likely reflects the fact that arid regions tend to be more poorly gauged ($\rho = -0.17$ between distance and catchment-mean humidity index).

Among the five Köppen-Geiger climate classes, the lowest KGE values were obtained for the arid class, and the highest for the temperate and cold classes (Figure 5), similar to previous large catchment sample studies (Beck et al., 2016; Beck, van Dijk, et al., 2017; Beck, Vergopolan, et al., 2017). The good performance in temperate regions is attributable to the relative simplicity of the hydrological response and the dense precipitation measurement network (Kidd et al., 2017; Schneider et al., 2014), while in cold regions it is likely attributable to the smoothly varying seasonal cycle of streamflow and the predictability of frontal weather systems (Beck, Vergopolan, et al., 2017; Ebert et al., 2007). The somewhat greater spread among KGE values for the tropical class (Figure 5) may reflect the large variability in precipitation measurement network density across the tropics (Kidd et al., 2017; Schneider et al., 2014). The slightly greater spread among regionalization KGE values than among uncalibrated KGE values (Figure 5) suggests that catchments for which the simulation performance was poor prior to model parameter optimization (e.g., due to model structural or forcing data deficiencies) derived less benefit overall from our regionalization approach. The substantial spread among KGE values obtained using local calibration for the polar class (Figure 5) likely reflects the inability of lumped models to simulate streamflow in mountainous, snowmelt-dominated catchments.

3.3. Spatial Patterns of the Regionalized Parameters

Figure 6 presents maps of four key HBV model parameters (BETA, FC, K2, and PERC; Table 2) derived using the new regionalization approach. Our maps vary according to landscape and climate characteristics for the entire global land surface including ungauged regions. Although many studies have optimized HBV parameters (e.g., Bárdossy, 2007; Beck et al., 2016; Booij, 2005; Hundecha & Bárdossy, 2004; Jin et al., 2009; Masih et al., 2010; Merz & Blöschl, 2004; Parajka et al., 2005, 2007; Seibert, 1999), only two have actually published maps of their optimized parameters to which we can compare our regionalized parameter maps (Beck et al., 2016; Merz & Blöschl, 2004). The maps of Merz and Blöschl (2004, their Figures 4–7), derived by calibrating HBV for 308 Austrian catchments individually, exhibit reasonable agreement with ours (Figure 6): For example, both exhibit lower FC (maximum soil moisture storage) in the northeastern lowlands of Austria and higher BETA (shape coefficient of recharge function) in the mountainous west. The maps of Beck et al. (2016, their Figure 4), derived by optimizing parameters for each catchment individually and transferring them to "similar" 0.5° grid cells globally, also exhibit reasonable agreement with ours

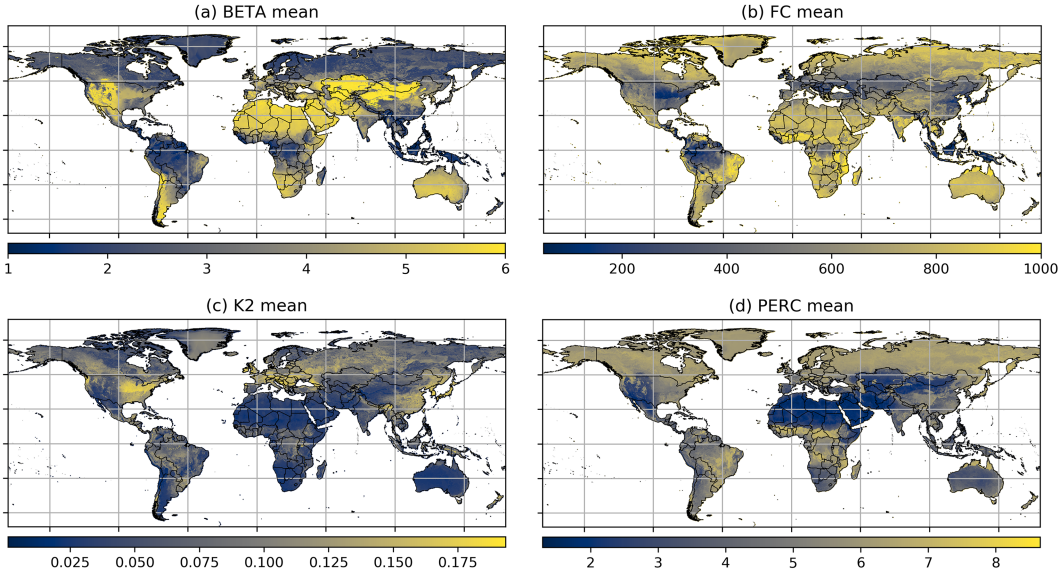

Figure 6. Global ensemble-mean maps of four key HBV model parameters (Table 2) derived using the new regionalization approach. The ensemble comprises 10 members obtained via tenfold cross validation.

(Figure 6). However, both studies optimized the parameters for the catchments individually, and therefore their parameter maps lack spatial coherence. Conversely, we optimized transfer equation coefficients in a spatially distributed fashion at high resolution for large groups of catchments jointly, resulting in parameter maps with high spatial coherence.

However, it can be difficult to explain the spatial patterns of the regionalized parameters, due to identifiability issues (Sorooshian & Gupta, 1983) and parameter interactions (Gupta & Sorooshian, 1983). Yet, the higher BETA in arid regions (Figure 6a) reduces the recharge rate to the groundwater stores, which is in line with global recharge assessments based on conceptual (Döll & Fiedler, 2008) and statistical (Mohan et al., 2018) models. The patterns in K2 (recession coefficient of lower zone; Figure 6c) can be explained by the sand content of the soil (Hengl et al., 2017), which is consistent with the notion that the permeability of the soil has a strong impact on baseflow (Hewlett, 1961; Price, 2011; Zecharias & Brutsaert, 1988). The lower PERC (maximum percolation to lower zone) in arid regions (Figure 6d) reduces the flow of water from the upper groundwater store (which has a quick outflow) to the lower groundwater store (which has a slow outflow) and hence results in more rapidly receding runoff, consistent with baseflow recession assessments using observed streamflow for the pantropics (Peña-Arancibia et al., 2010) and the globe (Beck et al., 2015; Beck, van Dijk, et al., 2013). The patterns in FC (Figure 6b) are less easily interpretable; the low FC in tropical regions is counterintuitive given the thick regolith cover (Bonell, 2005). It may reflect parameter compensation behavior within the model to reduce evaporation and increase runoff.

4. Conclusion

Previous regionalization approaches generally ignored the within-catchment variability in climate and landscape, did not optimize the model parameters for all catchments jointly, failed to evaluate the performance in independent validation catchments, and had a regional focus and thus questionable generalizability. To overcome these limitations, we introduced a novel regionalization approach and implemented it for a gridded version of the HBV model using a uniquely large database of daily streamflow data for 4,229 catchments worldwide. An ensemble of high-resolution (0.05°) parameter maps covering the entire land surface including ungauged regions was derived. Our findings can be summarized as follows:

1. The regionalized parameters yielded, for the independent validation catchments, a median daily KGE of 0.46. The median KGE improvement due to the regionalization (relative to uncalibrated parameters) for the independent validation catchments was 0.29, with improvements obtained for 88% of the independent validation catchments. Our scores compare favorably to those from previous continental- and global-scale studies, confirming the effectiveness of our approach in improving streamflow simulation performance.

2. The performance improvement due to the regionalized parameters did not depend on climate or topography. Substantial improvements were obtained even for independent validation catchments located far away from the catchments used for optimization, highlighting the value of the approach for poorly gauged regions. The streamflow simulation performance was worst in arid regions and best in temperate and cold regions, in agreement with several previous large catchment sample studies. The spread in performance was greatest among tropical catchments.
3. In contrast to conventional catchment-by-catchment calibration approaches, the new regionalization approach yields parameters that vary according to landscape and climate characteristics for the entire land surface including ungauged regions. The obtained parameter maps exhibit reasonable agreement with those from two previous studies. We were able to interpret the spatial patterns of most of the regionalized parameters based on hydrological process understanding.

Data Availability Statement

The HBV parameter maps and the optimized transfer equation coefficients can be downloaded online (via www.gloh2o.org/hbv). The majority of the streamflow, meteorological forcing, and predictor data can be obtained via the URLs listed in sections 2.2 and 2.4 and Table 1, respectively.

Acknowledgments

The following organizations are thanked for providing observed streamflow data: the United States Geological Survey (USGS), the Global Runoff Data Centre (GRDC), the Brazilian Agência Nacional de Águas, EURO-FRIEND-Water, the Water Survey of Canada (WSC), the Australian Bureau of Meteorology (BoM), and the Chilean Chilean Center for Climate and Resilience Research (CR2). We are thankful to Ross Woods for helpful comments on an earlier draft of this paper. The editor and two of the three anonymous reviewers are thanked for constructive comments which helped us to improve the paper. Hylke Beck was supported in part by the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM), under the auspices of UNESCO.

References

Abdulla, F. A., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology*, *197*(1–4), 230–257.

Abebe, N. A., Ogden, F. L., & Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *Journal of Hydrology*, *389*(3), 301–310.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*, 8792–8812. https://doi.org/10.1029/2018WR022606

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X*, *6*, 100049.

Alijanian, M., Rakhshandehroo, G. R., Mishra, A. K., & Dehghani, M. (2017). Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran. *International Journal of Climatology*, *37*, 4896–4914.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies—Chile dataset. *Hydrology and Earth System Sciences*, *22*(11), 5817–5846.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2019). Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. *Hydrology and Earth System Sciences Discussions*, *2019*, 1–34.

Bai, P., & Liu, X. (2018). Evaluation of five satellite-based precipitation products in two gauge-scarce basins on the Tibetan Plateau. *Remote Sensing*, *10*, 1316.

Bajracharya, S. R., Shrestha, M. S., & Shrestha, A. B. (2017). Assessment of high-resolution satellite rainfall estimation products in a streamflow model for flood prediction in the Bagmati basin, Nepal. *Journal of Flood Risk Management*, *10*(1), 5–16.

Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., et al. (2012). Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *Journal of Hydrology*, *466–467*(1), 37–46.

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, *11*, 703–710.

Bastola, S., Ishidaira, H., & Takeuchi, K. (2008). Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe. *Journal of Hydrology*, *357*(3), 188–206.

Beck, H. E., Bruijnzeel, L. A., van Dijk, A. I. J. M., McVicar, T. R., Scatena, F. N., & Schellekens, J. (2013). The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments. *Hydrology and Earth System Sciences*, *17*(7), 2613–2635.

Beck, H. E., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garret, C., Baez-Villanueva, O. M., Sheffield, J., et al. (2020). Bias correction of global precipitation climatologies using discharge observations from 9372 catchments. *Journal of Climate*, *33*, 1299–1315.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*, 207–224.

Beck, H. E., van Dijk, A. I. J. M., & de Roo, A. (2015). Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology*, *16*(4), 1478–1501.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, *21*(6), 2881–2903.

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

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). 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.

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., de Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., & Schellekens, J. (2013). Global patterns in baseflow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, *49*, 7843–7863. https://doi.org/10.1002/2013WR013918

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. M., 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.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(180214).

Becker, A., & Braun, P. (1999). Disaggregation, aggregation and spatial scaling in hydrological modelling. *Journal of Hydrology*, *217*(3), 239–252.

Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments (PhD thesis, SMHI Reports RHO No. 7), Swedish Meteorological and Hydrological Institute (SMHI), Norköping, Sweden.

Bergström, S. (1992). The HBV model—Its structure and applications (*SMHI Reports RH No. 4*). Norrköping, Sweden: Swedish Meteorological and Hydrological Institute (SMHI).

Beven, K. J. (1989). Changing ideas in hydrology—The case of physically-based models. *Journal of Hydrology*, *105*(1–2), 157–172.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36.

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (Eds.) (2013). *Runoff prediction in ungauged basins: Synthesis across processes, places and scales Edited by* Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. New York, US: Cambridge University Press.

Bonell, M. (2005). Runoff generation in tropical forests. In M. Bonell, & L. A. Bruijnzeel (Eds.), *Forests, water and people in the humid tropics* (pp. 314–406). Cambridge, UK: Cambridge University Press.

Booij, M. J. (2005). Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of Hydrology*, *303*(1–4), 176–198.

Boughton, W., & Chiew, F. (2007). Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model. *Environmental Modelling & Software*, *22*(4), 476–487.

Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., et al. (2009). Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. *Advances in Water Resources*, *32*(2), 129–146.

Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, *13*(3), 637–643.

Casson, D. R., Werner, M., Weerts, A., & Solomatine, D. (2018). Global re-analysis datasets to improve hydrological assessment and snow water equivalent estimation in a sub-Arctic watershed. *Hydrology and Earth System Sciences*, *22*(9), 4685–4697.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Part A*, *137*(656), 553–597.

Deelstra, J., Farkas, C., Engebretsen, A., Kværnø, S., Beldring, S., Olszewska, A., & Nesheim, L. (2010). Can we simulate runoff from agriculture dominated watersheds? Comparison of the DrainMod, SWAT, HBV, COUP and INCA models applied for the Skuterud catchment. *Bioforsk FOKUS*, *5*(6), 119–128.

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2015). The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models. *Hydrology and Earth System Sciences*, *19*(1), 275–291.

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*(6), 913–921.

Döll, P., & Fiedler, K. (2008). Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences*, *12*(3), 863–885.

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology*, *270*(1), 105–134.

Donohue, R. J., Roderick, M. L., & McVicar, T. R. (2007). On the importance of including vegetation dynamics in Budyko's hydrological model. *Hydrology and Earth System Sciences*, *11*(2), 983–995.

Driessen, T. L. A., Hurkmans, R. T. W. L., Terink, W., Hazenberg, P., Torfs, P. J. J. F., & Uijlenhoet, R. (2010). The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model. *Hydrology and Earth System Sciences*, *14*(4), 651–665.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., et al. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1), 3–17.

Duan, Q., Schaake, J., & Koren, V. (2001). A priori estimation of land surface model parameters. In V. Lakshmi, J. Albertson, & J. Schaake (Eds.), *Land surface hydrology, meteorology, and climate: Observations and modeling*(pp. 77–94), *Water Science and Application*. Washington, D.C., US: AGU.

Ebert, E. E., Janowiak, J. E., & Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, *88*(1), 47–64.

Essou, G. R. C., Arsenault, R., & Brissette, F. P. (2016). Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology*, *537*, 334–345.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, *91*(2), 621.

Fekete, B. M., & Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. In *Predictions in Ungauged Basins: PUB kick-off (proceedings of the PUB kick-off meeting held in Brasilia, 20–22 November 2002), IAHS publication no. 309*, pp. 129–136. Brazil.

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.

Filipova, V., & Leedal, D. (2018). A physical similarity approach to regionalisation using a global database of catchments. 20th EGU General Assembly, EGU 2018 https://ui.adsabs.harvard.edu/abs/2018EGUGA..20.8808F/abstract

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*, 2171–2175.

Gao, H., Birkel, C., Hrachowitz, M., Tetzlaff, D., Soulsby, C., & Savenije, H. H. G. (2019). A simple topography-driven and calibration-free runoff generation module. *Hydrology and Earth System Sciences*, *23*(2), 787–809.

Garambois, P. A., Roux, H., Larnier, K., Labat, D., & Dartus, D. (2015). Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology*, *525*, 383–399.

Gentine, P., D'Odorico, P., Litner, B. R., Sivandran, G., & Salvucci, G. (2012). Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. *Geophysical Research Letters*, *39*, L19404. https://doi.org/10.1029/2012GL053492

Gericke, O. J., & Smithers, J. C. (2014). Review of methods used to estimate catchment response time for the purpose of peak discharge estimation. *Hydrological Sciences Journal*, *59*(11), 1935–1971.
Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *370*(1–2), 80–91.
Gupta, V. K., & Sorooshian, S. (1983). Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined. *Water Resources Research*, *19*(1), 269–276.
Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P. (2013). Advances in automatic calibration of watershed models, *Calibration of watershed models* (pp. 9–28): American Geophysical Union (AGU). https://doi.org/10.1029/WS006p0009
Hannah, D. M., Demuth, S., Van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200.
Hargreaves, G. H. (1994). Defining and using reference evapotranspiration. *Journal of Irrigation and Drainage Engineering*, *120*(6), 1132–1139.
Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*, 1–40.
Heuvelmans, G., Muys, B., & Feyen, J. (2006). Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. *Journal of Hydrology*, *319*(1), 245–265.
Hewlett, J. D. (1961). Soil moisture as a source of baseflow from steep mountain watersheds (*Forest Experimental Station Research Paper*): US Forest Service Southeast.
Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J. (2018). Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, *566*, 595–606.
Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., et al. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, *58*(6), 1198–1255.
Huang, S., Eisner, S., Magnusson, J. O., Lussana, C., Yang, X., & Beldring, S. (2019). Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway. *Journal of Hydrology*, *577*, 23585.
Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1–4), 281–295.
Jin, X., Xu, C., Zhang, Q., & Chen, Y. D. (2009). Regionalization study of a conceptual hydrological model in Dongjiang basin, south China. *Quaternary International*, *208*(1–2), 129–137.
Jódar, J., Carpintero, E., Martos-Rosillo, S., Ruiz-Constán, A., Marín-Lechado, C., Cabrera-Arrabal, J. A., et al. (2018). Combination of lumped hydrological and remote-sensing models to evaluate water resources in a semi-arid high altitude ungauged watershed of Sierra Nevada (southern Spain). *Science of The Total Environment*, *625*, 285–300.
Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., & Westerberg, I. K. (2013). Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, *17*(7), 2845–2013.
Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78.
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.
Kling, H., & Gupta, H. (2009). On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. *Journal of Hydrology*, *373*(3), 337–351.
Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan Series I*, *93*, 5–48.
Kokkonen, T. S., Jakeman, A. J., Young, P. C., & Koivusalo, H. J. (2003). Predicting daily flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes*, *17*(11), 2219–2238.
Lehner, B. (2012). Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network (*Tech. Rep. No. 41*): Global Runoff Data Centre (GRDC), Federal Institute of Hydrology (BfG), Koblenz, Germany.
Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22.
Livneh, B., & Lettenmaier, D. P. (2013). Regional parameter estimation for the unified land model. *Water Resources Research*, *19*, 100–114. https://doi.org/10.1029/2012WR012220
Maisongrande, P., Duchemin, B., & Dedieu, G. (2004). VEGETATION/SPOT: An operational mission for the Earth monitoring; presentation of new standard products. *International Journal of Remote Sensing*, *25*(1), 9–14.
Masih, I., Uhlenbrook, S., Maskey, S., & Ahmad, M. D. (2010). Regionalization of a conceptual rainfall-runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran. *Journal of Hydrology*, *391*(1–2), 188–201.
Mathevet, T., Andréassian, C. M. V., & Perrin, C. (2006). A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, *Large sample basin experiments for hydrological model parameterization: Results of the Model Parameter Experiment-MOPEX* (pp. 211–219). Wallingford, UK: IAHS Press.
McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, *43*, W07301. https://doi.org/10.1029/2006WR005467
McIntyre, N. R., Lee, H., Wheater, H. S., & Young, A. R. (2004). Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments, *Complexity and integrated resources management*. Osnabrueck, Germany.
McIntyre, N. R., Lee, H., Wheater, H. S., Young, A., & Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, *41*, W12434. https://doi.org/10.1029/2005WR004289
McMillan, H., Freer, J., Pappenberger, F., Krueger, T., & Clark, M. (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, *24*(10), 1270–1284.
Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, *51*, 716–728. https://doi.org/10.1002/2014WR015820
Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, *287*(1–4), 95–123.
Merz, R., Parajka, J., & Blöschl, G. (2009). Scale effects in conceptual hydrological modeling. *Water Resources Research*, *45*, W09405. https://doi.org/10.1029/2009WR007872
Mishra, V., Shah, R., Azhar, S., Shah, H., Modi, P., & Kumar, R. (2018). Reconstruction of droughts in India using multiple land-surface models (1951–2015). *Hydrology and Earth System Sciences*, *22*(4), 2269–2284.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., et al. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*, 8020–8040. https://doi.org/10.1002/2017WR020401

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614.

Mohan, C., Western, A. W., Wei, Y., & Saft, M. (2018). Predicting groundwater recharge for varying land cover and climate conditions—A global meta-study. *Hydrology and Earth System Sciences*, *22*(5), 2689–2703.

Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In S. Hsu, K.-L. Sorooshian, E. Coppola, B. Tomassetti, M. Verdecchia, & G. Visconti (Eds.), *Hydrological modelling and the water cycle* (Vol. 63). Berlin, Heidelberg: Springer.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*, 209–223.

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., & Wood, E. F. (2001). Predicting the discharge of global rivers. *Journal of Climate*, *14*(15), 3307–3323.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*, W03413. https://doi.org/10.1029/2007WR006240

Parajka, J., Blöschl, G., & Merz, R. (2007). Regional calibration of catchment models: Potential for ungauged catchments. *Water Resources Research*, *43*, W06406. https://doi.org/10.1029/2006WR005271

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, *9*(3), 157–171.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins—Part 1: Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, *17*, 1783–1795.

Peel, M. C. (2009). Hydrology: Catchment vegetation and runoff. *Progress in Physical Geography*, *33*(6), 837–844.

Peña-Arancibia, J. L., Van Dijk, A. I. J. M., Mulligan, M., & Bruijnzeel, L. A. (2010). The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrology and Earth System Sciences*, *14*(11), 2193–2205.

Pilgrim, D. H., Chapman, T. G., & Doran, D. G. (1988). Problems of rainfall-runoff modelling in arid and semiarid regions. *Hydrological Sciences Journal*, *33*(4), 379–400.

Plesca, I., Timbe, E., Exbrayat, J. F., Windhorst, D., Kraft, P., Crespo, P., et al. (2012). Model intercomparison to explore catchment functioning: Results from a remote montane tropical rainforest. *Ecological Modelling*, *239*, 3–13.

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, *35*(4), 465–492.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., et al. (2016). Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology*, *17*(1), 287–307.

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering*, *18*(8), 958–975.

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, *198*(1), 69–97.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., & Chiew, F. H. S. (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research*, *45*, W10423. https://doi.org/10.1029/2008WR007248

Rouholahnejad-Freund, E., Fan, Y., & Kirchner, J. W. (2019). Global assessment of how averaging over land-surface heterogeneity affects modeled evapotranspiration rates. *Hydrology and Earth System Sciences Discussions*, *2019*, 1–19.

Sahlu, D., Moges, S. A., Nikolopoulos, E. I., Anagnostou, E. N., & Hailu, D. (2017). Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa. *Advances in Meteorology*, *2017*, 14.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., et al. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., & Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. *Journal of Hydrology*, *351*(1–2), 139–153.

Satgé, F., Ruelland, D., Bonnet, M.-P., Molina, J., & Pillco, R. (2019). Consistency of satellite-based precipitation products in space and over time compared with gauge observations and snow-hydrological modelling in the Lake Titicaca region. *Hydrology and Earth System Sciences*, *23*(1), 595–619.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., & Rudolf, B. (2014). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, *115*(1), 15–40.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, *98–99*(1–4), 279–293.

Seibert, J., & Vis, M. J. P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, *16*(9), 3315–3325.

Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. J. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, *32*(8), 1120–1125.

Singh, R., Archfield, S. A., & Wagener, T. (2014). Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments—A comparative hydrology approach. *Journal of Hydrology*, *517*, 985–996.

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, *17*(15), 3163–3170.

Sorooshian, S., & Gupta, V. K. (1983). Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness. *Water Resources Research*, *19*(1), 260–268.

Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang, S., Hanafin, J., & Nolan, P. (2008). The impacts of climate change on hydrology in Ireland. *Journal of Hydrology*, *356*(1–2), 28–45.

Te Linde, A. H., Aerts, J. C. J. H., Hurkmans, R. T. W. L., & Eberle, M. (2008). Comparing model performance of two rainfall-runoff models in the Rhine Basin using different atmospheric forcing data sets. *Hydrology and Earth System Sciences*, *12*(3), 943–957.

Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research*, *45*, W10438. https://doi.org/10.1029/2008WR007474

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., & Sawicz, K. (2013). Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution. *Hydrology and Earth System Sciences*, *17*, 2209–2217.

Troy, T. J., Wood, E. F., & Sheffield, J. (2008). An efficient calibration method for continental-scale land surface modeling. *Water Resources Research*, *44*, W09411. https://doi.org/10.1029/2007WR006513

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment*, *8*(2), 127–150.

Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., & Beck, H. E. (2013). Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resources Research*, *49*, 2729–2746. https://doi.org/10.1002/wrcr.20251

Van der Velde, Y., Lyon, S. W., & Destouni, G. (2013). Data-driven regionalization of river discharges and emergent land cover-evapotranspiration relationships across Sweden. *Journal of Geophysical Research: Atmospheres*, *118*, 2576–2587. https://doi.org/10.1002/jgrd.50224

Vandewiele, G. L., & Elias, A. (1995). Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of Hydrology*, *170*(1–4), 277–291.

Vereecken, H., Weihermüller, L., Assouline, S., Šimůnek, J., Verhoef, A., Herbst, M., et al. (2019). Infiltration from the pedon to global grid scales: An overview and outlook for land surface modeling. *Vadose Zone Journal*, *18*(1), 1–53.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., & Hattermann, F. (2015). Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth System Dynamics*, *6*(1), 17–43.

Viney, N. R., Vaze, J., Chiew, F. H. S., Perraud, J., Post, D. A., & Teng, J. (2009). Comparison of multi-model and multi-donor ensembles for regionalisation of runoff generation using five lumped rainfall-runoff models, 18th world imacs/modsim congress.

Vogel, R. M., & Kroll, C. N. (1996). Estimation of baseflow recession constants. *Water Resources Management*, *10*(4), 303–320.

Vogt, J., Soille, P., de Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., et al. (2007). A pan-European river and catchment database (*Reference Report*). Ispra, Italy: Joint Research Centre (JRC).

Widén-Nilsson, E., Halldin, S., & Xua, C. (2007). Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. *Journal of Hydrology*, *340*(1–2), 105–118.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, *117*, D03110. https://doi.org/10.1029/2011JD016051

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774.

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*, 5844–5853. https://doi.org/10.1002/2017GL072874

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M., & Jakeman, A. J. (1997). Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. *Water Resources Research*, *33*(1), 153–166.

Yilmaz, K. K., Vrugt, J. A., Gupta, H. V., & Sorooshian, S. (2010). Model calibration in watershed hydrology. In *Advances in data-based approaches for hydrologic modeling and forecasting* (pp. 53–105). World Scientific.

Young, A. R. (2006). Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *Journal of Hydrology*, *320*(1–2), 155–172.

Zecharias, Y. B., & Brutsaert, W. (1988). Recession characteristics of groundwater outflow and base flow from mountainous watersheds. *Water Resources Research*, *24*(10), 1651–1658.

Zelelew, M. B., & Alfredsen, K. (2012). Sensitivity-guided evaluation of the HBV hydrological model parameterization. *Journal of Hydroinformatics*, *15*(3), 967–990.

Zhang, Y., & Chiew, F. H. S. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, *45*, W07412. https://doi.org/10.1029/2008WR007504

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.

Zhang, D., Liu, X., Bai, P., & Li, X.-H. (2019). Suitability of satellite-based precipitation products for water balance simulations using multiple observations in a humid catchment. *Remote Sensing*, *11*, 151.

Zhang, Y., Viney, N., Frost, A., Oke, A., Brooks, M., Chen, Y., & Campbell, N. (2013). Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments (*Tech. Rep. No. EP113194*). Canberra, Australia.

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, *21*(3), 1769–1790.