Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*

HYLKE E. BECK AND AD DE ROO

Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy

ALBERT I. J. M. VAN DIJK

Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia

(Manuscript received 24 August 2014, in final form 15 December 2014)

ABSTRACT

Streamflow Q estimation in ungauged catchments is one of the greatest challenges facing hydrologists. Observed Q from 3000 to 4000 small-to-medium-sized catchments ($10\text{--}10\,000\text{ km}^2$) around the globe were used to train neural network ensembles to estimate Q characteristics based on climate and physiographic characteristics of the catchments. In total, 17 Q characteristics were selected, including mean annual Q , baseflow index, and a number of flow percentiles. Testing coefficients of determination for the estimation of the Q characteristics ranged from 0.55 for the baseflow recession constant to 0.93 for the Q timing. Overall, climate indices dominated among the predictors. Predictors related to soils and geology were relatively unimportant, perhaps because of their data quality. The trained neural network ensembles were subsequently applied spatially over the entire ice-free land surface, resulting in global maps of the Q characteristics (at 0.125° resolution). These maps possess several unique features: they represent observation-driven estimates, they are based on an unprecedentedly large set of catchments, and they have associated uncertainty estimates. The maps can be used for various hydrological applications, including the diagnosis of macroscale hydrological models. To demonstrate this, the produced maps were compared to equivalent maps derived from the simulated daily Q of four macroscale hydrological models, highlighting various opportunities for improvement in model Q behavior. The produced dataset is available online (<http://water.jrc.ec.europa.eu/GSCD>).

1. Introduction

Quantitative knowledge of streamflow Q and its spatial and temporal distribution is important for hydro-power production, drinking water supply, industrial uses, irrigation systems, aquatic habitats, recreation, and sediment and contaminant transport (Poff et al. 1997;

Nuttle 2002; Brauman et al. 2007). Perhaps the greatest obstacle to advancing current understanding is that the majority of Earth's land surface is ungauged or poorly gauged (Fekete and Vörösmarty 2007). The estimation of Q in ungauged and poorly gauged catchments has therefore been highlighted as one of the major challenges facing the hydrologic sciences today (Sivapalan 2003).

One way to estimate Q characteristics in ungauged catchments is to use physically based continuous rainfall-runoff models. These models are expected to provide accurate Q estimates for ungauged catchments owing to their physically based representation of the chief processes governing the water cycle, provided good quality data on precipitation and other governing factors are available. However, model parameters are typically difficult or impossible to measure at the model application scale (Beven 1989; Duan et al. 2001), and hence models still require calibration (Duan et al. 2006; Nasonova et al. 2009;

Denotes Open Access content.

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JHM-D-14-0155.s1>.

Corresponding author address: Hylke Beck, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy.
E-mail: hylke.beck@jrc.ec.europa.eu

DOI: 10.1175/JHM-D-14-0155.1

TABLE 1. The Q characteristics considered here.

Characteristic	Unit	Description and computation
BFI1	—	Baseflow index, defined as the ratio of long-term base flow to total Q (Smakhtin 2001), computed from daily Q data using the recursive digital filter of Van Dijk (2010) with the “window size” set to 5 days (cf. Peña-Arancibia et al. 2010).
BFI2	—	Baseflow index computed from daily Q data following the local-min method described in Pettyjohn and Henning (1979) and Sloto and Crouse (1996) with the “duration of surface runoff” N set to 5 days.
BFI3	—	Baseflow index computed from daily Q data using a 7-day moving min to derive base flow. This method is referred to as the 7-day sliding interval in Pettyjohn and Henning (1979) and Sloto and Crouse (1996).
BFI4	—	Baseflow index computed from daily Q data following the procedure described in Gustard et al. (1992), which takes the minima at 5-day nonoverlapping intervals and subsequently connects the valleys in this series of minima to generate base flow.
k	day^{-1}	Baseflow recession constant, defined as the rate of baseflow decay (Vogel and Kroll 1996), computed from daily Q data as described in Beck et al. (2013b), with the “window size” set to 5 days and days with zero flow ignored.
Q1, Q5, Q10, Q20, Q50, Q80, Q90, Q95, and Q99	mm day^{-1}	Daily flow percentiles (exceedance probability) computed from daily Q data. The number refers to the percentage of time that the flow is exceeded.
T50	—	The day of the water year marking the timing of the center of mass of Q (Stewart et al. 2005). The water year is defined in this study as the 12-month period October–September (April–March) in the Northern (Southern) Hemisphere. T50 can be computed from daily Q data but also monthly Q data linearly interpolated to daily values.
RC	—	Runoff coefficient is the ratio of mean annual Q to P . See Table 2 for details on the P data used. It can be computed from daily as well as monthly Q data.
QMEAN	mm yr^{-1}	Mean annual Q per unit area, which can be computed from daily as well as monthly Q data.

Rosero et al. 2011). Approaches exist that transfer model parameters from gauged to ungauged catchments based on regression, physical similarity, or spatial proximity (Kim and Kaluarachchi 2008; He et al. 2011), but so far these approaches have met with limited success (Wagener and Montanari 2011).

An alternative approach to gain knowledge about Q behavior in ungauged catchments is to use empirical models linking catchment attributes (related to climate, topography, land cover, soils, and/or geology) to indices that quantify various characteristics of the Q regime [e.g., Mazvimavi et al. 2005; Brandes et al. 2005; Detenbeck et al. 2005; Longobardi and Villani 2008; Van Dijk 2010; Peña-Arancibia et al. 2010; Krakauer and Temimi 2011; see Blöschl et al. (2013), Salinas et al. (2013), Hrachowitz et al. (2013), and Beck et al. (2013b) for reviews]. Based on observed Q , these models capture the Q behavior in a lumped way by integrating the effects of spatial heterogeneity in climate and physiography. However, such models have typically been based on a relatively small number of catchments (22–183 for the cited studies) and regional datasets to characterize climate and physiography and hence may have limited applicability outside the study region. Moreover, most regression models have not been applied to an independent set of catchments to test their generalization capability. In an effort to overcome these limitations, Beck et al. (2013b) produced global maps for two baseflow-related Q characteristics based on data from 3394 catchments around the globe.

The present study aims to produce global maps for a more comprehensive selection of Q characteristics. The subsidiary aims are threefold: 1) to investigate the relationships between Q characteristics and climate and physiographic characteristics at the global scale, 2) to produce global maps of Q characteristics, and 3) to compare the produced maps to equivalent maps derived from macroscale hydrological models to test the usefulness of observation-based maps in model evaluation and improvement.

2. Streamflow characteristics

Indices quantifying various characteristics of the Q regime have been widely used in ecological studies (Clausen and Biggs 2000; Olden and Poff 2003). Table 1 lists the 17 Q characteristics selected for the present analysis, including their computation from Q time series. Some of the selected Q characteristics are related to base flow (BFI1–4, Q95, and Q99), and others are related to peak flow (Q1), the shape of the flow duration curve (Q1–Q99), the mean water balance (RC and QMEAN), or flow timing (T50). Figure 1 gives an example of base flow computed using the four techniques considered here (BFI1–4). Note that base flow is defined in this study as the slowly varying portion of Q and thus includes all slow runoff components, including snow and ice melt and the outflow from surface-water bodies (cf. Hall 1968; Smakhtin 2001).

FIG. 1. Time series of observed Q from USGS catchment 50145395 (Rio Casei above Hacienda Casei, Puerto Rico; catchment area is 19 km^2) and base flow computed using the different techniques (BFI1–4).

3. Data

a. Climate and physiographic characteristics

Table 2 lists the climate and physiographic characteristics used as predictors of the Q characteristics. Among the 20 selected predictors, eight were related to climate, three to topography, five to land cover, one to geology, and three to soils. A similar selection of predictors was used by Beck et al. (2013b). However, we used regional data on precipitation P and air temperature TA for the United States, regional P data for New Zealand, and improved datasets for surface elevation $ELEV$, surface slope SLO , fraction covered by lakes and reservoirs fW , fraction of forest cover fTC , normalized difference vegetation index $NDVI$, and the soil-related predictors (content of sand $SAND$, content of silt $SILT$, and content of clay $CLAY$). Additionally, we added topographic wetness index TWI , permafrost abundance PF , and fraction covered by glaciers $GLAC$ as predictors. On the other hand, we removed gravel soil content $GRAV$, as it is not included in the SoilGrids1km dataset. Furthermore, to give drier catchments more weight in training datasets, we replaced mean annual precipitation by square-root-transformed mean annual precipitation P_{trans} and used the reciprocal of the humidity index, the aridity index AI . Last, snow water equivalent depth $SNOW$ was replaced with fraction of snow cover fS , which can be mapped more accurately and at higher spatial resolution (Gao et al. 2010). Most data used for the predictors have a resolution $\leq 1 \text{ km}$. For the catchment-scale estimation of the Q characteristics the full-resolution data were used, while for the computation of global maps the data were resampled to 0.125° using simple averaging with gaps filled by nearest-neighbor interpolation.

b. Observed streamflow

The observed Q data in this study were compiled from three sources. First, daily Q and catchment boundary data for 9169 U.S. stations that were part of the

Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAUGES II; Falcone et al. 2010), were downloaded from the U.S. Geological Survey (USGS) website (<http://water.usgs.gov>). Daily and monthly Q data for 1961 and 3357 stations, respectively, around the globe and associated catchment boundaries (Lehner 2012) were provided by the Global Runoff Data Centre (GRDC; <http://grdc.bafg.de>). Daily Q and associated catchment boundary data for 321 Australian stations compiled by Peel et al. (2000) were used to complete the dataset. This resulted in an initial dataset consisting of 14 808 catchments, of which the 3357 GRDC catchments with only monthly Q data were only considered for the Q characteristics insensitive to the shape of individual flow events ($T50$, RC , and $QMEAN$). Prior to the computation of $T50$, the monthly Q data were linearly interpolated to daily values. The analysis of $T50$ was restricted to northern catchments and regions with some snowfall (defined by $fS > 0.1$), which tend to have distinct seasonal flow patterns.

The following criteria were used to exclude unsuitable catchments from our analysis. First, in order to reduce anthropogenic influences, catchments were required to have $< 2\%$ classified as urban (using the “artificial areas” class of the map from GlobCover, version 2.3; Bontemps et al. 2011) or subject to irrigation (using the Global Irrigated Area Map; <http://www.iwmigiam.org>). Second, a minimum Q record length of 10 years (but not necessarily consecutive) was chosen to provide a reasonable compromise between data availability and robustness of the computed Q characteristics (cf. Kauffeldt et al. 2013). Third, to minimize the effects of channel routing (evaporation, leakage losses, and travel time delays), the catchment area was limited to $< 10\,000 \text{ km}^2$ (cf. Peña-Arancibia et al. 2010; Van Dijk et al. 2013). Fourth, to reduce sampling errors in the catchment-mean predictor values, the catchment area had to be $> 10 \text{ km}^2$. Finally, to reduce the number of disinformative catchments, all Q records were screened for artifacts and anthropogenic influences (diversions and/or impoundments), and U.S.

TABLE 2. The climate and physiographic characteristics selected to estimate the Q characteristics. The table is adapted from Beck et al. (2013b).

Predictor	Unit	Description	Calculation and data source	Resolution
Climate				
AI	—	Aridity index	Calculated as $AI = PET/P$, where P is the mean annual precipitation and PET is the mean annual potential evaporation. See P_{trans} and PET for data sources.	~1 km
P_{si}	—	Precipitation seasonality	Calculated following Walsh and Lawler (1981) as $P_{si} = P_{yr}^{-1} \sum P_m - P_{yr}/12 $, where P_{yr} and P_m are the mean annual and monthly precipitation, respectively, and the summation is over all months. See P_{trans} for data source.	~1 km
P_{trans}	—	Transformed mean annual precipitation	Computed according to $P_{trans} = \sqrt{P}$, where P is the mean annual precipitation derived from WorldClim (Hijmans et al. 2005), Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994) for the United States, and Tait et al. (2006) for New Zealand.	~1 km
PET	mm yr^{-1}	Mean annual potential evaporation	Calculated from monthly values derived following the temperature-based approach of Hargreaves et al. (1985). See TA for data source.	~1 km
PET_{si}	—	Potential evaporation seasonality	Calculated following Walsh and Lawler (1981) as $PET_{si} = PET_{yr}^{-1} \sum PET_m - PET_{yr}/12 $, where PET_{yr} and PET_m are the mean annual and monthly potential evaporation, respectively, and the summation is over all months. See PET for data source.	~1 km
CORR	—	Seasonal correlation between water supply and demand	Correlation coefficient calculated between monthly climate values of P and PET (Petersen et al. 2012). See P_{trans} and PET for data sources.	~1 km
TA	K	Mean annual air temperature	WorldClim (Hijmans et al. 2005) and PRISM (Daly et al. 1994) for the United States.	~1 km
PF	—	Permafrost abundance	National Snow and Ice Data Center vector map (Brown et al. 1997) with classes C, D, S, and I reclassified to permafrost abundances of 0.95, 0.70, 0.30, and 0.05, respectively.	~10 km
Topography				
TWI	—	Topographic wetness index	Global map of topographic wetness index (Marthews et al. 2015).	~500 m
SLO	°	Surface slope	Consultative Group for International Agricultural Research (CGIAR) Consortium for Spatial Information (CSI) Shuttle Radar Topography Mission (SRTM), version 4.1 (Farr et al. 2007), for lat < 60°N, GTOPO30 (http://lta.cr.usgs.gov/GTOPO30) for lat > 60°N.	~90 m, ~1 km
ELEV	m MSL	Surface elev	CSI SRTM, version 4.1 (Farr et al. 2007), for lat < 60°N, GTOPO30 for lat > 60°N.	~90 m, ~1 km
Land cover				
fW	—	Fraction covered by lakes and reservoirs	World Wildlife Fund (WWF) Global Lakes and Wetlands Database (GLWD) level 3 (Lehner and Döll 2004).	~1 km
fTC	—	Fraction of forest cover	Food and Agriculture Organization (FAO) Global Forest Resources Assessment (FRA) 2000 forest cover (FAO 2000).	~1 km
NDVI	—	Normalized difference vegetation index	Système Pour l'Observation de la Terre (SPOT) Vegetation (VGT) S10 10-day max value composites (www.vgt.vito.be), mean of 2005.	~1 km
fS	—	Fraction of snow cover	MODIS Aqua snow cover daily level 3 global climate modeling grid product (MYD10C1), version 5 (Hall et al. 2006), mean of 2003–13.	0.05°
GLAC	—	Fraction covered by glaciers	Randolph Glacier Inventory (RGI), version 3.2, glacier outlines.	~30–60 m
Geology and soils				
PERM	$\log_{10} \text{m}^2$	Permeability of geology	Global permeability map (Gleeson et al. 2011).	~1 km
SAND	%	Soil sand content	SoilGrids1km (Hengl et al. 2014) version April 2014, mean over all layers.	~1 km
SILT	%	Soil silt content	SoilGrids1km (Hengl et al. 2014) version April 2014, mean over all layers.	~1 km
CLAY	%	Soil clay content	SoilGrids1km (Hengl et al. 2014) version April 2014, mean over all layers.	~1 km

TABLE 3. For each Q characteristic, the transformation equations used to improve the normality of the data and the number of catchments that fulfilled the selection criteria.

Characteristic	Transformation equation	Inverse transformation equation	No. of catchments
BFI1	$BFI1_{trans} = BFI1^2$	$BFI1 = \sqrt{BFI1_{trans}}$	3104
BFI2	$BFI2_{trans} = BFI2$	$BFI2 = BFI2_{trans}$	3104
BFI3	$BFI3_{trans} = BFI3^2$	$BFI3 = \sqrt{BFI3_{trans}}$	3104
BFI4	$BFI4_{trans} = BFI4$	$BFI4 = BFI4_{trans}$	3104
k	$k_{trans} = \ln(k)$	$k = \exp(k_{trans})$	3104
Q1	$Q1_{trans} = \sqrt{Q1}$	$Q1 = Q1_{trans}^2$	3104
Q5	$Q5_{trans} = \sqrt{Q5}$	$Q5 = Q5_{trans}^2$	3104
Q10	$Q10_{trans} = \sqrt{Q10}$	$Q10 = Q10_{trans}^2$	3104
Q20	$Q20_{trans} = \sqrt{Q20}$	$Q20 = Q20_{trans}^2$	3104
Q50	$Q50_{trans} = \sqrt{Q50}$	$Q50 = Q50_{trans}^2$	3104
Q80	$Q80_{trans} = \sqrt{Q80}$	$Q80 = Q80_{trans}^2$	3104
Q90	$Q90_{trans} = \sqrt{Q90}$	$Q90 = Q90_{trans}^2$	3104
Q95	$Q95_{trans} = \sqrt{Q95}$	$Q95 = Q95_{trans}^2$	3104
Q99	$Q99_{trans} = \sqrt{Q99}$	$Q99 = Q99_{trans}^2$	3104
T50	$T50_{trans} = T50$	$T50 = T50_{trans}$	2439
RC	$RC_{trans} = \sqrt{RC}$	$RC = RC_{trans}^2$	4079
QMEAN	$QMEAN_{trans} = \sqrt{QMEAN}$	$QMEAN = QMEAN_{trans}^2$	4079

catchments flagged as “non-reference” in the GAUGES-II database were discarded. Table 3 lists the number of catchments that fulfilled the selection criteria for each Q characteristic. Figure 2 shows the locations of the catchments. All Q data were converted to millimeters per day using the provided catchment areas.

c. Modeled streamflow

We used modeled Q from four macroscale hydrological models for two purposes: first, to compare the performance in estimating the Q characteristics obtained by the neural networks against the performance obtained

by hydrological models, and second, to test whether the produced global maps of Q characteristics can be used for model evaluation and improvement.

The first model, HTESSEL-ERAIR, uses the Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) land surface model (LSM; Balsamo et al. 2009) driven by the ERA-Interim forcing (Dee et al. 2011) corrected using Global Precipitation Climatology Project (GPCP), version 2.1 (Adler et al. 2003; Balsamo et al. 2015). The data have a daily temporal and $\sim 0.7^\circ$ spatial resolution and were available for the period 1979–2010.

FIG. 2. Centroids of the catchments that fulfilled the selection criteria. Catchments with daily Q data are indicated in blue and catchments with only monthly Q data are in red. For Q characteristic T50, only the region with $fs > 0.1$ indicated in green was considered. The catchments with only monthly Q data were only used for the Q characteristics T50, RC, and QMEAN.

The second model, Noah–GLDAS, is based on the Noah LSM, version 2.7.1 (Schaake et al. 1996; Ek et al. 2003), driven by the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004). The data were obtained from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC; <http://disc.sci.gsfc.nasa.gov>) and cover the period from 2000 to present. The data have a 3-hourly temporal and 0.25° spatial resolution and were aggregated to daily means for this study.

The third model, PCR-GLOBWB–ERA1, uses the PCRaster Global Water Balance (PCR-GLOBWB) global hydrological model (GHM; Van Beek and Bierkens 2009) driven by ERA-Interim (Dee et al. 2011). The data have a daily temporal and 0.5° spatial resolution and were available for the period 2003–10.

The fourth and final model, W3RA–Princeton, is based on the World-Wide Water Resources Assessment (W3RA) GHM, version 1 (Van Dijk et al. 2013), driven by the Sheffield et al. (2006) meteorological dataset developed at Princeton University. The data have a daily temporal and 1° spatial resolution and were available for the period 1980–2010.

4. Methodology

a. Climate and physiographic controls of streamflow characteristics

Regression analysis was used to evaluate the strength and shape of the relationships between catchment values of the climate and physiographic characteristics and the transformed Q characteristics, providing clues to the controls on Q characteristics. Linear, exponential, and power functions were fitted by least squares and the function with the highest coefficient of determination R^2 was reported. Significance levels (or probability values) were not calculated as they were strongly correlated to the R^2 values, making it redundant to list both. In addition, most relationships were found to be highly significant, despite visual interpretation indicating no apparent relationship. Indeed, several studies have found that significance levels may be misleading when working with such large sample sizes (Royall 1986; Johnson 1999; Nicholls 2001).

b. Catchment- and global-scale estimation of streamflow characteristics

The procedure described in Beck et al. (2013b) with one modification was used to estimate the Q characteristics at the catchment and global scale. Briefly summarized, Beck et al. (2013b) trained neural network ensembles to estimate the Q characteristics BFI1 and k

based on climate and physiographic characteristics of the catchments. Each ensemble member consisted of a multilayer perceptron feed-forward neural network with one hidden layer comprising 30 neurons. Although Beck et al. (2013b) used ensembles comprising 10 members obtained using 10-fold cross validation, here we used ensembles comprising 20 members obtained using 20-fold cross validation in order to achieve more accurate estimation results. For each cross-validation iteration, the catchment set was partitioned into subsets of 80% for training, 15% for validation, and 5% for testing. The subdivision was random, although each catchment was used only once for testing. The validation subset of catchments was used to prevent overfitting, while the testing subset was used to independently evaluate the performance of the trained neural network. For training, the Levenberg–Marquardt algorithm was used in combination with the mean-square error performance function. Prior to the estimation, the values of the Q characteristics were transformed to make the catchment distribution of values resemble a normal distribution and to stretch important parts of the scale of measurement. After the estimation, the estimates were back-transformed. Table 3 lists the transformation equations used for the 17 Q characteristics considered here.

Global maps of the Q characteristics (0.125°) were obtained by spatially applying the trained neural network ensemble over the entire ice-free land surface, computing the median of the 20 estimates for each grid cell and back-transforming the result. Associated uncertainty maps were computed from the standard deviation of the 20 transformed estimates for each grid cell [see Beck et al. (2013b) for further details]. The ice-free land surface was defined using the following datasets. Ocean boundaries were taken from the Global Self-Consistent, Hierarchical, High-Resolution Shoreline (GSHHS) dataset (<http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html>). The World Wildlife Fund (WWF) Global Lakes and Wetlands Database (GLWD) Level 1 (Lehner and Döll 2004) was used to exclude large inland water bodies (>20 000 km²). The ice-covered portion of the land surface was defined by climate type EF (polar, frost) of the Köppen–Geiger climate type map (Peel et al. 2007; see Fig. S2.1 in the supplemental material).

To evaluate the value of the neural network ensembles, the obtained performance in estimating the Q characteristics was compared against the performance obtained by two LSMs (HTESSEL–ERAIR and Noah–GLDAS) and two GHMs (PCR-GLOBWB–ERA1 and W3RA–Princeton). For each model, total Q was computed by summing the surface and subsurface runoff contributions of the model. Thus, BFI values derived from the models do not necessarily reflect the fraction of total Q

originating from the subsurface compartments of the models. For each Q characteristic, only the catchments that had nonmissing values of the Q characteristics for all four models were considered. For k , catchments were rejected if one of the models produced flow events without sufficiently long recession periods, as this prevented the computation of k . For BFI1–4, k , Q1–Q99, and T50, catchments for which one of the models produced $Q_{\text{MEAN}} < 5 \text{ mm yr}^{-1}$ were rejected. The analysis of T50 was restricted to catchments with some snowfall (defined by $\text{fS} > 0.1$). Values for RC were not computed for the models since the P forcing data were not available for all models.

c. Example application

As an example application of the new global maps of Q characteristics, a comparison was made to equivalent maps derived from the daily Q estimates of four macroscale hydrological models (HTESSEL–ERAIR, Noah–GLDAS, PCR-GLOBWB–ERAIR, and W3RA–Princeton) in order to reveal deficiencies in the model or the forcing. Total Q was obtained by summing the surface and subsurface runoff contributions of the models, after which the Q characteristics were computed per grid cell. The model-based maps of the Q characteristics were resampled, together with the observation-based maps produced in this study, to a common 1° resolution using simple averaging. For each model and Q characteristic, spatial correlation coefficients R (indicative of the agreement in spatial variability), median differences D (indicative of the overall bias), and median absolute differences $|D|$ (indicative of the overall difference) were computed between the simulation- and observation-based maps. Significance levels were not reported as nearly all combinations yielded extremely significant probability values.

5. Results

a. Climate and physiographic controls of the streamflow characteristics

Table 4 lists R^2 values computed between catchment values of the Q characteristics and the catchment attributes (for associated scatterplots, see Figs. S1.1–S1.17 in the supplemental material). In terms of mean R^2 , the five most important predictors (ordered from highest to lowest mean R^2) were AI, P_{trans} , SLO, fS, and TWI, while the five least important predictors (again ordered from highest to lowest mean R^2) were GLAC, fW, P_{si} , SAND, and PERM. Overall, the predictors related to climate proved most informative, followed by the ones related to topography and land cover, while the predictors related to geology and soils were least

informative. The topography-related predictors TWI and SLO achieved almost identical R^2 values because of the strong correlation between the two predictors. Among the soil-related predictors, CLAY was the most informative. The most important predictors for BFI4 were PET, TA, ELEV, fS, and CLAY; and for k were AI, PET, and fS. Predictors identified as important for Q1–Q99 were AI, P_{trans} , and SLO. Important predictors for T50 were P_{trans} , TA, NDVI, and fS; for RC were AI, PET, and fS; and for Q_{MEAN} were AI, P_{trans} , SLO, and fS. Variables BFI1–4 achieved comparable R^2 results (see Figs. S1.1–S1.4 in the supplemental material) because of strong correlations between the BFI estimates.

b. Catchment-scale estimation of streamflow characteristics

Table 5 shows mean R^2 and RMSE values obtained by the 20 neural networks (one for each cross-validation iteration) for the training and testing subsets of catchments. The mean training R^2 values were excellent (≥ 0.90) for Q5, Q10, Q20, Q50, T50, and Q_{MEAN} ; good (0.80–0.89) for Q1, Q80, Q90, Q95, and RC; acceptable (0.70–0.79) for BFI1–4 and Q99; and modest (< 0.70) for k (Table 5; for associated scatterplots, see Fig. S2.2 in the supplemental material). Variables BFI1–4 achieved very similar training and testing R^2 values because of high correlations among the BFI estimates. The testing R^2 values were 0.02–0.14 lower than the training R^2 values (Table 5), suggesting a good generalization capability of the trained models. Table 6 lists the performance of the four macroscale hydrological models in estimating the Q characteristics. The testing R^2 values obtained by the neural networks (Table 5) substantially exceed the R^2 values obtained by the models (Table 6), confirming the usefulness of the neural networks.

c. Global-scale estimation of streamflow characteristics

Figures 3–5 show global maps of estimated and observed BFI4, T50, and Q_{MEAN} , respectively (for global maps of the other Q characteristics, see Figs. S2.3–S2.19 in the supplemental material). Visual comparison of the estimates and the observations revealed good agreement between estimated and observed BFI4, despite a slight tendency for the neural networks to underestimate high BFI4 and overestimate low BFI4 (Fig. 3). Particularly high BFI4 values were found for central Africa, the Canadian Shield in North America, the mountain ranges of the western United States, the Baltic Shield in northern Europe, and the eastern part of the Amazon basin (Fig. 3a). The T50 estimates and observations exhibited excellent agreement overall (Fig. 4). A pronounced north–south

TABLE 4. Values of R^2 computed between catchment values of Q characteristics and predictors (for associated scatterplots, see Figs. S1.1–S1.17 in the supplemental material). Various functions were considered, but only the R^2 of the best relationship is shown. Values of $R^2 < 0.05$ are indicated by an em dash. For each Q characteristic, the three highest R^2 values are shown in boldface. The type of relationship that achieved the best R^2 [linear (LIN), exponential (EXP), and power (POW)] is given. The plus and minus symbols indicate positive and negative relationships, respectively. The four BFI estimates produced very similar results; hence, only BFI4 is shown.

Predictor	BFI4	k	Q1	Q5	Q10	Q20	Q50	Q80	Q90	Q95	Q99	T50	RC	QMEAN
AI	0.06 ^{POW}	0.15 ^{LIN}	0.60 ^{EXP}	0.74 ^{EXP}	0.76 ^{EXP}	0.78 ^{EXP}	0.77 ^{EXP}	0.64 ^{EXP}	0.57 ^{EXP}	0.53 ^{EXP}	0.47 ^{EXP}	0.15 ^{POW} 0.08 ^{POW}	0.37 ^{EXP}	0.74 ^{EXP}
P_{si}	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P_{trans}	—	—	0.58 ^{POW}	0.53 ^{POW}	0.49 ^{LIN}	0.48 ^{LIN}	0.51 ^{LIN}	0.40 ^{LIN}	0.34 ^{LIN}	0.31 ^{LIN}	0.26 ^{LIN}	0.32 ^{EXP} 0.19 ^{LIN}	0.05 ^{LIN} 0.28 ^{POW}	0.54 ^{POW}
PET	0.18 ^{POW} 0.12 ^{POW}	0.12 ^{LIN} 0.08 ^{POW}	—	0.09 ^{POW}	0.12 ^{POW}	0.14 ^{POW}	0.11 ^{LIN}	0.11 ^{LIN}	0.11 ^{LIN}	0.11 ^{LIN}	0.11 ^{LIN}	0.29 ^{POW}	0.12 ^{POW}	—
PET _{si}	—	—	—	—	—	—	—	—	—	—	—	—	—	—
CORR	—	—	0.11 ^{LIN}	0.16 ^{LIN}	0.17 ^{EXP}	0.15 ^{EXP}	0.13 ^{LIN}	0.11 ^{EXP}	0.10 ^{EXP}	0.10 ^{EXP}	0.10 ^{LIN}	0.17 ^{LIN}	0.16 ^{POW}	0.15 ^{LIN}
TA	0.19 ^{LIN}	0.10 ^{LIN}	—	—	—	—	—	—	—	—	—	0.55 ^{LIN}	0.09 ^{EXP}	—
PF	—	0.05 ^{LIN}	—	—	—	—	—	—	—	—	—	0.41 ^{EXP}	0.26 ^{POW}	—
TWI	0.11 ^{EXP}	0.08 ^{LIN}	0.16 ^{EXP}	0.24 ^{EXP}	0.26 ^{EXP}	0.26 ^{EXP}	0.19 ^{EXP}	0.18 ^{EXP}	0.17 ^{EXP}	0.17 ^{EXP}	0.17 ^{EXP}	—	0.26 ^{POW}	0.24 ^{EXP}
SLO	0.11 ^{POW}	0.08 ^{POW}	0.18 ^{EXP}	0.26 ^{EXP}	0.28 ^{EXP}	0.28 ^{POW}	0.23 ^{EXP}	0.20 ^{EXP}	0.20 ^{EXP}	0.20 ^{EXP}	0.19 ^{EXP}	—	0.25 ^{EXP}	0.26 ^{EXP}
ELEV	0.17 ^{POW}	—	—	—	—	—	—	—	—	—	—	0.20 ^{LIN}	—	—
fW	0.07 ^{LIN}	0.06 ^{LIN}	—	—	—	—	—	—	—	—	—	—	—	—
fTC	0.07 ^{LIN}	0.07 ^{LIN}	0.10 ^{EXP}	0.11 ^{LIN}	0.11 ^{LIN}	0.10 ^{LIN}	0.09 ^{LIN}	0.08 ^{LIN}	0.07 ^{LIN}	0.07 ^{LIN}	0.06 ^{EXP}	—	—	0.08 ^{LIN}
NDVI	0.08 ^{POW}	—	0.08 ^{POW}	—	—	—	—	—	—	—	—	—	—	—
fS	0.28 ^{LIN}	0.15 ^{EXP}	—	0.09 ^{LIN}	0.12 ^{LIN}	0.11 ^{LIN}	0.07 ^{LIN}	0.08 ^{LIN}	0.09 ^{LIN}	0.10 ^{LIN}	0.11 ^{LIN}	0.51 ^{LIN} 0.45 ^{LIN}	0.10 ^{POW} 0.29 ^{EXP}	—
GLAC	—	—	—	0.05 ^{LIN}	0.07 ^{LIN}	0.09 ^{LIN}	—	—	—	—	—	0.06 ^{LIN}	0.11 ^{LIN}	—
PERM	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SAND	0.06 ^{LIN}	—	—	—	—	—	—	—	—	—	—	—	—	—
SILT	—	—	0.07 ^{POW}	0.09 ^{POW}	0.08 ^{LIN}	0.08 ^{LIN}	0.07 ^{LIN}	—	—	—	—	—	—	—
CLAY	0.21 ^{EXP}	0.11 ^{POW}	—	0.12 ^{EXP}	0.15 ^{EXP}	0.15 ^{EXP}	0.15 ^{EXP}	0.16 ^{EXP}	0.16 ^{EXP}	0.16 ^{EXP}	0.17 ^{EXP}	0.07 ^{LIN}	0.19 ^{EXP}	0.05 ^{POW}

TABLE 5. Mean R^2 and RMSE statistics obtained for the training and testing subsets of catchments for the 20 neural networks (one for each cross-validation iteration). Both statistics were calculated from transformed values of the Q characteristics; hence, the RMSE value is unitless. See Table 3 for the number of catchments used for each Q characteristic.

Characteristic	Mean R^2		Mean RMSE	
	Training	Testing	Training	Testing
BFI1	0.77	0.65	0.11	0.13
BFI2	0.78	0.64	0.09	0.12
BFI3	0.78	0.67	0.11	0.13
BFI4	0.76	0.65	0.10	0.13
k	0.66	0.55	0.34	0.39
Q1	0.84	0.80	0.56	0.65
Q5	0.90	0.85	0.32	0.41
Q10	0.91	0.88	0.25	0.32
Q20	0.93	0.89	0.19	0.25
Q50	0.91	0.87	0.14	0.19
Q80	0.86	0.79	0.13	0.17
Q90	0.83	0.75	0.13	0.16
Q95	0.80	0.70	0.13	0.16
Q99	0.75	0.66	0.12	0.15
T50	0.95	0.93	8.81	11.45
RC	0.85	0.77	0.11	0.13
QMEAN	0.91	0.88	3.21	3.91

gradient from late to early Q timing can be observed (Fig. 4a). Similarly, the QMEAN estimates did not show any clear regional biases (Fig. 5). Particularly high QMEAN was obtained for the Amazon, maritime Southeast Asia, the western flank of the Coast Mountains in North America, and the southern section of the Andes in Patagonia (Fig. 5a).

Table 7 lists mean values of the Q characteristics for the entire ice-free land surface and the major Köppen–Geiger climate types (for a map of the major Köppen–Geiger climate types, see Fig. S2.1 in the supplemental material). The lowest BFI values were found in arid regions (broad climate type B) and the highest were found in tropical and cold regions (broad climate types A and D, respectively; Table 7, Fig. 3a). The values of the different BFI estimates varied consistently among climate types—BFI1 always produced the highest values and BFI2 or BFI4 always produced the lowest values, while BFI3 produced values slightly lower than BFI1 (Table 7). The values of k showed little variability among climate types—only for arid regions (broad climate type B) was the baseflow recession markedly faster (Table 7). Values of Q1–Q99 and QMEAN also varied rather consistently among climate types, with higher values found for tropical climate types Af and Am (rain forest and monsoon, respectively) and polar climate type EF (frost), while lower values were found for arid regions (broad climate type B; Table 7). The means for Q99, Q95, and Q90 were (near) zero for many climate

TABLE 6. The range of R^2 computed between observed and simulated catchment values of the transformed Q characteristics. The observed values were derived from the Q records and the simulated values from total Q of the four macroscale hydrological models. The statistics for BFI1–4 were nearly identical; hence, only the BFI4 results are shown. Values for RC are not shown because the P forcing data were not available for all models.

Characteristic	R^2 range	No. of catchments
BFI4	0.00–0.05	2475
k	0.00–0.04	2164
Q1	0.21–0.34	2475
Q5	0.36–0.47	2475
Q10	0.42–0.52	2475
Q20	0.49–0.54	2475
Q50	0.38–0.57	2475
Q80	0.18–0.41	2475
Q90	0.12–0.34	2475
Q95	0.09–0.29	2475
Q99	0.05–0.24	2475
T50	0.48–0.59	2063
QMEAN	0.45–0.56	4072

types (Table 7), which is indicative of intermittent and ephemeral streams. The earliest T50 values were found for climate type BS (arid, steppe) and the latest for climate type ET (polar, tundra; Table 7). The lowest RC values were found in arid regions (broad climate type B) and the highest for polar climate type ET (Table 7). The mean RC value obtained for climate type ET is in fact higher than one (Table 7), probably because of P underestimation.

d. Example application

As an example application of the new global maps of Q characteristics, a comparison was made to equivalent maps derived from the daily Q estimates of four macroscale hydrological models. Figures 6–8 show global maps of the difference for transformed values of the Q characteristics BFI4, T50, and QMEAN, respectively (for the other Q characteristics, see Figs. S3.1–S3.13 in the supplemental material). Table 8 lists correlation coefficients, median differences, and median absolute differences for BFI4, k , Q1–Q99, T50, and QMEAN. HTESSEL–ERAIR and W3RA–Princeton yielded the highest R values for BFI4 (Table 8), meaning that these models most accurately simulate the spatial variability in the ratio of base flow to total Q . However, HTESSEL–ERAIR produces excess base flow over taiga and temperate regions (Fig. 6a), while W3RA–Princeton does so over tropical regions (Fig. 6d). Noah–GLDAS yielded the lowest R value (Table 8), producing insufficient base flow over tundra and taiga regions (Fig. 6b). PCR–GLOBWB–ERAIR performed moderately well in terms of R but appears to overestimate BFI4 over most of the land surface (Table 8, Fig. 6c). All

FIG. 3. Global maps of (a) estimated and (b) observed BFI computed using technique 4. The values in (a) represent the (back transformed) median estimates of the neural network ensembles. In (b), only gauged regions have values.

models produce insufficient base flow over the Rocky Mountains in North America (Fig. 6).

In terms of k , W3RA–Princeton achieved the highest R value (Table 8), meaning it most accurately simulates the spatial variability in baseflow recession rate. Nevertheless, it exhibits slightly too slow baseflow recessions overall, as indicated by the negative D value (Table 8). In contrast, Noah–GLDAS performed poorly, as indicated by the negative R value (Table 8). W3RA–Princeton performed best for the median and low flow percentiles (Q50–Q99), while Noah–GLDAS performed best for the high flow percentiles (Q1–Q10; Table 8). In terms of T50, all models exhibit a too early production of Q over most of the land surface considered here (Table 8, Fig. 7). However, PCR-GLOBWB–ERA1 performed markedly better than the other models, obtaining a D value of -3 days whereas the other models yielded D values from -11 to -14 days (Table 8).

PCR-GLOBWB–ERA1 demonstrated the best R value for QMEAN and thus best captures the spatial

variability in mean annual Q (Table 8, Fig. 8c). Noah–GLDAS produces insufficient Q over most of the land surface (Table 8, Fig. 8b), while W3RA–Princeton does so over most of the Northern Hemisphere with the exception of India (Table 8, Fig. 8d). HTESSEL–ERAIR performed moderately well in terms of QMEAN (Table 8, Fig. 8a). All models produce insufficient Q over the Rocky Mountains, the Andes, and the Himalayas (Fig. 8). The two GHMs (PCR-GLOBWB–ERA1 and W3RA–Princeton), which have Q simulation as one of their primary aims, did not outperform the two LSMs (HTESSEL–ERAIR and Noah–GLDAS; Table 8, Fig. 8).

6. Discussion

a. Climate and physiographic controls of the streamflow characteristics

The high R^2 values found for PET versus BFI4, TA versus BFI4, fS versus BFI4, AI versus k , PET versus k ,

FIG. 4. As in Fig. 3, but for timing of the center of mass of Q (T50). The analysis of T50 was restricted to areas with some snowfall.

and fS versus k (Table 4) reflect the dominant control of climate conditions on baseflow volume and recession rate (Price 2011). The positive relationship obtained between fS and BFI4 (Table 4) is due to the gradual release of snowmelt to Q , which results in slowly varying Q exhibiting high BFI4. The relationships obtained between SAND or CLAY and BFI4 (Table 4) correspond with the influence of soil composition on the infiltration-excess mechanism of runoff generation (Horton 1933; Wolock and McCabe 1999). Values of BFI1–4 showed very similar relationships with the predictors (Figs. S1.1–S1.4 in the supplemental material) because of high correlations among the BFI estimates. Using a different but partly overlapping set of 3394 catchments around the globe, Beck et al. (2013b) obtained results similar to those presented here (Figs. S1.1, S1.5 in the supplemental material). We refer to Beck et al. (2013b) for a review of multicatchment observational studies investigating the links between catchment attributes and BFI and k .

The Q characteristics Q1–Q99, RC, and QMEAN were found to be strongly related to AI (Table 4),

emphasizing the dominant influence of climate conditions on the overall water balance (cf. Schreiber 1904; Ol'dekop 1911; Pike 1964; Budyko 1974; Zhang et al. 2004). The RC exhibited a particularly strong (positive) relationship with fS (Table 4), presumably because of reductions in evaporation when the surface is snow covered (Berghuijs et al. 2014) and/or because of common P underestimations in mountainous environments caused by wind-induced undercatch and the elevation bias in gauge placement (Groisman and Legates 1994; Hijmans et al. 2005; Sevruk et al. 2009). The weakly negative relationships obtained here between QMEAN and the phase difference between precipitation and potential evaporation seasonalities CORR (Table 4) is in accord with several studies investigating the influence of CORR on the overall water balance (Milly 1994; Wolock and McCabe 1999; Potter et al. 2005). Values of fTC did not exhibit clear relationships with the Q characteristics (Table 4), which agrees with some multicatchment studies examining the spatial relationship between fTC and Q (e.g., Oudin et al. 2008; Peel et al. 2010; Van Dijk

FIG. 5. As in Fig. 3, but for QMEAN. Note the nonlinear color scale.

et al. 2012) as well as some meso- and large-scale catchment ($>1 \text{ km}^2$) studies examining the impact of temporal changes in fTC on Q [see review in Beck et al. (2013a)]. The lack of a relationship between NDVI and Q5–Q99 or QMEAN (Table 4) agrees with Hope and Bart (2012), who used observed Q from 30 South African catchments ($1\text{--}254 \text{ km}^2$) to examine whether QMEAN and 11 flow percentiles between Q5 and Q95 are linked to NDVI derived from the Moderate Resolution Imaging Spectroradiometer (MODIS).

Values of TA were found to be strongly negatively related to T50 (Table 4) because temperature represents the dominant control on snow accumulation and melt (Aguado et al. 1992; Stewart et al. 2004). The almost equally strong negative relationship obtained between NDVI and T50 (Table 4) is probably not due to the effect of vegetation on Q , but rather because of the correlation between TA and NDVI owing to the strong control exerted by temperature on vegetation productivity in northern regions (e.g., Jia et al. 2003). The negative relationship found between P_{trans} and T50 (Table 4) is also unlikely to be causal but reflects the strong positive

correlation between P_{trans} and TA in northern regions. Similarly, the positive relationship obtained between ELEV and T50 (Table 4) is presumably not causal but reflects the strong correlations between ELEV, air temperature, snowfall, and snow accumulation and melt (Daly et al. 2002; Leibowitz et al. 2012). A noncausal relationship does not necessarily mean that the predictor should be excluded from the estimation exercise, as it is possible that noncontrolling predictors compensate for the deficiencies of controlling predictors.

Overall, the climate indices showed the strongest relationships with the Q characteristics (Table 4). Among the predictors related to land cover, only fS was clearly related to multiple Q characteristics (Table 4). Among the topography-related predictors, SLO and TWI were related to several Q characteristics, while ELEV was found to be relatively unimportant (Table 4). Our results suggest that topography is not the dominant control on Q behavior at the global scale (cf. Devito et al. 2005). In general, low R^2 values were found for single predictors (Table 4), suggesting that multiple predictors have to be used to satisfactorily estimate the Q

TABLE 7. For the ice-free land surface and the major Köppen–Geiger climate types, mean values of the Q characteristics derived from the new maps. Figure S2.1 (in the supplemental material) shows a map of the major Köppen–Geiger climate types. Mean values were only computed if $>10\,000$ grid cells of 0.125° had nonmissing values. For T50, only the portion of the ice-free land surface with $fs > 0.1$ was considered.

Climate type	BF11 (—)	BF12 (—)	BF13 (—)	BF14 (—)	k (day ⁻¹)	Q1 (mm day ⁻¹)	Q5 (mm day ⁻¹)	Q10 (mm day ⁻¹)	Q20 (mm day ⁻¹)	Q50 (mm day ⁻¹)	Q80 (mm day ⁻¹)	Q90 (mm day ⁻¹)	Q95 (mm day ⁻¹)	Q99 (mm day ⁻¹)	T50 (—)	RC (—)	QMEAN (mm yr ⁻¹)
Ice-free land surface	0.61	0.38	0.56	0.42	0.15	7.02	3.34	2.22	1.33	0.49	0.23	0.17	0.12	0.08	223	0.34	365
Af: tropical, rain forest	0.77	0.63	0.74	0.60	0.06	23.79	12.88	9.33	6.13	3.00	1.69	1.22	0.88	0.57	—	0.60	1667
Am: tropical, monsoon	0.78	0.61	0.74	0.63	0.06	18.29	9.60	6.90	4.13	1.56	0.71	0.49	0.35	0.19	—	0.47	1080
Aw: tropical, savannah	0.70	0.51	0.66	0.53	0.09	8.85	3.91	2.49	1.40	0.48	0.21	0.15	0.11	0.06	—	0.24	384
BW: arid, desert	0.45	0.07	0.35	0.18	0.32	0.71	0.29	0.20	0.12	0.04	0.04	0.02	0.01	0.01	—	0.06	20
BS: arid, steppe	0.49	0.22	0.41	0.27	0.21	1.77	0.57	0.31	0.17	0.05	0.03	0.02	0.01	0.01	191	0.09	50
Cs: temperate, dry summer	0.63	0.43	0.55	0.43	0.11	7.11	3.04	1.98	1.15	0.40	0.15	0.10	0.08	0.05	—	0.30	333
Cw: temperate, dry winter	0.64	0.47	0.61	0.50	0.08	10.37	4.25	2.74	1.58	0.53	0.25	0.19	0.15	0.08	—	0.33	460
Cf: temperate, no dry season	0.59	0.42	0.55	0.41	0.10	11.68	4.92	3.27	2.01	0.81	0.36	0.25	0.19	0.12	—	0.39	593
Ds: cold, dry summer	0.74	0.58	0.72	0.59	0.07	6.66	3.85	2.76	1.63	0.44	0.19	0.15	0.12	0.09	228	0.65	383
Dw: cold, dry winter	0.66	0.43	0.60	0.45	0.09	5.08	2.64	1.68	0.91	0.20	0.05	0.04	0.04	0.02	251	0.42	221
Df: cold, no dry season	0.71	0.52	0.69	0.54	0.07	6.45	3.24	1.99	1.15	0.35	0.14	0.10	0.09	0.06	224	0.58	321
ET: polar, tundra	0.66	0.38	0.66	0.45	0.07	8.37	4.59	2.94	1.82	0.26	0.07	0.06	0.05	0.03	259	1.29	487

FIG. 6. The difference between $BFI4_{trans}$ derived from (a) HTESSEL-ERAIR, (b) Noah-GLDAS, (c) PCR-GLOBWB-ERAIR, and (d) W3RA-Princeton, and $BFI4_{trans}$ from this study (Fig. 3a). $BFI4$ was only computed for the models if simulated $QMEAN > 5 \text{ mm yr}^{-1}$. Red (blue) indicates regions where the models overestimate (underestimate) $BFI4$ relative to the map produced in this study.

characteristics. It bears mentioning that a low R^2 obtained in Table 4 could mean that the predictor under consideration is unimportant but also that the quality of the dataset is poor. For example, it is generally

recognized that geology and soils exert strong controls on Q characteristics (e.g., Farvolden 1963; Boorman et al. 1995; Bruijnzeel 2004), but weak or nonexistent relationships were found here (Table 4), probably

FIG. 7. The difference between $T50_{trans}$ derived from (a) HTESSEL-ERAIR, (b) Noah-GLDAS, (c) PCR-GLOBWB-ERAIR, and (d) W3RA-Princeton, and $T50_{trans}$ from this study (Fig. 4a). Only regions with at least some snowfall were considered. Values of $T50$ were only computed for the models if simulated $QMEAN > 5 \text{ mm yr}^{-1}$. Red (blue) indicates regions where the models produce too late (too early) $T50$ relative to the map produced in this study.

FIG. 8. The difference between $QMEAN_{trans}$ derived from (a) HTESSEL-ERAIR, (b) Noah-GLDAS, (c) PCR-GLOBWB-ERAIR, and (d) W3RA-Princeton, and $QMEAN_{trans}$ from this study (Fig. 5a). Red (blue) indicates regions where the models overestimate (underestimate) $QMEAN$ relative to the map produced in this study.

because of the limited quality and consistency of the global datasets (cf. Table 2). Despite the transformations applied to some of the Q characteristics, the relationships were generally highly nonlinear (see Figs. S1.1–S1.17 in the supplemental material), confirming the need for neural networks instead of linear models for the estimation of the Q characteristics.

b. Catchment-scale estimation of streamflow characteristics

The obtained R^2 values were from good to excellent for most Q characteristics and exceeded the R^2 values obtained by the macroscale hydrological models by a wide margin for all Q characteristics (Tables 5, 6). We compare our results to four recent reviews of studies estimating Q characteristics in ungauged catchments (Blöschl et al. 2013; Salinas et al. 2013; Hrachowitz et al. 2013; Beck et al. 2013b). The mean training R^2 values obtained for BFI1–4 (0.76–0.78; Table 5) are in the range of those obtained in other BFI estimation studies [e.g., Lacey and Grayson 1998; Neff et al. 2005; Van Dijk 2010; see review by Beck et al. (2013b)]. Although the mean training R^2 of 0.66 obtained for k was the lowest among the Q characteristics examined (Table 5), it is still in the upper range of training R^2 values obtained in other k estimation studies [e.g., Hughes 1997; Peña-Arancibia et al. 2010; Krakauer and Temimi 2011; see review by Beck et al. (2013b)].

The best way to quantitatively summarize the shape of the flow duration curve is not commonly agreed upon and depends on the application requirements. Some studies

used normalized flow percentiles (e.g., Smakhtin et al. 1997), while others used parameters of fitted probability distributions (e.g., Fennessey and Vogel 1990) or the slope of the flow duration curve (e.g., Yadav et al. 2007); we used a number of flow percentiles (Q1–Q99; Table 1). The mean training and testing R^2 values obtained here for Q95 (0.80 and 0.70, respectively; Table 5) were higher than those obtained in two regional studies (Gustard and Irving 1994; Laaha and Blöschl 2006). Using observed Q from 1530 European catchments, Gustard and Irving (1994) constructed a multivariate linear model incorporating the catchment proportions of nine soil classes to estimate Q95, yielding a training R^2 of 0.29. Using observed Q from 325 small Austrian catchments (7–963 km²), Laaha and Blöschl (2006) estimated transformed Q95 based on seven catchment attributes, achieving training and testing R^2 values of 0.62 and 0.57, respectively. In contrast to the previous two studies, Mazvimavi et al. (2005) and Hope and Bart (2012) examined multiple flow percentiles. Using observed Q from 52 Zimbabwean catchments (4–2630 km²), Mazvimavi et al. (2005) trained neural networks to estimate nine flow percentiles between Q10 and Q90 from two slope-related indices, drainage density, PET, and P . They obtained a training R^2 range of 0.64–0.92, which is lower than the range of training R^2 values obtained here (Table 5). Using observed Q from 30 South African catchments (1–254 km²), Hope and Bart (2012) estimated 11 flow percentiles between Q5 and Q95 from P , SLO, CLAY, SILT, and drainage density, yielding a training R^2 range

TABLE 8. For the four macroscale models, correlation coefficients, median differences, and median absolute differences computed between simulation- and observation-based maps of the Q characteristics. All statistics were calculated from transformed values of the Q characteristics; hence, D and $|D|$ are unitless. For each Q characteristic, the models yielding the highest Q value, the D value closest to zero, and the lowest $|D|$ value are listed in boldface. The statistics for BFI1–4 were nearly identical; hence, only the BFI4 results are shown. Values for RC are not shown because the P forcing data were not available for all models. For some grid cells, the models produced flow events without sufficiently long recession periods, preventing the computation of k .

Model-forcing	BFI4	k	Q1	Q5	Q10	Q20	Q50	Q80	Q90	Q95	Q99	T50	QMEAN
R													
HTESSEL-ERAIR	0.51	0.34	0.71	0.73	0.74	0.73	0.81	0.75	0.72	0.70	0.64	0.67	0.74
Noah-GLDAS	−0.16	−0.54	0.74	0.77	0.77	0.73	0.73	0.62	0.54	0.46	0.41	0.80	0.81
PCR-GLOBWB-ERAIR	0.17	0.03	0.67	0.73	0.76	0.78	0.79	0.70	0.68	0.66	0.59	0.72	0.86
W3RA-Princeton	0.53	0.48	0.62	0.69	0.75	0.76	0.86	0.87	0.85	0.83	0.79	0.48	0.79
D													
HTESSEL-ERAIR	−0.07	−1.02	− 0.08	−0.22	−0.16	− 0.06	0.01	0.00	0.00	0.00	0.00	−12.26	−0.78
Noah-GLDAS	−0.13	0.06	−0.30	−0.23	−0.24	−0.26	−0.21	−0.14	−0.12	−0.10	−0.07	−10.79	−3.28
PCR-GLOBWB-ERAIR	0.07	−1.31	−0.21	− 0.09	− 0.01	0.07	0.15	0.12	0.11	0.09	0.08	− 3.10	0.01
W3RA-Princeton	− 0.05	−0.70	−0.56	−0.48	−0.40	−0.23	−0.06	−0.02	−0.02	−0.02	−0.01	−14.45	−2.82
D													
HTESSEL-ERAIR	0.19	1.26	0.52	0.45	0.35	0.26	0.14	0.12	0.11	0.09	0.08	18.89	3.01
Noah-GLDAS	0.29	0.72	0.51	0.39	0.34	0.32	0.22	0.16	0.13	0.12	0.09	17.29	4.01
PCR-GLOBWB-ERAIR	0.19	1.40	0.61	0.41	0.30	0.23	0.21	0.19	0.17	0.16	0.15	14.37	3.18
W3RA-Princeton	0.15	0.76	0.80	0.56	0.46	0.33	0.15	0.10	0.08	0.08	0.07	20.57	4.34

of 0.45–0.82, which is lower than the range of training R^2 values found here (Table 5).

Mean training and testing R^2 values of 0.91 and 0.88, respectively, were obtained for QMEAN (Table 5). These numbers are considerably higher than the testing R^2 range of 0.55–0.70 obtained in the validation of QMEAN from 14 LSMs and six Budyko-type models against observed QMEAN from 150 large catchments ($>10\,000\text{ km}^2$) around the globe (Zhou et al. 2012). They also exceed the training R^2 range of 0.29–0.58 obtained by Arnell (1995), who used observed Q from ~ 3500 European catchments ($29\,000\text{--}160\,000\text{ km}^2$) to train a multivariate model incorporating P and PET; and the training R^2 of 0.85 obtained by Hope and Bart (2012), who used observed Q from 30 South African catchments ($1\text{--}254\text{ km}^2$) to estimate QMEAN from P , satellite-derived enhanced vegetation index, and SILT. However, the training R^2 values obtained here are similar to those obtained in three regional studies (Mazvimavi et al. 2005; Viglione et al. 2007; Duan et al. 2010). Based on observed Q from 52 Zimbabwean catchments ($4\text{--}2630\text{ km}^2$), Mazvimavi et al. (2005) yielded a training R^2 of 0.89 in the estimation of QMEAN based on P , two slope-related indices, and a geologic index. Viglione et al. (2007) used observed Q from 47 catchments in northwestern Italy to estimate natural log-transformed QMEAN based on catchment orientation, ELEV, and a radiational aridity index, yielding a training R^2 of 0.90. Duan et al. (2010) used observed Q from 11 Chinese catchments ($3322\text{--}53\,829\text{ km}^2$) to estimate QMEAN from P , PET, latitude and longitude, ELEV, wetland area, catchment shape, and catchment area, yielding a training R^2 of 0.87. However, most such studies did not report testing R^2 values, making it impossible to verify the generalization capability of the derived models. Studies estimating mean annual volumetric Q instead of QMEAN (e.g., Smakhtin et al. 1997; Vogel et al. 1999; Sanborn and Bledsoe 2006; Viglione et al. 2007) are not compared against here as the strong relationship with catchment area results in inflated R^2 values.

c. Global-scale estimation of streamflow characteristics

The global maps of Q characteristics developed here (Figs. 3–5) have some unique features. First, they were derived using a data-driven (top down) approach based on Q observations, rather than using a physically based (bottom up) process model (cf. Sivapalan et al. 2003; Babovic 2005). Being based on Q observations, the maps capture the spatially aggregated behavior without being affected by the limitations and assumptions of physically based models (Beven 1989; Duan et al. 2006). Second, an unprecedentedly large global catchment set spanning a

FIG. 9. Maps of the conterminous United States showing (a) estimated QMEAN from Gebert et al. (1987), (b) estimated QMEAN from this study, and (c) observed QMEAN from this study. Only values for the conterminous United States are shown. The maps have limits 25°–48°N, 125°–70°W and grid lines are at every 10° lat and 15° lon. Note the nonlinear color scale. The values in (b) represent the (back transformed) median estimates of the neural network ensembles.

broad range of physiographic and climate conditions was used to produce the maps, leading to more generally applicable results (cf. Andréassian et al. 2007; Gupta et al. 2014). Third, uncertainty estimates have been computed (Figs. S2.20–S2.36 in the supplemental material), which can be used to evaluate the reliability of the new maps.

The global QMEAN and BFI maps developed in this study can be compared to QMEAN and BFI maps for the conterminous United States produced by Gebert et al.

(1987) and Wolock (2003), respectively. The QMEAN map for the conterminous United States was based on inverse distance weighting (IDW) interpolation of observed QMEAN from 5951 catchments (Gebert et al. 1987). The map was available as a vector file and was rasterized to the 0.125° geographical grid using IDW interpolation to allow spatial comparison to the newly derived QMEAN map (Fig. 5a). Figure 9 shows the QMEAN maps from Gebert et al. (1987) and this study side by side. By and large, there is good large-scale agreement between the maps, although the map of Gebert et al. (1987) has a smoother appearance because of the use of IDW interpolation (Fig. 9). The spatial comparison yielded an R^2 of 0.89 based on transformed values. Considering only the regions where both maps have data, very similar mean QMEAN values of 243 mm yr⁻¹ for the map developed here and 238 mm yr⁻¹ for the map of Gebert et al. (1987) were obtained. The good correspondence between the maps probably reflects the relative ease with which QMEAN can be estimated based on climate and physiographic data.

The BFI map for the conterminous United States (1-km resolution) was produced by IDW interpolation of BFI values from 8249 catchments (Wolock 2003). The BFI was computed using the so-called BFI program (Wahl and Wahl 1995), which implements the base-flow separation procedure described by Gustard et al. (1992) used here for BFI4 (Table 1). After reprojecting the map of Wolock (2003) to the 0.125° geographical grid of the newly derived BFI4 map (Fig. 3a), the maps could be spatially compared. This comparison yielded an R^2 of 0.29, indicating somewhat poor agreement in terms of spatial patterns. There was, however, moderate agreement in terms of mean BFI4—considering only the regions where both maps have data, mean BFI4 values of 0.38 for the newly derived map and 0.43 for the map from Wolock (2003) were obtained. The low agreement between the maps emphasizes the difficulty in estimating BFI (cf. Tables 5 and 6).

Differences among the BFI1–4 values (Table 7) are caused by differences in the baseflow separation techniques used (Table 1). Figure 1 shows that the base flow derived using BFI1 rises the day after the Q starts rising, resulting in relatively high BFI values. BFI3 also produces relatively high BFI values as a short (7 days) moving minimum is used to compute base flow. On the other hand, BFI2 produces relatively low BFI values because of the use of a long (11 days) moving minimum. Relatively low BFI values are also produced by BFI4 as it computes minima at 5-day nonoverlapping intervals and subsequently connects the valleys in this series of minima. Base flow is defined in this study as the slowly varying portion of Q (Hall 1968; Smakhtin 2001), and in this sense the technique of choice is somewhat

subjective and should be guided by the application requirements.

Haddeland et al. (2011) provide estimates of mean QMEAN and RC for the land surface derived from six LSMs and five GHMs. The mean QMEAN based on the map produced here (Fig. 5a) was 365 mm yr^{-1} (Table 7), which is within the range of $290\text{--}457 \text{ mm yr}^{-1}$ obtained by Haddeland et al. (2011). The mean RC based on the map produced here (see Fig. S2.18 in the supplemental material) was 0.34 (Table 7), which is in the low end of the range of $0.33\text{--}0.52$ obtained by Haddeland et al. (2011). The relatively low mean RC found here may be due to the tendency of LSMs and GHMs to overestimate QMEAN in arid and semiarid regions (Haddeland et al. 2011; Trambauer et al. 2013).

The global maps of Q characteristics produced here should be useful for a wide range of large-scale hydrological applications, such as water resource assessments, catchment classification, groundwater recharge estimation, stream ecosystem assessments, and diagnosis and parameterization of hydrological models. Additionally, the produced uncertainty maps may be useful in the design of Q monitoring networks by guiding the placement of gauging stations. However, there are some caveats worth mentioning. First, the maps reflect unregulated flow conditions unaffected by urbanization and irrigation. Second, the maps are representative of flows in small-to-medium-sized catchments ($10\text{--}10\,000 \text{ km}^2$) and should not be used for larger catchments because of the increasing influence of channel routing effects at larger scales (i.e., evaporation and leakage losses and travel time delays). Third, the baseflow index estimates are not necessarily indicative of the portion of total Q originating from the groundwater—other potential sources include snow and ice melt and the outflow from surface-water bodies. Fourth, prior to computing the Q characteristics from hydrological model output, the surface and subsurface contributions of Q should be summed. Fifth, for many grid cells the runoff coefficient exceeds unity and/or the mean annual Q per unit area exceeds the WorldClim P . Such grid cells are especially common in mountainous and northern regions and may be indicative of P underestimation. Sixth, for regions characterized by multimodal seasonal flow distributions, such as the Pacific coast of northern North America, the timing of the center of mass of Q estimates may be less reliable and require more careful interpretation. Finally, the maps are most accurate in regions with dense Q monitoring networks, such as Europe and the United States, and less accurate in ungauged regions (Fig. 2). If a particular analysis requires a higher level of accuracy, the provided uncertainty maps may be used to mask out uncertain results, or one may restrict the analysis to gauged regions only using the gridded observational maps.

d. Example application

Daily Q estimates from four macroscale hydrological models were diagnosed as an example application of the global maps of Q characteristics produced here. By calculating the difference between the simulation- and observation-based maps of the (transformed) Q characteristics, a number of deficiencies in the Q behavior of the models have been brought to light (Figs. 6–8, Table 8, and Figs. S3.1–S3.13 in the supplemental material). Although there are a host of studies using observed Q to diagnose models in a regional context (e.g., Lohmann et al. 1998, 2004; Robock et al. 2003; Slater et al. 2007; Gusev et al. 2008; Feng and Houser 2008; Choi and Liang 2010; Rodell et al. 2011; Vano et al. 2012; Xia et al. 2012) and at a global scale (e.g., Oki et al. 1999; Döll et al. 2003; Decharme and Douville 2007; Decharme 2007; Zaitchik et al. 2010; Matera et al. 2010; Gosling and Arnell 2011; Haddeland et al. 2011; Zhou et al. 2012), these studies typically used observed Q from a small number of large catchments. The use of a small number of basins limits confidence in the results, while the large size of the basins prevents a spatially detailed analysis of the results and makes it more difficult to distinguish between deficiencies in the forcing, the land surface component, or the routing component. The global maps of Q characteristics produced here enable a spatially explicit diagnosis of models, and since the maps from this study were based on small-to-medium-sized catchments, the confounding influence of the routing component is minimized.

The results for the baseflow-related Q characteristics (BFI1–4, k , and Q99) depended mostly on the runoff schemes of the models and less on the meteorological forcing used. Our findings are in agreement with Slater et al. (2007), who used observed Q from four large (sub)Arctic catchments ($1\,680\,000\text{--}2\,440\,000 \text{ km}^2$) and found that Noah–GLDAS produces too much quickflow under (sub)Arctic conditions, likely because of excessive reduction of the soil infiltration capacity when the surface is frozen (Pan et al. 2003; Niu et al. 2011). The k performance was rather poor for all models (Table 8) because of the baseflow-related parameterizations of the models. The k values of HTESSEL and Noah depend on soil texture [Balsamo et al. (2009) and Schaake et al. (1996), respectively]. On the other hand, PCR-GLOBWB parameterizes k based on geology (Van Beek and Bierkens 2009), while W3RA parameterizes k using an AI-based empirical relationship taken from Peña-Arancibia et al. (2010). However, this study found weak or nonexistent relationships between k and soil texture (SAND, SILT, and CLAY), geology (PERM), or AI (Table 8) using the most comprehensive global datasets available (Table 2).

The poor k performance obtained by the models confirms the difficulty of estimating k (cf. Table 5; Van Dijk 2010; Peña-Arancibia et al. 2010; Beck et al. 2013b). It should be noted that erroneous BFI and/or k estimates by the models may become less noticeable downstream because of the effects of routing.

Our analysis of T50 was restricted to areas with at least some snowfall. The T50 performance depended on the snow and runoff schemes of the models but also on the meteorological forcing used. Our results for T50 (Table 8, Fig. 7) were confirmed in four previous studies. Balsamo et al. (2009) found an early bias in Q timing for HTESSEL using observed Q from the (sub)Arctic Yenisei catchment (2 580 000 km²). Lohmann et al. (2004) found a similar difference pattern in Q timing for Noah using observed Q from 1145 small-to-medium-sized U.S. catchments (23–10 000 km²). Slater et al. (2007) found a too early Q timing for Noah–GLDAS using observed Q from four large (sub)Arctic catchments (1 680 000–2 440 000 km²). Zaitchik et al. (2010) also found a too early timing of Q for Noah–GLDAS but for the globe based on observed Q from 66 catchments (19 000–4 600 000 km²) around the world. The early biases in Q timing suggest that the models poorly represent the processes affecting melt-water refreezing and/or that there are biases in the meteorological forcing.

The QMEAN performance of the models depended on both the model and the meteorological forcing used. Our results for PCR-GLOBWB are in line with Van Beek and Bierkens (2009), who found that the model accurately simulates the spatial variability in Q based on observed Q from 19 large catchments (66 000–6 915 000 km²) around the world. The pronounced QMEAN underestimation by W3RA–Princeton at northern latitudes suggests that improvements in the forcing or model may be required. All models with the exception of PCR-GLOBWB–ERA1 appeared to underestimate QMEAN at the global scale (Table 8, Fig. 8), while all models underestimated QMEAN over mountainous regions (Fig. 8). These findings point to P underestimation despite the gauge undercatch and orographic corrections applied to the forcings of HTESSEL–ERA1 and W3RA–Princeton (Sheffield et al. 2006; Adler et al. 2003; Balsamo et al. 2015). Such P underestimation appears to be common in the high-resolution P data used here (cf. Table 2) as well, as the RC exceeds unity over many northern and mountainous regions for both the observations and the estimates (Table 7 and Fig. S2.18 in the supplemental material).

Overall, the results suggest that the global maps of Q characteristics derived in this study can be employed to refine the parameterization or structure of the examined models in order to obtain better Q behavior (cf. Yadav et al. 2007; Zhang et al. 2008; Castiglioni et al. 2010;

Wagener and Montanari 2011; Lombardi et al. 2012; Pinheiro and Naghettini 2013).

7. Conclusions

In this study, observed Q from 3000 to 4000 small-to-medium-sized catchments (10–10 000 km²) around the globe were used to examine the relationships between catchment attributes (related to climate, topography, land cover, soils, and geology) and Q characteristics. We examined 17 Q characteristics, each quantifying a different aspect of the hydrograph. Additionally, neural network ensembles were used to produce global observation-based maps of Q characteristics, which may be useful for numerous large-scale hydrological applications. The main findings are as follows.

- 1) Climate indices exhibited the strongest relationships with the Q characteristics. In contrast, predictors related to soils and geology showed weak or nonexistent relationships with the Q characteristics, possibly because of data quality. The individual relationships between catchment attributes and Q characteristics were generally weak, suggesting the need for models incorporating multiple predictors to estimate Q characteristics. The four base flow indices (BFI1–4) were highly intercorrelated but showed systematic differences.
- 2) Neural networks were trained to estimate the Q characteristics from catchment attributes. The estimation performance was good to excellent for most Q characteristics and generally exceeded the performance achieved in previous (regional) studies. The trained neural network ensembles were subsequently applied spatially over the entire ice-free land surface to produce global maps of the Q characteristics, providing a unique observation-based global picture of the Q behavior.
- 3) The produced global maps of Q characteristics should be of particular value for the diagnosis of macroscale hydrological models. To test this, the newly produced maps were compared to equivalent maps derived from four macroscale hydrological models. This comparison revealed that, among other things, the models fail to simulate the baseflow recession rate, exhibit early bias in Q timing, and underestimate Q over mountain ranges.

The maps, including uncertainty estimates and observations, have been released as the Global Streamflow Characteristics Dataset (GSCD; freely available via <http://water.jrc.ec.europa.eu>).

Acknowledgments. We thank the Global Runoff Data Centre (GRDC) and the U.S. Geological Survey (USGS)

for providing most of the observed Q data. We gratefully acknowledge the WorldClim developers for making the global climate maps available. We would also like to thank Tom Gleeson (McGill University, Montreal, Canada) for providing the global permeability map; Toby Marthews (University of Oxford, Oxford, United Kingdom) for providing the global map of topographic wetness index; Frederiek Sperna Weiland (Deltares, Delft, the Netherlands) for providing the PCR-GLOBWB–ERA-Interim modeled Q data; and finally, the European Centre for Medium-Range Weather Forecasts (ECMWF; Reading, United Kingdom) for providing the HRES–ERA-Interim modeled Q data. The views expressed herein are those of the authors and do not necessarily reflect those of the European Commission.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.
- Aguado, E., D. Cayan, L. Riddle, and M. Roos, 1992: Climatic fluctuations and the timing of West Coast streamflow. *J. Climate*, **5**, 1468–1483, doi:10.1175/1520-0442(1992)005<1468:CFATTO>2.0.CO;2.
- Andréassian, V., J. Lerat, C. Loumagne, T. Mathevet, C. Michel, L. Oudin, and C. Perrin, 2007: What is really undermining hydrologic science today? *Hydrol. Processes*, **21**, 2819–2822, doi:10.1002/hyp.6854.
- Arnell, N. W., 1995: Grid mapping of river discharge. *J. Hydrol.*, **167**, 39–56, doi:10.1016/0022-1694(94)02626-M.
- Babovic, V., 2005: Data mining in hydrology. *Hydrol. Processes*, **19**, 1511–1515, doi:10.1002/hyp.5862.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeorol.*, **10**, 623–643, doi:10.1175/2008JHM1068.1.
- , and Coauthors, 2015: ERA-Interim/Land: A global land water resources dataset. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:10.5194/hess-19-389-2015.
- Beck, H. E., L. A. Bruijnzeel, A. I. J. M. van Dijk, T. R. McVicar, F. N. Scatena, and J. Schellekens, 2013a: The impact of forest regeneration on streamflow in 12 mesoscale humid tropical catchments. *Hydrol. Earth Syst. Sci.*, **17**, 2613–2635, doi:10.5194/hess-17-2613-2013.
- , A. I. J. M. van Dijk, D. G. Miralles, R. A. M. de Jeu, L. A. Bruijnzeel, T. R. McVicar, and J. Schellekens, 2013b: Global patterns in baseflow index and recession based on streamflow observations from 3394 catchments. *Water Resour. Res.*, **49**, 7843–7863, doi:10.1002/2013WR013918.
- Berghuijs, W. R., R. A. Woods, and M. Hrachowitz, 2014: A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nat. Climate Change*, **4**, 583–586, doi:10.1038/nclimate2246.
- Beven, K. J., 1989: Changing ideas in hydrology—The case of physically-based models. *J. Hydrol.*, **105**, 157–172, doi:10.1016/0022-1694(89)90101-7.
- Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije, Eds., 2013: *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*. Cambridge University Press, 490 pp.
- Bontemps, S., P. Defourny, and E. van Bogaert, 2011: GlobCover 2009, products description and validation report. Tech. rep., ESA GlobCover project, 53 pp. [Available online at http://epic.awi.de/31014/16/GLOBCOVER2009_Validation_Report_2-2.pdf.]
- Boorman, D. B., J. M. Hollist, and A. Lilly, 1995: Hydrology of soil types: A hydrologically based classification of the soils of the United Kingdom. IH Rep. 126, Institute of Hydrology, 137 pp. [Available online at <http://www.ceh.ac.uk/products/publications/documents/ih126hydrologyofsoiltypes.pdf>.]
- Brandes, D., J. G. Hoffmann, and J. T. Mangarillo, 2005: Base flow recession rates, low flows, and hydrologic features of small watersheds in Pennsylvania, USA. *J. Amer. Water Resour. Assoc.*, **41**, 1177–1186, doi:10.1111/j.1752-1688.2005.tb03792.x.
- Brauman, K. A., G. C. Daily, T. K. Duarte, and H. A. Mooney, 2007: The nature and value of ecosystem services: An overview highlighting hydrologic services. *Annu. Rev. Environ. Resour.*, **32**, 67–98, doi:10.1146/annurev.energy.32.031306.102758.
- Brown, J., O. J. Ferrians, J. A. Heginbottom, and E. S. Melnikov, 1997: Circum-Arctic map of permafrost and ground-ice conditions, version 2. National Snow and Ice Data Center, accessed 28 December 2012. [Available online at <http://nsidc.org/data/ggd318>.]
- Bruijnzeel, L. A., 2004: Hydrological functions of tropical forests: Not seeing the soil for the trees. *Agric. Ecosyst. Environ.*, **104**, 185–228, doi:10.1016/j.agee.2004.01.015.
- Budyko, M. I., 1974: *Climate and Life*. Academic Press, 508 pp.
- Castiglioni, S., L. Lombardi, E. Tot, A. Castellarin, and A. Montanari, 2010: Calibration of rainfall–runoff models in ungauged basins: A regional maximum likelihood approach. *Adv. Water Resour.*, **33**, 1235–1242, doi:10.1016/j.advwatres.2010.04.009.
- Choi, H. I., and X. Liang, 2010: Improved terrestrial hydrologic representation in mesoscale land surface models. *J. Hydrometeorol.*, **11**, 797–809, doi:10.1175/2010JHM1221.1.
- Clausen, B., and B. J. F. Biggs, 2000: Flow variables for ecological studies in temperate streams: Groupings based on covariance. *J. Hydrol.*, **237**, 184–197, doi:10.1016/S0022-1694(00)00306-1.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113, doi:10.3354/cr022099.
- Decharme, B., 2007: Influence of runoff parameterization on continental hydrology: Comparison between the Noah and the ISBA land surface models. *J. Geophys. Res.*, **112**, D19108, doi:10.1029/2007JD008463.
- , and H. Douville, 2007: Global validation of the ISBA sub-grid hydrology. *Climate Dyn.*, **29**, 21–37, doi:10.1007/s00382-006-0216-7.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137A**, 553–597, doi:10.1002/qj.828.
- Detenbeck, N. E., V. J. Brady, D. L. Taylor, V. M. Snarski, and S. L. Batterman, 2005: Relationship of stream flow regime in western Lake Superior basin to watershed type characteristics. *J. Hydrol.*, **309**, 258–276, doi:10.1016/j.jhydrol.2004.11.024.
- Devito, K., I. Creed, T. Gan, C. Mendoza, R. Petrone, U. Silins, and B. Smerdon, 2005: A framework for broad-scale classification of hydrologic response units on the Boreal Plain: Is topography the last thing to consider? *Hydrol. Processes*, **19**, 1705–1714, doi:10.1002/hyp.5881.

- Döll, P., F. Kaspar, and B. Lehner, 2003: A global hydrological model for deriving water availability indicators: Model tuning and validation. *J. Hydrol.*, **270**, 105–134, doi:[10.1016/S0022-1694\(02\)00283-4](https://doi.org/10.1016/S0022-1694(02)00283-4).
- Duan, L., T. Liu, X. Wang, Y. Luo, and L. Wu, 2010: Development of a regional regression model for estimating annual runoff in the Hailar River basin of China. *J. Water Resour. Prot.*, **2**, 934–943, doi:[10.4236/jwarp.2010.211111](https://doi.org/10.4236/jwarp.2010.211111).
- Duan, Q., J. Schaake, and V. Koren, 2001: *A priori* estimation of land surface model parameters. *Land Surface Hydrology, Meteorology, and Climate: Observations and Modeling*, V. Lakshmi, J. Albertson, and J. Schaake, Eds., Water Science and Application Series, Vol. 3, Amer. Geophys. Union, 77–94.
- , and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.*, **320**, 3–17, doi:[10.1016/j.jhydrol.2005.07.031](https://doi.org/10.1016/j.jhydrol.2005.07.031).
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:[10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296).
- Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador, 2010: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, **91**, 621–621, doi:[10.1890/09-0889.1](https://doi.org/10.1890/09-0889.1).
- FAO, 2000: Global Forest Resources Assessment 2000: Main report. FAO Forestry Paper 140, 479 pp. [Available online at [ftp://ftp.fao.org/docrep/fao/003/Y1997E/FRA%202000%20Main%20report.pdf](http://ftp.fao.org/docrep/fao/003/Y1997E/FRA%202000%20Main%20report.pdf).]
- Farr, T. G., and Coauthors, 2007: The Shuttle Radar Topography Mission. *Rev. Geophys.*, **45**, RG2004, doi:[10.1029/2005RG000183](https://doi.org/10.1029/2005RG000183).
- Farvolden, R. N., 1963: Geologic controls on ground-water storage and base flow. *J. Hydrol.*, **1**, 219–249, doi:[10.1016/0022-1694\(63\)90004-0](https://doi.org/10.1016/0022-1694(63)90004-0).
- Fekete, B. M., and C. J. Vörösmarty, 2007: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS Publ.*, **309**, 129–136.
- Feng, X., and P. Houser, 2008: An investigation of GSWP-2 Mississippi River basin surface water and energy budgets. *J. Geophys. Res.*, **113**, D15118, doi:[10.1029/2007JD009144](https://doi.org/10.1029/2007JD009144).
- Fennessey, N., and R. M. Vogel, 1990: Regional flow-duration curves for ungauged sites in Massachusetts. *J. Water Resour. Plann. Manage.*, **116**, 530–549, doi:[10.1061/\(ASCE\)0733-9496\(1990\)116:4\(530\)](https://doi.org/10.1061/(ASCE)0733-9496(1990)116:4(530)).
- Gao, Y., H. Xie, N. Lu, T. Yao, and T. Liang, 2010: Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra–Aqua MODIS and Aqua AMSR-E measurements. *J. Hydrol.*, **385**, 23–35, doi:[10.1016/j.jhydrol.2010.01.022](https://doi.org/10.1016/j.jhydrol.2010.01.022).
- Gebert, W. A., D. J. Graczyk, and W. R. Krug, 1987: Average annual runoff in the United States, 1951–80. Hydrologic Investigations Atlas HA-710, U.S. Geological Survey.
- Gleeson, T., L. Smith, N. Moosdorf, J. Hartmann, H. H. Dürr, A. H. Manning, L. P. H. van Beek, and A. M. Jellinek, 2011: Mapping permeability over the surface of the earth. *Geophys. Res. Lett.*, **38**, L02401, doi:[10.1029/2010GL045565](https://doi.org/10.1029/2010GL045565).
- Gosling, S. N., and N. W. Arnell, 2011: Simulating current global river runoff with a global hydrological model: Model revisions, validation, and sensitivity analysis. *Hydrol. Processes*, **25**, 1129–1145, doi:[10.1002/hyp.7727](https://doi.org/10.1002/hyp.7727).
- Groisman, P. Y., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227, doi:[10.1175/1520-0477\(1994\)075<0215:TAOUSP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2).
- Gupta, H. V., C. Perrin, R. Kumar, G. Blöschl, M. Clark, A. Montanari, and V. Andréassian, 2014: Large-sample hydrology: A need to balance depth with breadth. *Hydrol. Earth Syst. Sci.*, **18**, 463–477, doi:[10.5194/hess-18-463-2014](https://doi.org/10.5194/hess-18-463-2014).
- Gusev, E. M., O. N. Nasonova, L. Y. Dzhogan, and E. E. Kovalev, 2008: The application of the land surface model for calculating river runoff in high latitudes. *Water Resour.*, **35**, 171–184, doi:[10.1134/S009780780802005X](https://doi.org/10.1134/S009780780802005X).
- Gustard, A., and K. M. Irving, 1994: Classification of the low flow response of European soils. *IAHS Publ.*, **221**, 113–117.
- , A. Bullock, and J. M. Dixon, 1992: Low flow estimation in the United Kingdom. Tech. Rep. 108, Institute of Hydrology, Wallingford, United Kingdom, 292 pp.
- Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.*, **12**, 869–884, doi:[10.1175/2011JHM1324.1](https://doi.org/10.1175/2011JHM1324.1).
- Hall, D. K., V. V. Salomonson, and G. A. Riggs, 2006: MODIS/Aqua snow cover daily L3 global 0.05deg CMG, version 5. National Snow and Ice Data Center, accessed 11 November 2014, doi:[10.5067/EW53FPU9NAS6](https://doi.org/10.5067/EW53FPU9NAS6).
- Hall, F. R., 1968: Base-flow recessions—A review. *Water Resour. Res.*, **4**, 973–983, doi:[10.1029/WR004i005p00973](https://doi.org/10.1029/WR004i005p00973).
- Hargreaves, G. L., G. H. Hargreaves, and J. P. Riley, 1985: Irrigation water requirements for Senegal River basin. *J. Irrig. Drain. Eng.*, **111**, 265–275, doi:[10.1061/\(ASCE\)0733-9437\(1985\)111:3\(265\)](https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265)).
- He, Y., A. Bárdossy, and E. Zehe, 2011: A review of regionalisation for continuous streamflow simulation. *Hydrol. Earth Syst. Sci.*, **15**, 3539–3553, doi:[10.5194/hess-15-3539-2011](https://doi.org/10.5194/hess-15-3539-2011).
- Hengl, T., and Coauthors, 2014: SoilGrids1km—Global soil information based on automated mapping. *PLOS One*, **9**, e105992, doi:[10.1371/journal.pone.0105992](https://doi.org/10.1371/journal.pone.0105992); Corrigendum: **9**, e114788, doi:[10.1371/journal.pone.0114788](https://doi.org/10.1371/journal.pone.0114788).
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, **25**, 1965–1978, doi:[10.1002/joc.1276](https://doi.org/10.1002/joc.1276).
- Hope, A., and R. Bart, 2012: Synthetic monthly flow duration curves for the Cape Floristic region, South Africa. *Water SA*, **38**, 191–200, doi:[10.4314/wsa.v38i2.4](https://doi.org/10.4314/wsa.v38i2.4).
- Horton, R. E., 1933: The role of infiltration in the hydrological cycle. *Trans. Amer. Geophys. Union*, **14**, 446–460.
- Hrachowitz, M., and Coauthors, 2013: A decade of predictions in ungauged basins (PUB)—A review. *Hydrol. Sci. J.*, **58**, 1198–1255, doi:[10.1080/02626667.2013.803183](https://doi.org/10.1080/02626667.2013.803183).
- Hughes, G. O., 1997: An analysis of baseflow recession in the Republic of South Africa. M.S. thesis, Department of Agricultural Engineering, University of Natal, 205 pp. [Available online at <http://researchspace.ukzn.ac.za/xmlui/handle/10413/5416>.]
- Jia, G. J., H. E. Epstein, and D. A. Walker, 2003: Greening of Arctic Alaska, 1981–2001. *Geophys. Res. Lett.*, **30**, 2067, doi:[10.1029/2003GL018268](https://doi.org/10.1029/2003GL018268).
- Johnson, D. H., 1999: The insignificance of statistical significance testing. *J. Wildl. Manage.*, **63**, 763–772, doi:[10.2307/3802789](https://doi.org/10.2307/3802789).
- Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrol. Earth Syst. Sci.*, **17**, 2845–2013, doi:[10.5194/hess-17-2845-2013](https://doi.org/10.5194/hess-17-2845-2013).
- Kim, U., and J. J. Kaluarachchi, 2008: Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River basin, Ethiopia. *J. Hydrol.*, **362**, 39–56, doi:[10.1016/j.jhydrol.2008.08.016](https://doi.org/10.1016/j.jhydrol.2008.08.016).

- Krakauer, N. Y., and M. Temimi, 2011: Stream recession curves and storage variability in small watersheds. *Hydrol. Earth Syst. Sci.*, **15**, 2377–2389, doi:10.5194/hess-15-2377-2011.
- Laaha, G., and G. Blöschl, 2006: A comparison of low flow regionalisation methods—Catchment grouping. *J. Hydrol.*, **323**, 193–214, doi:10.1016/j.jhydrol.2005.09.001.
- Lacey, G., and R. Grayson, 1998: Relating baseflow to catchment properties in south-eastern Australia. *J. Hydrol.*, **204**, 231–250, doi:10.1016/S0022-1694(97)00124-8.
- Lehner, B., 2012: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network. Tech. Rep. 41, Global Runoff Data Centre, Federal Institute of Hydrology, Koblenz, Germany, 12 pp.
- , and P. Döll, 2004: Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.*, **296**, 1–22, doi:10.1016/j.jhydrol.2004.03.028.
- Leibowitz, S. G., P. J. Wigington Jr., R. L. Comeleo, and J. L. Ebersole, 2012: A temperature–precipitation-based model of thirty-year mean snowpack accumulation and melt in Oregon, USA. *Hydrol. Processes*, **26**, 741–759, doi:10.1002/hyp.8176.
- Lohmann, D., and Coauthors, 1998: The Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) phase 2(c) Red–Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes. *Global Planet. Change*, **19**, 161–179, doi:10.1016/S0921-8181(98)00046-0.
- , and Coauthors, 2004: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *J. Geophys. Res.*, **109**, D07S91, doi:10.1029/2003JD003517.
- Lombardi, L., E. Toth, A. Castellarin, A. Montanari, and A. Brath, 2012: Calibration of a rainfall–runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime. *Phys. Chem. Earth*, **42–44**, 77–84, doi:10.1016/j.pce.2011.05.013.
- Longobardi, A., and P. Villani, 2008: Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: Role of the catchment permeability index. *J. Hydrol.*, **355**, 63–75, doi:10.1016/j.jhydrol.2008.03.011.
- Marthews, T. R., S. J. Dadson, B. Lehner, S. Abele, and N. Gedney, 2015: A high-resolution global dataset of topographic index values for use in large-scale hydrological modelling. *Hydrol. Earth Syst. Sci.*, **19**, 91–104, doi:10.5194/hess-19-91-2015.
- Materia, S., P. A. Dirmeyer, Z. Guo, A. Alessandri, and A. Navarra, 2010: The sensitivity of simulated river discharge to land surface representation and meteorological forcings. *J. Hydrometeorol.*, **11**, 334–351, doi:10.1175/2009JHM1162.1.
- Mazvimavi, D., A. M. J. Meijerink, H. H. G. Savenije, and A. Stein, 2005: Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe. *Phys. Chem. Earth*, **30**, 639–647, doi:10.1016/j.pce.2005.08.003.
- Milly, P. C. D., 1994: Climate, soil water storage, and the average annual water balance. *Water Resour. Res.*, **30**, 2143–2156, doi:10.1029/94WR00586.
- Nasonova, O. N., Y. M. Gusev, and Y. E. Kovalev, 2009: Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: A case study using MOPEX materials. *J. Hydrometeorol.*, **10**, 1128–1150, doi:10.1175/2009JHM1083.1.
- Neff, B. P., S. M. Day, A. R. Piggott, and L. M. Fuller, 2005: Base flow in the Great Lakes basin. Scientific Investigations Rep. 2005-5217, USGS, 23 pp. [Available online at <http://pubs.usgs.gov/sir/2005/5217/>.]
- Nicholls, N., 2001: Commentary and analysis: The insignificance of significance testing. *Bull. Amer. Meteor. Soc.*, **82**, 981–986, doi:10.1175/1520-0477(2001)082<0981:CAATIO>2.3.CO;2.
- Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, doi:10.1029/2010JD015139.
- Nuttle, W., 2002: Taking stock of water resources. *Eos, Trans. Amer. Geophys. Union*, **83**, 513, doi:10.1029/2002EO000358.
- Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assessment of annual runoff from land surface model using Total Runoff Integration Pathways (TRIP). *J. Meteor. Soc. Japan*, **77**, 235–255.
- Ol'dekop, E. M., 1911: On evaporation from the surface of river basins (in Russian). Transactions on Meteorological Observations, Lur-evskogo, University of Tartu, Tartu, Estonia.
- Olden, J. D., and N. L. Poff, 2003: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Res. Appl.*, **19**, 101–121, doi:10.1002/rra.700.
- Oudin, L., V. Andréassian, J. Lerat, and C. Michel, 2008: Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments. *J. Hydrol.*, **357**, 303–316, doi:10.1016/j.jhydrol.2008.05.021.
- Pan, M., and Coauthors, 2003: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model-simulated snow cover extent. *J. Geophys. Res.*, **108**, 8850, doi:10.1029/2003JD003994.
- Peel, M. C., F. H. S. Chiew, A. W. Western, and T. A. McMahon, 2000: Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Rep. for the Australian National Land and Water Resources Audit, Centre for Environmental Applied Hydrology, University of Melbourne, Melbourne, Australia, 37 pp.
- , B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644, doi:10.5194/hess-11-1633-2007.
- , T. A. McMahon, and B. L. Finlayson, 2010: Vegetation impact on mean annual evapotranspiration at a global catchment scale. *Water Resour. Res.*, **46**, W09508, doi:10.1029/2009WR008233.
- Peña-Arancibia, J. L., A. I. J. M. Van Dijk, M. Mulligan, and L. A. Bruijnzeel, 2010: The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrol. Earth Syst. Sci.*, **14**, 2193–2205, doi:10.5194/hess-14-2193-2010.
- Petersen, T., N. Devineni, and A. Sankarasubramanian, 2012: Seasonality of monthly runoff over the continental United States: Causality and relations to mean annual and mean monthly distributions of moisture and energy. *J. Hydrol.*, **468–469**, 139–150, doi:10.1016/j.jhydrol.2012.08.028.
- Pettyjohn, W. A., and R. Henning, 1979: Preliminary estimate of ground-water recharge rates, related streamflow and water quality in Ohio. Water Resources Centre Project Completion Rep. 552, Ohio State University, 323 pp.
- Pike, J. G., 1964: The estimation of annual run-off from meteorological data in a tropical climate. *J. Hydrol.*, **2**, 116–123, doi:10.1016/0022-1694(64)90022-8.
- Pinheiro, V. B., and M. Naghettini, 2013: Calibration of the parameters of a rainfall–runoff model in ungauged basins using synthetic flow duration curves as estimated by regional analysis. *J. Hydrol. Eng.*, **18**, 1617–1626, doi:10.1061/(ASCE)JHE.1943-5584.0000737.
- Poff, N. L. R., D. Allan, M. B. Bain, J. R. Karr, and K. L. Prestegard, 1997: The natural flow regime. *BioScience*, **47**, 769–784, doi:10.2307/1313099.
- Potter, N. J., L. Zhang, P. C. D. Milly, T. A. McMahon, and A. J. Jakeman, 2005: Effects of rainfall seasonality and soil moisture

- capacity on mean annual water balance for Australian catchments. *Water Resour. Res.*, **41**, W06007, doi:10.1029/2004WR003697.
- Price, K., 2011: Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Prog. Phys. Geogr.*, **35**, 465–492, doi:10.1177/0309133311402714.
- Robock, A., and Coauthors, 2003: Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season. *J. Geophys. Res.*, **108**, 8846, doi:10.1029/2002JD003245.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, doi:10.1175/BAMS-85-3-381.
- , E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoin, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Processes*, **25**, 4082–4092, doi:10.1002/hyp.8369.
- Rosero, E., L. E. Gulden, and Z. Yang, 2011: Ensemble evaluation of hydrologically enhanced Noah-LSM: Partitioning of the water balance in high-resolution simulations over the Little Washita River experimental watershed. *J. Hydrometeorol.*, **12**, 45–64, doi:10.1175/2010JHM1228.1.
- Royall, R. M., 1986: The effect of sample size on the meaning of significance tests. *Amer. Stat.*, **40**, 313–315.
- Salinas, J. L., G. Laaha, M. Rogger, J. Parajka, A. Viglione, M. Sivapalan, and G. Blöschl, 2013: Comparative assessment of predictions in ungauged basins—Part 2: Flood and low flow studies. *Hydrol. Earth Syst. Sci.*, **17**, 2637–2652, doi:10.5194/hess-17-2637-2013.
- Sanborn, S. C., and B. P. Bledsoe, 2006: Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *J. Hydrol.*, **325**, 241–261, doi:10.1016/j.jhydrol.2005.10.018.
- Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. Mitchell, and F. Chen, 1996: Simple water balance model for estimating runoff at different spatial and temporal scales. *J. Geophys. Res.*, **101**, 7461, doi:10.1029/95JD02892.
- Schreiber, P., 1904: Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa. *Meteor. Z.*, **21**, 441–452.
- Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercomparisons. *Atmos. Res.*, **92**, 376–380, doi:10.1016/j.atmosres.2009.01.016.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, doi:10.1175/JCLI3790.1.
- Sivapalan, M., 2003: Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrol. Processes*, **17**, 3163–3170, doi:10.1002/hyp.5155.
- , G. Blöschl, L. Zhang, and R. Vertessy, 2003: Downward approach to hydrological prediction. *Hydrol. Processes*, **17**, 2101–2111, doi:10.1002/hyp.1425.
- Slater, A. G., T. J. Bohn, J. L. McCreight, M. C. Serreze, and D. P. Lettenmaier, 2007: A multimodel simulation of pan-Arctic hydrology. *J. Geophys. Res.*, **112**, G04S45, doi:10.1029/2006JG000303.
- Sloto, R. A., and M. Y. Crouse, 1996: HYSEP: A computer program for streamflow hydrograph separation and analysis. USGS Water-Resources Investigations Rep. 96-4040, 46 pp. [Available online at <http://water.usgs.gov/software/HYSEP/code/doc/hysep.pdf>.]
- Smakhtin, V., 2001: Low flow hydrology: A review. *J. Hydrol.*, **240**, 147–186, doi:10.1016/S0022-1694(00)00340-1.
- , D. A. Hughes, and E. Creuse-Naudin, 1997: Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrol. Sci. J.*, **42**, 919–936, doi:10.1080/02626669709492088.
- Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2004: Changes in snowmelt runoff timing in western North America under a ‘business as usual’ climate change scenario. *Climatic Change*, **62**, 217–232, doi:10.1023/B:CLIM.0000013702.22656.e8.
- , —, and —, 2005: Changes toward earlier streamflow timing across western North America. *J. Climate*, **18**, 1136–1155, doi:10.1175/JCLI3321.1.
- Tait, A., R. Henderson, R. Turner, and X. Zheng, 2006: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *Int. J. Climatol.*, **26**, 2097–2115, doi:10.1002/joc.1350.
- Trambauer, P., S. Maskeya, H. Winsemius, M. Werner, and S. Uhlenbrook, 2013: A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa. *Phys. Chem. Earth*, **66**, 16–26, doi:10.1016/j.pce.2013.07.003.
- Van Beek, L. P. H., and M. F. P. Bierkens, 2009: The global hydrological model PCR-GLOBWB: Conceptualization, parameterization and verification. Tech. rep., Utrecht University, 53 pp. [Available online at <http://vanbeek.geo.uu.nl/supinfo/vanbeekbierkens2009.pdf>.]
- Van Dijk, A. I. J. M., 2010: Climate and terrain factors explaining streamflow response and recession in Australian catchments. *Hydrol. Earth Syst. Sci.*, **14**, 159–169, doi:10.5194/hess-14-159-2010.
- , J. L. Peña-Arancibia, and L. A. Bruijnzeel, 2012: Land cover and water yield: Inference problems when comparing catchments with mixed land cover. *Hydrol. Earth Syst. Sci.*, **16**, 3461–3473, doi:10.5194/hess-16-3461-2012.
- , —, E. F. Wood, J. Sheffield, and H. E. Beck, 2013: Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resour. Res.*, **49**, 2729–2746, doi:10.1002/wrcr.20251.
- Vano, J. A., T. Das, and D. P. Lettenmaier, 2012: Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *J. Hydrometeorol.*, **13**, 932–949, doi:10.1175/JHM-D-11-069.1.
- Viglione, A., P. Claps, and F. Laio, 2007: Mean annual runoff estimation in north-western Italy. *Water Resources Assessment and Management under Water Scarcity Scenarios*, G. La Loggia, Ed., CDSU, 97–122.
- Vogel, R. M., and C. N. Kroll, 1996: Estimation of baseflow recession constants. *Water Resour. Manage.*, **10**, 303–320, doi:10.1007/BF00508898.
- , I. Wilson, and C. Daly, 1999: Regional regression models of annual streamflow for the United States. *J. Irrig. Drain. Eng.*, **125**, 148–157, doi:10.1061/(ASCE)0733-9437(1999)125:3(148).
- Wagener, T., and A. Montanari, 2011: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resour. Res.*, **47**, W06301, doi:10.1029/2010WR009469.
- Wahl, K. L., and T. L. Wahl, 1995: Determining the flow of comal springs at New Braunfels, Texas. *Proc. First Int. Conf. on Water Resources Engineering*, San Antonio, TX, American Society of Civil Engineers, 77–86.
- Walsh, R. P. D., and D. M. Lawler, 1981: Rainfall seasonality: Description, spatial patterns and change through time. *Weather*, **36**, 201–208, doi:10.1002/j.1477-8696.1981.tb05400.x.
- Wolock, D. M., 2003: Base-flow index grid for the conterminous United States. USGS Open-File Rep. 03-263. [Available

- online at <http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml>.]
- , and G. J. McCabe, 1999: Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Res.*, **11**, 149–159, doi:[10.3354/cr011149](https://doi.org/10.3354/cr011149).
- Xia, Y., and Coauthors, 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. *J. Geophys. Res.*, **117**, D03110, doi:[10.1029/2011JD016051](https://doi.org/10.1029/2011JD016051).
- Yadav, M., T. Wagener, and H. Gupta, 2007: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Adv. Water Resour.*, **30**, 1756–1774, doi:[10.1016/j.advwatres.2007.01.005](https://doi.org/10.1016/j.advwatres.2007.01.005).
- Zaitchik, B. F., M. Rodell, and F. Olivera, 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resour. Res.*, **46**, W06507, doi:[10.1029/2009WR007811](https://doi.org/10.1029/2009WR007811).
- Zhang, L., K. Hickel, W. R. Dawes, F. H. S. Chiew, A. W. Western, and P. R. Briggs, 2004: A rational function approach for estimating mean annual evapotranspiration. *Water Resour. Res.*, **40**, W02502, doi:[10.1029/2003WR002710](https://doi.org/10.1029/2003WR002710).
- , N. Potter, K. Hickel, Y. Zhang, and Q. Shao, 2008: Water balance modeling over variable time scales based on the Budyko framework—Model development and testing. *J. Hydrol.*, **360**, 117–131, doi:[10.1016/j.jhydrol.2008.07.021](https://doi.org/10.1016/j.jhydrol.2008.07.021).
- Zhou, X., Y. Zhang, Y. Wang, H. Zhang, J. Vaze, L. Zhang, Y. Yang, and Y. Zhou, 2012: Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *J. Hydrol.*, **470–471**, 269–279, doi:[10.1016/j.jhydrol.2012.09.002](https://doi.org/10.1016/j.jhydrol.2012.09.002).