Global maps of streamflow characteristics based on observations from many small catchments

HYLKE BECK * AND AD DE ROO

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

ALBERT VAN DIJK

Fenner School of Environment & Society, Australian National University, Canberra, ACT, Australia

ABSTRACT

Streamflow (Q) estimation in ungauged catchments is one of the greatest challenges facing hydrologists. We used observed Q from approximately 7500 small catchments (< 10000 km^2) around the globe to train neural network ensembles to estimate Q characteristics from 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. Training coefficients of determination for the estimation of the Q characteristics ranged from 0.56 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 the least important, perhaps due to data quality. The trained neural network ensembles were subsequently applied spatially over the entire ice-free land surface including ungauged regions, resulting in global maps of the Q characteristics (0.125° resolution). These maps possess several unique features: 1) they represent purely observation-driven estimates; 2) are based on an unprecedentedly large set of catchments; and 3) have associated uncertainty estimates. The maps can be used for various hydrological applications, including the diagnosis of macro-scale hydrological models. To demonstrate this, the produced maps were compared to equivalent maps derived from the simulated daily Q of four macro-scale hydrological models, highlighting various opportunities for improvement in model Q behavior. The produced dataset is available via `http://water.jrc.ec.europa.eu`.

1. Introduction

Quantitative knowledge of streamflow (Q) and its spatial and temporal distribution is important for hydropower production, drinking water supply, industrial uses, irrigation systems, aquatic habitats, recreation, and sediment and contaminant transport, among others (Poff et al. 1997; Nuttle 2002; Brauman et al. 2007). Perhaps the greatest obstacle to advancing current understanding is that the majority of the 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 reasonably 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; 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; Beck et al. 2013b for reviews). Being 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, these models were typically 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 were not applied to an independent set of catchments to evaluate the generalization capability. In an effort to overcome these limitations, Beck et al. (2013b) produced neural network models for two baseflow-related Q characteristics based on data from 3394 catchments around the globe (Beck et al. 2013b).

The present study aims to produce global maps for a more comprehensive selection of Q characteristics. The subsidiary aims are three-fold: 1) to investigate the relationships between

Q characteristics and climate and physiographic characteristics at global scale; 2) to produce global maps of Q characteristics; and 3) to compare the produced maps to equivalent maps derived from macro-scale hydrological models, to test their usefulness for 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 baseflow (BFI1–4, k, Q95, and Q99), others to peak flow (Q1), the shape of the flow duration curve (Q1–99), the mean water balance (RC and QMEAN), or flow timing (T50). Fig. 1 gives an example of baseflow computed using the four techniques considered here (BFI1–4). Note that baseflow 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).

3. Data

a. Climate and physiographic characteristics

Table 2 lists the climate and physiographic characteristics used as predictors of the Q characteristics. Among the 19 selected predictors, eight were related to climate, three to topography, three to land cover, one to geology, and four to soils. A similar selection of predictors was used by Beck et al. (2013b). However, we used regional P and T data for the USA, regional P data for New Zealand, CGIAR-CSI SRTM and GTOPO30 data for ELEV and SLO, WWF GLWD-3 data (Lehner and Döll 2004) for fW, FAO (2000) data for fTC, and SPOT-VGT S10 data for NDVI. Additionally, we used the Leibowitz et al. (2012) model to compute SNOW, added the predictor TWI, used transformed instead of untransformed mean annual precipitation (P_{trans}), and used the reciprocal of the humidity index (AI). All data used for the predictors have a resolution ≤ 1 km^2. 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.

b. Observed streamflow

The observed Q data in this study were compiled from three sources. First, daily Q and catchment boundary data for 9169 USA stations that were part of the Gages for Evaluating Streamflow (GAGES) II database (Falcone et al. 2010) were downloaded from the U.S. Geological Survey (USGS) website (`http://water.usgs.gov`). Daily and monthly Q data for, respectively, 1961 and 3357 stations around the globe and associated catchment boundaries (Lehner 2012) were requested from 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 data set. This resulted in an initial dataset consisting of 14 808 catchments, of which the 3357 GRDC catchments with only monthly Q data were considered only 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 at least some snowfall (defined by SNOW > 10 mm), which tend to have a pronounced Q seasonality.

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 GlobCover v2.3 map; Bontemps et al. 2011) or subject to irrigation (using the Global Irrigated Area Map; `http://www.iwmigiam.org`). Second, to obtain reliable estimates of the Q characteristics, the Q record length was required to be > 10 yr, not necessarily consecutive (cf. Kauffeldt et al. 2013). Third, to minimize the effects of channel routing (evaporation and leakage losses and travel time delays), the catchment area was limited to $< 10\,000$ 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 km^2. Fifth and finally, to avoid outliers from dominating the training, catchments with values of the Q characteristics outside the acceptable ranges listed in Table 3 were excluded. Determination of acceptable ranges for the Q characteristics was guided by inspection of histograms of catchment values of the Q characteristics. Table 3 lists the number of catchments that fulfilled the selection criteria for each Q characteristic. Fig. 2 shows the locations of the catchments. All Q data were converted to mm yr^{-1} using the provided catchment areas.

c. Modeled streamflow

As an example application of the global maps of Q characteristics produced here, a comparison was made to equivalent maps derived from four macro-scale hydrological models. The first model, HTESSEL+ERAIR, uses the Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land (HTESSEL) land surface scheme (LSM; Balsamo et al. 2009) driven by the ERA-Interim reanalysis forcing (Dee et al. 2011) corrected using Global Precipitation Climatology Project (GPCP) v2.1 data (Adler et al. 2003; Balsamo et al. 2013). The model has a daily temporal and $\sim$0.7° spatial resolution and was available for the period 1979–2010.

The second model, Noah+GLDAS, is based on the Noah LSM (v2.7.1; Schaake et al. 1996; Ek et al. 2003) driven by the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004). The data resolution. The data were obtained from the NASA Goddard Earth Science (GES) Data and Information Service Center (DISC; `http://disc.sci.gsfc.nasa.gov`) and cover the period 2000–present. The data have a 0.25° spatial

Q characteristic	Description and computation
BFI1 (−)	Baseflow index, defined as ratio of long-term baseflow 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 five days (cf. Peña-Arancibia et al. 2010).
BFI2 (−)	Baseflow index, computed from daily Q data following the local-minimum method described in Pettyjohn and Henning (1979) and Sloto and Crouse (1996), with the "duration of surface runoff" (N) set to five days.
BFI3 (−)	Baseflow index, computed from daily Q data following the sliding-interval method described in Pettyjohn and Henning (1979) and Sloto and Crouse (1996), with N set to three days.
BFI4 (−)	Baseflow index, computed from daily Q data following the procedure described in Institute of Hydrology (1980) and Gustard et al. (1992), which takes the minima at five-day non-overlapping intervals and subsequently connects the valleys in this series of minima to generate baseflow.
k (d^{-1})	Baseflow recession constant, defined as the rate of baseflow decay (Vogel and Kroll 1996). Computed from daily Q data following Van Dijk (2010), with the "window size" set to five days and days with zero flow ignored.
Q1, Q5, Q10, Q20, Q50, Q80, Q90, Q95, Q99 ($mm\ d^{-1}$)	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 sensu the United States Geological Survey (USGS) as the 12-month period from October to September in the Northern Hemisphere and April to March in the Southern Hemisphere. T50 can be computed from both daily Q data and monthly Q data linearly interpolated to daily values.
RC (−)	Runoff coefficient, the ratio of Q to mean annual precipitation (P). See Table 2 for details on the precipitation data used. Can be computed from daily and monthly Q data.
QMEAN ($mm\ yr^{-1}$)	Mean annual Q per unit area. Can be computed from daily and monthly Q data.

Table 1: The Q characteristics considered here.

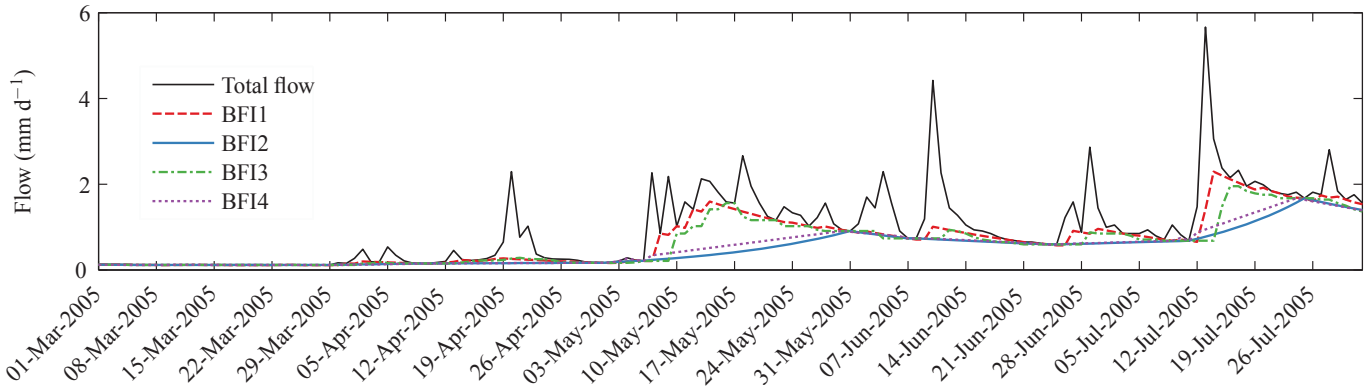

Figure 1: Time series of observed Q from United States Geological Survey (USGS) catchment no. 50145395 (Rio Casei above Hacienda Casei, Puerto Rico; 19 km^2) and baseflow computed using the different techniques (BFI1–4).

and 3-hourly temporal resolution, and were aggregated to daily means for this analysis.

The third model, PCR-GLOBWB+ERAI, was based on the PCRaster Global Water Balance (PCR-GLOBWB) global hydrological model (GHM; Van Beek and Bierkens 2009) driven by the ERA-Interim reanalysis dataset (Dee et al. 2011). The data have a daily temporal and 0.5° spatial resolution and were available for the period 2003–2010.

The fourth and final model, W3RA+Princeton, uses the World Wide Water Resources Assessment (W3RA) GHM (v1; 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 these may be misleading when

Predictor	Description	Calculation and data source	Resolution
Climate			
AI (−)	Aridity index	Calculated as: AI = PET/P, where P is the mean annual precipitation and PET 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: $P_{si} = P_{yr}^{-1} \sum \left\lvert P_m - P_{yr}/12 \right\rvert$, where P_{yr} and P_m are, respectively, the mean annual and monthly precipitation 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), PRISM (Daly et al. 1994) for the USA, 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: $PET_{si} = PET_{yr}^{-1} \sum \left\lvert PET_m - PET_{yr}/12 \right\rvert$, where PET_{yr} and PET_m are, respectively, the mean annual and monthly potential evaporation 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 USA.	∼1 km
SNOW (mm)	Mean snow-water equivalent depth	Calculated following Leibowitz et al. (2012). A maximum value of 200 mm was used to prevent extrapolation too far beyond the training data range. See P_{trans}, PET, and TA for data sources.	∼1 km
Topography			
TWI (−)	Mean topographic wetness index (Beven and Kirkby 1979)	Global map of topographic wetness index (Marthews et al. 2014).	∼500 m
SLO (°)	Mean surface slope	CGIAR-CSI SRTM v4.1 (Farr et al. 2007) for latitudes $< 60°$N, GTOPO30 (`http://lta.cr.usgs.gov/GTOPO30`) for latitudes $\geq 60°$N.	∼90 m, ∼1 km
ELEV (m asl)	Mean surface elevation	Idem.	∼90 m, ∼1 km
Land cover			
fW (−)	Fraction of lakes and reservoirs	WWF GLWD-3 (Lehner and Döll 2004).	∼1 km
fTC (−)	Fraction of forest	FAO FRA 2000 forest cover (FAO 2000).	∼1 km
NDVI (−)	Mean Normalized Difference Vegetation Index	SPOT-VGT S10 10-day maximum value composites (`http://www.vgt.vito.be`), mean of 2005.	∼1 km
Geology and soils			
PERM ($\log_{10}$ m^2)	Mean permeability of geology	Global permeability map (Gleeson et al. 2011).	∼1 km
GRAV (%)	Mean soil gravel content	HWSD v1.21 (FAO/IIASA 2012), mean of top- and subsoil values.	∼1 km
SAND (%)	Mean soil sand content	HWSD v1.21 (FAO/IIASA 2012), mean of top- and subsoil values, and STATSGO (Wolock et al. 2004) for the USA.	∼1 km
SILT (%)	Mean soil silt content	HWSD v1.21 (FAO/IIASA 2012), mean of top- and subsoil values.	∼1 km
CLAY (%)	Mean soil clay content	HWSD v1.21 (FAO/IIASA 2012), mean of top- and subsoil values, and ASRIS (Johnston et al. 2003) for Australia.	∼1 km

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

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) was used to derive global maps of the Q characteristics. Briefly summarized, Beck et al. (2013b) trained neural network ensembles comprising ten members derived from ten cross-validation iterations to estimate the Q characteristics BFI1 and k from climate and physiographic characteristics of the catchments. Prior to the estimation using the neural networks, the values of the Q characteristics were transformed to make the catchment distribution of values resemble a normal distribution, while the values were back-transformed after the estimation. Table 3 lists the transformation equations used for the 17 Q characteristics considered here. Each cross-validation iteration excluded a different testing subset of catchments consisting of 10 % of the complete catchment set, with the remaining subset of catchments further subdivided into training and validation subsets. 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. Global maps of the Q characteristics (0.125°) were subsequently obtained by applying the ten trained neural networks (one for each cross-validation iteration) to the entire ice-free land surface, computing the median of the ten estimates for each grid cell, and back transforming the result using the inverse transformation equations listed in Table 3. The ice-covered portion of the land surface was determined using climate type EF (polar, frost) of the Köppen-Geiger climate type map (Peel et al. 2007; Supplementary material Fig. S2.1). Associated uncertainty maps were computed from the standard deviation of the ten transformed estimates for each grid cell (see Beck et al. 2013b for further details).

c. Example application

As an example application of the new global maps of Q characteristics, a comparison was made to equivalent maps derived from daily Q estimates of four macro-scale hydrological models in order to reveal deficiencies in the model and the forcing. Two LSMs (HTESSEL+ERAIR and Noah+GLDAS) and two GHMs

Q characteristic	Transformation equation	Inverse transformation equation	Acceptable range	Number of catchments
BFI1	$\mathrm{BFI1}_{\mathrm{trans}} = \mathrm{BFI1}^2$	$\mathrm{BFI1} = \sqrt{\mathrm{BFI1}_{\mathrm{trans}}}$	0–1	7479
BFI2	$\mathrm{BFI2}_{\mathrm{trans}} = \mathrm{BFI2}$	$\mathrm{BFI2} = \mathrm{BFI2}_{\mathrm{trans}}$	0–1	7479
BFI3	$\mathrm{BFI3}_{\mathrm{trans}} = \mathrm{BFI3}^2$	$\mathrm{BFI3} = \sqrt{\mathrm{BFI3}_{\mathrm{trans}}}$	0–1	7479
BFI4	$\mathrm{BFI4}_{\mathrm{trans}} = \mathrm{BFI4}$	$\mathrm{BFI4} = \mathrm{BFI4}_{\mathrm{trans}}$	0–1	7479
k	$k_{\mathrm{trans}} = \ln(k)$	$k = \exp(k_{\mathrm{trans}})$	0.01–0.41 d^{-1}	7412
Q1	$\mathrm{Q1}_{\mathrm{trans}} = \sqrt{\mathrm{Q1}}$	$\mathrm{Q1} = \mathrm{Q1}^2_{\mathrm{trans}}$	0–46.24 mm d^{-1}	7390
Q5	$\mathrm{Q5}_{\mathrm{trans}} = \sqrt{\mathrm{Q5}}$	$\mathrm{Q5} = \mathrm{Q5}^2_{\mathrm{trans}}$	0–22.09 mm d^{-1}	7398
Q10	$\mathrm{Q10}_{\mathrm{trans}} = \sqrt{\mathrm{Q10}}$	$\mathrm{Q10} = \mathrm{Q10}^2_{\mathrm{trans}}$	0–15.21 mm d^{-1}	7389
Q20	$\mathrm{Q20}_{\mathrm{trans}} = \sqrt{\mathrm{Q20}}$	$\mathrm{Q20} = \mathrm{Q20}^2_{\mathrm{trans}}$	0–10.24 mm d^{-1}	7383
Q50	$\mathrm{Q50}_{\mathrm{trans}} = \sqrt{\mathrm{Q50}}$	$\mathrm{Q50} = \mathrm{Q50}^2_{\mathrm{trans}}$	0–4.41 mm d^{-1}	7373
Q80	$\mathrm{Q80}_{\mathrm{trans}} = \sqrt{\mathrm{Q80}}$	$\mathrm{Q80} = \mathrm{Q80}^2_{\mathrm{trans}}$	0–2.56 mm d^{-1}	7420
Q90	$\mathrm{Q90}_{\mathrm{trans}} = \sqrt{\mathrm{Q90}}$	$\mathrm{Q90} = \mathrm{Q90}^2_{\mathrm{trans}}$	0–1.69 mm d^{-1}	7399
Q95	$\mathrm{Q95}_{\mathrm{trans}} = \sqrt{\mathrm{Q95}}$	$\mathrm{Q95} = \mathrm{Q95}^2_{\mathrm{trans}}$	0–1.44 mm d^{-1}	7404
Q99	$\mathrm{Q99}_{\mathrm{trans}} = \sqrt{\mathrm{Q99}}$	$\mathrm{Q99} = \mathrm{Q99}^2_{\mathrm{trans}}$	0–1.12 mm d^{-1}	7408
T50	$\mathrm{T50}_{\mathrm{trans}} = \mathrm{T50}$	$\mathrm{T50} = \mathrm{T50}_{\mathrm{trans}}$	1–366	5196
RC	$\mathrm{RC}_{\mathrm{trans}} = \sqrt{\mathrm{RC}}$	$\mathrm{RC} = \mathrm{RC}^2_{\mathrm{trans}}$	0–1.69	8366
QMEAN	$\mathrm{QMEAN}_{\mathrm{trans}} = \sqrt{\mathrm{QMEAN}}$	$\mathrm{QMEAN} = \mathrm{QMEAN}^2_{\mathrm{trans}}$	0–3600 mm yr^{-1}	8437

Table 3: For each Q characteristic, the transformation equations used to improve the normality of the data, the acceptable range used to remove outliers, and the number of catchments that fulfilled the selection criteria.

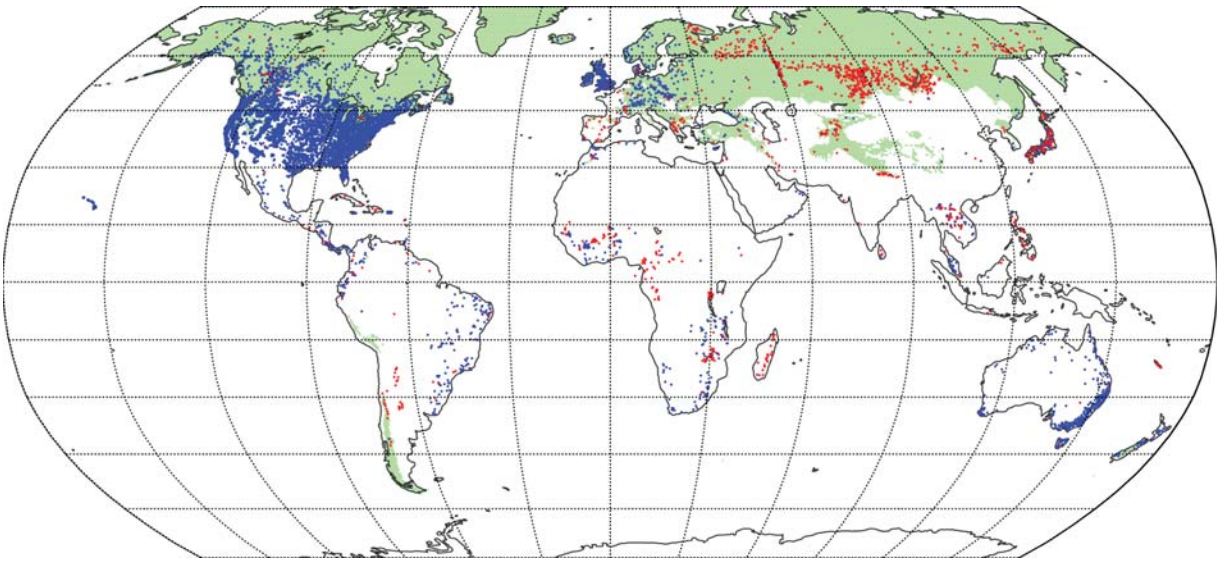

Figure 2: Centroids of the catchments used here. Catchments with daily Q data are indicated in blue while catchments with only monthly Q data are indicated in red. For Q characteristic T50 only the region with SNOW > 10 mm indicated in green was considered. The catchments with only monthly Q data were only used for the Q characteristics T50, RC, and QMEAN. Note that the exact catchment set differs slightly depending on the Q characteristic. Throughout the study the global maps have limits 75°N–75°S and 180°W–180°E, with grid lines every 15°.

(PCR-GLOBWB+ERAI and W3RA+Princeton) were evaluated. 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. Thus, BFI values derived from the models do not necessarily reflect the fraction of total Q originating from the subsurface compartments of the models.

The simulation-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 Supplementary material Figs. S1.1–17). In terms of mean R^2, the five most important predictors (ordered from highest to lowest mean R^2) were AI, P_{trans}, SNOW, PET, and SLO, while the five least important predictors (again ordered from highest to lowest mean R^2) were P_{si}, SILT, fW, PERM, and GRAV. Overall, the predictors related to climate proved most valuable, followed by the ones related to topography and land cover, while the predictors related to geology and soils were least useful. Among the soil-related predictors, SAND and CLAY were the most informative. The geologic index PERM was not related to any of the Q characteristics. The most important predictors for BFI4 were PET, TA, SNOW, ELEV, and CLAY and for k were AI, PET, and SNOW. Predictors identified as important for Q1–99 were AI, P_{trans}, and SNOW. Important predictors for T50 were P_{trans}, TA, and NDVI, for RC were AI, PET, and SNOW, and for QMEAN were AI, P_{trans}, SNOW, and SLO. The topography-related predictors TWI and SLO achieved almost identical R^2 values due to the strong intercorrelation between the two predictors. BFI1–4 achieved comparable R^2 results (Supplementary material Figs. S1.1–4) due to strong intercorrelations between the BFI estimates.

b. Catchment-scale estimation of streamflow characteristics

Table 5 shows mean R^2 and RMSE values obtained by the ten 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 T50, good (0.80–0.90) for Q1, Q5, Q10, Q20, Q50, RC, and QMEAN, acceptable (0.70–0.80) for BFI1–4, Q80, Q90, and Q95, but somewhat poor (< 0.70) for k and Q99 (Table 5; for associated scatterplots see Supplementary material Fig. S2.2). BFI1–4 achieved very similar training and testing R^2 values due to high intercorrelations among the BFI estimates. The testing R^2 values were 0.02–0.12 lower than the training R^2 values (Table 5), suggesting a good generalization capability of the trained models.

Table 6 shows the range of R^2 values computed between catchment values of the Q characteristics derived from observed Q and the four macro-scale hydrological models. The testing R^2 values obtained by the neural networks (Table 5) substantially exceeded the R^2 values obtained by the models (Table 6), demonstrating the usefulness of the neural networks.

Q	Mean R^2		Mean RMSE	
characteristic	Training	Testing	Training	Testing
BFI1	0.74	0.67	0.11	0.13
BFI2	0.72	0.66	0.10	0.11
BFI3	0.75	0.70	0.11	0.12
BFI4	0.73	0.66	0.11	0.12
k	0.56	0.44	0.37	0.41
Q1	0.80	0.77	0.49	0.56
Q5	0.84	0.81	0.31	0.37
Q10	0.84	0.83	0.26	0.30
Q20	0.85	0.83	0.21	0.24
Q50	0.85	0.82	0.15	0.18
Q80	0.78	0.73	0.13	0.16
Q90	0.75	0.68	0.13	0.16
Q95	0.72	0.65	0.13	0.15
Q99	0.66	0.60	0.12	0.14
T50	0.93	0.90	10.10	12.49
RC	0.81	0.75	0.09	0.11
QMEAN	0.86	0.84	3.21	3.81

Table 5: For the Q characteristics, mean coefficient of determination (R^2) and root mean square error (RMSE) statistics obtained for the training and testing subsets of catchments for the ten neural networks (one for each cross-validation iteration). Both statistics were calculated from transformed catchment values of the Q characteristics, hence the RMSE value is unitless. See Table 3 for the number of catchments used for each Q characteristic.

c. Global-scale estimation of streamflow characteristics

Figs. 3, 4, and 5 show global maps of estimated and observed BFI4, T50, and QMEAN, respectively (for global maps of the other Q characteristics see Supplementary material Figs. S2.3–19). Visual comparison of the estimates and the observations revealed acceptable agreement between estimated and observed BFI4, although the neural networks appeared to underestimate high observed BFI4 values (> 0.80) in eastern South America, over the southern section of the Great Rift Valley in Africa, over the Mountain Ranges of the Western USA, and over the Alps in Europe (Fig. 3). Particularly high BFI4 values were found over the Canadian Shield in North America and the Congo Basin in west Africa (Fig. 3a). Excellent agreement was obtained for T50, although the estimates seemed to be slightly too early over the northern high latitudes (Fig. 4). A pronounced north-south gradient from late to early Q timing can be observed (Fig. 4a). QMEAN did not exhibit any clear regional biases (Fig. 5). Par-

Predictor	BFI4	k	Q1	Q5	Q10	Q20	Q50	Q80	Q90	Q95	Q99	T50	RC	QMEAN
AI	–	$\mathbf{0.11^{+}_{L}}$	$\mathbf{0.57^{-}_{E}}$	$\mathbf{0.71^{-}_{E}}$	$\mathbf{0.74^{-}_{E}}$	$\mathbf{0.75^{-}_{E}}$	$\mathbf{0.69^{-}_{E}}$	$\mathbf{0.55^{-}_{E}}$	$\mathbf{0.48^{-}_{E}}$	$\mathbf{0.44^{-}_{E}}$	$\mathbf{0.38^{-}_{E}}$	0.12^{+}_{P}	$\mathbf{0.50^{-}_{E}}$	$\mathbf{0.74^{-}_{E}}$
P_{si}	–	–	0.08^{-}_{E}	0.07^{-}_{E}	0.06^{-}_{P}	0.06^{-}_{P}	0.09^{-}_{P}	0.06^{-}_{P}	0.06^{-}_{P}	–	–	0.12^{+}_{P}	–	0.06^{-}_{E}
P_{trans}	–	–	$\mathbf{0.56^{+}_{P}}$	$\mathbf{0.54^{+}_{L}}$	$\mathbf{0.51^{+}_{L}}$	$\mathbf{0.50^{+}_{L}}$	$\mathbf{0.46^{+}_{L}}$	$\mathbf{0.34^{+}_{L}}$	$\mathbf{0.28^{+}_{L}}$	$\mathbf{0.25^{+}_{L}}$	$\mathbf{0.20^{+}_{L}}$	$\mathbf{0.34^{-}_{E}}$	0.16^{+}_{L}	$\mathbf{0.57^{+}_{P}}$
PET	0.18^{-}_{L}	$\mathbf{0.07^{+}_{L}}$	–	0.06^{-}_{P}	0.09^{-}_{P}	0.10^{-}_{P}	0.09^{-}_{L}	0.10^{-}_{L}	0.10^{-}_{L}	0.10^{-}_{P}	0.10^{-}_{P}	0.28^{-}_{L}	$\mathbf{0.22^{-}_{P}}$	–
PET_{si}	0.10^{+}_{P}	0.05^{-}_{P}	–	–	–	–	–	–	–	–	–	0.25^{+}_{P}	0.08^{+}_{P}	–
CORR	0.05^{-}_{L}	–	0.07^{-}_{L}	0.09^{-}_{L}	0.10^{-}_{E}	0.08^{-}_{E}	0.06^{-}_{E}	–	–	–	–	–	–	0.09^{-}_{L}
TA	$\mathbf{0.21^{-}_{P}}$	0.05^{+}_{L}	–	–	–	–	–	–	–	–	–	$\mathbf{0.56^{-}_{L}}$	0.10^{-}_{L}	–
SNOW	$\mathbf{0.21^{+}_{L}}$	$\mathbf{0.08^{-}_{E}}$	0.07^{+}_{L}	$\mathbf{0.19^{+}_{E}}$	$\mathbf{0.22^{+}_{E}}$	$\mathbf{0.21^{+}_{L}}$	$\mathbf{0.16^{+}_{L}}$	$\mathbf{0.18^{+}_{L}}$	$\mathbf{0.18^{+}_{L}}$	$\mathbf{0.18^{+}_{L}}$	$\mathbf{0.19^{+}_{E}}$	0.16^{+}_{P}	$\mathbf{0.27^{+}_{E}}$	0.17^{+}_{L}
TWI	0.13^{-}_{E}	–	0.07^{-}_{E}	0.14^{-}_{P}	0.15^{-}_{P}	0.13^{-}_{P}	0.09^{-}_{L}	0.10^{-}_{L}	0.10^{-}_{L}	0.10^{-}_{P}	0.10^{-}_{P}	0.09^{-}_{P}	0.16^{-}_{P}	0.16^{-}_{E}
SLO	0.12^{+}_{P}	0.05^{-}_{E}	0.08^{+}_{E}	0.14^{+}_{E}	0.15^{+}_{E}	0.14^{+}_{E}	0.10^{+}_{P}	0.11^{+}_{E}	0.10^{+}_{P}	0.11^{+}_{E}	0.11^{+}_{E}	0.09^{+}_{P}	0.16^{+}_{E}	$\mathbf{0.17^{+}_{E}}$
ELEV	$\mathbf{0.19^{+}_{L}}$	–	–	–	–	–	–	–	–	–	–	0.31^{+}_{L}	–	–
fW	0.05^{+}_{L}	–	–	–	–	–	–	–	–	–	–	–	–	–
fTC	0.10^{+}_{L}	0.06^{-}_{L}	0.08^{+}_{E}	0.13^{+}_{E}	0.15^{+}_{L}	0.14^{+}_{L}	0.13^{+}_{L}	0.11^{+}_{L}	0.10^{+}_{E}	0.10^{+}_{L}	0.09^{+}_{E}	–	0.10^{+}_{L}	0.11^{+}_{E}
NDVI	0.07^{-}_{E}	–	$\mathbf{0.14^{+}_{L}}$	0.07^{+}_{L}	0.07^{+}_{L}	0.07^{+}_{L}	0.08^{+}_{L}	–	–	–	–	$\mathbf{0.52^{-}_{L}}$	–	0.09^{+}_{L}
PERM	–	–	–	–	–	–	–	–	–	–	–	–	–	–
GRAV	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SAND	0.15^{+}_{L}	0.07^{-}_{E}	–	–	–	–	–	–	0.05^{+}_{L}	0.05^{+}_{L}	0.05^{+}_{L}	0.11^{+}_{P}	–	–
SILT	0.05^{-}_{L}	–	–	–	–	–	–	–	–	–	–	–	–	–
CLAY	0.18^{-}_{E}	0.05^{+}_{E}	–	–	–	–	–	0.06^{-}_{E}	0.06^{-}_{E}	0.06^{-}_{E}	0.07^{-}_{E}	0.17^{-}_{P}	0.05^{-}_{E}	–

Table 4: Coefficient of determination (R^2) values computed between catchments values of the Q characteristics and predictors (for associated scatterplots see Supplementary material Figs. S1.1–17). Various functions were considered, but only the R^2 of the best relationship is shown. R^2 values < 0.05 are indicated by a dash. For each Q characteristic, the three highest R^2 values are shown in bold font. The letters indicate the type of relationship that achieved the best R^2 (L, linear; E, exponential; and P, power). The plus and minus symbols indicate positive and negative relationships, respectively. The four BFI estimate produced very similar results, hence only BFI4 is shown.

ticularly high QMEAN was obtained for the Amazon, maritime Southeast Asia, the western flank of the Coast Mts. 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 (see Supplementary material Fig. S2.1 for a map of the major Köppen-Geiger climate types). The lowest BFI values were found in arid regions (broad climate type B) and the highest in tropical and cold regions (broad climate types A and D, respectively; Table 7 and Fig. 3a). The values of the different BFI estimates varied consistently amongst climate types—BFI1 always produced the highest values, BFI2 or BFI4 always the lowest, whilst BFI3 produced values slightly lower than BFI1 (Table 7). k showed little variability amongst climate types—only for arid regions (broad climate type B) was the baseflow recession markedly faster (Table 7). Q1–99 and QMEAN also varied rather consistently amongst amongst climate types, with higher values found for tropical climate types Af and Am (rainforest and monsoon, respectively) and for 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 types (Table 7), which is indicative of intermittent and ephemeral streams. The earliest T50 values were found in temperate regions (broad climate type C) and the latest in polar regions (broad climate type E; Table 7). The lowest RC values were found in arid regions (broad climate type B) and the highest in polar regions (broad climate type E; Table 7).

d. Example application

As an example application of the new global maps of Q characteristics, a comparison was made to equivalent maps derived from daily Q estimates of four macro-scale hydrological models. Figs. 6, 7, and 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 Supplementary material Figs. S3.1–13). Table 8 lists correlation coefficients (R), median differences (D), and median absolute differences ($|D|$) for BFI4, k, Q1–99, 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 baseflow to total Q. However, HTESSEL+ERAIR produces excess baseflow over taiga and temperate regions (Fig. 6a), whilst W3RA+Princeton does so over tropical regions (Fig. 6d). Noah+GLDAS yielded the lowest

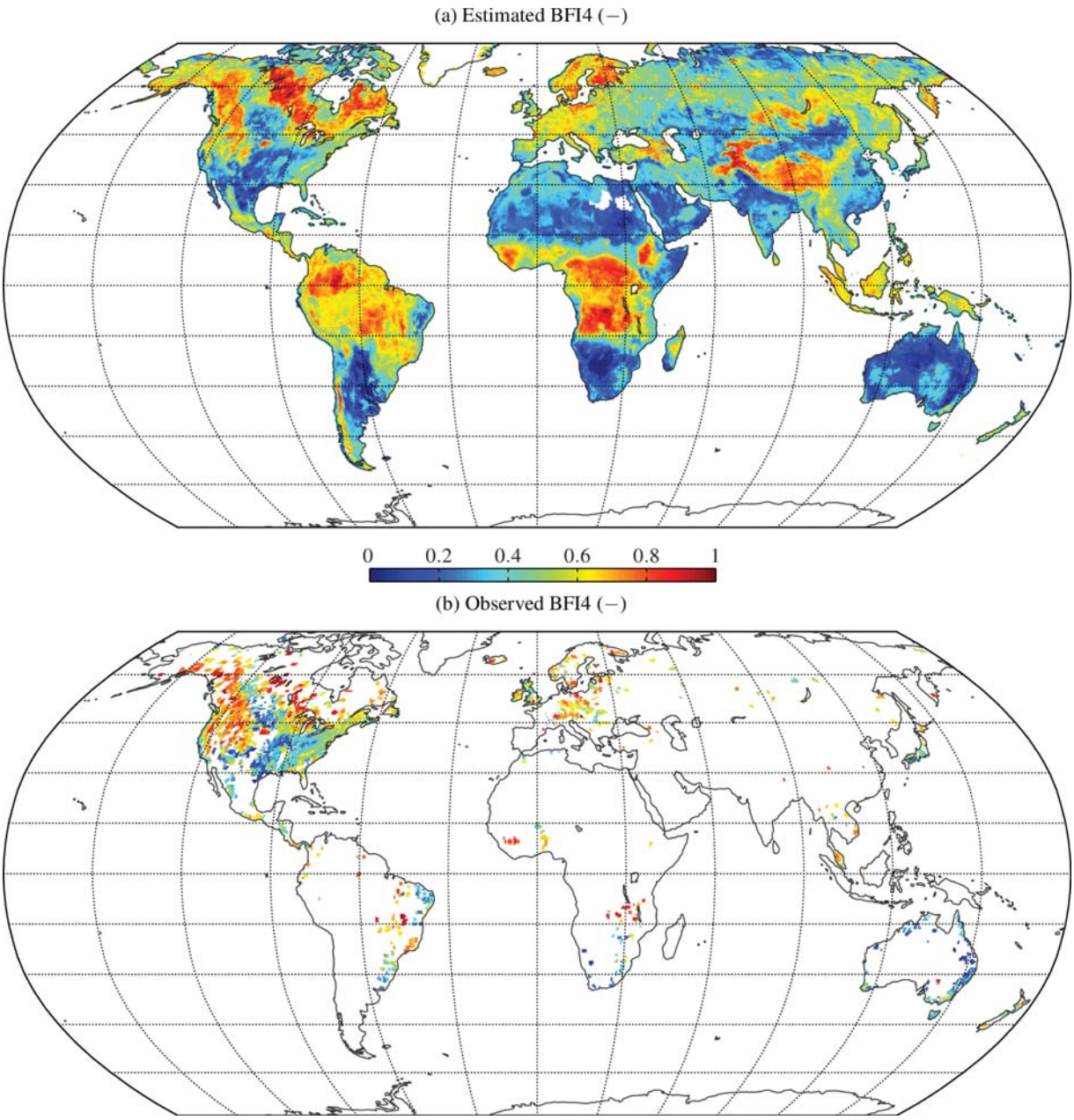

Figure 3: Global maps of (a) estimated and (b) observed baseflow index computed using technique 4 (BFI4). The values in (a) represent the (back transformed) median estimates of the neural network ensembles. In (b) only gauged regions have values.

R value (Table 8), producing insufficient baseflow over tundra and taiga regions (Fig. 6b). PCR-GLOBWB+ERAI performed moderately well in terms of *R* but appears to overestimate BFI4 over most of the land surface (Table 8 and Fig. 6c). All models produce insufficient baseflow over the Rocky Mts. 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 negative *R* value (Table 8). W3RA+Princeton performed best for the median and low flow percentiles (Q50–99), whereas Noah+GLDAS and PCR-GLOBWB+ERAI performed best for the high flow percentiles (Q1–20; 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 and Fig. 7). However, PCR-GLOBWB+ERAI performed markedly

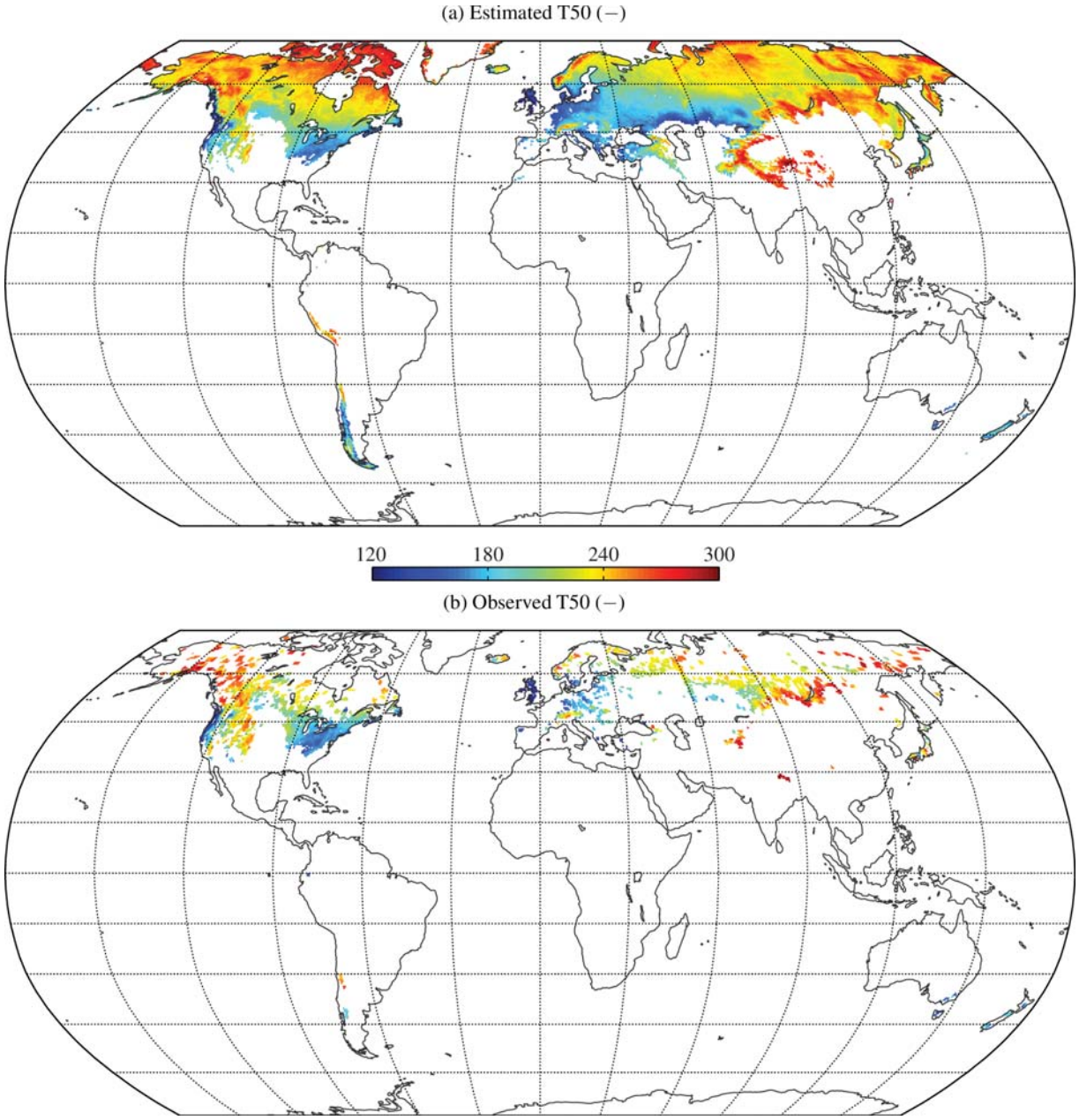

Figure 4: Global maps of (a) estimated and (b) observed timing of the center of mass of Q (T50). The values in (a) represent the (back transformed) median estimates of the neural network ensembles. In (b) only gauged regions have values. The analysis of T50 was restricted to areas with at least some snowfall.

better than the other models, obtaining a D value of -4 days whereas the other models yielded D values of -13 to -15 days (Table 8).

PCR-GLOBWB+ERAI demonstrated the best R value for QMEAN and thus best captures the spatial variability in mean annual Q (Table 8 and Fig. 8c). Noah+GLDAS produces insufficient Q over most of the land surface (Table 8 and Fig. 8b), while W3RA+Princeton does so over most of the Northern Hemisphere with the exception of India (Table 8 and Fig. 8d). HTESSEL+ERAIR performed moderately well in terms of QMEAN (Table 8 and Fig. 8a). All models produce insufficient Q over the Rocky Mts., the Andes, and the Himalayas (Fig. 8). The two GHMs (PCR-GLOBWB+ERAI 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 and Fig. 8).

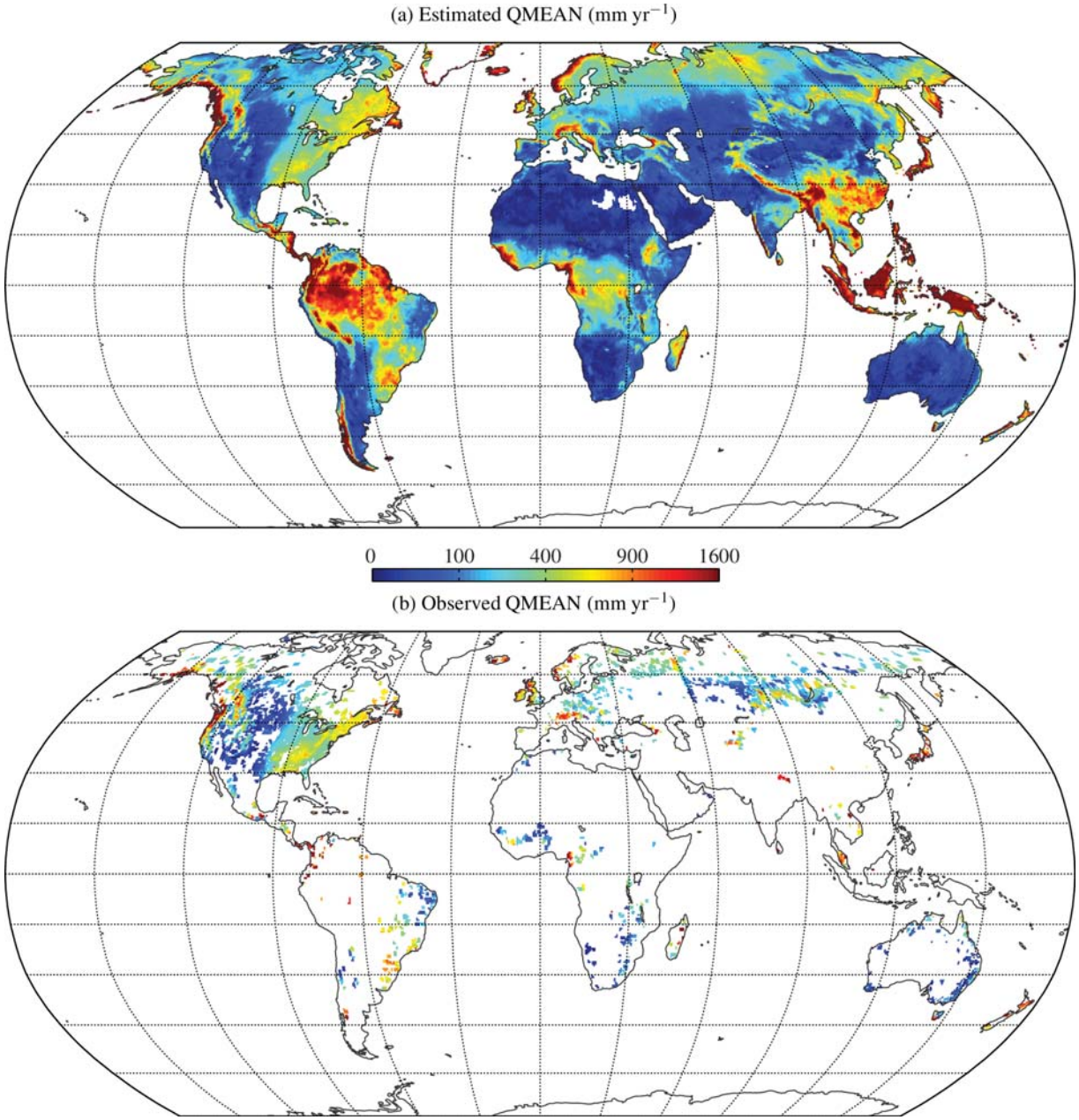

Figure 5: Global maps of (a) estimated and (b) observed mean annual Q (QMEAN). The values in (a) represent the (back transformed) median estimates of the neural network ensembles. In (b) only gauged regions have values. Note the non-linear scale.

6. Discussion

a. Climate and physiographic controls of the streamflow characteristics

The high R^2 values found for TA and SNOW vs. BFI4 and AI, PET, and SNOW vs. k (Table 4) reflects the dominant control of climate conditions on baseflow volume and recession rate (Price 2011). The positive relationship obtained between SNOW 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 assumed influence of soil composition on the infiltration-excess mechanism of runoff generation (Horton 1933; Wolock and McCabe 1999). BFI1–4 showed very similar relationships with the predictors (Supplementary material Figs. S1.1–4) due to high intercorrelations 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, although we found a lower R^2 for

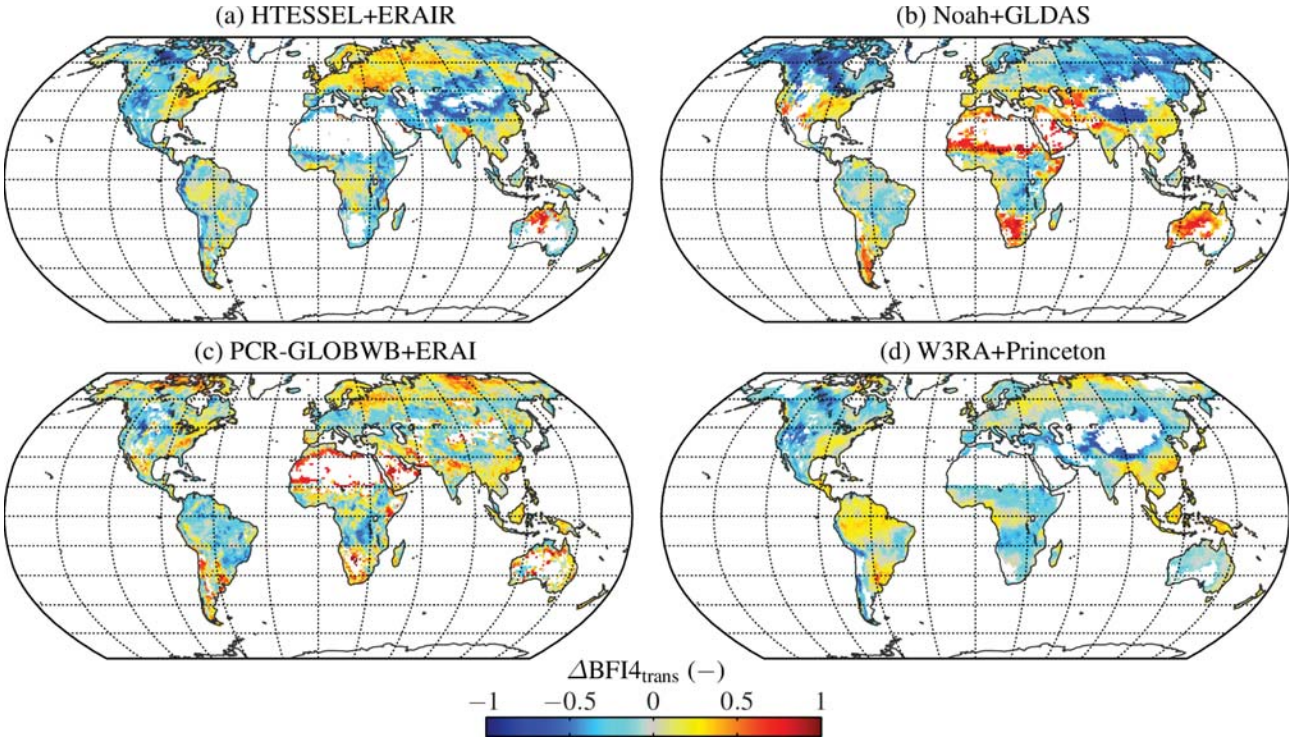

Figure 6: The difference between transformed baseflow index computed using technique 4 ($BFI4_{trans}$) derived from (a) HTESSEL+ERAIR, (b) Noah+GLDAS, (c) PCR-GLOBWB+ERAI, and (d) W3RA+Princeton, and $BFI4_{trans}$ from this study (Fig. 3a). BFI4 was only computed for the models if the simulated QMEAN was > 5 mm yr^{-1}. Red (blue) indicates regions where the models overestimate (underestimate) BFI4 relative to the map produced in this study.

fW vs. BFI1, a higher R^2 for CLAY vs. BFI1, and overall weaker relationships for k (Supplementary material Figs. S1.1 and S1.5). We refer to Beck et al. (2013b) for a review of multi-catchment observational studies investigating the links between catchments attributes and BFI and k.

The Q characteristics Q1–99, RC, and QMEAN were found to be strongly related to the climate index 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 SNOW (Table 4), presumably due to reductions in evaporation when the surface is snow covered (Berghuijs et al. 2014) and/or due to common P underestimation 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 weak yet 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). fTC did not exhibit clear relationships with the Q characteristics (Table 4), which agrees with some multi-catchment studies examining the spatial relationship between fTC and Q (e.g., Oudin et al. 2008; Peel et al. 2010) as well as some meso- and large-scale catchment (> 1 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 Q5–99 or QMEAN (Table 4) agrees with Hope and Bart (2012), who used observed Q from 30 South African catchments (1–254 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).

TA was found to be strongly negatively related to T50 (Table 4) because temperature represents the dominant control on the snowmelt contribution to Q (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 due to 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 positive 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 reflecting the strong intercorrelations between ELEV, air temperature, snowfall, and snow accumulation and melt (Daly et al. 2002; Leibowitz et al.

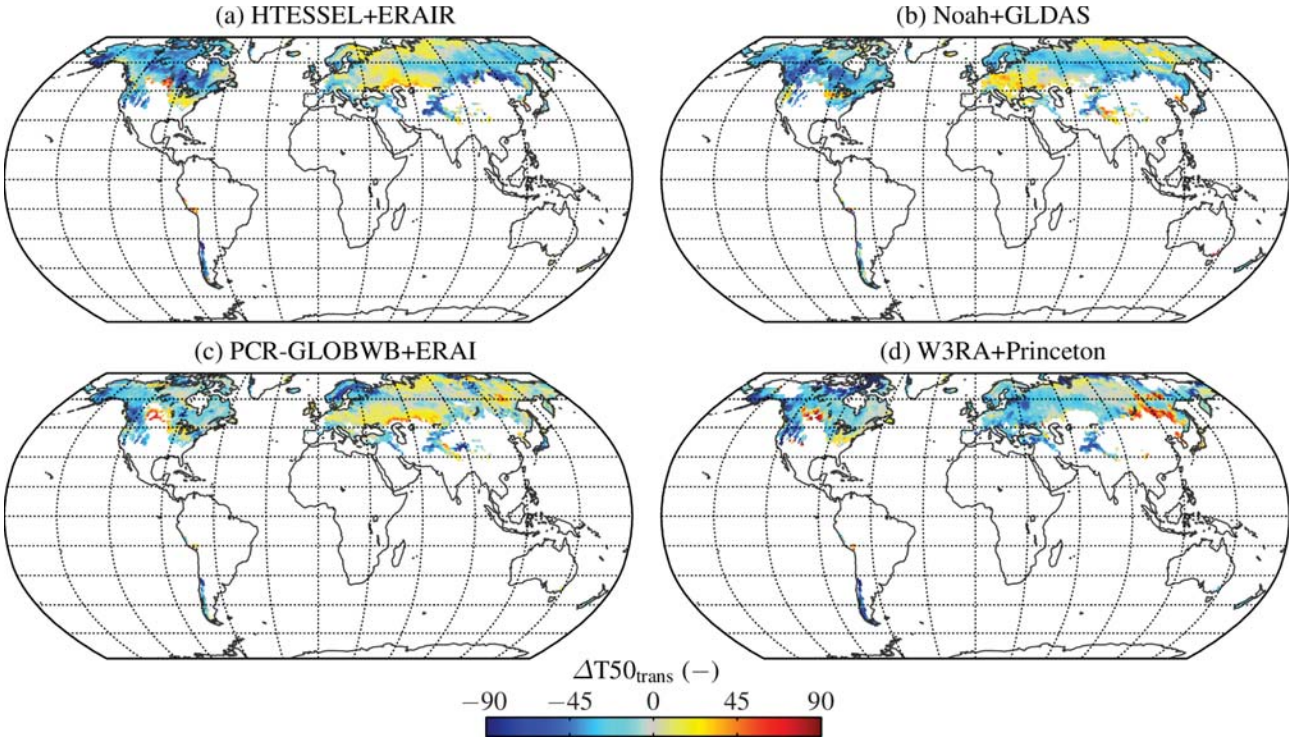

Figure 7: The difference between timing of the center of mass of Q ($T50_{trans}$) derived from (a) HTESSEL+ERAIR, (b) Noah+GLDAS, (c) PCR-GLOBWB+ERAI, and (d) W3RA+Princeton, and $T50_{trans}$ from this study (Fig. 4a). Only regions with at least some snowfall were considered. T50 was only computed for the models if the simulated QMEAN was > 5 mm yr^{-1}. Red (blue) indicates regions where the models overestimate (underestimate) T50 relative to the map produced in this study.

2012). A non-causal relationship does not necessarily mean that the predictor should be excluded from the estimation exercise, as it is possible that non-controlling predictors compensate for the deficiencies of controlling predictors.

Overall, the climate indices showed the strongest relationships with the Q characteristics (Table 4). The predictors related to land cover (fW, fTC, and NDVI) were not strongly related to any of the Q characteristics, with the exception of the (non causal) relationship between NDVI and T50 (Table 4). The topography-related predictors (TWI, SLO, and ELEV) were found to be relatively unimportant (Table 4). However, a weak (positive) relationship was found between SLO and QMEAN (Table 4), presumably because steeper slopes exhibit faster lateral drainage leaving less water available for evaporation (cf. Zecharias and Brutsaert 1988; Mazvimavi et al. 2005; Price 2011). On the whole, our results suggest that topography is not the dominant control on Q behavior (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 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 and non-existent relationships were found here (Table 4), probably due to the quality of the global datasets (cf. Table 2).

b. Catchment-scale estimation of streamflow characteristics

The global maps of Q characteristics developed here 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 broad range of physiographic and climate conditions was used to produce the maps, which should have led to more generally applicable results (cf. Andréassian et al. 2007; Gupta et al. 2014). Third and last, uncertainty estimates have been computed, which can be used to evaluate the reliability of the new maps.

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

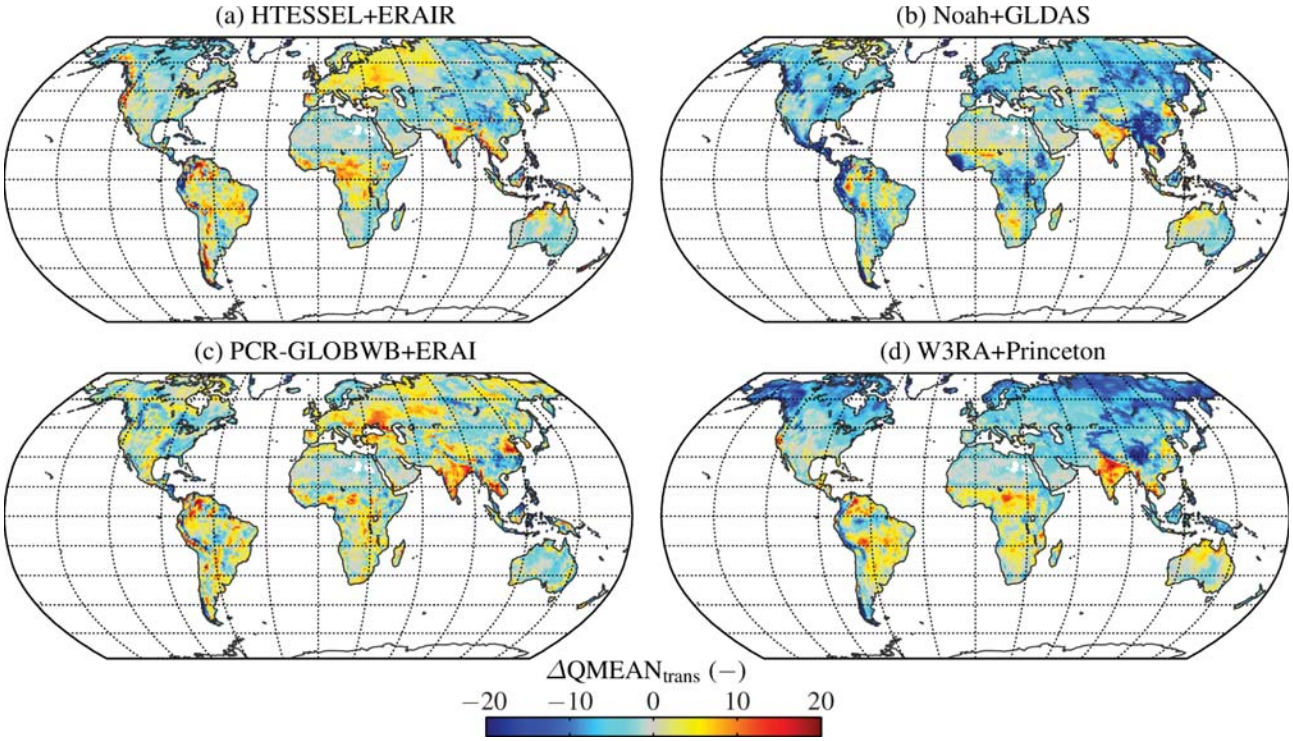

Figure 8: The difference between transformed mean annual Q ($QMEAN_{trans}$) derived from (a) HTESSEL+ERAIR, (b) Noah+GLDAS, (c) PCR-GLOBWB+ERAI, and (d) W3RA+Princeton, and $QMEAN_{trans}$ from this study (Fig. 5a). Red (blue) indicates regions where the models overestimate (underestimate) QMEAN relative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

Model+forcing	BFI4	k	Q1	Q5	Q10	Q20	Q50	Q80	Q90	Q95	Q99	T50	QMEAN
						Correlation coefficient (R)							
HTESSEL+ERAIR	0.41	0.30	0.69	0.72	0.74	0.75	0.81	0.76	0.72	0.70	0.65	0.72	0.73
Noah+GLDAS	−0.06	−0.47	**0.72**	**0.74**	0.75	0.73	0.71	0.58	0.46	0.41	0.33	**0.81**	0.80
PCR-GLOBWB+ERAI	0.19	0.07	0.66	0.73	**0.77**	**0.79**	0.78	0.71	0.68	0.64	0.58	0.76	**0.86**
W3RA+Princeton	**0.43**	**0.41**	0.57	0.68	0.72	0.74	**0.83**	**0.83**	**0.80**	**0.79**	**0.74**	0.56	0.79
						Median difference (D)							
HTESSEL+ERAIR	−0.07	−1.08	**−0.04**	−0.20	−0.15	−0.07	**0.00**	**−0.00**	**−0.01**	**−0.01**	**−0.00**	−14.59	−1.11
Noah+GLDAS	−0.12	**−0.01**	−0.29	−0.20	−0.23	−0.24	−0.21	−0.15	−0.13	−0.13	−0.09	−13.11	−3.30
PCR-GLOBWB+ERAI	0.05	−1.36	−0.18	**−0.05**	**−0.00**	**0.06**	0.13	0.12	0.09	0.07	0.08	**−4.30**	**−0.02**
W3RA+Princeton	**−0.05**	−0.70	−0.51	−0.42	−0.33	−0.22	−0.06	−0.03	−0.03	−0.04	−0.02	−14.96	−3.40
						Median absolute difference (|D|)							
HTESSEL+ERAIR	0.21	1.33	0.55	0.47	0.37	0.28	**0.15**	0.13	0.11	0.10	0.09	19.25	3.07
Noah+GLDAS	0.26	**0.75**	**0.55**	0.41	0.36	0.32	0.23	0.17	0.15	0.13	0.10	18.47	4.01
PCR-GLOBWB+ERAI	0.19	1.44	0.62	**0.39**	**0.29**	**0.23**	0.19	0.18	0.17	0.16	0.14	**12.68**	**3.02**
W3RA+Princeton	**0.16**	0.78	0.76	0.54	0.42	0.34	0.16	**0.11**	**0.09**	**0.09**	**0.07**	19.25	4.66

Table 8: For the four models, correlation coefficients (R), median differences (D), and median absolute differences ($|D|$) computed between simulation and observation-based maps of the Q characteristics. All statistics were calculated from transformed values of the Q characteristics, hence the D and $|D|$ values 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 bold font. The statistics for BFI1–4 were nearly identical, hence only the BFI4 results are shown. RC is not shown since 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.

Q characteristic	R^2 range	Number of catchments
BFI4	0.00–0.04	6189
k	0.00–0.03	5378
Q1	0.15–0.32	6113
Q5	0.26–0.42	6119
Q10	0.34–0.47	6110
Q20	0.42–0.51	6105
Q50	0.30–0.48	6093
Q80	0.12–0.32	6136
Q90	0.07–0.26	6114
Q95	0.04–0.23	6119
Q99	0.02–0.18	6123
T50	0.27–0.52	4519
QMEAN	0.47–0.53	8421

Table 6: For the Q characteristics, the range of coefficient of determination (R^2) values computed between transformed catchment values of the Q characteristics derived from observed Q and simulated Q from the four macro-scale hydrological models. The statistics for BFI1–4 were nearly identical, hence only the BFI4 results are shown. For each Q characteristic, only the catchments which had non-missing values of the Q characteristics for all four models were considered. For some catchments, the models produced flow events without sufficiently long recession periods, preventing the computation of k. RC is not shown since the P forcing data were not available for all models. The analysis of T50 was restricted to catchments with at least some snowfall. BFI4, k, Q1–99, and T50 were only computed for the models if the simulated QMEAN was > 5 mm yr^{-1}.

(0.72–0.75; 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 Beck et al. 2013b). Although the mean training R^2 of 0.56 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 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), others 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), while we used a number of flow percentiles (Q1–99; Table 1). The mean training and testing R^2 values obtained here for Q95 (0.72 and 0.65, 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 multi-variate

Table 7: For the ice-free land surface and the major Köppen-Geiger climate types, mean values of the Q characteristics derived from the newly produced maps. Supplementary material Fig. S2.1 shows a map of the major Köppen-Geiger climate types. Values were only listed if > 1000 0.125° grid cells had a value. For T50 only regions with at least some snowfall were considered.

Climate type	BFI1 (−)	BFI2 (−)	BFI3 (−)	BFI4 (−)	k (d^{-1})	Q1	Q5	Q10	Q20	Q50	Q80	Q90	Q95	Q99	T50 (−)	RC (−)	QMEAN (mm yr^{-1})
										(mm d^{-1})							
Ice-free land surface	0.63	0.43	0.59	0.45	0.13	6.33	2.94	1.92	1.15	0.42	0.19	0.13	0.11	0.06	223	0.30	349
Af: tropical, rainforest	0.74	0.58	0.69	0.64	0.06	16.26	9.12	6.68	4.74	2.26	1.12	0.75	0.62	0.33	–	0.52	1468
Am: tropical, monsoon	0.73	0.57	0.69	0.61	0.07	14.96	7.54	5.31	3.17	1.26	0.51	0.36	0.28	0.14	–	0.44	985
Aw: tropical, savannah	0.69	0.53	0.66	0.54	0.10	7.95	3.40	2.19	1.19	0.43	0.17	0.13	0.11	0.06	–	0.23	365
BW: arid, desert	0.49	0.22	0.48	0.28	0.21	0.60	0.31	0.13	0.10	0.03	0.03	0.01	0.01	0.01	235	0.04	26
BS: arid, steppe	0.55	0.32	0.48	0.33	0.19	1.88	0.60	0.31	0.20	0.05	0.03	0.02	0.02	0.01	202	0.09	59
Cs: temperate, dry summer	0.63	0.47	0.58	0.46	0.11	6.41	2.67	1.75	0.97	0.32	0.14	0.09	0.07	0.04	172	0.26	317
Cw: temperate, dry winter	0.65	0.50	0.62	0.50	0.10	8.25	3.32	2.13	1.21	0.45	0.22	0.17	0.14	0.08	262	0.29	404
Cf: temperate, no dry season	0.57	0.40	0.50	0.39	0.11	11.13	4.73	2.98	1.80	0.69	0.30	0.21	0.17	0.10	167	0.38	584
Ds: cold, dry summer	0.76	0.62	0.74	0.64	0.08	6.41	3.43	2.45	1.50	0.43	0.18	0.15	0.12	0.08	228	0.55	382
Dw: cold, dry winter	0.73	0.55	0.67	0.56	0.08	7.38	3.69	2.23	1.30	0.30	0.10	0.08	0.05	0.03	251	0.46	271
Df: cold, no dry season	0.71	0.51	0.67	0.52	0.08	6.62	3.14	2.05	1.15	0.34	0.14	0.11	0.09	0.06	223	0.52	321
ET: polar, tundra	0.63	0.40	0.59	0.42	0.08	7.59	3.53	2.37	1.14	0.21	0.08	0.07	0.06	0.04	264	0.84	379
EF: polar, frost	0.60	0.47	0.55	0.44	0.08	13.59	7.48	4.45	2.51	0.29	0.07	0.05	0.05	0.01	272	1.25	875

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 km^2), 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 km^2), 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 similar to the range of training R^2 values obtained here (Table 5). Using observed Q from 30 South African catchments (1–254 km^2), 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 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, respectively, 0.86 and 0.84 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 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 $\sim$3500 European catchments (29 000–160 000 km^2) to train a multi-variate 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 (1–254 km^2) to estimate QMEAN from P, mean satellite-derived Enhanced Vegetation Index, and SILT. However, the training R^2 values obtained here are lower than 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 (4–2630 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 (3322–53 829 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 QMEAN and BFI maps developed in this study can be compared to QMEAN and BFI maps for the conterminous USA produced by Gebert et al. (1987) and Wolock (2003), respectively. The QMEAN map for the conterminous USA 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° using IDW interpolation to allow spatial comparison to the newly derived QMEAN map. Fig. 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 due to 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 236 mm yr^{-1} for the map developed here and 242 mm yr^{-1} for the map of Gebert et al. (1987) were obtained. The good correspondence between the maps is likely due to overlap in the catchment sets used to derive the maps.

The BFI map for the conterminous USA (1-km resolution) was produced by IDW interpolation of BFI values from 8249 catchments (Wolock 2003). BFI was computed using the so-called BFI Program (Wahl and Wahl 1995), which implements the baseflow separation procedure described in Institute of Hydrology (1980) and 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. The comparison yielded an R^2 of 0.53, indicating moderate agreement in terms of spatial patterns, likely attributable to the moderate performance achieved in the estimation of BFI4 (Table 5). There was, however, good agreement in terms of mean BFI4—considering only the regions where both maps have data, mean BFI4 values of 0.41 for the newly derived map and 0.43 for the map from Wolock (2003) were obtained.

Differences between the BFI1–4 values (Table 7) are caused by differences in the baseflow separation techniques used (Table 1). Fig. 1 shows that the baseflow 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 (seven days) moving minimum is used to compute baseflow. On the other hand, BFI2 produces relatively low BFI values due of the use of a long (11 days) moving minimum. BFI4 also produces relatively low BFI values as it computes minima at five-day non-overlapping intervals and subsequently connects the valleys in this series of minima. Baseflow 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) provides estimates of mean QMEAN and RC for the land surface derived from from six LSMs and five GHMs. The mean QMEAN based on the map produced here

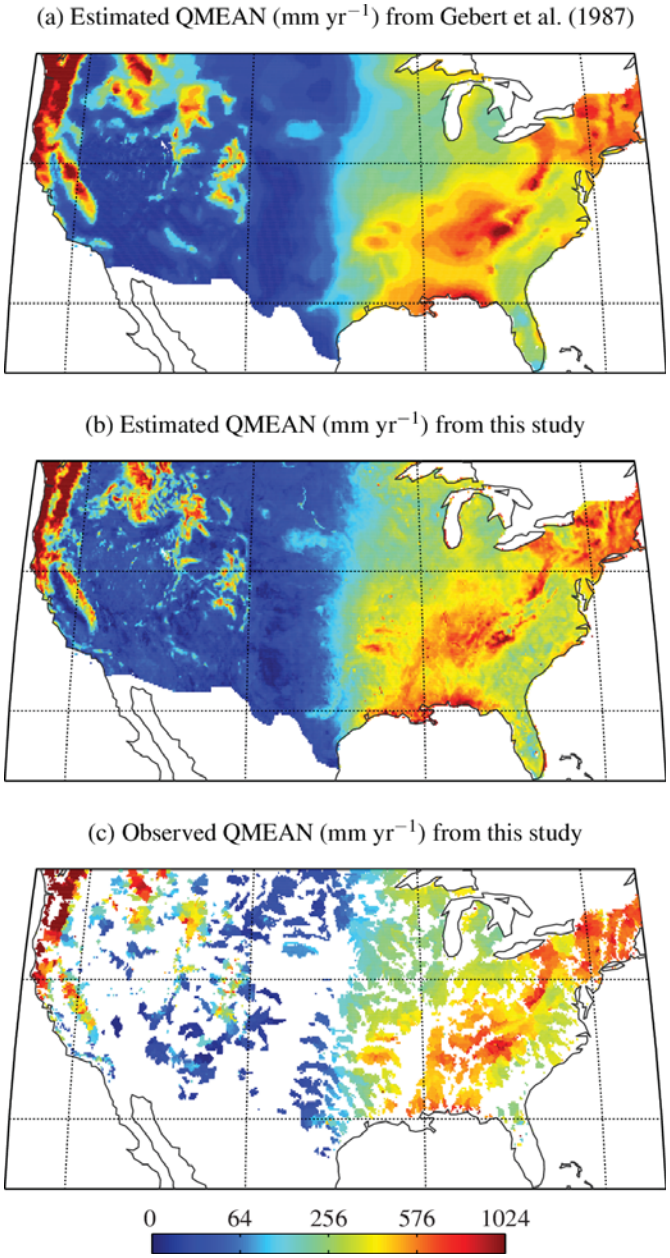

Figure 9: Maps of the conterminous USA showing (a) estimated mean annual Q (QMEAN) from Gebert et al. (1987), (b) estimated QMEAN from this study, and (c) observed QMEAN from this study. The values in (b) represent the (back transformed) median estimates of the neural network ensembles. In (a) and (b) only values for the conterminous USA are shown. In panel (c) only gauged regions within the conterminous USA are shown. Note the non-linear color scale. The maps have limits 25–48°N and 125–70°W and grid lines at every 10° latitude and 15° longitude.

(Fig. 5a) was 349 mm yr^{-1} (Table 7), which is within the range of 290 to 457 mm yr^{-1} obtained by Haddeland et al. (2011). The mean RC based on the map produced here (see Supplementary material Fig. S2.18) was 0.30 (Table 7), which is below the range of 0.33 to 0.52 obtained by Haddeland et al. (2011). The lower RC found here appears to 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 continuous rainfall-runoff models. The maps may even prove useful for continuous rainfall-runoff models having spatial resolutions down to 1 km that are anticipated in the near future (Wood et al. 2011), since the maps have a resolution limited only by the resolution of the input data used. However, there are some caveats worth mentioning. First, the maps reflect flows under natural, unregulated conditions, as catchments with extensive irrigation or urbanization have been excluded. Second, the maps are not representative of flows in catchments $> 10\,000$ km^2 due to the increasing influence of channel routing effects at larger scales (i.e., evaporation and leakage losses and travel time delays). Hence, the maps should not be directly used to estimate the Q from continents into the oceans, for example. Third and last, the maps are most accurate in regions with dense Q monitoring networks, such as Europe and the USA, 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 macro-scale 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, 7, and 8, Table 8, and Supplementary material Figs. S3.1–13). Although there are a host of studies using observed Q to diagnose models in a regional context (e.g., Lohmann et al. 1998; Robock et al. 2003; Lohmann et al. 2004; 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 global scale (e.g., Oki et al. 1999; Döll et al. 2003; Decharme and Douville 2007; Decharme 2007; Zaitchik et al. 2010; Materia 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 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 (1 680 000–2 440 000 km^2) and found that Noah+GLDAS produces too much quickflow under (sub-)Arctic conditions, likely due to 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) due to the baseflow-related parameterizations of the models. HTESSEL and Noah produce k values that 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 aridity index (AI) based empirical relationship taken from Peña-Arancibia et al. (2010). However, this study found weak or non-existent relationships between k and soil texture (GRAV, 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 due to 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 and 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 km^2). Lohmann et al. (2004) found a similar difference pattern in Q timing for Noah using observed Q from 1145 small USA catchments (23–10 000 km^2). 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 km^2). 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 km^2) 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. All models with the exception of PCR-GLOBWB+ERAI appeared to underestimate QMEAN at the global scale (Fig. 8 and Table 8). 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 km^2) around the world. All models appeared to underestimate QMEAN over mountainous regions (Fig. 8), despite the orographic corrections applied to the forcing by HTESSEL+ERAIR and W3RA+Princeton (Sheffield et al. 2006; Adler et al. 2003; Balsamo et al. 2013). The pronounced QMEAN underestimation by W3RA+Princeton at northern latitudes suggest that improvements in the model structure and/or parameterization are required.

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 2012).

7. Conclusion

In this study, observed Q from approximately 7500 small catchments ($< 10\,000$ km^2) around the globe were used to examine the relationships between catchment attributes (related to climate, topography, land cover, soils, and geology) and Q characteristics. In total 17 Q characteristics were examined, 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:

i. Climate indices exhibited the strongest relationships with the Q characteristics. In contrast, predictors related to soils and geology showed weak and non-existent relationships with the Q characteristics, possibly due to 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 BFI estimates were highly intercorrelated but showed systematic differences.

ii. Neural networks were trained to estimate the Q characteristics from catchment attributes. The estimation performance was excellent for the Q timing (T50) but somewhat poor for the baseflow recession constant (k) and the 99th percentile flow (Q99). The trained neural network ensembles were subsequently applied to 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.

iii. The produced global maps of Q characteristics should be of particular value for the diagnosis of macro-scale hydrological models. To test this, the newly produced maps were compared to equivalent maps derived from four macro-scale 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.

The Global Runoff Data Centre (GRDC) and the U.S. Geological Survey (USGS) are thanked for providing most of the observed Q data. The WorldClim developers are thanked for making available the global climate maps. We also would like to thank Tom Gleeson (McGill University, Montreal, Canada) for providing the global permeability map, Toby Marthews (University of Oxford, Oxford, UK) for providing the global map of topographic wetness index, Frederiek Sperna Weiland (Deltares, Delft, The Netherlands), for providing the PCR-GLOBWB+ERAI modeled Q data, and last but not least, the European Centre for Medium-Range Weather Forecasts (ECMWF; Reading, UK) for providing the HTESSEL+ERAIR 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., et al., 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, **4 (6)**, 1147–1167.

Aguado, E., D. Cayan, L. Riddle, and M. Roos, 1992: Climatic fluctuations and the timing of west coast streamflow. *Journal of Climate*, **5 (12)**, 1468–1483.

Andréassian, V., J. Lerat, C. Loumagne, T. Mathevet, C. Michel, L. Oudin, and C. Perrin, 2007: What is really undermining hydrologic science today? *Hydrological Processes*, **21 (20)**, 2819–2822.

Arnell, N. W., 1995: Grid mapping of river discharge. *Journal of Hydrology*, **167 (1–4)**, 39–56.

Babovic, V., 2005: Data mining in hydrology. *Hydrological Processes*, **19**, 1511–1515.

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. *Journal of Hydrometeorology*, **10 (3)**, 623–643.

Balsamo, G., et al., 2013: ERA-Interim/Land: a global land water resources dataset. *Hydrology and Earth System Sciences Discussions*, **10**, 14 705–14 745.

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 meso-scale humid tropical catchments. *Hydrology and Earth System Sciences*, **17 (7)**, 2613–2635.

Beck, H. E., 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 Resources Research*, **49 (12)**, 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. *Nature Climate Change*, **4**, 583–586, doi:10.1038/nclimate2246.

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

Beven, K. J. and M. J. Kirkby, 1979: A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, **24 (1)**, 43–69.

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, New York, US.

Bontemps, S., P. Defourny, and E. van Bogaert, 2011: GlobCover 2009, products description and validation report. Tech. rep., ESA GlobCover project, available at: `http://ionia1.esrin.esa.int`.

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. Tech. Rep. 126, Institute of Hydrology, Wallingford, UK. URL `http://www.ceh.ac.uk/products/publications/hydrologypublications.html`.

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. *Journal of the American Water Resources Association*, **41 (5)**, 1177–1186.

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. *Annual Review of Environment and Resources*, **32**, 67–98, doi:10.1146/annurev.energy.32.031306.102758.

Bruijnzeel, L. A., 2004: Hydrological functions of tropical forests: not seeing the soil for the trees. *Agriculture, Ecosystems and Environment*, **104 (1)**, 185–228.

Budyko, M. I., 1974: *Climate and life*. Academic Press, New York.

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. *Advances in Water Resources*, **33 (10)**, 1235–1242.

Choi, H. I. and X. Liang, 2010: Improved terrestrial hydrologic representation in mesoscale land surface models. *Journal of Hydrometeorology*, **11 (3)**, 797–809.

Clausen, B. and B. J. F. Biggs, 2000: Flow variables for ecological studies in temperate streams: groupings based on covariance. *Journal of Hydrology*, **237**, 184–197.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **22**, 99–113.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33 (2)**, 140–158.

Decharme, B., 2007: Influence of runoff parameterization on continental hydrology: Comparison between the Noah and the ISBA land surface models. *Journal of Geophysical Research*, **112**, D19 108, doi:10.1029/2007JD008463.

Decharme, B. and H. Douville, 2007: Global validation of the ISBA sub-grid hydrology. *Climate Dynamics*, **29 (1)**, 21–37.

Dee et al., D. P., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Part A*, **137 (656)**, 553–597.

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. *Journal of Hydrology*, **309**, 258–276.

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? *Hydrological Processes*, **19 (8)**, 1705–1714.

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

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. *Journal of Water Resource and Protection*, **2 (11)**, 934–943.

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., AGU, Washington, D.C., US, No. 3 in Water Science and Application, 77–94.

Duan, Q., et al., 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, **320 (1)**, 3–17.

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. *Journal of Geophysical Research*, **108 (D22)**, 8851.

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 (2)**, 621–621.

FAO, 2000: Food and Agriculture Organization (FAO) Forest Resource Assessment (FRA) forest cover. URL `www.fao.org/forestry/32203/en/`.

FAO/IIASA, F., 2012: Harmonized World Soil Database (version 1.2). Tech. rep., FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Farr, T. G., et al., 2007: The Shuttle Radar Topography Mission. *Reviews of Geophysics*, **45 (2)**, doi:10.1029/2005RG000183.

Farvolden, R. N., 1963: Geologic controls on ground-water storage and base flow. *Journal of Hydrology*, **1 (3)**, 219–249.

Fekete, B. M. and C. J. Vörösmarty, 2007: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB Kick-off meeting held in Brasilia, 20–22 November 2002), IAHS publication no. 309*, 129–136, URL `http://iahs.info/redbooks/a309/309015.pdf`.

Feng, X. and P. Houser, 2008: An investigation of GSWP-2 Mississippi River basin surface water and energy budgets. *Journal of Geophysical Research*, **113**, D15 118, doi:10.1029/2007JD009144.

Fennessey, N. and R. M. Vogel, 1990: Regional flow-duration curves for ungauged sites in Massachusetts. *Journal of Water Resources Planning and Management*, **116 (4)**, 530–549.

Gebert, W. A., D. J. Graczyk, and W. R. Krug, 1987: Average annual runoff in the United States, 1951–80. Tech. Rep. HA-710, Hydrologic Investigations Atlas, U.S. Department of the Interior, 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. *Geophysical Research Letters*, **38 (L02401)**, doi: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. *Hydrological Processes*, **25 (7)**, 1129–1145.

Groisman, P. Y. and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society*, **72 (2)**, 215–227.

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. *Hydrology and Earth System Sciences*, **18**, 463–477.

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 Resources*, **35 (2)**, 171–184.

Gustard, A., A. Bullock, and J. M. Dixon, 1992: Low flow estimation in the united kingdom. Tech. Rep. 108, Institute of Hydrology, Wallingford, UK.

Gustard, A. and K. M. Irving, 1994: Classification of the low flow response of European soils. *FRIEND: Flow Regimes from International Experimental and Network Data (Proceedings of the Braunschweig Conference, October 1993), IAHS publication no. 221*, P. Seuna, A. Gustard, N. W. Arnell, and G. A. Cole, Eds., 113–117.

Haddeland, I., et al., 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *Journal of Hydrometeorology*, **12 (5)**, 869–884.

Hall, F. R., 1968: Base-flow recessions—a review. *Water Resources Research*, **4 (5)**, 973–983.

Hargreaves, G. L., G. H. Hargreaves, and J. P. Riley, 1985: Irrigation water requirements for seegal river basin. *Journal of Irrigation and Drainage Engineering*, **111 (3)**, 265–275.

He, Y., A. Bárdossy, and E. Zehe, 2011: A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, **15**, 3539–3553.

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. *International Journal of Climatology*, **25 (15)**, 1965–1978.

Hope, A. and R. Bart, 2012: Synthetic monthly flow duration curves for the Cape Floristic Region, South Africa. *Water SA*, **38 (2)**, 191–200.

Horton, R. E., 1933: The role of infiltration in the hydrological cycle. *Trans. Am. Geophys. Union*, **14**, 446–460.

Hrachowitz, M., et al., 2013: A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, **58 (6)**, 1198–1255.

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, Pietermaritzburg, South Africa, URL `http://researchspace.ukzn.ac.za/xmlui/handle/10413/5416`.

Institute of Hydrology, 1980: Low flow studies. Tech. Rep. 1, Institute of Hydrology, Wallingford, UK.

Jia, G. J., H. E. Epstein, and D. A. Walker, 2003: Greening of arctic Alaska, 1981–2001. *Geophysical Research Letters*, **30 (20)**, 2067, doi:10.1029/2003GL018268.

Johnson, D. H., 1999: The insignificance of statistical significance testing. *The Journal of Wildlife Management*, **63 (3)**, 763–772.

Johnston, R. M., et al., 2003: ASRIS: the database. *Australian Journal of Soil Research*, **41 (6)**, 1021–1036.

Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, **17 (7)**, 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. *Journal of Hydrology*, **362 (1–2)**, 39–56.

Krakauer, N. Y. and M. Temimi, 2011: Stream recession curves and storage variability in small watersheds. *Hydrology and Earth System Sciences*, **15 (7)**, 2377–2389.

Laaha, G. and G. Blöschl, 2006: A comparison of low flow regionalisation methods—catchment grouping. *Journal of Hydrology*, **323 (1–4)**, 193–214.

Lacey, G. and R. Grayson, 1998: Relating baseflow to catchment properties in south-eastern Australia. *Journal of Hydrology*, **204 (1–4)**, 231–250.

Lehner, B., 2012: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network. Tech. Rep. 41, Global Runoff Data Centre (GRDC), Federal Institute of Hydrology (BfG), Koblenz, Germany.

Lehner, B. and P. Döll, 2004: Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296 (1–4)**, 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. *Hydrological Processes*, **26 (5)**, 741–759.

Lohmann, D., et al., 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 and Planetary Change*, **19 (1)**, 161–179.

Lohmann, D., et al., 2004: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *Journal of Geophysical Research*, **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. *Physics and Chemistry of the Earth, Parts A/B/C*, **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. *Journal of Hydrology*, **355 (1–4)**, 63–75.

Marthews, T. R., S. J. Dadson, B. Lehner, S. Abele, and N. Gedney, 2014: A high-resolution global dataset of topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences Discussions*, **11**, 6139–6166.

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. *Journal of Hydrometeorology*, **11 (2)**, 334–351.

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. *Physics and Chemistry of the Earth, Parts A/B/C*, **30 (11)**, 639–647.

Milly, P. C. D., 1994: Climate, soil water storage, and the average annual water balance. *Water Resources Research*, **30 (7)**, 2143–2156.

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. *Journal of Hydrometeorology*, **10 (5)**, 1128–1150.

Neff, B. P., S. M. Day, A. R. Piggott, and L. M. Fuller, 2005: Base flow in the Great Lakes Basin. Scientific Investigations Report 2005-5217, U.S. Geological Survey, Reston, VA, US. URL `http://pubs.usgs.gov/sir/2005/5217/`.

Nicholls, N., 2001: Commentary and analysis: the insignificance of significance testing. *Bulletin of the American Meteorological Society*, **82 (5)**, 981–986.

Niu, G.-Y., et al., 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **116 (D12)**, doi:10.1029/2010JD015139.

Nuttle, W., 2002: Taking stock of water resources. *Eos, Transactions American Geophysical Union*, **83 (45)**, 513.

Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assesment of annual runoff from land surface model using Total Runoff Integration Pathways (TRIP). *Journal of the Meteorological Society of Japan*, **77 (1)**, 235–255.

Ol'dekop, E. M., 1911: Ob isparenii s poverknosti rechnykh basseinov (On evaporation from the surface of river basins). *Transactions on Meteorological Observations, University of Tartu 4*.

Olden, J. D. and N. L. Poff, 2003: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, **19 (2)**, 101–121.

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. *Journal of Hydrology*, **357 (3–4)**, 303–316, doi:10.1016/j.jhydrol.2008.05.021.

Pan, M., et al., 2003: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. evaluation of model-simulated snow cover extent. *Journal of Geophysical Research*, **108 (D22)**, 8850.

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. *Hydrology and Earth System Sciences*, **14 (11)**, 2193–2205.

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. Report prepared for the Australian National Land and Water Resources Audit. Centre for Environmental Applied Hydrology, University of Melbourne, Australia.

Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, **11**, 1633–1644, doi:10.5194/hess-11-1633-2007.

Peel, M. C., T. A. McMahon, and B. L. Finlayson, 2010: Vegetation impact on mean annual evapotranspiration at a global catchment scale. *Water Resources Research*, **46**, W09 508.

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. *Journal of Hydrology*, **468–469**, 139–150, doi:10.1016/j.jhydrol.2012.08.028.

Pettyjohn, W. A. and R. Henning, 1979: Preliminary estimate of groundwater recharge rates, related streamflow and water quality in Ohio. Water Resources Centre Project Completion Report 552, Ohio State University.

Pike, J. G., 1964: The estimation of annual run-off from meteorological data in a tropical climate. *Journal of Hydrology*, **2 (2)**, 116–123.

Pinheiro, V. B. and M. Naghettini, 2012: Calibration of the parameters of a rainfall-runoff model in ungauged basins using synthetic flow duration curves as estimated by regional analysis. *Journal of Hydrologic Engineering (in press)*, doi:10.1061/(ASCE)HE.1943-5584.0000737.

Poff, N. L. R., D. Allan, M. B. Bain, J. R. Karr, and K. L. Prestegaard, 1997: The natural flow regime. *BioScience*, **47 (11)**, 769–784.

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 Resources Research*, **41 (6)**, doi:10.1029/2004WR003697.

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

Robock, A., et al., 2003: Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season. *Journal of Geophysical Research*, **108 (D22)**, 8846.

Rodell, M., E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoing, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrological Processes*, **25 (26)**, 4082–4092.

Rodell, M., et al., 2004: The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85 (3)**, 381–394.

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. *Journal of Hydrometeorology*, **12 (1)**, 45–64.

Royall, R. M., 1986: The effect of sample size on the meaning of significance tests. *The American Statistician*, **40 (4)**, 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. *Hydrology and Earth System Sciences*, **17**, 2637–2652.

Sanborn, S. C. and B. P. Bledsoe, 2006: Predicting streamflow regime metrics for ungauged streamsin Colorado, Washington, and Oregon. *Journal of Hydrology*, **325 (1–4)**, 241–261.

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. *Journal of Geophysical Research*, **101 (3)**, 7461.

Schreiber, P., 1904: Über die Beziehungen zwischen dem Niederschlag und der Wasseruhrung der Flusse in Mitteleuropa. *Meteorologisches Zeitschrift*, **21 (1)**, 441–452.

Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercomparisons. *Atmospheric Research*, **92 (3)**, 376–380.

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. *Journal of Climate*, **19 (13)**, 3088–3111.

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

Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy, 2003: Downward approach to hydrological prediction. *Hydrological Processes*, **17 (11)**, 2101–2111.

Slater, A. G., T. J. Bohn, J. L. McCreight, M. C. Serreze, and D. P. Lettenmaier, 2007: A multimodel simulation of pan-Arctic hydrology. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, **112 (G4)**, doi:10.1029/2006JG000303.

Sloto, R. A. and M. Y. Crouse, 1996: HYSEP: a computer program for streamflow hydrograph separation and analysis. Water-Resources Investigations Report 96-4040, U.S. Geological Survey. URL `water.usgs.gov/software/HYSEP/code/doc/hysep.pdf`.

Smakhtin, V. U., 2001: Low flow hydrology: a review. *Journal of Hydrology*, **240 (3–4)**, 147–186.

Smakhtin, V. Y., D. A. Hughes, and E. Creuse-Naudin, 1997: Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrological Sciences Journal*, **42 (6)**, 919–936.

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 (1–3)**, 217–232.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2005: Changes toward earlier streamflow timing across western North America. *Journal of Climate*, **18 (8)**, 1136–1155.

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. *International Journal of Climatology*, **26 (14)**, 2097–2115.

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. *Physics and Chemistry of the Earth*, **66**, 16–26.

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. URL `http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf`.

Van Dijk, A. I. J. M., 2010: Climate and terrain factors explaining streamflow response and recession in Australian catchments. *Hydrology and Earth System Sciences*, **14 (1)**, 159–169.

Van Dijk, A. I. J. M., J. L. Peña-Arancibia, 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 Resources Research*, **49 (5)**, 2729–2746.

Vano, J. A., T. Das, and D. P. Lettenmaier, 2012: Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, **13 (3)**, 932–949.

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*, CDSU, 97–122.

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

Vogel, R. M., I. Wilson, and C. Daly, 1999: Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, **125 (3)**, 148–157.

Wagener, T. and A. Montanari, 2011: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, **47 (6)**, doi:10.1029/2010WR009469.

Wahl, K. L. and T. L. Wahl, 1995: Determining the flow of comal springs at New Braunfels, Texas. *Texas Water '95, American Society of Civil Engineers, August 16-17*, San Antonio, Texas, 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.

Wolock, D. M., 2003: Base-flow index grid for the conterminous United States, Open-File Report 03–263, U.S. Geological Survey. URL `http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml`.

Wolock, D. M. and G. J. McCabe, 1999: Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Research*, **11**, 149–159.

Wolock, D. M., T. C. Winter, and G. McMohan, 2004: Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental Management*, **34 (1)**, 71–88.

Wood, E. F., et al., 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, **47 (5)**, doi:10.1029/2010WR010090.

Xia, Y., et al., 2012: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, **117**, D03 110, doi:10.1029/2011JD016048.

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

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 Resources Research*, **46 (6)**, doi:10.1029/2009WR007811.

Zecharias, Y. B. and W. Brutsaert, 1988: The influence of basin morphology on groundwater outflow. *Water Resources Research*, **24 (10)**, 1645–1650.

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 Resources Research*, **40**, W02 502.

Zhang, L., 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. *Journal of Hydrology*, **360 (1–4)**, 117–131.

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. *Journal of Hydrology*, **470–471**, 269–279.