# Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments

Hylke E. Beck,[a] Eric F. Wood,[a] Tim R. McVicar,[b,c] Mauricio Zambrano-Bigiarini,[d,e] Camila Alvarez-Garreton,[e,f] Oscar M. Baez-Villanueva,[g,h] Justin Sheffield,[i] and Dirk N. Karger[j]

[a] *Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey*
[b] *CSIRO Land and Water, Black Mountain, Canberra, Australia*
[c] *Australian Research Council Centre of Excellence for Climate Extremes, Canberra, Australia*
[d] *Department of Civil Engineering, Universidad de La Frontera, Temuco, Chile*
[e] *Center for Climate and Resilience Research, Santiago, Chile*
[f] *Institute of Conservation, Biodiversity and Territory, Universidad Austral de Chile, Valdivia, Chile*
[g] *Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Technology Arts Sciences TH Köln, Cologne, Germany*
[h] *Faculty of Spatial Planning, TU Dortmund University, Dortmund, Germany*
[i] *School of Geography and Environmental Science, University of Southampton, Southampton, United Kingdom*
[j] *Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

## ABSTRACT

We introduce a set of global high-resolution (0.05°) precipitation ($P$) climatologies corrected for bias using streamflow ($Q$) observations from 9372 stations worldwide. For each station, we inferred the "true" long-term $P$ using a Budyko curve, which is an empirical equation relating long-term $P$, $Q$, and potential evaporation. We subsequently calculated long-term bias correction factors for three state-of-the-art $P$ climatologies [the "WorldClim version 2" database (WorldClim V2); Climatologies at High Resolution for the Earth's Land Surface Areas, version 1.2 (CHELSA V1.2 ); and Climate Hazards Group Precipitation Climatology, version 1 (CHPclim V1)], after which we used random-forest regression to produce global gap-free bias correction maps for the $P$ climatologies. Monthly climatological bias correction factors were calculated by disaggregating the long-term bias correction factors on the basis of gauge catch efficiencies. We found that all three climatologies systematically underestimate $P$ over parts of all major mountain ranges globally, despite the explicit consideration of orography in the production of each climatology. In addition, all climatologies underestimate $P$ at latitudes >60°N, likely because of gauge undercatch. Exceptionally high long-term correction factors (>1.5) were obtained for all three $P$ climatologies in Alaska, High Mountain Asia, and Chile—regions characterized by marked elevation gradients, sparse gauge networks, and significant snowfall. Using the bias-corrected WorldClim V2, we demonstrated that other widely used $P$ datasets (GPCC V2015, GPCP V2.3, and MERRA-2) severely underestimate $P$ over Chile, the Himalayas, and along the Pacific coast of North America. Mean $P$ for the global land surface based on the bias-corrected WorldClim V2 is 862 mm yr$^{-1}$ (a 9.4% increase over the original WorldClim V2). The annual and monthly bias-corrected $P$ climatologies have been released as the Precipitation Bias Correction (PBCOR) dataset, which is available online (http://www.gloh2o.org/pbcor/).

---

ꝺ Denotes content that is immediately available upon publication as open access.

🖉 Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-19-0332.s1.

*Corresponding author address*: Hylke E. Beck, hylke.beck@gmail.com

## 1. Introduction

Information about the spatiotemporal distribution of precipitation $P$ is crucial for numerous scientific, commercial, and operational applications (e.g., Tapiador et al. 2012; Kucera et al. 2013; Kirschbaum et al. 2017). Such information can be obtained from three main data sources: 1) satellites, 2) reanalyses, and 3) $P$ gauges (Strangeways 2007; Sun et al. 2018). However, all three sources show reduced accuracy in mountainous and

snowfall-dominated regions (Roe 2005; Wrzesien et al. 2019). Satellite retrievals are confounded by surface snow and ice (Kidd et al. 2012; Cao et al. 2018) and there are major challenges associated with the detection of snowfall (Levizzani et al. 2011; Skofronick-Jackson et al. 2015). Reanalyses [e.g., JRA-55 (Kobayashi et al. 2015), MERRA-2 (Reichle et al. 2017), and ERA5 (Hersbach et al. 2018)] rely on uncertain parameterizations, and their spatial resolutions may be too coarse ($\geq 0.25°$) to adequately represent orographic $P$ (Skamarock 2004; Ménégoz et al. 2013; Liu et al. 2018). Precipitation gauge networks tend to be sparse or nonexistent and topographically biased toward low elevations in mountainous regions (Briggs and Cogley 1996; Schneider et al. 2014; Kidd et al. 2017), and gauges can underestimate snowfall by up to 90% because of wind-induced undercatch (Groisman and Legates 1994; Sevruk et al. 2009; Rasmussen et al. 2012).

Observations of streamflow $Q$ have revealed that existing $P$ datasets—including those explicitly considering gauge undercatch and orographic effects—underestimate $P$ in mountainous and snow-dominated regions as a result of the aforementioned limitations (e.g., Oki et al. 1999; Fekete et al. 2004; Biemans et al. 2009; Kauffeldt et al. 2013; Beck et al. 2015, 2017a; Immerzeel et al. 2015; Prein and Gobiet 2017; Alvarez-Garreton et al. 2018; Ghatak et al. 2018). Recently, several global high-resolution $P$ climatologies have been developed, such as WorldClim, version 2 (WorldClim V2; Fick and Hijmans 2017); Climatologies at High Resolution for the Earth's Land Surface Areas, version 1.2 (CHELSA V1.2; Karger et al. 2017); and the Climate Hazards Group Precipitation Climatology, version 1 (CHPclim V1; Funk et al. 2015). These datasets have been specifically designed to provide highly accurate climatic $P$ estimates over the entire land surface by combining disparate data sources. However, the gauge data used to derive WorldClim V2 and CHPclim V1 have not been explicitly corrected for undercatch, and, despite the explicit consideration of orography in the production of each climatology, comparisons with $Q$ observations suggest that CHPclim V1 still underestimates orographic $P$ (Beck et al. 2017b). These underestimations are concerning because mountainous and snow-dominated regions contribute a large share of the world's population with freshwater (Viviroli and Weingartner 2004; Viviroli et al. 2007).

Only a few, mostly regional, studies have inferred or corrected $P$ estimates using $Q$ observations (Adam et al. 2006; Salo 2006; Weingartner et al. 2007; Lundquist et al. 2009; Valery et al. 2009; Henn et al. 2015, 2016, 2018; Le Moine et al. 2015; Koppa and Gebremichael 2017). Several studies focused on the Sierra Nevada mountain range of California and demonstrated the value of $Q$ observations for improving individual storm estimates through visual time series comparison (Lundquist et al. 2009) and for inferring long-term $P$ through hydrological modeling (Henn et al. 2015, 2016, 2018). Two studies focused on the European Alps: Weingartner et al. (2007) inferred long-term $P$ from $Q$ observations using a water balance approach, while Le Moine et al. (2015) used $Q$ observations to enhance the interpolation of station-based $P$ and air temperature data. Salo (2006) and Adam et al. (2006) were the first to recognize the value of the Budyko curve (Donohue et al. 2007; Wang et al. 2016)—a parsimonious first-order empirical equation relating long-term actual evaporation $E$, $P$, and potential evaporation $E_p$—for inferring $P$ from $Q$ observations. Salo (2006) used just two Russian catchments and inferred $P$ at annual and long-term time scales, while Adam et al. (2006) used 357 catchments worldwide and inferred only long-term $P$. Adam et al. (2006) represents the only global study thus far. The usefulness of the Budyko curve was further demonstrated by Valery et al. (2009), who corrected long-term $P$ in Swedish and Swiss catchments using $Q$ observations, and more recently, by Koppa and Gebremichael (2017), who illustrated how the Budyko curve can be used in conjunction with $Q$ observations to identify systematic biases in both $P$ and $E_p$ data for catchments in the conterminous United States.

Here, we quantify the magnitude of $P$ underestimation globally using a Budyko curve in combination with an unprecedented database of $Q$ observations from 9372 stations worldwide. As baseline, we use the three aforementioned high-resolution $P$ climatologies (i.e., WorldClim V2, CHELSA V1.2, and CHPclim V1). Additionally, we use random-forest (RF) regression to derive global gap-free bias correction maps for each $P$ climatology. Finally, we present and discuss improved estimates of mean $P$ over the land surface.

## 2. Data and methods

### a. Observed streamflow and derived runoff

We used an initial database of daily and monthly observed $Q$ and catchment boundaries for 21 940 stations across the globe. The database was compiled from seven national and international sources (listed in descending order of number of catchments):

1) the U.S. Geological Survey (USGS) National Water Information System (NWIS; https://waterdata.usgs.gov/nwis) and GAGES-II database (Falcone et al. 2010; 9180 catchments),
2) the Global Runoff Data Centre (GRDC; http://grdc.bafg.de; Lehner 2012; 4628 catchments),
3) the HidroWeb portal of the Brazilian Agência Nacional de Águas (http://www.snirh.gov.br/hidroweb; 3029 catchments),
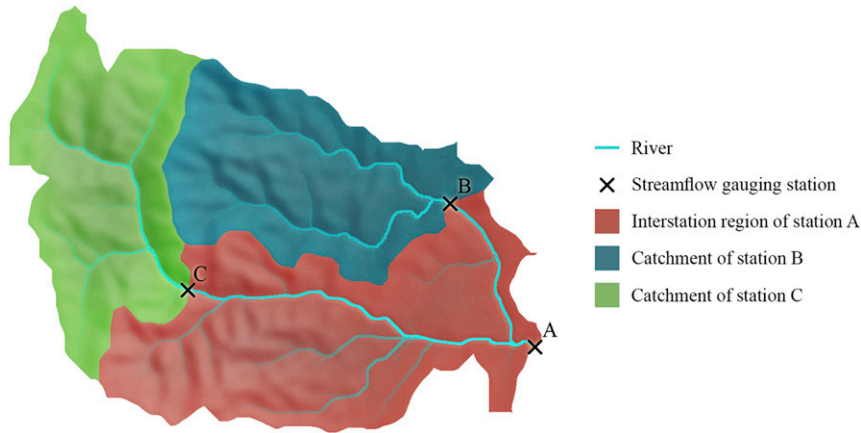
FIG. 1. Example of an interstation region in red. The interstation region of station A consists of the catchment of station A excluding the catchments of nested upstream stations B and C.

4) the European Water Archive of the European Flow Regimes from International Experimental and Network Data (EURO-FRIEND-Water; http://ne-friend.bafg.de) and the Catchment Characterisation and Modelling–Joint Research Centre database (CCM2-JRC; Vogt et al. 2007; 2260 catchments),

5) the Water Survey of Canada Hydrometric Data (HYDAT; https://www.canada.ca/en/environment-climate-change; 1479 catchments),

6) the Australian Bureau of Meteorology (BoM; http://www.bom.gov.au/waterdata; Zhang et al. 2013; 776 catchments), and

7) the Chilean Center for Climate and Resilience Research (CR2; http://www.cr2.cl/datos-de-caudales/) and Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS-CL; Alvarez-Garreton et al. 2018; 516 catchments).

For each station, monthly $Q$ values were calculated by averaging the daily values if $\geq 26$ daily values were available, annual $Q$ values were calculated by averaging the monthly values if $\geq 11$ monthly values were available, and mean annual $Q$ values were calculated by averaging the annual values if $\geq 5$ annual values were available (not necessarily consecutive). We subsequently calculated the mean annual runoff $R$ generated in the "interstation" region [the catchment area of the station excluding, if needed, the catchment area(s) of nested upstream catchment(s), following Fekete et al. (2002); see Fig. 1 herein] as follows:

$$R_i = \frac{Q_i - \sum_{k=1}^{n} Q_k}{1000 \left( A_i - \sum_{k=1}^{n} A_k \right)}, \quad (1)$$

where $R_i$ is the mean annual runoff depth (mm yr$^{-1}$) for the interstation region of station $i$, $Q$ denotes mean annual volumetric streamflow (m$^3$ yr$^{-1}$), $A$ represents catchment area (km$^2$), and the summation is over nested upstream stations $k = 1, 2, \ldots, n$. Since stations located close to each other on the same river reach yield $R$ values that reflect systematic bias in the $Q$ measurements rather than the actual $R$, stations with interstation area $<15\%$ of the total catchment area were excluded. Additionally, we omitted stations with catchment or interstation areas of less than 200 km$^2$, to ensure that each station covers at least two grid cells of the inputs (Table 1). In total, 9372 stations of the initial set of 21 940 stations passed both criteria. The 10th-, 50th-, and 90th-percentile interstation areas of the remaining stations are 281, 929, and 8372 km$^2$, respectively.

### b. Precipitation inference from observed runoff

The Budyko framework is encapsulated by empirical first-order equations relating long-term $E$ to the water supply (expressed by $P$) and demand (expressed by $E_p$) at catchment scales (Donohue et al. 2007; Wang et al. 2016). For each interstation region, we inverted the Budyko function to infer long-term $P$ from long-term $R$ and long-term $E_p$ by assuming that 1) $E = P - R$, 2) long-term water storage changes are negligible, and 3) $P$ is the only input of water, and $R$ and $E$ are the only outputs. A large number of functional forms of the Budyko curve with slightly varying shapes have been proposed since the early twentieth century (e.g., Ol'dekop 1911; Pike 1964; Budyko 1974; Fu 1981; Choudhury 1999; Zhang et al. 2001; Porporato et al. 2004). We adopted the widely used Fu (1981) curve (Fig. 2a) given by

TABLE 1. Overview of gridded precipitation $P$ and potential evaporation $E_p$ climatologies used in this study.

| Short name | Full name and details | Temporal span | Spatial resolution | Reference |
|---|---|---|---|---|
| | *P* datasets | | | |
| WorldClim V2 | WorldClim version 2 (http://www. worldclim.org) | 1970–2000 | 0.0083° | Fick and Hijmans (2017) |
| CHELSA V1.2 | Climatologies at High Resolution for the Earth's Land Surface Areas version 1.2 (http://chelsa-climate.org) | 1979–2013 | 0.0083° | Karger et al. (2017) |
| CHPclim V1 | Climate Hazards Group's Precipitation Climatology version 1 (http:// chg.geog.ucsb.edu/data/CHPclim/) | 1980–2009 | 0.05° | Funk et al. (2015) |
| | $E_p$ datasets | | | |
| WC-HG | Calculated using Hargreaves (1994) from WorldClim V1 (Hijmans et al. 2005) min and max daily temperature (https://cgiarcsi.community/data/ global-aridity-and-pet-database/) | 1960–90 | 0.0083° | Trabucco et al. (2008) |
| WFDEI-HG | Calculated using Hargreaves (1994) from downscaled Water and Global Change (WATCH) Forcing Data Method Applied to ERA-Interim Data (WFDEI) data (Weedon et al. 2014) | 1979–2010 | 0.1° | Sperna Weiland et al. (2015) |
| WFDEI-PT | Calculated using Priestley and Taylor (1972) from downscaled WFDEI data | 1979–2010 | 0.1° | Sperna Weiland et al. (2015) |
| WFDEI-PM | Calculated using Penman–Monteith (Monteith 1965) from downscaled WFDEI data | 1979–2010 | 0.1° | Sperna Weiland et al. (2015) |

$$\frac{E}{P} = 1 + \frac{E_p}{P} - \left[ 1 + \left( \frac{E_p}{P} \right)^w \right]^{1/w}, \qquad (2)$$

where $E$ is the mean annual actual evaporation, $P$ is the mean annual precipitation, $E_p$ is the mean2 annual potential evaporation (all in millimeters per year), and $w$

is an empirical parameter that represents catchment characteristics (unitless).

The $w$ parameter exhibits substantial variability across catchments due to differences in, among other factors, $P$ and $E_p$ seasonality, $P$ phase, vegetation cover, soil moisture capacity, and topography (Milly 1994; Potter
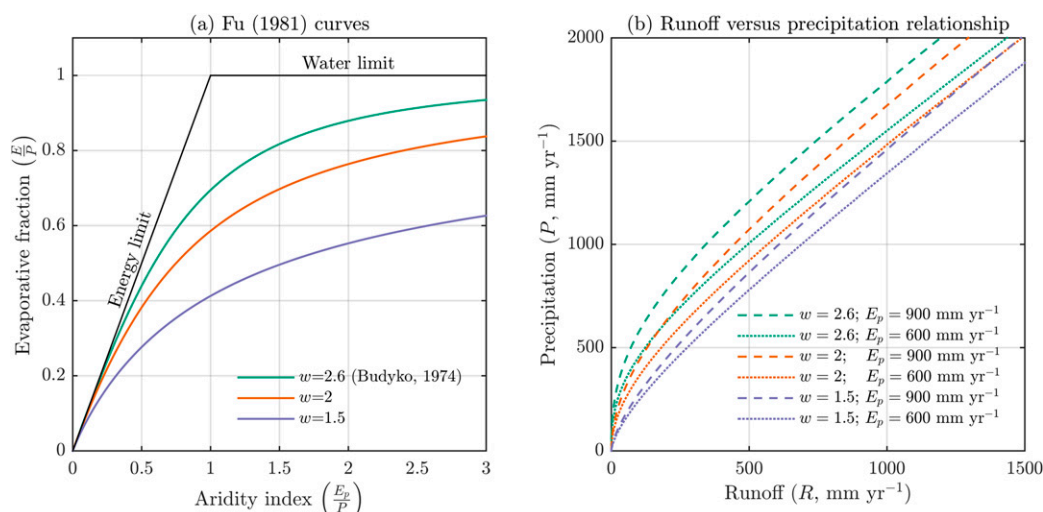


FIG. 2. (a) Fu (1981) curves [Eq. (2)] for different $w$ values. The energy limit corresponds to $E = E_p$, whereas the water limit corresponds to $E = P$. (b) Relationships for mean annual runoff $R$ vs precipitation $P$, obtained by numerically solving the Fu (1981) equation for different $w$ and potential evaporation $E_p$ values.

TABLE 2. Predictors used in the RF regression models to estimate $P$ bias correction factors.

| Predictor | Description | Resolution | Data source |
|---|---|---|---|
| Elev | Surface elevation (m above mean sea level) | 90 m | Multi-Error-Removed Improved Terrain (MERIT; Yamazaki et al. 2017) |
| GradN | North component of the smoothed elevation gradient (m) | 90 m | See Elev |
| GradE | East component of the smoothed elevation gradient (m) | 90 m | See Elev |
| $f$snow | Long-term fraction of total $P$ falling as snow (calculation provided in appendix A) | 1 km | WorldClim V2 (Fick and Hijmans 2017) |
| CldCov | Cloud cover frequency | 1 km | Wilson and Jetz (2016) |
| Lat | Geographical latitude (°) | — | — |
| Lon | Geographical longitude (°) | — | — |

et al. 2005; Donohue et al. 2010a; Xu et al. 2013; Wang and Tang 2014; Padrón et al. 2017; de Lavenne and Andréassian 2018). Although the default value for $w$ is 2.6 [which yields the Budyko (1974) curve; see Fig. 2a herein] this value is probably too high for our purpose given the $E$ reduction observed in snow-dominated environments (Berghuijs et al. 2014; Zhang et al. 2015; Padrón et al. 2017). Xu et al. (2013) and Padrón et al. (2017) obtained $w$ values of around 1–2 in the global northern regions, but they used $P$ data lacking gauge undercatch corrections, resulting in too low $w$ values. Therefore, as a compromise, we considered, for each interstation region, 10 random values for $w$ drawn from a normal distribution with mean 2 and standard deviation 1 (truncated at a lower limit of 1.2). Assuming an aridity index (ratio between long-term $E_p$ and $P$) of 1, the spread in $w$ values introduces a variance in the inferred mean annual $P$ of 19% (Fig. 2b).

The inferred mean annual $P$ values will also be affected by uncertainties in the $E_p$ and $R$ estimates. To account for uncertainty in the $E_p$ estimates (Weiß and Menzel 2008; Donohue et al. 2010b; Fisher et al. 2011), we considered four $E_p$ datasets derived using various $E_p$ formulations and meteorological datasets (Table 1; Fig. 2b). To account for uncertainty in the $R$ estimates, following Di Baldassarre and Montanari (2009) and Kiang et al. (2018), we assumed an error of 25% and drew 10 random values for $R$ from a normal distribution with mean $R$ and standard deviation $0.25 \times R$, introducing a variance in the inferred mean annual $P$ of 25% (Fig. 2b).

For each interstation region, we inferred mean annual $P$ for all 400 possible combinations of $w$ parameters (10 options), $E_p$ estimates (4 options), and $R$ estimates (10 options) by numerically solving Eq. (2). From this distribution, we used the median to provide a "best estimate" mean annual $P$, and the 25th and 75th percentiles to represent upper and lower uncertainty boundaries, respectively.

### c. Estimation of bias correction factors

For each of the 9372 interstation regions, we calculated long-term bias correction factors by dividing the inferred observation-based mean annual $P$ estimates (generated per section 2b) by the interstation-mean estimates from the state-of-the-art $P$ climatologies (WorldClim V2, CHELSA V1.2, and CHPclim V1; Table 1). We subsequently used RF regression models (Breiman 2001; Svetnik et al. 2003) to derive global gap-free bias correction maps for each $P$ climatology. RF regression models are nonparametric ensemble machine learning algorithms that grow a "forest" of individual regression trees. Advantages of RF regression models are their high predictive accuracy, computational efficiency, ability to model nonlinear relationships, and insensitivity to overfitting (Criminisi et al. 2012; Reichstein et al. 2019). RF regression models have been successfully applied in several previous $P$-related studies (e.g., Ibarra-Berastegi et al. 2011; Kühnlein et al. 2014; Baez-Villanueva et al. 2020). For RF training, interstation regions with bias correction factors > 5 were considered to be erroneous and were discarded. Interstation regions with bias correction factors < 1 were included, to avoid the trained model yielding estimates greater than 1 everywhere. The RF regression models have four parameters to specify: 1) the number of regression trees (set at 100); 2) the leaf size (i.e., the minimum number of observations per node; set at 5); 3) the number of variables to select at random for each decision split (set at one-third of the number of predictors; see Table 2); and 4) the number of observations randomly withheld for each tree (i.e., the "out of bag" portion; set at one-third of the total number of observations).

Seven predictors were incorporated in the RF regression models (Table 2). Predictors were selected based on their expected relationship with orographic $P$ and gauge undercatch. To highlight large-scale orographic features, the elevation predictor (Elev) was smoothed using a Gaussian kernel with a 10-km radius

following recommendations by Hutchinson (1998) and Smith et al. (2003). North and east components of the elevation gradient (GradN and GradE, respectively) were included to account for $P$ enhancement on windward flanks and $P$ suppression on the lee (Roe 2005). The snowfall fraction predictor ($f$snow) was included as snowfall is subject to greater gauge undercatch (Yang et al. 2005; Sevruk et al. 2009; Rasmussen et al. 2012). The cloud cover predictor (CldCov) was incorporated because cloudiness is associated with $P$ (Fick and Hijmans 2017). Latitude and longitude were also included, to enable region-specific behavior. To train the RF models, we calculated mean predictor values for each interstation region.

For each $P$ climatology, the predictor importance was estimated by random permutation of the out-of-bag values of each predictor. For each tree, the difference in mean square error (MSE) before and after permuting the predictor was calculated. The mean MSE difference over all trees was calculated and normalized with the standard deviation over all trees, yielding the permutation importance of the predictor in question. The permutation importance values are unitless due to the normalization and will not sum up to one. When interpreting the permutation importance values, one should focus on the relative differences among predictors rather than the absolute values.

The trained RF regression models were subsequently applied using global 0.05° maps of the predictors, yielding gap-free gridded long-term bias correction maps for each $P$ climatology covering the entire global land surface. For this purpose, the predictor maps were upscaled from their native resolution (Table 2) to 0.05° using bilinear resampling. The obtained bias correction factors were truncated at a lower limit of 1, as bias correction factors of <1 cannot be confidently attributed to $P$ overestimation, given the prevalence of factors that can reduce $Q$ compared to Budyko's fundamental assumption of a steady state (e.g., anthropogenic water use, channel transmission losses, and reservoir evaporation; Pilgrim et al. 1988; Vörösmarty and Sahagian 2000). Conversely, bias correction factors of >1 can confidently be attributed to $P$ underestimation, given the relative paucity of factors that can increase $Q$ in comparison with Budyko's estimate (perhaps only accelerated glacier melt and interbasin water transfers; Lutz et al. 2015; Emanuel et al. 2015). This truncation is not a major concern as $P$ is more likely to be under than overestimated due to gauge undercatch (Groisman and Legates 1994; Sevruk et al. 2009; Rasmussen et al. 2012) and the low-elevation bias in gauge placement (Briggs and Cogley 1996; Schneider et al. 2014; Kidd et al. 2017).

The delayed response of $Q$ to $P$ (Gericke and Smithers 2014; Liu et al. 2016) prevents us from inferring

monthly bias correction factors from monthly $Q$ data. We therefore calculated monthly climatological correction factors by disaggregating the gridded long-term correction factors derived for each $P$ climatology based on gauge catch ratios according to

$$CF_m = \left(\frac{1}{CR_m} - 1\right)\frac{CF_{yr} - 1}{\frac{1}{CR_{yr}} - 1} + 1, \qquad (3)$$

where $CF_m$ is the monthly bias correction factor, $CF_{yr}$ is the long-term bias correction factor, $CR_m$ and $CR_{yr}$ are the monthly and long-term gauge catch ratio, respectively (both ranging from 0 to 1; appendix B), and $m$ denotes month $(1, 2, \ldots, 12)$.

## 3. Results and discussion

### a. Precipitation inference from observed runoff

Annual $P$ bias correction factors were calculated from $Q$ observations for 9372 interstation regions using three high-resolution $P$ climatologies as baseline (WorldClim V2, CHELSA V1.2, and CHPclim V1). Since the results were similar for the three climatologies, we only present and discuss results for WorldClim V2 (see the online supplemental material for the other climatologies). Figure 3 shows the median inferred bias correction factors and the percentage of inferred bias correction factors >1 for the interstation regions, indicating the likelihood of underestimated $P$. Median correction factors were >1, >1.5, and >2 for 41.3%, 9.0%, and 2.8% of the interstation regions, respectively (Fig. 3a). The proportion of interstation regions with a very high likelihood of underestimated $P$ (defined when >90% of the bias correction factors of the previously introduced 400 possible combinations for each interstation region are greater than 1). was 10.4% (Fig. 3b). Parts of all major mountain ranges across the globe appear to be affected by substantial $P$ underestimation (Fig. 3), despite the explicit consideration of orography in the production of each climatology. Additionally, regions > 60°N consistently showed $P$ underestimation (Fig. 3), due at least partly to a lack of correction for gauge undercatch (Sevruk et al. 2009; Rasmussen et al. 2012). Correction factors were particularly high (>1.5) for interstation regions in Alaska, High Mountain Asia, and Chile (Figs. 3 and 4a). These regions exhibit marked elevation gradients, sparse gauge networks, and substantial snowfall: all factors that tend to favor $P$ underestimation (Adam and Lettenmaier 2003; Adam et al. 2006; Fick and Hijmans 2017; Kidd et al. 2017). Conversely, relatively well-gauged mountain ranges such

(a) Median observation-based bias correction factors



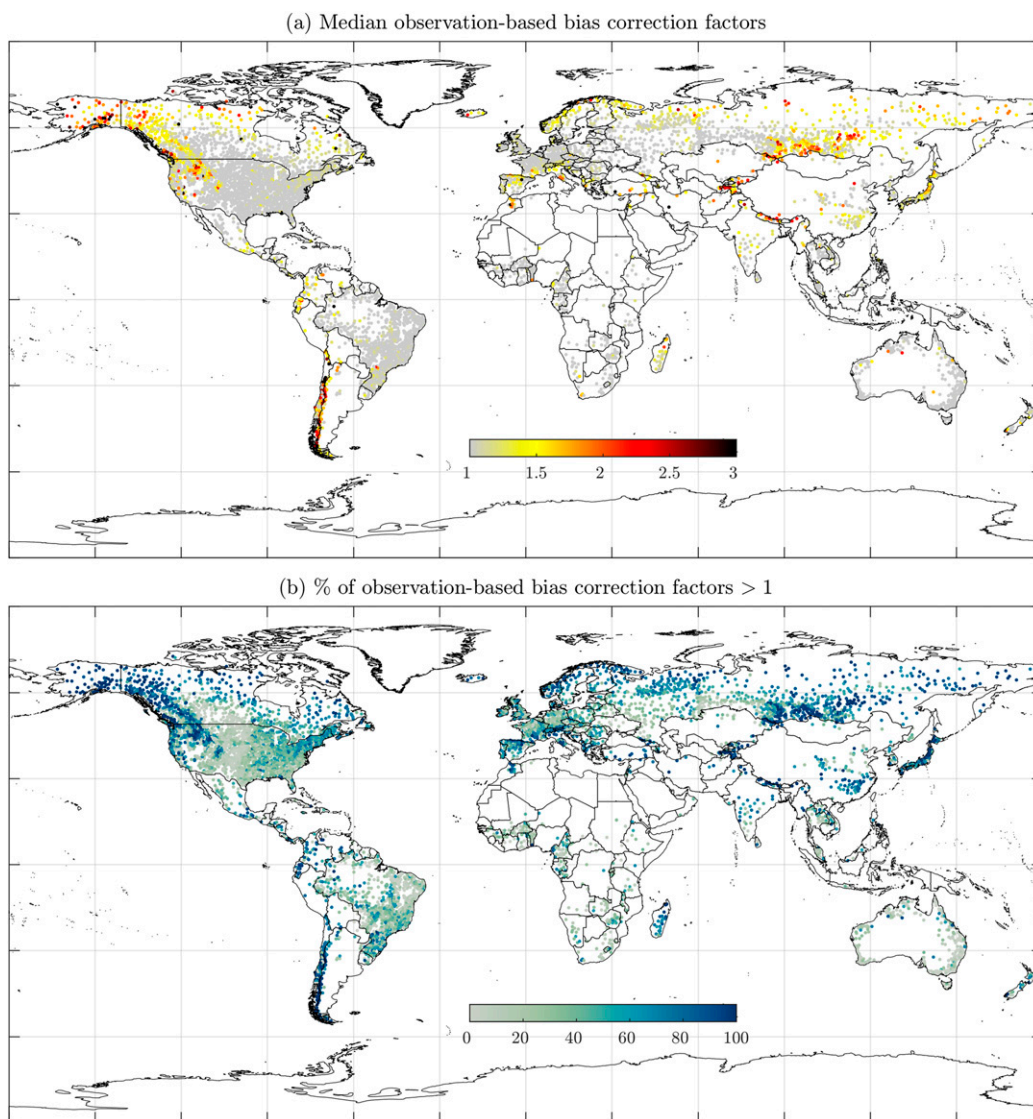(b) % of observation-based bias correction factors > 1



FIG. 3. (a) WorldClim V2 $P$ bias correction factors inferred for the interstation regions ($N = 9372$) using the Fu (1981) curve [see Eq. (2)] from streamflow $Q$ observations and potential evaporation $E_p$ estimates. For each interstation region, we calculated the median of the 400 bias correction factors inferred using different combinations of $w$ (10 options), $E_p$ (4 options), and $R$ (10 options). (b) The percentage of inferred bias correction factors that are greater than 1 for each interstation region, indicating the likelihood of underestimated $P$. In (a) and (b), each dot represents the centroid of an interstation region. Figures S1 and S2 in the online supplemental material present equivalent figures for CHELSA V1.2 and CHPclim V1, respectively.

as the U.S. Appalachians, the European Alps, and the Scandinavian Kjolen exhibit more moderate correction factors ranging from 1 to 1.5 (Fig. 3a), as they are only primarily affected by gauge undercatch.

These correction factors accord with numerous studies that evaluated hydrological model simulations and concluded that $P$ was underestimated (e.g., Oki et al. 1999; Fekete et al. 2004; Biemans et al. 2009; Beck et al. 2017a; Prein and Gobiet 2017; Ghatak et al. 2018). Additionally, our correction factors are consistent with

Kauffeldt et al. (2013), who obtained runoff coefficient (RC; ratio of long-term $R$ to $P$) values >1 for a large number of catchments in Alaska, and with Alvarez-Garreton et al. (2018), who obtained RC values >1 for numerous Andean catchments in Chile. Both studies tested multiple (four) state-of-the-art gridded $P$ datasets with similar results. Our findings also agree with Beck et al. (2015), who obtained RC values >1 using the WorldClim V1 $P$ dataset (Hijmans et al. 2005) for most regions showing $P$ underestimation here (Fig. 3a; see
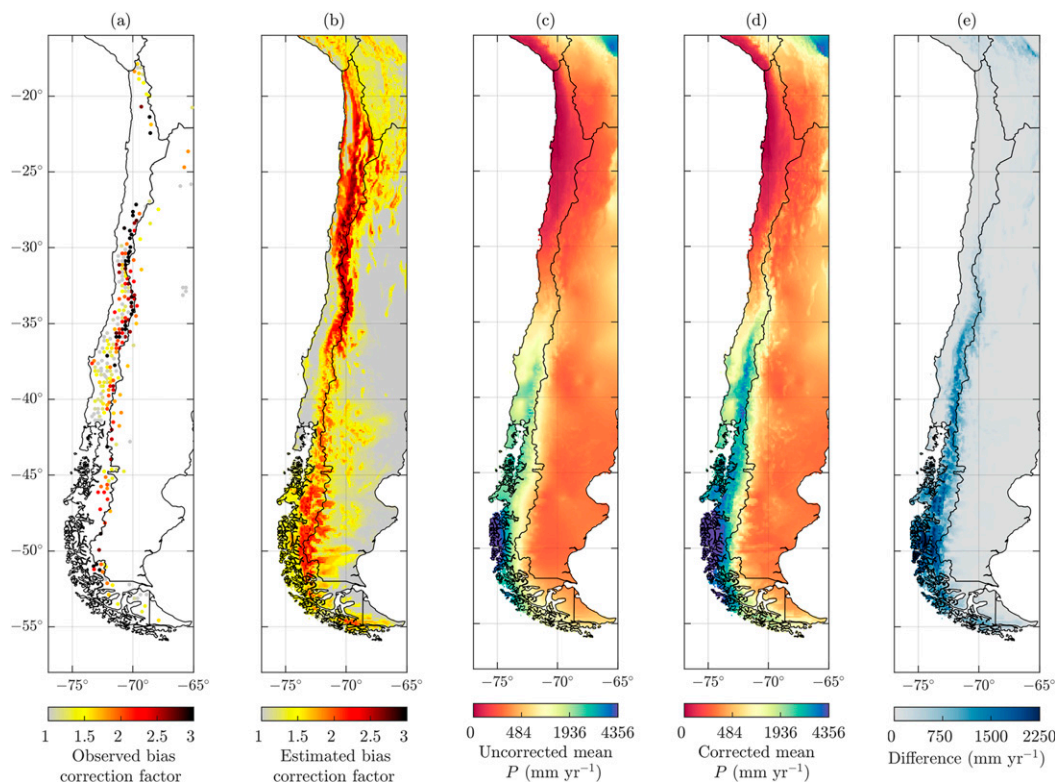
FIG. 4. WorldClim V2 $P$ bias correction results for Chile. (a) Median bias correction factors inferred using Fu (1981) [see Eq. (2) herein] from $Q$ observations and $E_p$ estimates. Each dot represents the centroid of an inter-station region. (b) Bias correction factors estimated using the trained RF regression model. (c) The original uncorrected WorldClim V2 mean $P$, and (d) the corrected WorldClim V2 mean $P$. (e) Corrected minus original WorldClim V2 mean $P$.

also Figs. S1a and S2a in the online supplemental material). Furthermore, our results are consistent with studies that inferred $P$ from $Q$ observations for the Sierra Nevada of California (Henn et al. 2015, 2016, 2018) and for Sweden and Switzerland (Valery et al. 2009), as well as with studies using numerical weather models to confirm that $P$ is severely underestimated across western North America (Wrzesien et al. 2019), Chile (Favier et al. 2009; Garreaud et al. 2016), and the Himalayas (Li et al. 2017; Bonekamp et al. 2018).

### b. Estimation of bias correction factors

RF regression models were trained to globally estimate the observation-based $P$ bias correction factors reported in section 3a. The results were again similar for the three $P$ climatologies, and therefore, we only present figures for WorldClim V2 (see the online supplemental material for the other climatologies). We obtained a training Pearson correlation coefficient $r$ of 0.91 and slightly lower validation $r$ of 0.77 (Figs. 5a and 5b, respectively), suggesting that the employed RF regression model and predictors are effective at estimating $P$ bias correction factors. However,

high values ($>$2) were underestimated (Figs. 5a and 5b); this is characteristic of RF regression models and is due to the averaging of different trees (e.g., Baccini et al. 2004; Kühnlein et al. 2014; Baez-Villanueva et al. 2020). The relatively small difference between the training and validation $r$ values suggests that the trained models generalize reasonably well to unseen data. Longitude (Lon) emerged as the most important predictor (Fig. 5c) because of the distinctly higher bias correction factors in some longitude ranges (Fig. 3a). Elevation (Elev) emerged as the second-most important predictor (Fig. 5c), reflecting the strong influence of orography on $P$ patterns (Roe 2005). The trained RF models were subsequently applied using global 0.05° predictor maps to generate gridded high-resolution long-term bias correction factors for the entire land surface (Fig. 5d). The RF-based maps of correction factors correspond well with the observation-based correction factors (Fig. 3a), although the underestimation of high values is again evident. Figure 4 presents WorldClim V2 results for Chile, highlighting the fine spatial detail of the produced bias correction map and the $>$1000 mm yr$^{-1}$ of $P$ underestimation over the Andes between 35° and
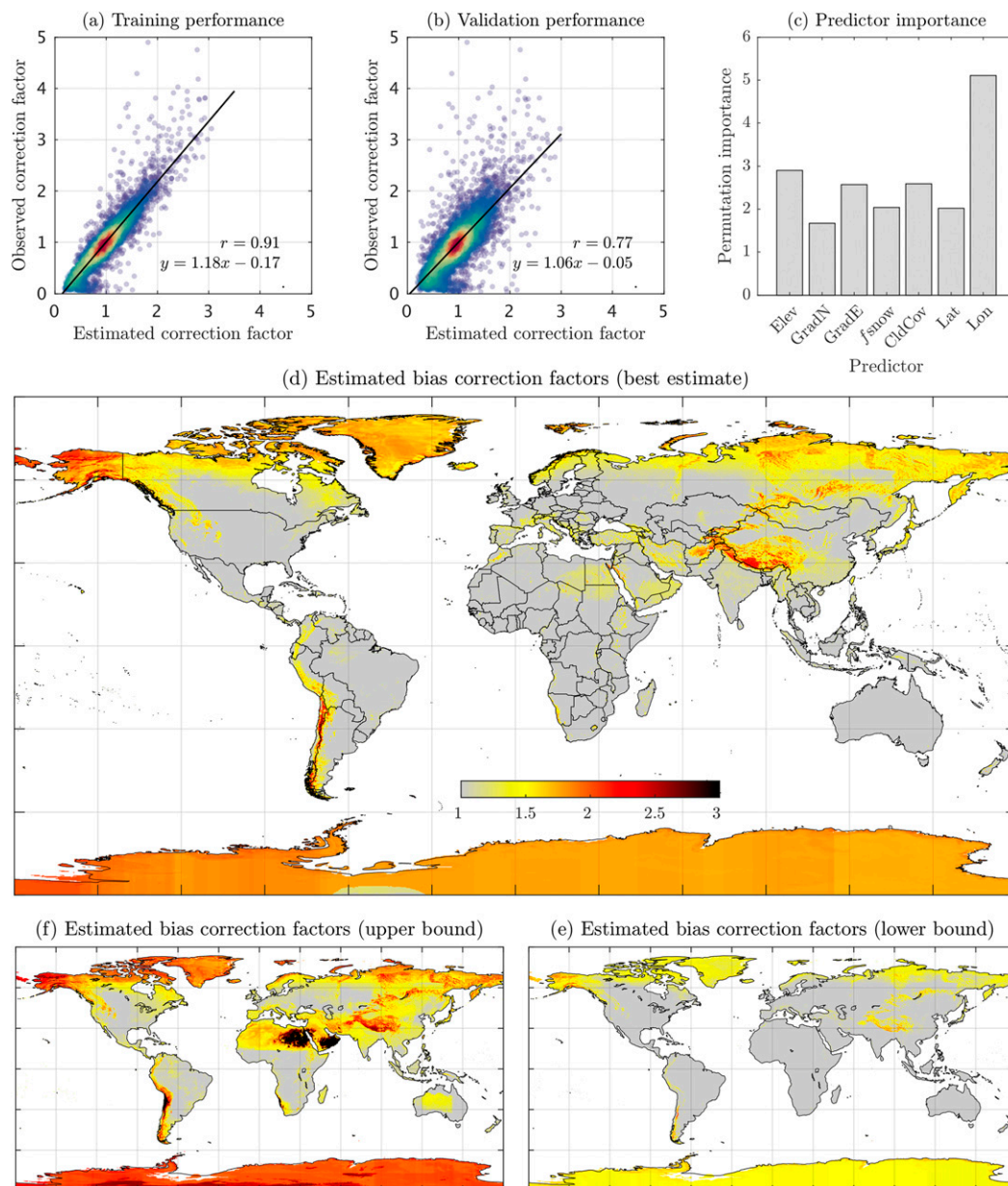
FIG. 5. Scatterplots between RF- and observation-based bias correction factors for (a) the so-called in-bag training data and (b) the out-of-bag validation data. The WorldClim V2 $P$ climatology was used as baseline. Each data point represents an interstation region ($N = 9372$); $r$ denotes the Pearson correlation coefficient. (c) The importance of the predictors incorporated in the RF model. Also shown are bias correction factors obtained by global application of the trained RF model [(d) best estimate, (e) lower uncertainty bound, and (f) upper uncertainty bound]. Figures S3 and S4 in the online supplemental material present equivalent figures for CHELSA V1.2 and CHPclim V1, respectively.

56°S (Favier et al. 2009; Garreaud et al. 2016; Alvarez-Garreton et al. 2018).

Adam and Lettenmaier (2003) estimated gauge undercatch correction factors globally by interpolation of correction factors computed using gauge-type-specific equations from observations of $P$, air temperature, and wind speed from 7878 stations. The map of Adam and Lettenmaier (2003, their Fig. 6) exhibits reasonable agreement with our map (Fig. 5d). Both show high correction factors at high latitudes as a result of wind-induced gauge undercatch and a clear discrepancy between Alaska and Canada due to the lower catch
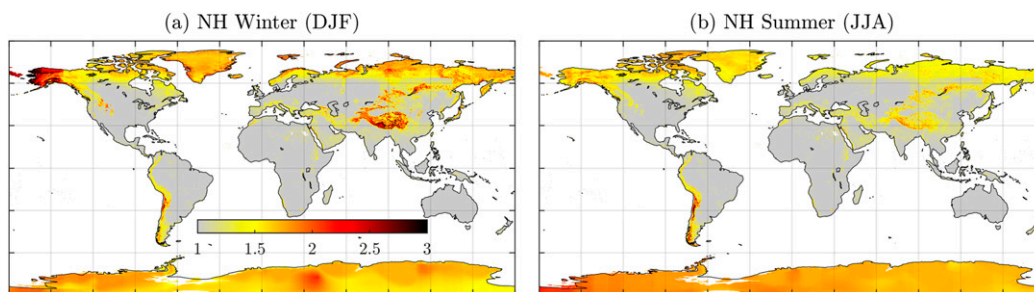
FIG. 6. Northern Hemisphere (a) winter (December–February) and (b) summer (June–August) bias correction factors for the WorldClim V2 $P$ climatology, calculated by averaging the monthly bias correction factors.

efficiency of the National Weather Service 8-in. (∼20 cm) gauge used in the United States relative to the Nipher gauge used in Canada (Scaff et al. 2015). However, the map of Adam and Lettenmaier (2003) lacks fine spatial detail and fails to represent several major mountain ranges (e.g., the Andes, the Himalayas, and the central Asian mountains) because of the sparseness of the station networks.

Similar to the present study, Adam et al. (2006) derived $P$ bias correction factors globally from $Q$ observations using a Budyko curve. Despite the use of a different baseline (the University of Delaware gauge-based interpolated $P$ dataset; monthly 0.5° resolution; Willmott and Matsuura 2001), the bias correction map of Adam et al. (2006, their Fig. 12) exhibits broad agreement with our maps (Fig. 5d and also Figs. S3d and S4d of the online supplemental material). However, their approach is subject to three limitations: 1) they used $Q$ observations from only 524 large catchments (>10 000 km$^2$), whereas we used $Q$ observations from 9372 medium- to large-sized interstation regions (>200 km$^2$ of which 1016 were >10 000 km$^2$); 2) their Budyko curve was optimized using "low relief" catchments, which may not be representative of mountainous and snow-dominated catchments (Xu et al. 2013; Berghuijs et al. 2014); and 3) they excluded several important mountain ranges from the correction domain, such as the U.S. Appalachians and the Russian Urals.

Immerzeel et al. (2015) used glacier mass balance data to infer long-term $P$ for the Hindu Kush region in Asia. They obtained bias correction factors of around 2 on average and up to 10 using the gauge-based APHRODITE $P$ dataset (daily 0.25° resolution; Yatagai et al. 2012), which lacks explicit gauge undercatch corrections, as their baseline. This is in good agreement with our estimates (Fig. 5d and also Figs. S3d and S4d of the online supplemental material); noting that our baseline $P$ datasets have already had orographic correction applied in their production, which is very likely why our maximum bias correction factor is 3 whereas Immerzeel et al. (2015) had similar values approaching 10.

Figure 6 presents climatic bias correction factors for Northern Hemisphere winter and summer, calculated by disaggregating the RF-based long-term correction factors (Fig. 5d) based on gauge catch efficiencies (appendix B). Large differences between summer and winter were found in regions with significant snowfall, emphasizing the importance of using the monthly correction factors rather than the long-term ones for subannual applications. Our maps exhibit substantially more detail than similar maps from previous studies derived by interpolation of correction factors based on $P$, air temperature, and wind speed observations from sparse and unevenly distributed measurement networks (e.g., Legates and Willmott 1990; Adam and Lettenmaier 2003; Schneider et al. 2017). However, our winter correction factors may be on the low side in the northern conterminous United States, possibly suggesting that the $w$ parameter [Eq. (2)] is too low or that $E_p$ (Table 1) is underestimated in this region.

### c. Revision of global land mean P estimates

Figure 7 presents the corrected WorldClim V2 mean annual $P$ map derived herein and the difference with 1) the original WorldClim V2 climatology (monthly 1-km resolution; Fick and Hijmans 2017); 2) GPCC V2015 (monthly 0.5° resolution; Schneider et al. 2017); 3) GPCP V2.3 (monthly 2.5° resolution; Adler et al. 2018); and 4) MERRA-2 (hourly 0.625° resolution; Reichle et al. 2017). The original WorldClim V2 has not been corrected for gauge undercatch. In contrast, both GPCC V2015 and GPCP V2.3 include explicit gauge undercatch corrections. At high latitudes, MERRA-2 is based on reanalysis output, negating the need for gauge undercatch corrections, while at middle and low latitudes MERRA-2 is corrected using GPCP V2.1 (monthly 2.5° resolution; Adler et al. 2003). The corrected WorldClim V2 exhibits substantially higher $P$ (>1000 mm yr$^{-1}$) than the other $P$ datasets over Chile, East Greenland, parts of Antarctica, the Himalayas, and along the Pacific coast of North America (Figs. 7b–e). The $P$ difference over
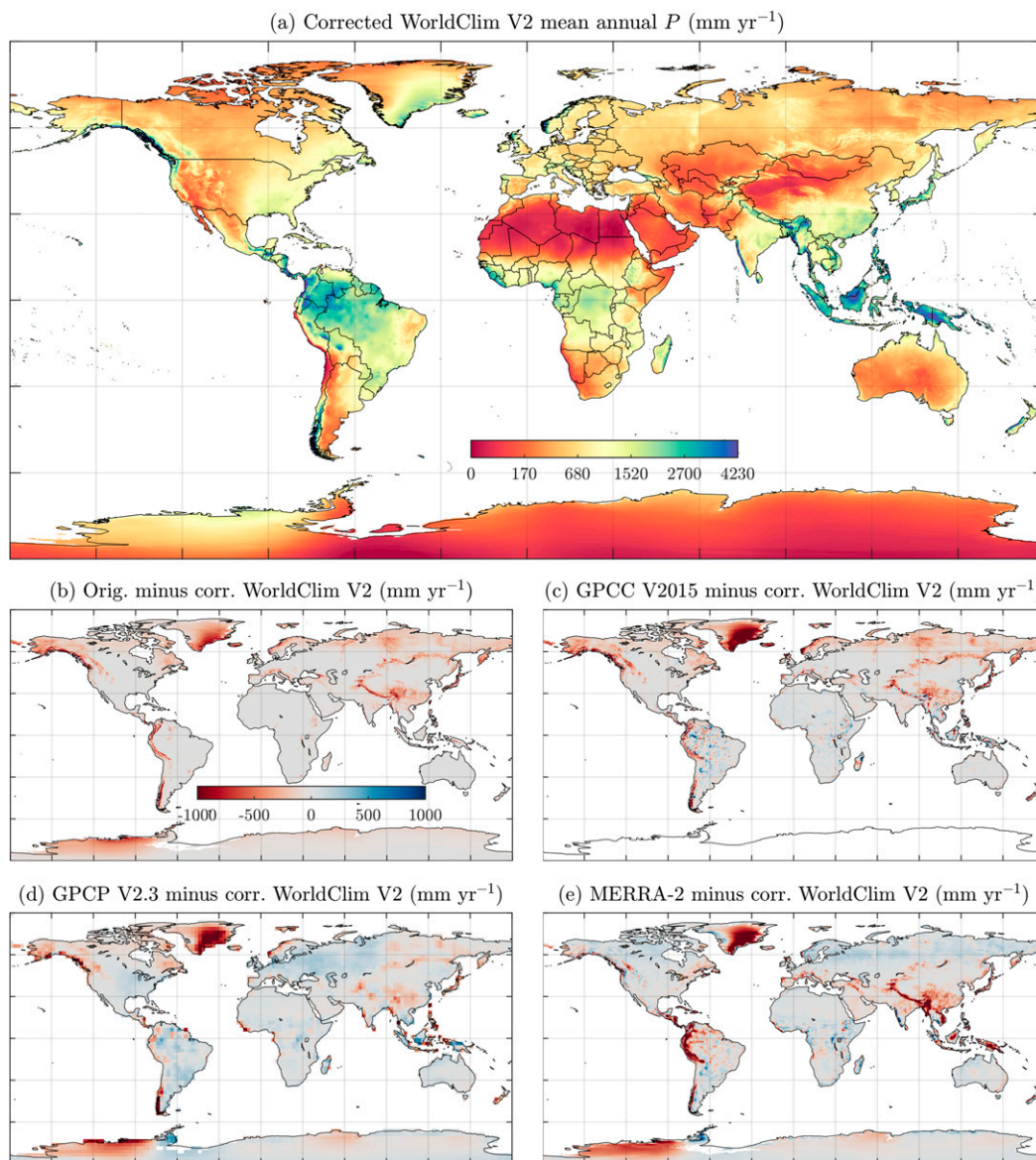
(a) Corrected WorldClim V2 mean annual $P$ (mm yr$^{-1}$)



(b) Orig. minus corr. WorldClim V2 (mm yr$^{-1}$)



(c) GPCC V2015 minus corr. WorldClim V2 (mm yr$^{-1}$)



(d) GPCP V2.3 minus corr. WorldClim V2 (mm yr$^{-1}$)



(e) MERRA-2 minus corr. WorldClim V2 (mm yr$^{-1}$)



FIG. 7. (a) Mean annual $P$ from the bias-corrected WorldClim V2 climatology, and differences relative to (b) the original WorldClim V2, (c) GPCC V2015, (d) GPCP V2.3, and (e) MERRA-2. The data represent 1970–2000 for both WorldClim V2 maps, 1951–2000 for GPCC V2015, 1979–2013 for GPCP V2.3, and 1980–2017 for MERRA-2. To calculate the differences, the corrected WorldClim V2 dataset was upscaled to the native resolution of each $P$ dataset using bilinear resampling.

East Greenland and parts of Antarctica should be interpreted with caution because of a lack of $Q$ (Fig. 3a) and $P$ (Fick and Hijmans 2017, their Fig. 2) observations in these areas. The differences over the other regions were derived from various quantities of local $Q$ and $P$ observations and are therefore more conclusive.

Mean $P$ for the global land surface (excluding Antarctica) based on the corrected WorldClim V2 map is 862 mm yr$^{-1}$ as compared with 788 mm yr$^{-1}$ for the original WorldClim V2—amounting to a 9.4% increase. Adam and Lettenmaier

(2003) and Adam et al. (2006) corrected the University of Delaware $P$ dataset (Willmott and Matsuura 2001) for gauge undercatch and orographic effects and obtained mean $P$ increases for the land surface of 11.7% and 6.2%, respectively. The combined correction (obtained by multiplying these two figures expressed as fractions, noting that the percentage increases are additional to the original value so 11.7% becomes 0.117 + 1.0 = 1.117 and the same for 6.2% to become 1.117 × 1.062 = 1.186) yields a total increase of 18.6%, considerably higher than our 9.4%

estimate, probably because WorldClim V2 has already, to a certain degree, been corrected for orographic effects and is based on markedly more $P$ gauges than the University of Delaware dataset (34 542 vs 1870–16 360 depending on the year) and thus likely is more accurate. Mean $P$ estimates for the global land surface (excluding Antarctica) for GPCC V2015, GPCP V2.3, and MERRA-2 are 793, 853, and 785 mm yr$^{-1}$, respectively (Figs. 7c–e). The GPCP V2.3 estimate is similar to ours because of higher rainfall amounts in GPCP V2.3 over most of the land surface, reflecting the applied gauge undercatch correction (Legates and Willmott 1990; Fig. 7d).

Several studies have used gravity anomaly measurements from the Gravity Recovery and Climate Experiment (GRACE) satellite pair to infer winter $P$ (e.g., Swenson 2010; Behrangi et al. 2017, 2018; Robinson and Clark 2019). Swenson (2010), for example, inferred $P$ using GRACE over high-latitude North America and Eurasia and concluded that the undercatch correction applied by GPCP V2.0 (monthly 2.5° resolution; Adler et al. 2003) is generally too high, in line with our results for northern central Eurasia (Fig. 5d). More recently, Behrangi et al. (2018) inferred winter $P$ for the same region using GRACE and found a better agreement with GPCP V2.3 (monthly 2.5° resolution; Adler et al. 2018) than with GPCC Full Data Reanalysis V7 (monthly 0.5° resolution; Schneider et al. 2017), in reasonable agreement with our findings (Figs. 7c and 7d, respectively). GRACE can only be used to infer $P$ under freezing conditions, when $E$ and $R$ are low, and is thus in some sense complementary to our $Q$-based approach to infer $P$. A limitation of the GRACE-based approach is the lack of fine spatial detail due to GRACE's large footprint size (~400 km; Rodell et al. 2009). Combining GRACE as an additional constraint to higher-resolution Budyko-based $P$ bias correction modeling could reveal new insights and warrants further investigation.

## 4. Conclusions

Our findings can be summarized as follows:

1) The "true" long-term annual $P$ was inferred using an unprecedented database of $Q$ observations from 9372 stations worldwide. Bias correction factors were subsequently calculated for three high-resolution global $P$ climatologies (WorldClim V2, CHELSA V1.2, and CHPclim V1). For all three climatologies, correction factors were > 1 over parts of nearly all major mountain ranges globally. Correction factors were particularly high (>1.5) for Alaska, High Mountain Asia, and Chile. More moderate correction factors (1–1.5) were obtained for mountain ranges with dense $P$ gauge networks such as the U.S. Appalachians, the European Alps, and the Scandinavian Kjolen.
2) RF regression models were trained to regionalize the observation-based $P$ bias correction factors to the entire global land surface for the three $P$ climatologies. We obtained training $r$ values of 0.91–0.92 and slightly lower validation $r$ values of 0.75–0.78, suggesting that the models are effective at predicting the correction factors, although they underestimate high values (>2). The trained RF models were subsequently applied using global 0.05° predictor maps to yield detailed gap-free bias correction maps for each $P$ climatology. Monthly climatological bias correction factors were calculated by disaggregating the long-term bias correction factors using gauge-type-specific catch ratios.
3) Mean annual $P$ for the global land surface (excluding Antarctica) based on the corrected WorldClim V2 map is 862 mm yr$^{-1}$, amounting to a 9.4% increase over the original WorldClim V2. This increase is substantially less than the total increase of 18.6% based on two previous studies using a gauge-based interpolated $P$ dataset as baseline (Adam and Lettenmaier 2003; Adam et al. 2006), probably because WorldClim V2 has already been corrected for topographic effects, whereas their $P$ dataset had not. Other widely used $P$ datasets (GPCC V2015, GPCP V2.3, and MERRA-2) underestimate $P$ by >1000 mm yr$^{-1}$ over Chile, the Himalayas, and along the Pacific coast of North America.

Our findings underscore the need to exercise caution when using gridded $P$ datasets—whether derived from gauge, satellite, or reanalysis data, whether corrected for gauge undercatch or not, and whether high or low spatial resolution—in mountainous and snow-dominated regions. The bias-corrected $P$ climatologies derived in this study can be useful for numerous purposes, including, among others, hydrological model simulations, water resources assessments, exploring spatial $P$ variations, evaluating $P$ datasets, and validating climate model outputs. However, the bias correction factors should be interpreted with caution at subcatchment scales and in regions with few or no $Q$ gauges (Fig. 3). Additionally, it should be kept in mind that the bias correction factors are affected by the $E_p$ estimates as well as by our choice of $w$ parameter (Fig. 2). The annual and monthly bias-corrected climatologies are freely available online (http://www.gloh2o.org/pbcor/) as the Precipitation Bias Correction (PBCOR) dataset.

# APPENDIX A

## Snowfall Fraction

The monthly fraction of total $P$ falling as snow $f\mathrm{snow}_m$ was calculated following Legates and Bogart (2009) as

$$f\mathrm{snow}_m = (1 + 2.147 \times 1.5315^{T_{\mathrm{avg}\_m}})^{-1}, \qquad (\mathrm{A}1)$$

where $T_{\mathrm{avg}\_m}$ (°C) is monthly mean air temperature (from WorldClim V2) and $m$ denotes month $(1, 2, \ldots, 12)$. The long-term annual snowfall fraction $f\mathrm{snow}$ was calculated as

$$f\mathrm{snow} = \frac{\sum\limits_{m=1}^{12} (P_m f\mathrm{snow}_m)}{P_{\mathrm{yr}}}, \qquad (\mathrm{A}2)$$

where $P_m$ (mm month$^{-1}$) is mean monthly precipitation (from WorldClim V2) and $P_{\mathrm{yr}}$ (mm yr$^{-1}$) is mean annual precipitation (from WorldClim V2).

# APPENDIX B

## Gauge Catch Ratios

Monthly climatological maps of gauge catch ratios (CRs; ranging from 0 to 1) were calculated using gauge-type-specific undercatch correction equations for rainfall from Legates (1988) and for snowfall from Goodison et al. (1998), based on monthly mean wind speed, air temperature, and water vapor pressure data from WorldClim V2. The wind speed was limited to $6.5\,\mathrm{m\,s}^{-1}$ to avoid extrapolating the equations beyond the data used for the fitting. For each month, the snowfall and rainfall CRs were weighted according to the snowfall fraction (appendix A). The long-term gauge CR was

subsequently calculated for each of the three $P$ climatologies following

$$\mathrm{CR}_{\mathrm{yr}} = P_{\mathrm{base\_yr}} \Big/ \sum_{m=1}^{12} \frac{P_{\mathrm{base}\_m}}{\mathrm{CR}_m}, \qquad (\mathrm{B}1)$$

where $\mathrm{CR}_m$ is the monthly climatological catch ratio, $P_{\mathrm{base}\_m}$ (mm month$^{-1}$) is the baseline (uncorrected) mean monthly precipitation, and $P_{\mathrm{base\_yr}}$ (mm yr$^{-1}$) is the baseline (uncorrected) mean annual precipitation.

## REFERENCES

Adam, J. C., and D. P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, **108**, 4257, https://doi.org/10.1029/2002JD002499.

——, E. A. Clark, D. P. Lettenmaier, and E. F. Wood, 2006: Correction of global precipitation products for orographic effects. *J. Climate*, **19**, 15–38, https://doi.org/10.1175/JCLI3604.1.

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

——, and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, https://doi.org/10.3390/atmos9040138.

Alvarez-Garreton, C., and Coauthors, 2018: The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies—Chile dataset. *Hydrol. Earth Syst. Sci.*, **22**, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018.

Baccini, A., M. A. Friedl, C. E. Woodcock, and R. Warbington, 2004: Forest biomass estimation over regional scales using multisource data. *Geophys. Res. Lett.*, **31**, L10501, https://doi.org/10.1029/2004GL019782.

Baez-Villanueva, O. M., and Coauthors, 2020: RF-MEP: A novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sens. Environ.*, in press.

Beck, H. E., A. de Roo, and A. I. J. M. van Dijk, 2015: Global maps of streamflow characteristics based on observations from several thousand catchments. *J. Hydrometeor.*, **16**, 1478–1501, https://doi.org/10.1175/JHM-D-14-0155.1.

——, A. I. J. M. van Dijk, A. de Roo, E. Dutra, G. Fink, R. Orth, and J. Schellekens, 2017a: Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.*, **21**, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017.

——, ——, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017b: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, https://doi.org/10.5194/hess-21-589-2017.

Behrangi, A., A. S. Gardner, J. T. Reager, and J. B. Fisher, 2017: Using GRACE to constrain precipitation amount over cold mountainous basins. *Geophys. Res. Lett.*, **44**, 219–227, https://doi.org/10.1002/2016GL071832.

——, ——, ——, ——, D. Yang, G. J. Huffman, and R. F. Adler, 2018: Using GRACE to estimate snowfall accumulation and assess gauge undercatch corrections in high latitudes. *J. Climate*, **31**, 8689–8704, https://doi.org/10.1175/JCLI-D-18-0163.1.

Berghuijs, W. R., R. A. Woods, and M. Hrachowitz, 2014: A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nat. Climate Change*, **4**, 583–586, https://doi.org/10.1038/nclimate2246.

Biemans, H., R. W. A. Hutjes, P. Kabat, B. J. Strengers, D. Gerten, and S. Rost, 2009: Effects of precipitation uncertainty on discharge calculations for main river basins. *J. Hydrometeor.*, **10**, 1011–1025, https://doi.org/10.1175/2008JHM1067.1.

Bonekamp, P. N. J., E. Collier, and W. W. Immerzeel, 2018: The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. *J. Hydrometeor.*, **19**, 1565–1581, https://doi.org/10.1175/JHM-D-17-0212.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Briggs, P. R., and J. G. Cogley, 1996: Topographic bias in mesoscale precipitation networks. *J. Climate*, **9**, 205–218, https://doi.org/10.1175/1520-0442(1996)009<0205:TBIMPN>2.0.CO;2.

Budyko, M. I., 1974: *Climate and Life*. Academic Press, 507 pp.

Cao, Q., T. H. Painter, W. R. Currier, J. D. Lundquist, and D. P. Lettenmaier, 2018: Estimation of precipitation over the OLYMPEX domain during winter 2015/16. *J. Hydrometeor.*, **19**, 143–160, https://doi.org/10.1175/JHM-D-17-0076.1.

Choudhury, B. J., 1999: Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *J. Hydrol.*, **216**, 99–110, https://doi.org/10.1016/S0022-1694(98)00293-5.

Criminisi, A., J. Shotton, and E. Konukoglu, 2012: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations Trends Comput. Graphics Vision*, **7**, 81–227, https://doi.org/10.1561/0600000035.

de Lavenne, A., and V. Andréassian, 2018: Impact of climate seasonality on catchment yield: A parameterization for commonly-used water balance formulas. *J. Hydrol.*, **558**, 266–274, https://doi.org/10.1016/j.jhydrol.2018.01.009.

Di Baldassarre, G., and A. Montanari, 2009: Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.*, **13**, 913–921, https://doi.org/10.5194/hess-13-913-2009.

Donohue, R. J., M. L. Roderick, and T. R. McVicar, 2007: On the importance of including vegetation dynamics in Budyko's hydrological model. *Hydrol. Earth Syst. Sci.*, **11**, 983–995, https://doi.org/10.5194/hess-11-983-2007.

——, ——, and ——, 2010a: Can dynamic vegetation information improve the accuracy of Budyko's hydrological model? *J. Hydrol.*, **390**, 23–34, https://doi.org/10.1016/j.jhydrol.2010.06.025.

——, T. R. McVicar, and M. L. Roderick, 2010b: Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *J. Hydrol.*, **386**, 186–197, https://doi.org/10.1016/j.jhydrol.2010.03.020.

Emanuel, R. E., J. J. Buckley, P. V. Caldwell, S. G. McNulty, and G. Sun, 2015: Influence of basin characteristics on the effectiveness and downstream reach of interbasin water transfers: Displacing a problem. *Environ. Res. Lett.*, **10**, 124005, https://doi.org/10.1088/1748-9326/10/12/124005.

Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador, 2010: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, **91**, 621, https://doi.org/10.1890/09-0889.1.

Favier, V., M. Falvey, A. Rabatel, E. Praderio, and D. López, 2009: Interpreting discrepancies between discharge and precipitation in high-altitude area of Chile's Norte Chico region (26°–32°S). *Water Resour. Res.*, **45**, W02424, https://doi.org/10.1029/2008WR006802.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles*, **16**, 1042, https://doi.org/10.1029/1999GB001254.

——, ——, J. O. Roads, and C. J. Willmott, 2004: Uncertainties in precipitation and their impacts on runoff estimates. *J. Climate*, **17**, 294–304, https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2.

Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, https://doi.org/10.1002/joc.5086.

Fisher, J. B., R. J. Whittaker, and Y. Malhi, 2011: ET come home: Potential evapotranspiration in geographical ecology. *Global Ecol. Biogeogr.*, **20**, 1–18, https://doi.org/10.1111/j.1466-8238.2010.00578.x.

Fu, B. P., 1981: On the calculation of the evaporation from land surface. *Chin. J. Atmos. Sci.*, **5**, 23–31.

Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak, 2015: A global satellite-assisted precipitation climatology. *Earth Syst. Sci. Data*, **7**, 275–287, https://doi.org/10.5194/essd-7-275-2015.

Garreaud, R., M. Falvey, and A. Montecinos, 2016: Orographic precipitation in coastal southern Chile: Mean distribution, temporal variability, and linear contribution. *J. Hydrometeor.*, **17**, 1185–1202, https://doi.org/10.1175/JHM-D-15-0170.1.

Gericke, O. J., and J. C. Smithers, 2014: Review of methods used to estimate catchment response time for the purpose of peak discharge estimation. *Hydrol. Sci. J.*, **59**, 1935–1971, https://doi.org/10.1080/02626667.2013.866712.

Ghatak, D., B. Zaitchik, S. Kumar, M. A. Matin, B. Bajracharya, C. Hain, and M. Anderson, 2018: Influence of precipitation forcing uncertainty on hydrological simulations with the NASA South Asia Land Data Assimilation System. *Hydrology*, **5**, 57, https://doi.org/10.3390/hydrology5040057.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation intercomparison. World Meteorological Organization Tech. Rep. WMO/TD-872, 212 pp.

Groisman, P. Ya., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227, https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2.

Hargreaves, G. H., 1994: Defining and using reference evapotranspiration. *J. Irrig. Drain. Eng.*, **120**, 1132–1139, https://doi.org/10.1061/(ASCE)0733-9437(1994)120:6(1132).

Henn, B., M. P. Clark, D. Kavetski, and J. D. Lundquist, 2015: Estimating mountain basin-mean precipitation from streamflow using Bayesian inference. *Water Resour. Res.*, **51**, 8012–8033, https://doi.org/10.1002/2014WR016736.

——, ——, ——, B. McGurk, T. H. Painter, and J. D. Lundquist, 2016: Combining snow, streamflow, and precipitation gauge observations to infer basin-mean precipitation. *Water Resour. Res.*, **52**, 8700–8723, https://doi.org/10.1002/2015WR018564.

——, ——, ——, A. J. Newman, M. Hughes, B. McGurk, and J. D. Lundquist, 2018: Spatiotemporal patterns of precipitation inferred from streamflow observations across the Sierra Nevada mountain range. *J. Hydrol.*, **556**, 993–1012, https://doi.org/10.1016/j.jhydrol.2016.08.009.

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ECMWF ERA Report Series No. 27, 63 pp., https://doi.org/10.21957/tkic6g3wm.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, **25**, 1965–1978, https://doi.org/10.1002/joc.1276.

Hutchinson, M. F., 1998: Interpolation of rainfall data with thin plate smoothing splines—Part I: Two dimensional smoothing of data with short range correlation. *J. Geogr. Info. Decision Anal.*, **2**, 139–151.

Ibarra-Berastegi, G., J. Saénz, A. Ezcurra, A. Elías, J. Diaz Argandoña, and I. Errasti, 2011: Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression. *Hydrol. Earth Syst. Sci.*, **15**, 1895–1907, https://doi.org/10.5194/hess-15-1895-2011.

Immerzeel, W. W., N. Wanders, A. F. Lutz, J. M. Shea, and M. F. P. Bierkens, 2015: Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrol. Earth Syst. Sci.*, **19**, 4673–4687, https://doi.org/10.5194/hess-19-4673-2015.

Karger, D. N., and Coauthors, 2017: Climatologies at high resolution for the earth's land surface areas. *Sci. Data*, **5**, 170122, https://doi.org/10.1038/SDATA.2017.122.

Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrol. Earth Syst. Sci.*, **17**, 2845–2857, https://doi.org/10.5194/hess-17-2845-2013.

Kiang, J. E., and Coauthors, 2018: A comparison of methods for streamflow uncertainty estimation. *Water Resour. Res.*, **54**, 7149–7176, https://doi.org/10.1029/2018WR022708.

Kidd, C., P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K.-L. Hsu, and D. Braithwaite, 2012: Intercomparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeor.*, **13**, 67–83, https://doi.org/10.1175/JHM-D-11-042.1.

——, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface *is* covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Kirschbaum, D. B., and Coauthors, 2017: NASA's remotely sensed precipitation: A reservoir for applications users. *Bull. Amer. Meteor. Soc.*, **98**, 1169–1184, https://doi.org/10.1175/BAMS-D-15-00296.1.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, https://doi.org/10.2151/JMSJ.2015-001.

Koppa, A., and M. Gebremichael, 2017: A framework for validation of remotely sensed precipitation and evapotranspiration based on the Budyko hypothesis. *Water Resour. Res.*, **53**, 8487–8499, https://doi.org/10.1002/2017WR020593.

Kucera, P. A., E. E. Ebert, F. J. Turk, V. Levizzani, D. Kirschbaum, F. J. Tapiador, A. Loew, and M. Borsche, 2013: Precipitation from space: Advancing Earth system science. *Bull. Amer. Meteor. Soc.*, **94**, 365–375, https://doi.org/10.1175/BAMS-D-11-00171.1.

Kühnlein, M., T. Appelhans, B. Thies, and T. Nauss, 2014: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI. *Remote Sens. Environ.*, **141**, 129–143, https://doi.org/10.1016/j.rse.2013.10.026.

Legates, D. R., 1988: A climatology of global precipitation. Ph.D. dissertation, University of Delaware, 91 pp.

——, and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected global precipitation. *Int. J. Climatol.*, **10**, 111–127, https://doi.org/10.1002/joc.3370100202.

——, and T. A. Bogart, 2009: Estimating the proportion of monthly precipitation that falls in solid form. *J. Hydrometeor.*, **10**, 1299–1306, https://doi.org/10.1175/2009JHM1086.1.

Lehner, B., 2012: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network. Federal Institute of Hydrology (BfG) Global Runoff Data Centre Tech. Rep. 41, 12 pp., https://www.bafg.de/GRDC/EN/02_srvcs/24_rprtsrs/report_41.pdf?__blob=publicationFile.

Le Moine, N., F. Hendrickx, J. Gailhard, R. Garçon, and F. Gottardi, 2015: Hydrologically aided interpolation of daily precipitation and temperature fields in a mesoscale Alpine catchment. *J. Hydrometeor.*, **16**, 2595–2618, https://doi.org/10.1175/JHM-D-14-0162.1.

Levizzani, V., S. Laviola, and E. Cattani, 2011: Detection and measurement of snowfall from space. *Remote Sens.*, **3**, 145–166, https://doi.org/10.3390/rs3010145.

Li, L., D. J. Gochis, S. Sobolowski, and M. D. S. Mesquita, 2017: Evaluating the present annual water budget of a Himalayan headwater river basin using a high-resolution atmosphere-hydrology model. *J. Geophys. Res.*, **122**, 4786–4807, https://doi.org/10.1002/2016JD026279.

Liu, Q., T. R. McVicar, Z. Yang, R. J. Donohue, L. Liang, and Y. Yang, 2016: The hydrological effects of varying vegetation characteristics in a temperate water-limited basin: Development of the dynamic Budyko-Choudhury-Porporato (dBCP) model. *J. Hydrol.*, **543**, 595–611, https://doi.org/10.1016/j.jhydrol.2016.10.035.

Liu, Z., Y. Liu, S. Wang, X. Yang, L. Wang, M. H. A. Baig, W. Chi, and Z. Wang, 2018: Evaluation of spatial and temporal performances of ERA-Interim precipitation and temperature in mainland China. *J. Climate*, **31**, 4347–4365, https://doi.org/10.1175/JCLI-D-17-0212.1.

Lundquist, J., B. Huggett, H. Roop, and N. Low, 2009: Use of spatially distributed stream stage recorders to augment rain gages by identifying locations of thunderstorm precipitation and distinguishing rain from snow. *Water Resour. Res.*, **45**, W00D25, https://doi.org/10.1029/2008WR006995.

Lutz, A. F., W. W. Immerzeel, A. B. Shrestha, and M. F. P. Bierkens, 2015: Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Climate Change*, **5**, 587–592, https://doi.org/10.1038/NCLIMATE2237.

Ménégoz, M., H. Gallée, and H. W. Jacobi, 2013: Precipitation and snow cover in the Himalaya: From reanalysis to regional climate simulations. *Hydrol. Earth Syst. Sci.*, **17**, 3921–3936, https://doi.org/10.5194/hess-17-3921-2013.

Milly, P. C. D., 1994: Climate, soil water storage, and the average annual water balance. *Water Resour. Res.*, **30**, 2143–2156, https://doi.org/10.1029/94WR00586.

Monteith, J. L., 1965: Evaporation and the environment. *Symp. Soc. Exp. Biol.*, **19**, 245–269.

Oki, T., T. Nishimura, and P. Dirmeyer, 1999: Assesment of annual runoff from land surface model using Total Runoff Integration Pathways (TRIP). *J. Meteor. Soc. Japan*, **77**, 235–255, https://doi.org/10.2151/jmsj1965.77.1B_235.

Ol'dekop, E. M., 1911: Ob isparenii s poverkhnosti rechnykh basseinov (On evaporation from the surface of river basins). *Trans. Meteor. Obs. Univ. Tartu*, **4**, 200.

Padrón, R. S., L. Gudmundsson, P. Greve, and S. I. Seneviratne, 2017: Large-scale controls of the surface water balance over land: Insights from a systematic review and meta-analysis. *Water Resour. Res.*, **53**, 9659–9678, https://doi.org/10.1002/2017WR021215.

Pike, J. G., 1964: The estimation of annual run-off from meteorological data in a tropical climate. *J. Hydrol.*, **2**, 116–123, https://doi.org/10.1016/0022-1694(64)90022-8.

Pilgrim, D. H., T. G. Chapman, and D. G. Doran, 1988: Problems of rainfall-runoff modelling in arid and semiarid regions. *Hydrol. Sci. J.*, **33**, 379–400, https://doi.org/10.1080/02626668809491261.

Porporato, A., E. Daly, and I. Rodriguez-Iturbe, 2004: Soil water balance and ecosystem response to climate change. *Amer. Nat.*, **164**, 625–632, https://doi.org/10.1086/424970.

Potter, N. J., L. Zhang, P. C. D. Milly, T. A. McMahon, and A. J. Jakeman, 2005: Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. *Water Resour. Res.*, **41**, W06007, https://doi.org/10.1029/2004WR003697.

Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.*, **37**, 305–327, https://doi.org/10.1002/joc.4706.

Priestley, C. H. B., and R. J. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.

Rasmussen, R. M., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, https://doi.org/10.1175/JCLI-D-16-0570.1.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Remote Sens. Environ.*, **566**, 195–204, https://doi.org/10.1038/S41586-019-0912-1.

Robinson, E. L., and D. B. Clark, 2019: Using GRACE to derive corrections to precipitation data sets and improve modelled snow mass at high latitudes. *Hydrol. Earth Syst. Sci. Discuss.*, https://doi.org/10.5194/HESS-2019-109.

Rodell, M., I. Velicogna, and J. S. Famiglietti, 2009: Satellite-based estimates of groundwater depletion in India. *Nature*, **460**, 999–1002, https://doi.org/10.1038/nature08238.

Roe, G. H., 2005: Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, **33**, 645–671, https://doi.org/10.1146/annurev.earth.33.092203.122541.

Salo, Y. A., 2006: Method for correction of annual precipitation records using the water balance approach. *Hydrol. Res.*, **37**, 433–440, https://doi.org/10.2166/nh.2006.024.

Scaff, L., D. Yang, Y. Li, and E. Mekis, 2015: Inconsistency in precipitation measurements across the Alaska–Yukon border. *Cryosphere*, **9**, 2417–2428, https://doi.org/10.5194/tc-9-2417-2015.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, https://doi.org/10.1007/s00704-013-0860-x.

——, P. Finger, A. Meyer-Christoffer, E. Rustemeier, M. Ziese, and A. Becker, 2017: Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, **8**, 52, https://doi.org/10.3390/atmos8030052.

Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercomparisons. *Atmos. Res.*, **92**, 376–380, https://doi.org/10.1016/j.atmosres.2009.01.016.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

Skofronick-Jackson, G., and Coauthors, 2015: Global Precipitation Measurement Cold Season Precipitation Experiment (GCPEX): For measurement's sake, let it snow. *Bull. Amer. Meteor. Soc.*, **96**, 1719–1741, https://doi.org/10.1175/BAMS-D-13-00262.1.

Smith, R. B., Q. Jiang, M. G. Fearon, P. Tabary, M. Dorninger, J. D. Doyle, and R. Benoit, 2003: Orographic precipitation and air mass transformation: An Alpine example. *Quart. J. Roy. Meteor. Soc.*, **129**, 433–454, https://doi.org/10.1256/qj.01.212.

Sperna Weiland, F., P. Lopez, A. I. J. M. van Dijk, and J. Schellekens, 2015: Global high-resolution reference potential evaporation. *Proc. 21st Int. Congress on Modelling and Simulation*, Gold Coast, Australia, Modelling and Simulation Society of Australia and New Zealand, 2548–2554.

Strangeways, I., 2007: *Precipitation: Theory, Measurement and Distribution*. Cambridge University Press, 290 pp.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation datasets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

Svetnik, V., A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, 2003: Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.*, **43**, 1947–1958, https://doi.org/10.1021/ci034160g.

Swenson, S., 2010: Assessing high-latitude winter precipitation from global precipitation analyses using GRACE. *J. Hydrometeor.*, **11**, 405–420, https://doi.org/10.1175/2009JHM1194.1.

Tapiador, F. J., and Coauthors, 2012: Global precipitation measurement: Methods, datasets and applications. *Atmos. Res.*, **104–105**, 70–97, https://doi.org/10.1016/j.atmosres.2011.10.021.

Trabucco, A., R. J. Zomer, D. A. Bossio, O. van Straaten, and L. V. Verchot, 2008: Climate change mitigation through afforestation/reforestation: A global analysis of hydrologic impacts with four case studies. *Agric. Ecosyst. Environ.*, **126**, 81–97, https://doi.org/10.1016/j.agee.2008.01.015.

Valery, A., V. Andréassian, and C. Perrin, 2009: Inverting the hydrological cycle: When streamflow measurements help assess altitudinal precipitation gradients in mountain areas. *New Approaches to Hydrological Prediction in Data-Sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention)*, IAHS Publ. 333, 281–286.

Viviroli, D., and R. Weingartner, 2004: The hydrological significance of mountains: From regional to global scale. *Hydrol. Earth Syst. Sci.*, **8**, 1017–1030, https://doi.org/10.5194/hess-8-1017-2004.

——, H. H. Dürr, B. Messerli, M. Meybeck, and R. Weingartner, 2007: Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resour. Res.*, **43**, W07447, https://doi.org/10.1029/2006WR005653.

Vogt, J., and Coauthors, 2007: A pan-European river and catchment database. Joint Research Centre Reference Rep. EUR 22920 EN, 119 pp., https://ec.europa.eu/jrc/en/publication/reference-reports/pan-european-river-and-catchment-database.

Vörösmarty, C. J., and D. Sahagian, 2000: Anthropogenic disturbance of the terrestrial water cycle. *BioScience*, **50**, 753–765, https://doi.org/10.1641/0006-3568(2000)050[0753:ADOTTW]2.0.CO;2.

Wang, C., S. Wang, B. Fu, and L. Zhang, 2016: Advances in hydrological modelling with the Budyko framework: A review. *Progress in Physical Geography: Earth and Environment*, **40**, 409–430, https://doi.org/10.1177/0309133315620997.

Wang, D., and Y. Tang, 2014: A one-parameter Budyko model for water balance captures emergent behavior in Darwinian hydrologic models. *Geophys. Res. Lett.*, **41**, 4569–4577, https://doi.org/10.1002/2014GL060509.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, https://doi.org/10.1002/2014WR015638.

Weingartner, R., D. Viviroli, and B. Schädler, 2007: Water resources in mountain regions: A methodological approach to assess the water balance in a highland-lowland-system. *Hydrol. Processes*, **21**, 578–585, https://doi.org/10.1002/hyp.6268.

Weiß, M., and L. Menzel, 2008: A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. *Adv. Geosci.*, **18**, 15–23, https://doi.org/10.5194/adgeo-18-15-2008.

Willmott, C. J., and K. Matsuura, 2001: Terrestrial air temperature and precipitation: monthly and annual time series (1950–1999). Accessed 1 April 2019, http://climate.geog.udel.edu~climate/html_pages/download.html.

Wilson, A. M., and W. Jetz, 2016: Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biol.*, **14**, e1002415, https://doi.org/10.1371/journal.pbio.1002415.

Wrzesien, M. L., M. T. Durand, and T. M. Pavelsky, 2019: A reassessment of North American river basin cool-season precipitation: Developments from a new mountain climatology dataset. *Water Resour. Res.*, **55**, 3502–3519, https://doi.org/10.1029/2018WR024106.

Xu, X., W. Liu, B. R. Scanlon, L. Zhang, and M. Pan, 2013: Local and global factors controlling water-energy balances within the Budyko framework. *Geophys. Res. Lett.*, **40**, 6123–6129, https://doi.org/10.1002/2013GL058324.

Yamazaki, D., and Coauthors, 2017: A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.*, **44**, 5844–5853, https://doi.org/10.1002/2017GL072874.

Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.*, **32**, L19501, https://doi.org/10.1029/2005GL024057.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1.

Zhang, D., Z. Cong, G. Ni, D. Yang, and S. Hu, 2015: Effects of snow ratio on annual runoff within the Budyko framework. *Hydrol. Earth Syst. Sci.*, **19**, 1977–1992, https://doi.org/10.5194/hess-19-1977-2015.

Zhang, L., W. R. Dawes, and G. R. Walker, 2001: Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resour. Res.*, **37**, 701–708, https://doi.org/10.1029/2000WR900325.

Zhang, Y., N. Viney, A. Frost, A. Oke, M. Brooks, Y. Chen, and N. Campbell, 2013: Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments. CSIRO Land and Water Tech. Rep. EP113194, 115 pp.