Global Reach-Level 3-Hourly River Flood Reanalysis (1980–2019)

Yuan Yang, Ming Pan, Peirong Lin, Hylke E. Beck, Zhenzhong Zeng, Dai Yamazaki, Cédric H. David, Hui Lu, Kun Yang, Yang Hong, and Eric F. Wood

ABSTRACT: Better understanding and quantification of river floods for very local and “flashy” events calls for modeling capability at fine spatial and temporal scales. However, long-term discharge records with a global coverage suitable for extreme events analysis are still lacking. Here, grounded on recent breakthroughs in global runoff hydrology, river modeling, high-resolution hydrography, and climate reanalysis, we developed a 3-hourly river discharge record globally for 2.94 million river reaches during the 40-yr period of 1980–2019. The underlying modeling chain consists of the VIC land surface model (0.05°, 3-hourly) that is well calibrated and bias corrected and the RAPID routing model (2.94 million river and catchment vectors), with precipitation input from MSWEP and other meteorological fields downscaled from ERA5. Flood events (above 2-yr return) and their characteristics (number, spatial distribution, and seasonality) were extracted and studied. Validations against 3-hourly flow records from 6,000+ gauges in CONUS and daily records from 14,000+ gauges globally show good modeling performance across all flow ranges, good skills in reconstructing flood events (high extremes), and the benefit of (and need for) subdaily modeling. This data record, referred as Global Reach-Level Flood Reanalysis (GRFR), is publicly available at <https://www.reachhydro.org/home/records/grfr>.

KEYWORDS: Rivers; Flood events; Hydrology; Hydrologic models; Risk assessment

<https://doi.org/10.1175/BAMS-D-20-0057.1>

Corresponding author: Ming Pan, mpan@princeton.edu

In final form 3 June 2021

©2021 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

AFFILIATIONS: **Y. Yang**—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, and State Key Laboratory of Hydrosience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China; **Pan**—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, and Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California; **Lin**—Institute of Remote Sensing and GIS, School of Earth and Space Sciences, Peking University, Beijing, China; **Beck**—Joint Research Centre, European Commission, Ispra, Italy; **Zeng**—School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China; **Yamazaki**—Institute of Industrial Science, The University of Tokyo, Tokyo, Japan; **David**—Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California; **Lu and K. Yang**—Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies, Tsinghua University, Beijing, China; **Hong**—School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, Oklahoma; **Wood**—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

Floods are among the most frequent and costly natural disasters (Wu et al. 2014). According to a United Nations report (UNISDR 2015), floods accounted for 43% of occurrences of natural disasters during 1995–2015. The EM-DAT international disaster database (<http://www.emdat.be>) indicates that fluvial floods (Willner et al. 2018; Beevers et al. 2020) contributed to two-thirds of flood events in the past two decades. Fluvial flood is the extreme upper tail of river hydrology that can happen at different scales. Some river floods are “flashy” and can occur within a very short time (e.g., hours), leaving less time for action (Lin et al. 2018a), and some may take longer to develop. Some may affect a small local area only and some can travel down to flood a large flat area. Fluvial flood also has wide geographic distributions globally. These variabilities make it hard to capture and study fluvial floods. The lack of global long-term fine-scale (in both time and space) river discharge (Q) records has been the major hurdle for large-scale flood research (Alfieri et al. 2020; He et al. 2020).

There are large gaps in river gauge data around the world (GRDC 2020) as most gauges are concentrated in developed areas like North America and Europe. The other observational alternative is remote sensing, though the first satellite mission dedicated to discharge estimation will not launch until 2022 (Gleason and Smith 2014; Biancamaria et al. 2016; Durand et al. 2016; Yang et al. 2019). As a result, various hydrologic models for runoff calculation (Duan et al. 1992; Wood et al. 1992; Moradkhani and Sorooshian 2008; Beven 2011) and river models for channel routing (e.g., Lohmann et al. 1998; Oki and Sud 1998; Beighley et al. 2009; Bates et al. 2010; David et al. 2011; Yamazaki et al. 2011; Getirana et al. 2012; Li et al. 2013) are employed to estimate river discharge given surface meteorological inputs. Continuous efforts have been made by the modeling community to quantify river discharge globally with increasingly finer spatial and temporal resolutions (Wood et al. 2011). Such efforts started as soon as global land surface modeling (LSM, for runoff calculation) became possible, even though the spatial resolution used to be low, e.g., 2° (Nijssen et al. 2001). Later efforts like the National Aeronautics and Space Administration (NASA) Global Land Data Assimilation System (GLDAS; Rodell et al. 2004) and Land Information System (Kumar et al. 2006; Peters-Lidard et al. 2007) brought the global LSM resolution to about 0.25° , and then even finer at 0.0833° (Sutanudjaja et al. 2018), 0.0625° (Li et al. 2013), and 0.05° (Beck et al. 2020a). Following these developments, operational river monitoring services emerged. For example, the European Commission’s Copernicus Emergency Management Service (CEMS) Global Flood Awareness System (GloFAS; <http://www.globalfloods.eu/>) provides operational river discharge monitoring and disaster warning services (Alfieri et al. 2013; Hirpa et al. 2018). GloFAS also created long-term river discharge reanalysis products (Alfieri et al. 2020; Harrigan et al. 2020).

NASA's Global Flood Monitoring System (GFMS) (Wu et al. 2012; Wu et al. 2014) also provides real-time monitoring and reanalysis of river discharge globally. GloFAS reanalysis (Alfieri et al. 2020) is based on the European Centre for Medium-Range Weather Forecasts's (ECMWF) latest atmospheric reanalysis (ERA5; Hersbach et al. 2018) and operates at 0.1° and daily resolution. GFMS is largely based on NASA's satellite rainfall products and operates at 0.125° and 3-hourly resolution.

The above discharge modeling efforts are all based on gridded river networks where the modeling units are interconnected rectangular grid pixels. Such an approach has been very reliable while more recent large-scale studies (e.g., Beighley et al. 2011; David et al. 2011; Getirana et al. 2012; Gochis et al. 2015; Mizukami et al. 2016; Lin et al. 2018b; O'Loughlin et al. 2020) started to favor vector river networks for efficiency and geometric fidelity reasons. Large gridded pixels can lead to inaccurate flow paths and drainage area, while it is expensive to substantially increase resolution. Progress has been made to upscale fine networks while preserving river features at coarser pixels to save computations (Olivera et al. 2002; Yamazaki et al. 2009; Wu et al. 2011). With vector river networks, the modeling units are irregularly shaped catchments and river reaches derived from fine DEM. A river "reach" is defined here as the river segment between two neighboring confluences (or channel head and confluence) (Linke et al. 2019), and its size depends on the channelization threshold (25 km² in this study) and is normally in kilometers to tens of kilometers. This better preserves the flow geometry without adding too many modeling units (Mizukami et al. 2016). The vector-to-river mapping is easier than pixel-to-river mapping, too, as multiple rivers may exist in the same pixel. Recent global studies with vector networks have illustrated their efficiency at high resolutions, e.g., the global HydroROUT study in Lehner and Grill (2013). Yamazaki et al. (2013) demonstrated that the Catchment-based Macro-scale Floodplain (CaMa-Flood) model with a vector-based river network (400-km² channelization threshold) was about 60% faster than with a 15-arc-min gridded network. And Qiao et al. (2019) showed the vector Routing Application for Parallel Computation of Discharge (RAPID) model (David et al. 2011, 2016) has comparable performance to GloFAS, but calculates discharge at higher resolutions. Also, global, high-quality, high-resolution river hydrography data have been derived (Allen and Pavelsky 2018; Yamazaki et al. 2019; Lin et al. 2020), making it convenient to parameterize routing models with vector networks globally (Lin et al. 2019).

Following these developments in the community, Lin et al. (2019) produced the first reach-level naturalized daily river discharge reanalysis (discharge without anthropogenic influences), the Global Reach-Level A Priori Discharge Estimates for SWOT (GRADES), over 2.94 million river reaches globally for 1979–2014. GRADES relies on the Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 2 (Beck et al. 2019a) and Climate Forecasting System Reanalysis (CFSR; Dee et al. 2014; Saha et al. 2014) for inputs and the Multi-Error-Removed-Improved-Terrain (MERIT) Hydro (Yamazaki et al. 2019) for hydrography. The hydrologic modeling in GRADES was performed at 0.25° and daily resolution. Similar to the GloFAS-ERA5 discharge dataset (Alfieri et al. 2020), GRADES puts a strong focus on parameter calibration (Yang et al. 2019) and bias correction of the hydrologic model at the global scale, which are performed against a set of machine-learning (ML)-derived, global runoff characteristics (i.e., runoff percentiles; Beck et al. 2015).

Both GRADES (Lin et al. 2019) and GloFAS-ERA5 (Alfieri et al. 2020) discharge reanalyses were well validated globally but not specifically tested for the high extremes. River floods are the extreme high tail of the discharge climatology and can be very rapid and local; therefore, flood analysis requires finer model resolutions in both space and time. This study will inherit the basic modeling framework established in GRADES, make considerable efforts to upgrade the spatial and temporal resolutions from 0.25° daily to 0.05° 3-hourly, extend the temporal coverage (1980–2019), and carry out flood-specific (high extremes) analysis at subdaily time

scales. Small basins will also be included in validations to better understand the skills for local and rapid events.

Global modeling framework

The global 3-hourly (0.05° runoff + vector river reaches) modeling framework consists of a series of state-of-the-art datasets and modeling techniques (Fig. 1). It is based upon the GRADES legacy but includes substantial upgrades on the spatial and temporal resolutions, coverage, and inputs to better simulate the extremes and serve the new purpose of flood analysis (Table 1). The hydrography has also been updated with minor fixes (Yamazaki et al. 2019).

Forcing inputs. The main forcing, precipitation, comes from MSWEP version 2.2 (Beck et al. 2019a). MSWEP is a global dataset (1979–present) with a 3-hourly and 0.1° resolution. It ingests a wide range of data sources (76,747 in situ gauges, see Fig. 2a, four satellite products, and two reanalysis products), makes distributional bias corrections, as well as correction of systematic terrestrial biases using river discharge observations from 13,762 stations. MSWEP compares better with Stage IV gauge–radar data (Lin and Mitchell 2005) than other global precipitation datasets (Beck et al. 2017, 2019b).

Fig. 1. Global subdaily modeling framework.

Other meteorological fields (surface air temperature, pressure, incoming shortwave and longwave radiation, humidity, and wind speed) come from ERA5 (Hersbach et al. 2018), the latest climate reanalysis dataset produced by ECMWF (data provided at hourly 0.25° resolution). ERA5 shows large improvements over its predecessor ERA-Interim (Dee et al. 2011), and has been used in global hydrological applications (Alfieri et al. 2020).

We downscaled forcing fields to 0.05° using fine-scale elevation as the correction factor (Pan et al. 2016) for temperature (fixed lapse rate of $-6.5^{\circ}\text{C km}^{-1}$), surface pressure (hydrostatic), longwave radiation (lapse adjustment for radiative temperature), humidity (lapse adjustment for saturated vapor pressure), wind speed (bilinear interpolation), and precipitation (bilinear interpolation). No orographic adjustment on precipitation was made and the impact on mountainous areas was partially addressed with MSWEP's built-in correction against long-term runoff observations (Beck et al. 2019a).

Runoff modeling. The Variable Infiltration Capacity (VIC; Liang et al. 1994; Liang et al. 1996) LSM is used for runoff modeling. VIC has been widely implemented in large-scale studies (Mitchell et al. 2004; Wu et al. 2014; Mizukami et al. 2017; Lin et al. 2019; He et al. 2020).

Table 1. Comparison of the modeling frameworks used in GRADES (Lin et al. 2019) and GRFR.

	GRADES (Lin et al. 2019)	GRFR (this study)
Focus	Discharge climatology	Extreme events (floods)
Spatial resolution	0.25°	0.05°
Temporal resolution	Daily	3-hourly
Coverage	1980–2014	1980–2019
Forcing	MSWEP V2 + CFSR	MSWEP V2.2 + ERA5

Fig. 2. (a) Density of in situ precipitation gauges in MSWEP (number per 10^4 km^2) and (b) training river gauges in GSCD.

To reduce model biases, we performed grid-level parameter calibration and bias correction (postprocessing) against ML-derived, global runoff characteristic maps from the Global Streamflow Characteristics Dataset (GSCD; Beck et al. 2015). GSCD includes 17 runoff statistics (mean flow, flow percentiles, baseflow indices, and timing of half flow volume of the year). Using artificial neural networks, GSCD first established a relationship between observed runoff signatures and 20 climatic and physiographic predictors over 4,000+ catchments with little or no influences of water management activities (Fig. 2b), and then derived the runoff signature globally.

Instead of calibrating VIC against limited gauge observations, we calibrated three VIC parameters (infiltration curve shape, second soil layer thickness, and nonlinear baseflow threshold) at 0.25° grid pixel level, using the baseflow index, mean runoff, and runoff percentiles (Q10 and Q90) from the GSCD (Yang et al. 2019), and then resampled the parameters to 0.05° by bilinear interpolation.

To reduce the remaining model biases after calibration, we applied the sparse cumulative distribution function (CDF) matching bias correction approach (Lin et al. 2019), which corrects the VIC runoff biases against nine runoff percentiles (Q1, Q5, Q10, Q20, Q50, Q80, Q90, Q95, and Q99) from the GSCD. This technique is similar to traditional CDF matching (Reichle and Koster 2004; Liu et al. 2011), except that no full CDF is available but some sparse percentile values. Correction ratios were calculated for the nine percentiles and gaps in between were filled with ratios from log-linear interpolation. The correction ratios were calculated at 0.25° level and applied to 0.05° pixels.

The model was spun up for the 10-yr period of 1969–78 forced with downscaled Princeton Global Forcing (PGF; Sheffield et al. 2006) which has an original resolution of 0.25°. The MSWEP–ERA5-driven simulation covers 1979–2019.

River routing. We implemented RAPID (David et al. 2011), a river routing model that uses a matrix-based version of the Muskingum method to calculate the flow and volume of water for each reach on a river network. RAPID is vector-based and well parallelized for large-scale applications. To best represent river reaches at their true geometry and location in RAPID, a vector hydrography dataset was recently derived from the Multi-Error-Removed Improved-Terrain (MERIT) DEM and flow directions (Yamazaki et al. 2017, 2019), which is called MERIT Basins (Lin et al. 2019, 2021). MERIT DEM is a currently high-quality hydrologically conditioned global DEM (including above 60°N) at 3 arc-s (~90 m). A 25-km² upstream drainage area is applied to delineate the river reaches and catchments and the segment of a river line between two confluence points is treated as a river reach. The vector hydrography contains ~2.94 million river reaches and corresponding catchments with derived properties (e.g., slope, length, connectivity). The total river length in MERIT Basins is 2.6×10^7 km and the median (mean) length is 6.8 (9.2) km. Compared with earlier hydrography datasets used in global routing studies (e.g., Yamazaki et al. 2011; Wu et al. 2012; Li et al. 2013; Sutanudjaja et al. 2018), MERIT Basins used here shows significant improvements in channel representation of the location, shape, sinuosity, slope and river length (Lin et al. 2019, 2021). Here, RAPID is run using a 30-min time step and forced with 3-hourly lateral inflows from VIC runoff. The parameters of the Muskingum method k and x are determined according to Lin et al. (2019).

Comparison to local precipitation-driven simulations. Model prediction errors are a combined result of model parameterization uncertainty and forcing (precipitation) errors, which are often hard to separate (Yilmaz et al. 2010). To better understand the sensitivity of the model skills to precipitation quality, we ran the model with a better-quality ground-based local product Stage IV (Lin and Mitchell 2005) over the heavily gauged conterminous

United States (CONUS) during 2002–19 and compared it to the default MSWEP-forced results and observed discharge (Fig. 3). The 4-km hourly Stage IV includes 150+ radars and 5,500+ rain gauges. The comparisons among MSWEP-forced, Stage IV–forced, and observed discharges will also help us understand the performance gap between global and local forcing products and where and how to improve the framework. Note that Stage IV has errors, too, which can be substantial over mountainous areas.

Skill assessment

We collected two sets of discharge observations over the globe and CONUS for the skill assessments.

Globally we compiled daily records over 21,000+ river gauges from multiple sources (Beck et al. 2020b). We screened the gauges for 1) gauges located <500 m from closest reach and 2) ≥ 3 years of valid data during the validation period 1980–2017 (1979 for routing model spinup, and no data for 2018–19). The selected gauges (14,000+) cover a range of hydroclimates, stream orders, and basin size. Given the 90-m DEM derived hydrography, we are able to include headwater gauges with very small drainage areas (29% of them <250 km²). In contrast to many previous global studies that validated their results primarily over large basins (Wu et al. 2014; Li et al. 2015; Alfieri et al. 2020), the inclusion of small basins inevitably poses much greater challenges to the modeling system which need to have higher resolution and skill.

Over CONUS, we compiled 3-hourly data over 9,000+ river gauges up to 2019 from the U.S. Geological Survey (USGS). After a similar screening, 6,000+ gauges match the river reaches being simulated and come with sufficient data for skill analysis.

We focused on two aspects of discharge simulations: 1) general skill assessment across all ranges of discharge values and 2) event-based flood analysis (extremely high values only). Since we want to evaluate the model skill for very rapid and local flood events, the event-based analysis will be conducted only over CONUS where subdaily observations are available. Note that the event analysis requires more continuous observations and only USGS gauges with more than 85% valid data are used (1,740 gauges).

General discharge skill assessment. Here we calculated the Kling–Gupta efficiency (KGE) and its three component statistics: correlation coefficient (CC), relative bias (RB), and variability ratio (VR). RB and VR measure the static errors in the first moment (relative bias in mean) and second moment (relative bias in variability) of the value distribution, and CC measures the dynamic errors (temporal coherence). KGE is the Euclidean distance from the optimal point of CC = 1, RB = 0, and VR = 1 (Gupta et al. 2009; Kling et al. 2012) thus an “overall” skill metric. We encourage readers to pay more attention to CC, RB, and VR individually as they provide much more useful information on specific types of errors. Appendix A provides the equations for the metrics. These skill metrics will be calculated at 3-hourly, daily, and monthly scales whenever the observations are available.

Figure 4 shows the skill metrics for 1980–2017 at 14,000+ gauges globally. Good CC values (top row), 0.5–0.9 (daily) and 0.7–1 (monthly), are seen in most parts of the world except for complex terrains like the central CONUS, southern Africa and eastern Europe.

Fig. 3. Skill assessment and comparison to local forcing-driven simulations. Stage IV–forced/MSWEP–forced Q is the simulated discharge using Stage IV/MSWEP as precipitation input.

Fig. 4. Skill metrics for simulated discharge. (a),(b) Correlation coefficient (CC); (c),(d) relative bias (RB), (e),(f) variability ratio (VR); and (g),(h) Kling–Gupta efficiency (KGE) for (left) daily and (right) monthly scales.

The mean flow (second row) is also well simulated, with about 44% (71%) gauges having RB within $\pm 20\%$ ($\pm 50\%$). Large overestimations (e.g., $RB > 100\%$) are mostly seen in arid regions like the central CONUS, southern Africa, and eastern Brazil, where a small absolute error (e.g., $0.1 \text{ m}^3 \text{ s}^{-1}$) leads to a large relative error. Large overestimation can also be seen in other regions like Niger basin, Congo basin, and Australia. The model reproduces the flow variability (third row) over most areas, but tends to underestimate in the central CONUS and Australia. Static errors (RB, VR) do not change much from daily to monthly due to the bias corrections. The overall KGE skill is greater than 0.5 over 27% (daily) and 50% (monthly) of the sites, hence indicating quality simulations. Its pattern looks similar to CC except for a few places of good CC but strong biases (Chile, Africa, and Australia). The KGE values are slightly better than Lin et al. (2019) and the pattern, even with much more small gauge basins and longer period, is similar to previous global studies (Alfieri et al. 2020; Arheimer et al. 2020).

Overall, the model works better in humid regions than arid regions, where runoff overestimation often occurs (e.g., Xia et al. 2012). Monthly skills are better than the daily, mostly thanks to significantly better CC (sites with $CC > 0.5$ increased from 64% to 90%). Further improvement in daily dynamics would need effort in both forcing and modeling, e.g., the Muskingum parameterization tends to have slow flood wave propagations and human regulations (dam and reservoir operations) can significantly alter daily dynamics.

The same general assessment was made over CONUS using 3-hourly data during 2003–19 (Stage IV period minus 1-yr RAPID spinup) (Fig. 5). The model performs well at the

Fig. 5. MSWEP-forced discharge skill metrics [(a) CC, (b) RB, (c) VR, (d) KGE] over CONUS and (e) Stage IV-forced discharge KGE for (left) 3-hourly, (center) daily, and (right) monthly scales.

3-hourly scale. Except for central CONUS, 3-hourly CC [Fig. 5a(1)] is high in most regions (>0.5 over 74% of the gauges). This is consistent with previous CONUS-wide studies such as Oubeidillah et al. (2014), Mizukami et al. (2017), and Yang et al. (2019). Again, the monthly CC [Fig. 5a(3), >0.5 over 91%] is still higher than both daily and 3-hourly CC. For the same sake of bias corrections, static errors (RB and VR) do not change significantly from 3-hourly to longer time scales, except for slightly worse VR in the southwest, Great Plains, and Appalachians at shorter time scales. The KGE pattern is consistent with previous studies as well with slightly better skills at monthly than 3-hourly (>0.5 over 58% and 36%).

During the precipitation sensitivity experiment, the Stage IV-forced discharge was compared to the MSWEP-forced during 2003–19. Beck et al. (2019b) showed that MSWEP is comparable to Stage IV at daily scale. The bottom two rows of Fig. 5 show the KGEs of MSWEP-forced and Stage IV-forced discharge are very similar across all time scales. The three component statistics (not shown) are also very close. Figure 6 shows the histograms of their 3-hourly skill metrics, which further confirm their similar performance. The main difference between the two is the lack of radar information in MSWEP, but that did not obviously degrade the skills. This suggests local radars provide limited added value over heavily gauged areas like CONUS and the room for further improvement lies more with better precipitation estimations over mountains (where both products have poor skills), better model parameterizations (floodplains, reservoirs, etc.) and calibrations.

Event-based flood reconstruction skill assessment. We extracted flood events from 3-hourly discharge time series and evaluated the model performance. Flood events are defined as flows that exceed the 2-yr return annual maximum 3-hourly flow

Fig. 6. Histogram of (a) CC, (b) RB, (c) VR, and (d) KGE for 3-hourly Stage IV–forced Q and MSWEP–forced Q .

(Carpenter et al. 1999; Reed et al. 2007; Wu et al. 2012; Toth 2016;). The 2-yr return annual maximum was calculated for every river reach by fitting the log Pearson type-III (LP3) distribution (Chow 1988), which is recommended by United States federal agencies for flood frequency analysis. Two neighboring (in time) events were considered independent events only when they are ≥ 2 days apart. We also require flood flow to be $\geq 10 \text{ m}^3 \text{ s}^{-1}$ to filter out tiny streams (Wu et al. 2014).

To minimize the influence of systematic errors, the simulated and observed flood thresholds were calculated separately with respect to their own distributions. More than 24,000 events were so extracted from USGS observations, Stage IV–forced discharge, and MSWEP–forced discharge, respectively. To verify the expected advantage the of 3-hourly simulation against daily simulation (recovering higher peaks at shorter time scales), we compared the 3-hourly flood peak values with corresponding daily peak values. This comparison was done for USGS

Fig. 7. Relative increase in 3-hourly flood peak values with respect to corresponding daily peak values calculated from (a) USGS observations, (b) Stage IV–forced Q , and (c) MSWEP–forced Q . The color map shows the scatter density.

observations, Stage IV–forced simulations and MSWEP–forced simulations (Fig. 7). Both the observed and simulated data show that 3-hourly peaks can be much greater than daily peaks (>100% increase) over smaller rivers and averaged out over large rivers. The simulations can well capture the observed increase in flood peaks from daily to 3-hourly scale (very slight underestimation). Results here suggest that subdaily modeling offers significant benefits for small-to-medium river flood assessment.

To study the flood detection capability, three categorical event verification metrics were calculated from events so extracted: 1) probability of detection (POD), 2) false alarm rate (FAR), and 3) critical success index (CSI). See appendix B for their definitions.

Figure 8 shows the three categorical metrics. Good skills (high POD, low FAR, and high CSI) are seen in the east and west (dominated by large-scale tropical or monsoon systems), while lesser skills in the central and worst in the Rocky Mountains where challenges exist in both precipitation (including Stage IV) and runoff calibration.

Stage IV–forced discharge and MSWEP–forced discharge show very similar categorical skills with a median POD 0.46, FAR 0.53, and CSI 0.30. The differences between the two are in the range of ± 0.1 (right column in Fig. 8). These metrics do not seem great or perfect, but they are calculated from the 3-hourly flood peaks over all rivers in CONUS with no filtering for dams, reservoirs, etc. Given that, we consider it reasonably skillful. On the other hand, using a small flood threshold (2-yr return) will lead to much more events and thus lower skill metrics. Further studies are needed to assess the skill for large flood events (Huang et al. 2021). For the hit events, we calculated three detailed event metrics: 1) relative peak flow error, 2) relative rising time error, and 3) relative volume error, to assess the model skills more specifically (see appendix C). Since we adopted different simulated and observed flood thresholds, the duration of the simulated and observed floods might be different for a certain flood. To minimize the influence of duration and threshold, both the relative peak flow error and relative volume error are normalized by the flood threshold.

Figure 9 shows the median of detailed event metrics. Both Stage IV–forced and MSWEP–forced discharge show small peak errors (within $\pm 20\%$ over 65% of the gauges). Peak underestimation ($< -30\%$) occurs primarily in drier areas (Texas and Arizona). Peak overestimation is rare. Flood rising time is generally biased too long, i.e., too slow to reach the peak. Twenty-three percent of the gauges show large overestimation, and floods in these regions are mostly flashy (Saharia et al. 2017). The challenges on very flashy events come

Fig. 8. Categorical flood event skills. (a) POD, (b) FAR, and (c) CSI for (left) Stage IV forced, (center) MSWEP forced, and (right) their difference. Teal color in the difference maps means the MSWEP forced underperforms Stage IV forced.

Fig. 9. Detailed event skill metrics for hit events. The median of (a) relative peak flow error, (b) relative rising time error, and (c) relative volume error for (left) Stage IV forced and (right) MSWEP forced.

from temporal resolution (3-hourly), rainfall data, and the fixed-celerity Muskingum routing. Flood volume errors are generally small (similar to peak errors). There is little difference between MSWEP forced and Stage IV forced.

Global reach-level 3-hourly river flood reanalysis

The final product of this study is a global reach-level 3-hourly database of flood records for 1980–2019 and we now refer to it as Global Reach-level Flood Reanalysis (GRFR). Here we show some basic characteristics of this product.

Flood events. Based on the MSWEP-forced discharge, 3-hourly flood thresholds were calculated using the log LP3 distribution globally (Fig. 10a) then flood events are extracted (Fig. 10b). Flood thresholds increase with the stream order and exceed $20,000 \text{ m}^3 \text{ s}^{-1}$ over major rivers like Mackenzie, Mississippi, Amazon, Nile, Yangtze, and Lena (0.3% of global river reaches).

Figure 10b shows the number of global flood events during 1980–2019. Compared with the Global Archive of Large Flood Events for 1985–2019 maintained by the Dartmouth Flood Observatory (DFO, <http://floodobservatory.colorado.edu/Archives/ArchiveNotes.html>) (Fig. 10c), which is compiled from a wide variety of news and governmental sources and is one of the most comprehensive flood databases (Karnes and Brakenridge 1996), our flood database is able to reproduce vast majority of regions with relative high flood frequency, such as southeastern China, Malay Archipelago, and Central America. Our database also identifies several regions with high number of flood events but not highlighted by DFO, such as the sparsely populated thus less reported Amazon and Congo basin. Global flood activeness generally follows the hydroclimate, e.g., more than 40 floods in tropical rain forests during the 40 years. Areas with heavy summer storms and spring snowmelt can also be seen. Overall, 53% of the rivers have 25–35 floods in 40 years. Rivers with >50 events are rare (3.3%).

Fig. 10. (a) Flood threshold (2-yr return) for river reaches of stream order ≥ 4 and threshold $\geq 100 \text{ m}^3 \text{ s}^{-1}$, (b) number of flood events during 1980–2019 for threshold $\geq 100 \text{ m}^3 \text{ s}^{-1}$, and (c) geographic centers of floods in the Dartmouth Flood Observatory (DFO) archive, 1985–2019.

Flood seasonality. Flood seasonality provides important information for management and adaptation practices and offers insights into the dominant flood-generating mechanisms (Villarini 2016; Blöschl et al. 2017; Ye et al. 2017; Hall and Blöschl 2018; Berghuijs et al. 2019; Collins 2019; Dickinson et al. 2019). Most existing studies are based on the annual maximum flows in gauge records, which restricts our understanding to a single peak per year at the regional scales (e.g., Villarini 2016; Hall and Blöschl 2018). We calculated two seasonality metrics globally: 1) mean date of flood occurrence $\bar{D}$ and 2) seasonal concentration index R (higher values for concentrated flood seasons and lower values for spread-out seasonal distributions). See appendix D for their definitions.

The mean seasonality is quantified by $\bar{D}$ (Fig. 11a) and it shows some clear features. Low-latitude tropical regions show a gradient in $\bar{D}$ that tracks the wet summer season from June–August in Northern Hemisphere to December–February in the Southern Hemisphere.

Fig. 11. Flood seasonality. (a) Mean date of flood occurrence $\bar{D}$ (units: month of year; 1–12 for January–December, respectively) and (b) seasonal concentration index R .

Floods in high latitudes and mountains (e.g., Rockies) are tied to snowmelt in the spring and summer (Parajka et al. 2009; Berghuijs et al. 2016; Villarini 2016; Berghuijs et al. 2019). Monsoon also stands out in places like the Indian Peninsula and Indo-China (summer) and the U.S. Northwest coast (winter). Complexity arises with factors like land–sea interactions. For example, the flood season in Europe transitions from December–January in coastal areas to April–May in the interior because of increasing continentality (away from the Atlantic) (Blöschl et al. 2017; Hall and Blöschl 2018). Floods in China are tied to the southeast monsoon generated by the western Pacific subtropical high (WPSH). As a result, the seasonality of floods in China shifts from south to north as the WPSH moves northward.

Seasonal concentration index R is shown in Fig. 11b. Overall, the seasonality is strong globally with $R > 0.9$ over most areas. Relatively small R values are found mainly in the southeastern CONUS, the Mediterranean, Paraná, and equatorial regions, with $R < 0.4$.

Summary and conclusions

Building upon a series of recent breakthroughs in long-term climate model reanalysis, optimal precipitation estimation, quantifying hydrology over ungauged basins, global high-resolution hydrography, hydrologic model calibration, hydrological model bias correction, and river routing based on vector river network, we developed a 3-hourly river discharge record globally for 2.94 million river reaches during the 40-yr period of 1980–2019. The modeling chain consists of the VIC land surface model (0.05°, 3-hourly) and the RAPID routing model (2.94 million river and catchment vectors) with precipitation input from MSWEP and other meteorological fields downscaled from ERA5. VIC model calibration and bias correction (postprocessing) were conducted against the GSCD database. Flood events

(above 2-yr return) and their detailed characteristics were extracted and studied. This data record, referred as Global Reach-Level Flood Reanalysis (GRFR), is publicly available at <https://www.reachhydro.org/home/records/grfr>.

Validation of GRFR against 3-hourly flow records from 6,000+ gauges in CONUS and daily records from 14,000+ gauges globally shows good general skills of discharge estimates across all flow values, as well as good skills in reconstructing flood events (high extremes). The effort to enable subdaily high-resolution modeling is expected to offer added values in capturing subdaily dynamics that may come with local and flashy events. The comparison between 3-hourly peaks and daily peaks highlights the significant benefit of (and need for) subdaily modeling for flood extremes. Precipitation sensitivity analysis suggests that, over well-gauged areas, the global product MSWEP performs very close to local product (Stage IV). Simple analysis on the GRFR revealed global patterns of flood thresholds, frequency, and seasonality.

Several caveats are worth noting and for future improvement. First, the fixed-celerity Muskingum routing is known to have deficiencies in flat areas (no backwater) biased slow flood rise and recession. Besides, the framework lacks explicit floodplain parameterization to capture the altered flow dynamics or inundation. Dynamic wave models, with proper floodplain parameterization may be more suitable for flat areas like the Amazon (Yamazaki et al. 2011; Miguez-Macho and Fan 2012; de Paiva et al. 2013). Second, no lake or reservoirs or human regulations were considered and their influences on flow (Zajac et al. 2017) are ignored. As human water management (e.g., reservoirs, irrigation, and water divisions) can significantly alter the flow, caution is needed with GRFR in highly managed regions. Third, the quality and reliability GRFR varies with river gauge availability. We have validated the data over more locations than most studies, but the gaps are still obvious. It is also hard to generalize the evaluation results from gauged to ungauged rivers. Over ungauged areas, the density of rain gauges and GSCD training basins (Fig. 2) can provide some qualitative assessment. We encourage additional assessment wherever possible. Finally, no data assimilation (Pan et al. 2009; Pan and Wood 2013) was performed though it is referred as a “reanalysis.”

Nevertheless, the GRFR database represents a pioneering effort on global reach-level flood reanalysis and may offer new opportunities for global flood studies in terms of baseline data and potential research pathways. For example, its global long-term coverage helps extending flood characterization studies to less gauged areas like parts of Africa and South America. It also helps assessing global flood exposure or vulnerability and how that has changed over time. With its coverage and spatial–temporal resolution, GRFR can serve as a benchmark for future large-scale flood analysis. And compared with GRADES, GRFR’s 3-hourly resolution can better help river-observing satellite missions to develop their discharge algorithms (Durand et al. 2016).

Acknowledgments. Yuan Yang was supported by the National Natural Science Foundation of China (Grant No. 7146101701). Cedric H. David was supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the U.S. National Aeronautics and Space Administration (NASA). The following organizations are thanked for providing river discharge data: the U.S. Geological Survey (USGS), the Global Runoff Data Centre (GRDC), the Brazilian Agência Nacional de Águas, EURO-FRIEND-Water, the European Commission Joint Research Centre (JRC), the Water Survey of Canada (WSC), the Australian Bureau of Meteorology (BoM), and the Chilean Centro de Ciencia del Clima y la Resiliencia (CR2).

Appendix A: Kling–Gupta efficiency (KGE)

KGE and its three component statistics are defined as

$$KGE = 1 - \sqrt{(CC - 1)^2 + RB^2 + (VR - 1)^2}, \quad (A1)$$

$$CC = \frac{\text{cov}(Q_o, Q_m)}{\sigma_{Q_o} \sigma_{Q_m}}, \quad (A2)$$

$$RB = \frac{\mu_{Q_m} - \mu_{Q_o}}{\mu_{Q_o}} \times 100\%, \quad (A3)$$

$$VR = \frac{\sigma_{Q_m} / \mu_{Q_m}}{\sigma_{Q_o} / \mu_{Q_o}}, \quad (A4)$$

where Q_o and Q_m are observed and simulated discharge and μ , σ , and $\text{cov}(\cdot)$ are the mean, standard deviation, and covariance.

Appendix B: Categorical event verification metrics

Three categorical event verification metrics were calculated based on the number of hit, miss, and false alarm cases. If the simulated flood event overlaps within a ± 3 -h buffer around the observed event, we consider it a true positive (TP or “hit”). No overlap leads to a false positive (FP or false alarm) or false negative (FN or miss) or both. When an observed long event overlaps with multiple simulated ones, only the simulated event with the highest peak is considered a hit and others false alarms. Similarly, when a simulated long event overlaps with multiple observed ones, only the observed event with the highest peak is considered a hit and others misses. Then, we calculated POD, FAR, and CSI as

$$POD = TP / (TP + FN), \quad (B1)$$

$$FAR = FP / (TP + FP), \quad (B2)$$

$$CSI = TP / (TP + FN + FP). \quad (B3)$$

Appendix C: Detailed event metrics

For a given “hit” event, we calculated 1) relative peak flow error (ΔQ_p), 2) relative rising time error (Δt_p), and 3) relative volume error (ΔV_k) through

$$\Delta Q_p = \frac{Q_p^m / th_m - Q_p^o / th_o}{Q_p^o / th_o} \times 100\%, \quad (C1)$$

$$\Delta t_p = \frac{[t(Q_p^m) - t_1] - [t(Q_p^o) - tt_1]}{tt_2 - tt_1 + 1} \times 100\%, \quad (C2)$$

$$\Delta V_k = \frac{\sum_{t_1}^{t_2} Q_t^m / th_m / (t_2 - t_1 + 1) - \sum_{tt_1}^{tt_2} Q_{tt}^o / th_o / (tt_2 - tt_1 + 1)}{\sum_{tt_1}^{tt_2} Q_{tt}^o / th_o / (tt_2 - tt_1 + 1)} \times 100\%, \quad (C3)$$

where Q_p^m and Q_p^o are simulated and observed peak flow ($\text{m}^3 \text{s}^{-1}$), th_m and th_o are simulated and observed flood threshold ($\text{m}^3 \text{s}^{-1}$), Q^m and Q^o are simulated and observed discharge ($\text{m}^3 \text{s}^{-1}$), t_1 (tt_1) and t_2 (tt_2) are the start and end time of a simulated (observed) flood event, $t(Q_p^m)$ and $t(Q_p^o)$ are simulated and observed flood peak time.

Appendix D: Seasonality metrics

To calculate the mean date of flood occurrence $\bar{D}$, we first converted the day of the year (DOY) of a flood event D (center date of the event, DOY 366 reset to 365) to a phase angle θ in $[0, 2\pi]$ and found its corresponding complex value U on the unit circle on complex plane:

$$\theta = (D - 0.5)/365 \times 2\pi, \quad (\text{D1})$$

$$U = \exp(\theta i), \quad (\text{D2})$$

where i is the imaginary unit. Then, we averaged U across all events at a river reach to obtain the mean event location $\bar{U}$ on the complex plane. Note that $\bar{U}$ may no longer sit on the unit circle, i.e., $|\bar{U}| \leq 1$. And the more spread out the events are across the year (unit circle) the smaller $|\bar{U}|$, and $|\bar{U}| = 1$ only if all events occurred on the same DOY. We define the $|\bar{U}|$ value as the seasonal concentration index R that goes from 0 (no seasonality) to 1 (strongest seasonality):

$$R = |\bar{U}|. \quad (\text{D3})$$

The phase angle of $\bar{U}$ and corresponding mean DOY of events are

$$\bar{\theta} = \arg(\bar{U})/i, \quad (\text{D4})$$

$$\bar{D} = \bar{\theta}/(2\pi) \times 365 + 0.5. \quad (\text{D5})$$

References

- Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS – Global Ensemble Streamflow Forecasting and Flood Early Warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>.
- , V. Lorini, F. A. Hirpa, S. Harrigan, E. Zsoter, C. Prudhomme, and P. Salamon, 2020: A global streamflow reanalysis for 1980–2018. *J. Hydrol. X*, **6**, 100049, <https://doi.org/10.1016/j.hydroa.2019.100049>.
- Allen, G. H., and T. M. Pavelsky, 2018: Global extent of rivers and streams. *Science*, **361**, 585–588, <https://doi.org/10.1126/science.aat0636>.
- Arheimer, B., R. Pimentel, K. Isberg, L. Crochemore, J. C. M. Andersson, A. Hasan, and L. Pineda, 2020: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrol. Earth Syst. Sci.*, **24**, 535–559, <https://doi.org/10.5194/hess-24-535-2020>.
- Bates, P. D., M. S. Horritt, and T. J. Fawcett, 2010: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *J. Hydrol.*, **387**, 33–45, <https://doi.org/10.1016/j.jhydrol.2010.03.027>.
- 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. Hydrometeorol.*, **16**, 1478–1501, <https://doi.org/10.1175/JHM-D-14-0155.1>.
- , and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, <https://doi.org/10.5194/hess-21-6201-2017>.
- , E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019a: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- , and Coauthors, 2019b: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, <https://doi.org/10.5194/hess-23-207-2019>.
- , M. Pan, P. Lin, J. Seibert, A. I. J. M. van Dijk, and E. F. Wood, 2020a: Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *J. Geophys. Res. Atmos.*, **125**, e2019JD031485, <https://doi.org/10.1029/2019JD031485>.
- , E. F. Wood, T. R. McVicar, M. Zambrano-Bigiarini, C. Alvarez-Garreton, O. M. Baez-Villanueva, J. Sheffield, and D. N. Karger, 2020b: Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *J. Climate*, **33**, 1299–1315, <https://doi.org/10.1175/JCLI-D-19-0332.1>.
- Beevers, L., L. Collet, G. Aitken, C. Maravat, and A. Visser, 2020: The influence of climate model uncertainty on fluvial flood hazard estimation. *Nat. Hazards*, **104**, 2489–2510, <https://doi.org/10.1007/s11069-020-04282-4>.
- Beighley, R. E., and Coauthors, 2011: Comparing satellite derived precipitation datasets using the Hillslope River Routing (HRR) model in the Congo River Basin. *Hydrol. Processes*, **25**, 3216–3229, <https://doi.org/10.1002/hyp.8045>.
- Beighley, R., K. Eggert, T. Dunne, Y. He, V. Gummadi, and K. Verdin, 2009: Simulating hydrologic and hydraulic processes throughout the Amazon River Basin. *Hydrol. Processes*, **23**, 1221–1235, <https://doi.org/10.1002/hyp.7252>.
- Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan, 2016: Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.*, **43**, 4382–4390, <https://doi.org/10.1002/2016GL068070>.
- , S. Harrigan, P. Molnar, L. J. Slater, and J. W. Kirchner, 2019: The relative importance of different flood-generating mechanisms across Europe. *Water Resour. Res.*, **55**, 4582–4593, <https://doi.org/10.1029/2019WR024841>.
- Beven, K. J., 2011: *Rainfall-Runoff Modelling: The Primer*. John Wiley & Sons, 488 pp.
- Biancamaria, S., D. P. Lettenmaier, and T. M. Pavelsky, 2016: The SWOT mission and its capabilities for land hydrology. *Surv. Geophys.*, **37**, 307–337, <https://doi.org/10.1007/s10712-015-9346-y>.
- Blöschl, G., and Coauthors, 2017: Changing climate shifts timing of European floods. *Science*, **357**, 588–590, <https://doi.org/10.1126/science.aan2506>.
- Carpenter, T. M., J. A. Sperflage, K. P. Georgakakos, T. Sweeney, and D. L. Fread, 1999: National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems. *J. Hydrol.*, **224**, 21–44, [https://doi.org/10.1016/S0022-1694\(99\)00115-8](https://doi.org/10.1016/S0022-1694(99)00115-8).
- Chow, V. T., 1988: *Applied Hydrology*. McGraw-Hill, 572 pp.
- Collins, M. J., 2019: River flood seasonality in the Northeast United States: Characterization and trends. *Hydrol. Processes*, **33**, 687–698, <https://doi.org/10.1002/hyp.13355>.
- David, C. H., D. R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout, 2011: River network routing on the NHDPlus Dataset. *J. Hydrometeorol.*, **12**, 913–934, <https://doi.org/10.1175/2011JHM1345.1>.
- , J. S. Famiglietti, Z. L. Yang, F. Habets, and D. R. Maidment, 2016: A decade of RAPID—Reflections on the development of an open source geoscience code. *Earth Space Sci.*, **3**, 226–244, <https://doi.org/10.1002/2015EA000142>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- , M. Balmaseda, G. Balsam, R. Engelen, A. J. Simmons, and J. N. Thepaut, 2014: Toward a consistent reanalysis of the climate system. *Bull. Amer. Meteor. Soc.*, **95**, 1235–1248, <https://doi.org/10.1175/BAMS-D-13-00043.1>.
- de Paiva, R. C. D., D. C. Buarque, W. Collischonn, M.-P. Bonnet, F. Frappart, S. Calmant, and C. A. Bulhões Mendes, 2013: Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resour. Res.*, **49**, 1226–1243, <https://doi.org/10.1002/wrcr.20067>.
- Dickinson, J. E., T. M. Harden, and G. J. McCabe, 2019: Seasonality of climatic drivers of flood variability in the conterminous United States. *Sci. Rep.*, **9**, 15321, <https://doi.org/10.1038/s41598-019-51722-8>.
- Duan, Q. Y., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015–1031, <https://doi.org/10.1029/91WR02985>.
- Durand, M., and Coauthors, 2016: An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resour. Res.*, **52**, 4527–4549, <https://doi.org/10.1002/2015WR018434>.
- Getirana, A. C. V., A. Boone, D. Yamazaki, B. Decharme, F. Papa, and N. Mognard, 2012: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin. *J. Hydrometeorol.*, **13**, 1641–1665, <https://doi.org/10.1175/JHM-D-12-021.1>.
- Gleason, C. J., and L. C. Smith, 2014: Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proc. Natl. Acad. Sci. USA*, **111**, 4788–4791, <https://doi.org/10.1073/pnas.1317606111>.
- Gochis, D., W. Yu, and D. Yates, 2015: The WRF-Hydro model technical description and user's guide, version 3.0. NCAR Tech. Doc., 120 pp., http://www.ral.ucar.edu/projects/wrf_hydro/.
- GRDC, 2020: GRDC stations with monthly data. Global Runoff Data Centre, accessed 2 October 2020, https://www.bafg.de/GRDC/EN/Home/homepage_node.html.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hall, J., and G. Blöschl, 2018: Spatial patterns and characteristics of flood seasonality in Europe. *Hydrol. Earth Syst. Sci.*, **22**, 3883–3901, <https://doi.org/10.5194/hess-22-3883-2018>.
- Harrigan, S., and Coauthors, 2020: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, **12**, 2043–2060, <https://doi.org/10.5194/essd-12-2043-2020>.
- He, X., M. Pan, Z. Wei, E. F. Wood, and J. Sheffield, 2020: A global drought and flood catalogue from 1950 to 2016. *Bull. Amer. Meteor. Soc.*, **101**, E508–E535, <https://doi.org/10.1175/BAMS-D-18-0269.1>.
- Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ERA Rep. Series 27, ECMWF, 65 pp., <https://doi.org/10.21957/tkic6g3wm>.

- Hirpa, F. A., P. Salamon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter, and S. J. Dadson, 2018: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.*, **566**, 595–606, <https://doi.org/10.1016/j.jhydrol.2018.09.052>.
- Huang, Z., H. Wu, R. F. Adler, G. Schumann, J. J. Gourley, A. Kettner, and N. Nanding, 2021: Multi-sourced flood inventories over the contiguous United States for actual and natural conditions. *Bull. Amer. Meteor. Soc.*, **102**, E1133–E1149, <https://doi.org/10.1175/BAMS-D-20-0001.1>.
- Karnes, D., and G. R. Brakenridge, 1996: The Dartmouth Flood Observatory: An electronic research tool and archive for investigations of extreme flood events. *Geosci. Inf. Soc. Proc.*, **27**, 31–36, <https://hdl.handle.net/1969.1/156084>.
- Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424–425**, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>.
- Kumar, S. V., and Coauthors, 2006: Land information system: An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415, <https://doi.org/10.1016/j.envsoft.2005.07.004>.
- Lehner, B., and G. Grill, 2013: Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrol. Processes*, **27**, 2171–2186, <https://doi.org/10.1002/hyp.9740>.
- Li, H., M. S. Wigmosta, H. Wu, M. Huang, Y. Ke, A. M. Coleman, and L. R. Leung, 2013: A physically based runoff routing model for land surface and earth system models. *J. Hydrometeorol.*, **14**, 808–828, <https://doi.org/10.1175/JHM-D-12-015.1>.
- , L. R. Leung, A. Getirana, M. Huang, H. Wu, Y. Xu, J. Guo, and N. Voisin, 2015: Evaluating global streamflow simulations by a physically based routing model coupled with the community land model. *J. Hydrometeorol.*, **16**, 948–971, <https://doi.org/10.1175/JHM-D-14-0079.1>.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *J. Geophys. Res.*, **99**, 14415–14428, <https://doi.org/10.1029/94JD00483>.
- , E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global Planet. Change*, **13**, 195–206, [https://doi.org/10.1016/0921-8181\(95\)00046-1](https://doi.org/10.1016/0921-8181(95)00046-1).
- Lin, P., L. J. Hopper Jr., Z.-L. Yang, M. Lenz, and J. W. Zeitler, 2018a: Insights into hydrometeorological factors constraining flood prediction skill during the May and October 2015 Texas Hill country flood events. *J. Hydrometeorol.*, **19**, 1339–1361, <https://doi.org/10.1175/JHM-D-18-0038.1>.
- , Z.-L. Yang, D. J. Gochis, W. Yu, D. R. Maidment, M. A. Somos-Valenzuela, and C. H. David, 2018b: Implementation of a vector-based river network routing scheme in the community WRF-Hydro modeling framework for flood discharge simulation. *Environ. Modell. Software*, **107**, 1–11, <https://doi.org/10.1016/j.envsoft.2018.05.018>.
- , and Coauthors, 2019: Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resour. Res.*, **55**, 6499–6516, <https://doi.org/10.1029/2019WR025287>.
- , M. Pan, G. H. Allen, R. P. de Frasson, Z. Zeng, D. Yamazaki, and E. F. Wood, 2020: Global estimates of reach-level bankfull river width leveraging big data geospatial analysis. *Geophys. Res. Lett.*, **47**, e2019GL086405, <https://doi.org/10.1029/2019GL086405>.
- , M. Pan, E. F. Wood, D. Yamazaki, and G. H. Allen, 2021: A new vector-based global river network dataset accounting for variable drainage density. *Sci. Data*, **8**, 28, <https://doi.org/10.1038/s41597-021-00819-9>.
- Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.
- Linke, S., B. Lehner, C. O. Dallaire, J. Ariwi, and M. Thieme, 2019: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Sci. Data*, **6**, 283, <https://doi.org/10.1038/s41597-019-0300-6>.
- Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. Evans, 2011: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 425–436, <https://doi.org/10.5194/hess-15-425-2011>.
- Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier, 1998: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol. Sci. J.*, **43**, 131–141, <https://doi.org/10.1080/02626669809492107>.
- Miguez-Macho, G., and Y. Fan, 2012: The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. *J. Geophys. Res.*, **117**, D15113, <https://doi.org/10.1029/2012JD017539>.
- Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, <https://doi.org/10.1029/2003JD003823>.
- Mizukami, N., and Coauthors, 2016: MizuRoute version 1: A river network routing tool for a continental domain water resources applications. *Geosci. Model Dev.*, **9**, 2223–2238, <https://doi.org/10.5194/gmd-9-2223-2016>.
- , M. P. Clark, A. J. Newman, A. W. Wood, E. D. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego, 2017: Towards seamless large-domain parameter estimation for hydrologic models. *Water Resour. Res.*, **53**, 8020–8040, <https://doi.org/10.1002/2017WR020401>.
- Moradkhani, H., and S. Sorooshian, 2008: General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. *Hydrological Modelling and the Water Cycle*, S. Sorooshian et al., Eds., Water Science and Technology Library, Vol. 63., Springer, 1–24, https://doi.org/10.1007/978-3-540-77843-1_1.
- Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *J. Climate*, **14**, 3307–3323, [https://doi.org/10.1175/1520-0442\(2001\)014<3307:PTDOGR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2).
- Oki, T., and Y. Sud, 1998: Design of Total Runoff Integrating Pathways (TRIP)—A global river channel network. *Earth Interact.*, **2**, [https://doi.org/10.1175/1087-3562\(1998\)002<0001:DOTRIP>2.3.CO;2](https://doi.org/10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2).
- Olivera, F., Lear, M. S., Famiglietti, J. S., and Asante, K., 2002: Extracting low-resolution river networks from high-resolution digital elevation models. *Water Resour. Res.*, **38**, 1231, <https://doi.org/10.1029/2001WR000726>.
- O'Loughlin, F. E., J. Neal, G. J. P. Schumann, E. Beighley, and P. D. Bates, 2020: A LISFLOOD-FP hydraulic model of the middle reach of the Congo. *J. Hydrol.*, **580**, 124203, <https://doi.org/10.1016/j.jhydrol.2019.124203>.
- Oubeidillah, A. A., S. C. Kao, M. Ashfaq, B. S. Naz, and G. Tootle, 2014: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrol. Earth Syst. Sci.*, **18**, 67–84, <https://doi.org/10.5194/hess-18-67-2014>.
- Pan, M., and E. F. Wood, 2013: Inverse streamflow routing. *Hydrol. Earth Syst. Sci.*, **17**, 4577–4588, <https://doi.org/10.5194/hess-17-4577-2013>.
- , —, D. B. McLaughlin, D. Entekhabi, and L. Luo, 2009: A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment. *J. Hydrometeorol.*, **10**, 794–806, <https://doi.org/10.1175/2009JHM1088.1>.
- , X. Cai, N. W. Chaney, D. Entekhabi, and E. F. Wood, 2016: An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophys. Res. Lett.*, **43**, 9662–9668, <https://doi.org/10.1002/2016GL069964>.
- Parajka, J., S. Kohnová, R. Merz, J. Szolgay, K. Hlavčová, and G. Blöschl, 2009: Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria. *Hydrol. Sci. J.*, **54**, 456–473, <https://doi.org/10.1623/hysj.54.3.456>.
- Peters-Lidard, C. D., and Coauthors, 2007: High-performance earth system modeling with NASA/GSFC's Land Information System. *Innovations Syst. Software Eng.*, **3**, 157–165, <https://doi.org/10.1007/s11334-007-0028-x>.
- Qiao, X., E. J. Nelson, D. P. Ames, Z. Li, and M. A. Matin, 2019: A systems approach to routing global gridded runoff through local high-resolution stream networks for flood early warning systems. *Environ. Modell. Software*, **120**, 104501, <https://doi.org/10.1016/j.envsoft.2019.104501>.
- Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged

- locations. *J. Hydrol.*, **337**, 402–420, <https://doi.org/10.1016/j.jhydrol.2007.02.015>.
- Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, <https://doi.org/10.1029/2004GL020938>.
- Rodell, M., and Coauthors, 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Saharia, M., P.-E. Kirstetter, H. Vergara, J. J. Gourley, Y. Hong, and M. Giroud, 2017: Mapping flash flood severity in the United States. *J. Hydrometeor.*, **18**, 397–411, <https://doi.org/10.1175/JHM-D-16-0082.1>.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, <https://doi.org/10.1175/JCLI3790.1>.
- Sutanudjaja, E. H., and Coauthors, 2018: PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geosci. Model Dev.*, **11**, 2429–2453, <https://doi.org/10.5194/gmd-11-2429-2018>.
- Toth, E., 2016: Estimation of flood warning runoff thresholds in ungauged basins with asymmetric error functions. *Hydrol. Earth Syst. Sci.*, **20**, 2383–2394, <https://doi.org/10.5194/hess-20-2383-2016>.
- UNISDR, 2015: The human cost of weather related disasters 1995–2005. UN Office for Disaster Risk Reduction, 30 pp., https://www.unisdr.org/2015/docs/climatechange/COP21_WeatherDisastersReport_2015_FINAL.pdf.
- Villarini, G., 2016: On the seasonality of flooding across the continental United States. *Adv. Water Resour.*, **87**, 80–91, <https://doi.org/10.1016/j.advwatres.2015.11.009>.
- Willner, S. N., C. Otto, and A. Levermann, 2018: Global economic response to river floods. *Nat. Climate Change*, **8**, 594–598, <https://doi.org/10.1038/s41558-018-0173-2>.
- Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, 1992: A land-surface hydrology parameterization with subgrid variability for general-circulation models. *J. Geophys. Res.*, **97**, 2717–2728, <https://doi.org/10.1029/91JD01786>.
- , and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.*, **47**, W05301, <https://doi.org/10.1029/2010WR010090>.
- Wu, H., J. S. Kimball, N. Mantua, and J. Stanford, 2011: Automated upscaling of river networks for macroscale hydrological modeling. *Water Resour. Res.*, **47**, W03517, <https://doi.org/10.1029/2009WR008871>.
- , H. Li, M. Huang, L. R. Leung, and R. F. Adler, 2012: A new global river network database for macroscale hydrologic modeling. *Water Resour. Res.*, **48**, W09701, <https://doi.org/10.1029/2012WR012313>.
- , R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.*, **50**, 2693–2717, <https://doi.org/10.1002/2013WR014710>.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, <https://doi.org/10.1029/2011JD016048>.
- Yamazaki, D., T. Oki, and S. Kanae, 2009: Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map. *Hydrol. Earth Syst. Sci.*, **13**, 2241–2251, <https://doi.org/10.5194/hess-13-2241-2009>.
- , S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, **47**, W04501, <https://doi.org/10.1029/2010WR009726>.
- , G. A. M. D. Almeida, and P. D. Bates, 2013: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resour. Res.*, **49**, 7221–7235, <https://doi.org/10.1002/wrcr.20552>.
- , and Coauthors, 2017: A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.*, **44**, 5844–5853, <https://doi.org/10.1002/2017GL072874>.
- , D. Ikeshima, J. Sosa, P. D. Bates, G. H. Allen, and T. M. Pavelsky, 2019: MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.*, **55**, 5053–5073, <https://doi.org/10.1029/2019WR024873>.
- Yang, Y., M. Pan, H. E. Beck, C. K. Fisher, R. E. Beighley, S.-C. Kao, Y. Hong, and E. F. Wood, 2019: In quest of calibration density and consistency in hydrologic modeling: Distributed parameter calibration against streamflow characteristics. *Water Resour. Res.*, **55**, 7784–7803, <https://doi.org/10.1029/2018WR024178>.
- Ye, S., H.-Y. Li, L. R. Leung, J. Guo, Q. Ran, Y. Demissie, and M. Sivapalan, 2017: Understanding flood seasonality and its temporal shifts within the contiguous United States. *J. Hydrometeor.*, **18**, 1997–2009, <https://doi.org/10.1175/JHM-D-16-0207.1>.
- Yilmaz, K. K., R. F. Adler, Y. Tian, Y. Hong, and H. F. Pierce, 2010: Evaluation of a satellite-based global flood monitoring system. *Int. J. Remote Sens.*, **31**, 3763–3782, <https://doi.org/10.1080/01431161.2010.483489>.
- Zajac, Z., B. Revilla-Romero, P. Salamon, P. Burek, F. A. Hirpa, and H. Beck, 2017: The impact of lake and reservoir parameterization on global streamflow simulation. *J. Hydrol.*, **548**, 552–568, <https://doi.org/10.1016/j.jhydrol.2017.03.022>.