AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR027794

Combining Optical Remote Sensing, McFLI Discharge Estimation, Global Hydrologic Modeling, and Data Assimilation to Improve Daily Discharge Estimates Across an Entire Large Watershed

Yuta Ishitsuka[1], Colin J. Gleason[1], Mark W. Hagemann[1,2], Edward Beighley[3], George H. Allen[4,5], Dongmei Feng[1], Peirong Lin[6], Ming Pan[6], Konstantinos Andreadis[1], and Tamlin M. Pavelsky[4]

[1]Department of Civil and Environmental Engineering, University of Massachusetts Amherst, Amherst, MA, USA, [2]School of Earth Sciences, The Ohio State University, Columbus, OH, USA, [3]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA, [4]Department of Geological Sciences, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA, [5]Department of Geography, Texas A&M University, College Station, TX, USA, [6]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

Key Points:

- ~1 million satellite-derived river widths are used to estimate river discharge via the BAM algorithm
- Daily basin-wide discharge estimation is improved by addition of remote sensing
- Our approach is a potential direct analog for future SWOT observations

Correspondence to:
Y. Ishitsuka,
yishitsuka@umass.edu

Citation:
Ishitsuka, Y., Gleason, C. J., Hagemann, M. W., Beighley, E., Allen, G. H., Feng, D., et al. (2021). Combining optical remote sensing, McFLI discharge estimation, global hydrologic modeling, and data assimilation to improve daily discharge estimates across an entire large watershed. *Water Resources Research*, *56*, e2020WR027794. https://doi.org/10.1029/2020WR027794

Received 1 MAY 2020
Accepted 22 NOV 2020

Abstract Remote sensing has gained attention as a novel source of primary information for estimating river discharge, and the Mass-conserved Flow Law Inversion (McFLI) approach has successfully estimated river discharge in ungauged basins solely from optical satellite data. However, McFLI currently suffers from two major drawbacks: (1) existing optical satellites lead to temporally and spatially sparse discharge estimates and (2) because of the assumptions required, McFLI cannot guarantee downstream flow continuity. Hydrological modeling has neither drawback, yet model accuracy is frequently limited by a lack of discharge observations. We therefore combine McFLI and models in a data assimilation framework applicable globally. We establish a daily "ungauged" baseline model for 28,998 reaches of the Missouri river basin forced by recently published global runoff data, which we do not calibrate. We estimate discharge via McFLI using ~1 million width measurements made from 12,000 Landsat scenes and assimilate McFLI into the model before validating at 403 USGS gauges. Results show that assimilated discharges did not impair already accurate baseline flows and achieved median improvements of 28% normalized root mean square error, 0.50 Nash–Sutcliffe efficiency (NSE), and 0.23 Kling–Gupta efficiency where baseline performance was poor (defined as baseline negative NSE, 225/403 reaches). We ultimately improved flows at 92% of these originally poorly modeled gauges, even though Landsat images only provide McFLI discharges at 1.5% of reaches and 26% of simulated days. Our results suggest that the combination of McFLI and state-of-the-art hydrology models can improve flow estimations in ungauged basins globally.

1. Introduction

Advances in computing power, geospatial data availability, and hydrologic theory have prompted visions for a future in which we move toward "predicting everywhere water resources state variables and outcomes, in short, the hydrologic, socioeconomic, and environmental consequences that are important to governments, industries, and individuals (Brown et al., 2015, p. 6619)." A particular hydrologic focus is often placed on predicting and understanding river discharge, a key component of surface water availability for human and ecosystem use and the integrator of all hydrologic processes in a watershed. This important quantity is best measured in the field or automatically monitored via gauging stations. However, the information these gauges provide is expensive and time consuming to maintain and often politically sensitive in international river basins, resulting in a gradual decline in globally available gauge data (Gleason & Hamdan, 2015; Hannah et al., 2011; Vörösmarty et al., 2010). This decline is doubly troubling as these gauges are also the main data employed to calibrate and validate hydrologic models used to predict water resources (Gleason & Durand, 2020).

Global hydrologic models (GHMs) have seen recent advancements in the representation of hydrologic and hydraulic physics (Bates & De Roo, 2000; Fan et al., 2019; Li et al., 2013; Yamazaki et al., 2011), in its

modules for water management and human water use (Sutanudjaja et al., 2018; Zajac et al., 2017), and in model resolution (Li et al., 2013; Sutanudjaja et al., 2018). As these models continue to advance, they need improved primary observations at the global scale for calibration and validation. These are unlikely to come from expanded access to in situ gauge data (Gleason & Hamdan, 2015).

Remote sensing offers one possible way to provide needed primary data about the global hydrologic system. Examples include use of the GRACE satellites to track changes in the Earth's gravity anomaly and thus infer groundwater changes (e.g., Ehalt Macedo et al., 2019; Lee et al., 2011; Rodell et al., 2009, 2018; Syed et al., 2009), use of satellites to measure river and landscape variables in conjunction with gauges to establish "virtual gauges" or other empirical means of estimating flow (e.g., Bjerklie et al., 2003, 2005; Bjerklie et al., 2018; Brakenridge et al., 2007; Paris et al., 2016; Pavelsky, 2014; Sichangi et al., 2016; Tarpanelli et al., 2013, 2019), use of satellites to measure hydrologic parameters in order to calibrate/assimilate into a hydrologic or hydraulic model (e.g., Andreadis & Lettenmaier, 2006; Emery et al., 2018; Reichle et al., 2014), and use of satellite-observed river height, width, and slope at cross sections in a mass-conserved reach in the Mass-conserved Flow Law Inversion (McFLI) paradigm that does not require gauge information (e.g., Durand et al., 2016; Garambois & Monnier, 2015; Gleason et al., 2017; Gleason & Smith, 2014; Hagemann et al., 2017). McFLI methods use satellite observations of river hydraulics (e.g., width, height, or slope in all combinations thereof) to estimate river discharge given a flow law (e.g., Manning's equation, hydraulic geometry). These algorithms assume that short reaches of rivers are mass conserved at the instant they are observed via satellite and thereby better constrain an ill-posed system of equations to estimate unknown parameters and ultimately discharge in the chosen flow law (Gleason & Durand, 2020; Gleason et al., 2017).

McFLI methods do not need in situ data to estimate discharge (although they improve markedly with gauge data; Durand et al., 2016), but they are temporally and spatially limited by satellite overpass geometry and sensor specific limitations (e.g., cloud cover). A new satellite mission has been a catalyst for development of McFLI methods: the NASA/CNES/CSA/UKSA Surface Water and Ocean Topography (SWOT) mission (Biancamaria et al., 2016; Durand et al., 2010). This Ka-band radar will measure surface water extent, elevation, and slope globally with a 3–21-day repeat time (i.e., 1–7 observation(s) available in one 21-day orbit cycle, although some areas of the earth are explicitly not sensed) depending on latitude (Biancamaria et al., 2016; Pavelsky, 2014). Several successful McFLI algorithms have been demonstrated with simulated SWOT data (Durand et al., 2016; Garambois & Monnier, 2015; Hagemann et al., 2017), yet McFLI approaches have only been demonstrated from currently extant satellite data by using optical satellites to measure river width combined with the at-many-stations hydraulic geometry (AMHG) theory of river behavior (Bonnema et al., 2016; Feng et al., 2019; Gleason et al., 2014, 2018; Gleason & Hamdan, 2015; Gleason & Smith, 2014). Despite these promising advances using optical satellites, even an ensemble of the latest generation optical satellites and SWOT (i.e., SWOT/Landsat/Sentinel-2) would still have a primary observation at best every ~2–3 days at some latitudes. Some cubesats, including SkySat/Maxar operated by Planet, are available as well to provide high spatial resolution data almost every day, but these data are not free of charge and previous work has shown some issues in river width measurement from available four-band sensors (Feng et al., 2019). While deploying McFLI on a compiled fusion data set would represent a major advance in river remote sensing, it would still fall short of providing primary observations for daily or subdaily estimates of discharge as often needed by water resource managers.

The assumption of mass conservation inherent to McFLI poses another challenge for global discharge estimation. Satellite data collected over a river reach provide observations of fluvial parameters simultaneously at all locations in the reach. McFLI algorithms use these observations to estimate unknown parameters of flow lows (e.g., Manning's equation, hydraulic geometry) that produce a conserved discharge in that reach. To solve the underconstrained system of equations, we must assume mass conservation at the instant of observation, because as without the assumption of mass conservation, the problem is too ill posed to solve (e.g., regularization is needed, Durand et al., 2016; Gleason et al., 2017). This means that we assume uniform discharge across all cross sections within a target reach (here, we have multiple cross-sectional observations of river width within a reach). This assumption creates a set of equations with unknown parameters that vary by cross section that share a spatially constant but temporally variable discharge in each reach. The problem remains ill posed even with the conservation assumption but the solution space is better controlled, although equally likely solutions to such a system can be wildly different hydrologically (i.e.,

equifinality in an underconstrained system). As a consequence, discharge estimates in consecutive reaches made from McFLI algorithms could produce results that violate basic principles of hydrology, and flow discontinuities that do not obey downstream continuity in adjacent reaches or false indications of water withdrawals are not detectable by current McFLI practice.

In contrast, GHMs can provide daily discharge estimates that obey downstream mass conservation. However, as discussed above, GHMs require extensive input data to improve their continued accuracy in some basins, and their accuracy is fundamentally tied to calibration data. Furthermore, most GHMs assume natural flow and do not consider the impact from human activities such as dam operations. In these cases, satellite data can provide observations of real river states in human-impacted rivers (e.g., water levels and river widths responding to actual, rather than natural, flows downstream of a dam). Significant effort has been poured into fusion of models and various remotely sensed observations (e.g., Andreadis & Lettenmaier, 2006; Andreadis et al., 2007; Durand et al., 2008; Beighley et al., 2011; Biancamaria et al., 2011; Yoon et al., 2012; Yoon & Beighley, 2015; Munier et al., 2015; Revilla-Romero et al., 2016; Emery et al., 2018, 2020; Oubanas et al., 2018; Häfliger et al., 2019; Oaida et al., 2019; Stampoulis et al., 2019), but the effectiveness of McFLI has never been addressed in this context. The establishment of a McFLI algorithm capable of working from optical data alone using a Bayesian framework (BAM, the Bayesian At Many-stations Hydraulic Geometry Manning algorithm, Hagemann et al., 2017) has made McFLI more attractive to integrate with GHMs via data assimilation. Note that we use the terms BAM and McFLI interchangeably in this study, although we favor the term McFLI when discussing generalized ideas/results. McFLI methods hold tremendous promise for remote sensing of discharge in ungauged settings, yet these methods will not be adopted or adapted by water resource managers until they can provide hydrologically consistent (and sensible) estimates of flow at least at daily time scales and ideally everywhere in a basin. Thus, combining McFLI and GHMs seems a promising avenue toward achieving these goals.

We therefore developed and validated a framework for estimating daily discharge across all major rivers in an entire river basin by combining the latest advances in GHMs and optical remote sensing in the Google Earth Engine (GEE) platform. We chose the 1.4 M km^2 Missouri river basin (hereafter, the Missouri) for demonstration and estimated daily flows at 28,998 reaches for an 8-year period from 2002 to 2010. Our goal is to ultimately address the question: *does the addition of Landsat-based McFLI discharges improve discharge estimation everywhere in a river basin?* We seek to overcome the inherently limited spatial and temporal resolutions of satellite data and the primary data limitations of GHMs to produce discharges that are hydrologically consistent, advancing the science of river discharge estimation.

2. Study Area and Experiment Design

2.1. Study Area and Basin Characterization

We developed and validated our approach for a large and hydrologically complex basin: the Missouri. The Missouri drains an area of ~1.4 M km^2, a significant portion of the continental United States and peak nonflood flows at the basin outlet routinely top 10,000 m^3/s. By contrast, the headwaters of the Missouri include rivers like the visible-from-Landsat Musselshell River, which has gauged flows (USGS No. 06130500) on the order of 1 m^3/s despite being wider than 100 m at times. In addition, the Missouri basin is highly regulated and the country's largest cascade of dams significantly alters the natural hydrograph. Accordingly, providing discharge estimates that obey routing laws and correctly predict flows simultaneously at headwaters and the outlet presents a stern hydrologic challenge.

Daily discharges from 403 USGS gauges in our basin from 1984 to 2010 were obtained via the *DataRetrieval* R package published by the USGS (http://usgs-r.github.io/dataRetrieval/#citing-dataretrieval). The novelty of McFLI is that we can estimate discharge solely from satellites, even in ungauged basins. To prove the effectiveness of McFLI and GHM assimilation in these basins, we assumed that the Missouri was ungauged, and we only used these gauge data for validation. Note that McFLI can improve its accuracy with prior information of flows from gauges (Durand et al., 2016; Feng et al, 2019), but we do not consider this scenario here to reflect our "ungauged" version of the Missouri. Similarly, our eventual modeling work described later is all uncalibrated, preserving our assertion of scalability.

2.2. GEE Processing

We developed an automated remote-sensing image processing procedure to compute multitemporal, cross-sectional river width measurements from Landsat imagery in GEE. We analyzed over 12,000 Landsat TM, ETM+, and OLI scenes to produce nearly 1 million unique river width measurements at ~2,700 discrete river locations from 1984 to 2017. We used the Global River Widths from Landsat (GRWL) Database (Allen, Pavelsky, et al., 2018; Allen & Pavelsky, 2015, 2018) to determine which rivers were wide enough for the AMHG algorithm to yield suitable results from 30-m-resolution Landsat imagery. We set a minimum width of 120 m, which represents rivers 3–4 Landsat pixels wide and allows for detection of width changes visible from these 30-m-resolution images. The GRWL database contains 7,874 km of rivers wider than 120 m in the Missouri river basin. We extracted these rivers and split each GRWL river segment (the reach between river network nodes as defined by GRWL) into approximately 40 evenly spaced nodes per reach for width measurement. This extraction created 74 river reaches with an average width-sampling site spacing of 2.9 km, resulting in a total of 2,723 width-sampling sites across the Missouri basin. At each of these nodes, we generated a vector cross section oriented perpendicular to the river centerline by first calculating the bearing of the GRWL centerline within a 5-pixel window along the centerline and then fixing an orthogonal cross section to this centerline bearing (Figure 1c). This orthogonal represents a static cross section from which we derive river widths compatible with the BAM algorithm (Section 2.3).

We then measured 33 years of water inundation extent along each of the river cross-section vectors using data from Landsat TM, ETM+, and OLI in GEE. We first identified Landsat scenes with cloud cover <10% that intersected the river cross-section vectors. For each of the 12,339 Landsat scenes that met these criteria, we classified surface water using the Modified Normalized Difference Water Index (MNDWI; Xu, 2006) and classified clouds and shadows using the FMask algorithm (Zhu & Woodcock, 2012). We then converted the orthogonal cross-section vectors to polygons by adding a 15-m buffer to each cross section, thus establishing a one-pixel-wide area to extract water surfaces. Within each of these newly buffered cross-sectional polygons, we calculated the area of water, cloud, and shadow for each classified Landsat TM, ETM+, and OLI scene. We then calculated river width by dividing the water area in each polygon by the orthogonal polygon width of 30 m, producing 997,453 nonzero river width measurements. Of these measurements, 68,804 contained clouds and/or shadows along the cross-section vector and were excluded from the analysis. The methods developed here are fully automated and could be implemented at the global scale. Figure 1 shows the river network and example width extractions for this study.

The fully automated GEE process described above efficiently generated nearly 1 million width measurements. However, despite careful planning to avoid obvious errors, we performed manual QA/QC of the resulting widths to eliminate some known problems. We assessed the cross sections used to generate widths rather than individual classifications or individual width measurements for efficiency's sake, as the 2,700 total cross sections could be evaluated manually in a reasonable amount of time. We eliminated those cross sections behind dams (980 cross sections) as there is no evidence that BAM is effective behind dams. As this was a pilot study, we manually applied this correction, yet this filter could easily be applied automatically in future. We also eliminated those cross sections that extended beyond river boundaries into neighboring water bodies (these created a large positive bias by including water outside the channel/not orthogonal to flow direction: 947 cross sections). A total of 773 cross sections in 40 reaches across 28,998 modeling reaches remained after this quality control, which erred toward conservatism to include only cross sections where width measurements are likely to be highly accurate. As a result, we had 784/3,015 days (26%) when a BAM observation was available somewhere across the 40 BAM reaches. Improvements to our GEE processing are needed to retain a higher degree of inputs automatically, and recent work has moved in this direction (e.g., Yang et al., 2020).

2.3. Estimating Discharge with BAM

BAM is a Bayesian McFLI algorithm that estimates discharge using either Manning's equation (when water surface elevation is available; Manning, 1891) or AMHG (when only widths are available, Gleason & Smith, 2014; C. J. Gleason & Wang, 2015) as a flow law. We used AMHG as an internal algorithm in BAM, as our primary data in this study are optical images. BAM was developed by Hagemann et al. (2017), and

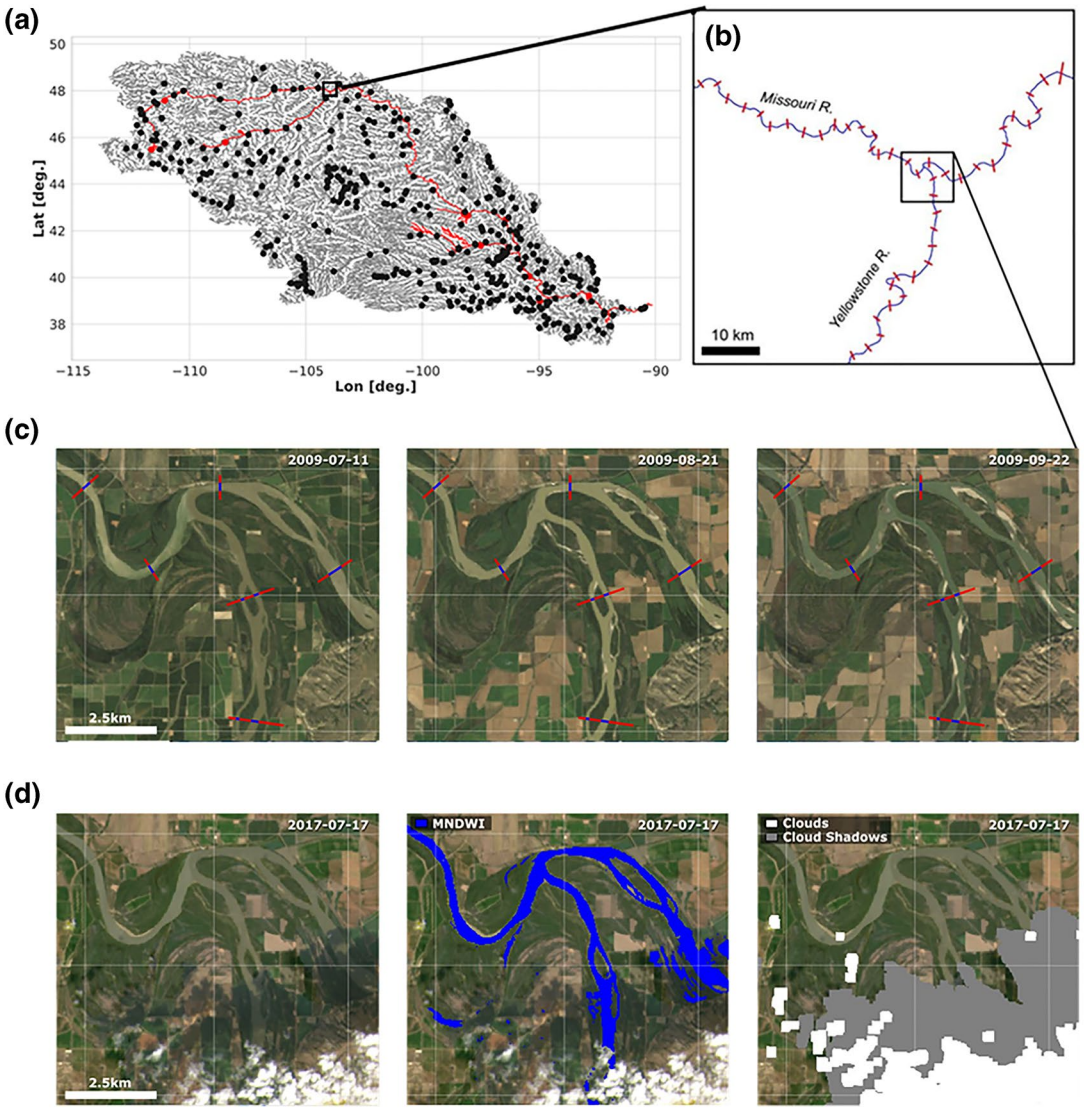

Figure 1. Automatic extraction of river widths from Landsat imagery. (a) Missouri basin with equally spaced river cross sections derived from the GRWL database (red), validation gauges (black), and 28,998 river reaches (light gray). (b) The Missouri-Yellowstone confluence with river centerline in blue and cross sections in red. (c) Time series of three example Landsat images above the Missouri-Yellowstone confluence showing classified water at each cross section in blue. (d) Classification of water (MNDWI, center) and clouds/cloud shadows (right). GRWL, Global River Widths from Landsat; MNDWI, Modified Normalized Difference Water Index.

a detailed description can be found therein. Here, we summarize the algorithm and note key inputs and parameters as deployed in this study. Readers are reminded that BAM is one of a handful of recently published McFLI methods (see Gleason et al., 2017, for description and Gleason & Durand, 2020 for a thorough review).

BAM estimates AMHG parameters and the resulting discharge at each cross section within a mass-conserved reach via Bayesian inference as a posterior distribution that accounts for both prior beliefs and information contained in the observed widths. A likelihood function representing the AMHG equation modifies a set of prior distributions to produce this posterior, and estimates are drawn as samples from the posterior using the Hamiltonian Monte-Carlo method. Thus, the "priors" of AMHG parameters and flow are important, as these guide the solution in tandem with observations (Feng et al., 2019). BAM requires priors (probability distributions) of flow, width–discharge exponents (hydraulic geometry b values), and the AMHG-specific parameters Q_c and w_c (Gleason & Wang, 2015). Q_c and w_c are defined as the intersection point of

all AHG curves at a connected reach fundamental to AMHG. These two values are shared across all reaches, which leads us to the basic AMHG equation:

$$w_c = aQ_c^b$$

where a and b are hydraulic geometry values at each cross section within a given reach. Readers may refer to Gleason and Wang (2015) for more description.

Our interest here is for BAM to improve flows in ungauged basins. We thus used state-of-the-art hydrologic reconstruction data (Global Reach-level A priori Discharge Estimates; GRADES; Lin et al., 2019) and the MERIT Hydro (Yamazaki et al., 2019) global hydrographic database to provide off-the-shelf priors for BAM and a baseline daily discharge for all 28,998 reaches in our study area. These are globally available and consistent data sets, and our experiment setup and model configuration are described next.

We use the output of the GEE process described in Section 2.2 to provide repeated width measurements at fixed cross sections within mass-conserved reaches, and it is this data set that drives BAM. We limit discharge estimation to the 8-year time period 2002–2010 and assume this reasonably preserves the assumption of a constant channel hydraulic geometry inherent in BAM, as channel geometric changes that would invalidate BAM might be expected over the entire 35-year time domain of the original GEE processing. After this step, our input data were prepared, and we ran BAM on a computing cluster using the BAM code version 0.1.2 at https://github.com/markwh/bamr. Ultimately, the GEE and BAM procedures resulted in 1,906 independent discharge estimates made at 40 reaches. These flows are generated independently of one another and are only made during a Landsat overpass, which is approximately twice per month for any given reach. Thus, these results alone do not achieve our goal of daily predictions of discharge that obey network continuity.

2.4. Routing Model

Our goal is to produce daily flows in all reaches that obey conservation laws, and for this we use the Hillslope River Routing (HRR) model to route flows throughout the basin (Beighley et al., 2009). The model has two routing components, where hillslope runoff (surface and subsurface) is routed laterally to the channels and integrated along the river network based on channel routing. The model solves the former process via an explicit kinematic wave and the latter via the Muskingum–Cunge method. HRR requires inputs of subcatchment water excess to the channel, catchment-averaged slopes, basic channel properties for Muskingum parameters like channel width, length, and slope derived from remote sensing, and a network topology to define how each reach interacts with the entire network. Some of those properties (i.e., channel width, length, slope, and network topology) were already derived by Lin et al. (2019) for MERIT HYDRO, and we derived catchment-averaged slopes from the same source. To deploy HRR, we use runoff estimates from GRADES as a lateral input of this model.

HRR has two empirical parameters: a hillslope roughness coefficient and a channel Manning's n. Since this study does not have gauge data available for calibration of these parameters, we directly imposed parameters on each reach. Channel friction (i.e., Manning's n) is a dominant control on discharge in a routing model, and simply assigning a global value would lead to inaccurate results. We used a recently published data set of time-varying river hydraulics at 1,600 stations across 155 rivers in the United States (Brinkerhoff et al., 2019) to assign n to our reaches. These hydraulics are derived from USGS measurements and represent high-quality field data for river width, depth, slope, and, importantly, Manning's n. Brinkerhoff et al. found that there are trends between slope and flow resistance for these data, and we use this logic to assign Manning's n via channel slope. We therefore binned slopes into five equal classes (i.e., quantiles) and then took the median of the n per slope bin in Brinkerhoff et al.'s training data and assigned it to all reaches in our study area falling within that slope bin. Given the variety of rivers in Brinkerhoff et al.'s data set (which includes rivers in the Missouri basin), we believe that these data can be used in our ungauged scenario. For each hillslope, we set the hillslope roughness coefficient to the default value (0.7) in the model. This global value will likely introduce error into the routing results, as the heterogeneity of landscapes throughout this vast basin are not represented. We could have used land cover from the National Land Cover Database

(e.g., Wickham et al., 2014) to assign hillslope frictions to each subcatchment, but the literature assigning hillslope roughness to land cover is much less developed than channel friction, and we believe introducing such a scheme would introduce more error than it prevented. Further, the portability of such a method to other areas with ill-defined landcover is unknown. Therefore, in absence of calibration data or large data sets of hillslope friction we can relate to our rivers a priori, we impose this global value and accept that it will introduce some error into our results. We have not directly included the effect of reservoirs in our routing for this study, although we have directly observed reservoir-affected reaches through Landsat images. Reservoirs were treated as river reaches with very low slope (e.g., Allen, David, et al. 2018), which approximates their lake-like function but does not account for human management decisions. As mentioned before, however, incorporation of satellite images can indirectly treat this problem by merging the actual state of rivers into models.

2.5. Input Forcing and Baseline Estimation

One of our key questions for this study is whether Landsat data can complement and improve state-of-the-art model estimates of discharge. Therefore, we started from the most recently published global hydrologic reconstruction by Lin et al. (2019), where runoff simulations were provided by running the Variable Infiltration Capacity model (Liang et al., 1994) together with recently published precipitation inputs (Beck et al., 2019) as well as extensive model calibration/bias correction with a set of machine learning-derived global runoff signature maps (Beck et al., 2015). GRADES is globally available and merges many existing data sets to constrain runoff as much as possible globally. While there exist other data sets of a similar kind, GRADES represents the most spatially explicit and recently updated publicly available data. Models like these form our best understanding of the global water cycle, and this product represents the state of the art in status quo "off-the shelf" estimates of global runoff, upon which potential gains from remote sensing can be investigated. Readers are referred to Lin et al. (2019) for details on GRADES, as further describing their substantial effort here is well outside the scope of this manuscript.

We used GRADES runoff and routed it through HRR as our data assimilation scheme (Section 2.6) interacts within HRR sequentially and allows us to incorporate hillslope delay; the data assimilation updates a model state when BAM discharge is available, and HRR forwards it from the updated state. Thus, we route water excess across hillslopes and then through the channels to arrive at discharge. Hereafter, we call this routed GRADES runoff as the "baseline," as it is serving as the status quo in this manuscript. We remind readers that we are routing this globally available runoff product without calibration of routing parameters.

2.6. Integrating BAM and GHMs

There are multiple ways of adding BAM flows to the globally forced HRR representation of the Missouri, ranging from direct insertion into the routing scheme to more sophisticated data assimilation schemes that consider uncertainties in both BAM and the original baseline. We first attempted direct insertion of McFLI discharges into the routing model; that is, we ignored uncertainties and accepted BAM flows as true. In the direct insertion method, we replaced daily mean discharge values with BAM flows. While BAM is inherently an instantaneous flow for the exact instant a Landsat overpass was made, for the sake of simplicity we treated BAM flows as daily representative discharge in this study, following previous literature (e.g., Durand et al., 2016). This was computationally efficient and produced slightly improved discharges, but it was not a satisfying solution to the problem because McFLI accuracy is sometimes poor and is also variable in space and time (Durand et al., 2016; Feng et al., 2019). In the case of BAM–AMHG, the algorithm is most accurate when widths are sufficiently variable with respect to discharge (Feng et al., 2019). By directly inserting BAM, we were not considering the quality of the signal passed to the routing model. Furthermore, BAM uncertainty is spatiotemporally dynamic, and a direct insertion method cannot account for these changing accuracies and uncertainties.

Data assimilation techniques, by contrast, are built specifically to account for explicit uncertainties in space and time. There are various algorithms developed to implement data assimilation, and here we used the local ensemble transform Kalman filter (LETKF; Hunt et al., 2007). The LETKF was designed to reduce

computational costs of inverting large matrices which often appear in global models by only applying the analysis equations within a specified spatial patch. The LETKF localizes matrices into small chunks (local patches) and adopts a sequential root filter, thus significantly reducing computational cost, especially in large domains. Although the original implementation of LETKF in climate models assumes a grid system, we adapted the algorithm for a hydrological network application (similar to Revel et al., 2019). As rivers have network systems, flow information is only shared in hydrologically connected reaches, and it is thus misleading to construct a local patch without considering network information. Here, the algorithm tracks stream paths based on the network topology to construct local patches of hydrologically connected areas (5,000 km^2) for each reach. In this way, we effectively propagated observational information (i.e., BAM discharge) to hydrologically connected areas, rather than by Euclidean distance.

Given the spatial sparseness of our observations (BAM discharge), we included only discharge in our state vector for assimilation. Note that in our routing scheme (Muskingum–Cunge per HRR), discharge is a prognostic variable; therefore, the assimilated state was used as an initial condition for a next integration after a filtering step. The temporal availability of BAM discharges varies across BAM reaches due to satellite overpass geometry and weather conditions, thus at a given date assimilation happens at a fraction of all BAM reaches simultaneously. We assumed the underlying discharge distribution was log-normal, consistent with BAM's explicit Bayesian formulation. We can thus derive both the assimilated posterior mean and standard deviation from an underlying Gaussian distribution by the nature of log-normal flows.

Uncertainty in observations are derived directly from the BAM algorithm as it explicitly generates uncertainties via Bayesian inference. Background uncertainties were represented by generating an ensemble of model realizations around the baseline (used as the mean). Although we acknowledge additional sources of uncertainty such as model physics, parameters, and boundary conditions (Cloke et al., 2011; Her et al., 2019), we chose the input forcing (i.e., GRADES runoff) as our main source of uncertainty following previous studies (e.g., Andreadis et al., 2007). Assuming that interannual variability was larger than or equal to the true uncertainty, 20 perturbed inputs were generated by the following steps. We first shuffled runoff randomly by year for each ensemble, resulting in a runoff forcing data set with 20 ensemble members composed of GRADES runoffs but in random temporal order. Thus, we have 20 members for every reach and timestep. We then calculated ensemble means and ratio of each ensemble member to the mean. Finally, we multiply these ratios with the baseline to generate ensembles with same variance as the shuffled years but centered on the original values from GRADES. We tested multiple methods of generating uncertainty for our routed GRADES runoff, including parametric error models using runoff characteristic uncertainty from Beck et al. (2015) or using other predetermined parameters, but these alternate methods had too little uncertainty for effective assimilation.

2.7. Methods Summary

To summarize what we have detailed above, our study proceeded as follows. (1) We used GRADES runoff as an input to an uncalibrated routing model for testing remote-sensing improvements in discharge estimation. (2) We used MERIT HYDRO to derive a river network and HRR routing model parameters for each of our three cases. (3) We independently estimated river width from Landsat for ~1 M cross sections, and then quality controlled these widths and converted remaining measurements to discharge using the BAM Algorithm. (4) We generated an uncertainty for our forcing runoff via an open-loop ensemble, and then (5) we combined BAM discharges and the GHM together via the LETKF data assimilation algorithm. Nowhere have we invoked calibration data, and all 403 available gauges in this study were all held for validation.

Finally, we discuss how we will evaluate our results. Evaluation is not as straightforward as in previous McFLI studies. For example, Durand et al. (2016) propose a set of metrics used to evaluate estimated discharge produced without prior knowledge against a known hydrograph, and this suite produces a full picture of how well discharge has been estimated in that reach. This approach works on a river-by-river study with single reaches used for analysis (as in Durand et al., 2016), but in this study, we have 403 validation gauges. We are interested in total performance for the basin, but results are more accurate at some gauges than at others. We do have a gauge at the outlet that reflects the entire system, but focusing

Table 1
Equations for the Metrics Used in This Study, Where Q^{sim} Is Predicted Discharge, Q^{obs} Is Observed Discharge, σ Is Standard Deviation, and $\overline{Q}$ Is Mean Discharge

Metrics	Equation
NRMSE	$\frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(Q_t^{sim}-Q_t^{obs}\right)^2}}{\overline{Q^{obs}}}$
NSE	$1-\frac{\sum_{t=1}^{N}\left(Q_t^{sim}-Q_t^{obs}\right)^2}{\sum_{t=1}^{mdl}\left(Q_t^{obs}-\overline{Q^{obs}}\right)^2}$
KGE	$1-\sqrt{(r-1)^2+\left(\frac{\sigma^{sim}}{\sigma^{obs}}-1\right)^2+\left(\frac{\overline{Q^{sim}}}{\overline{Q^{obs}}}-1\right)^2}$

only on the outlet could ignore large errors or excellent performance at smaller upstream sites. Therefore, we present a summary of the normalized root mean square error (NRMSE), Nash–Sutcliff efficiency (NSE), and Kling–Gupta efficiency (KGE) metrics across all reaches, using different means of summarizing to give as complete a picture of discharge estimation as possible. We used the posterior ensemble mean from the output of the LETKF as a representative value for comparison of our ensemble members following the previous studies (e.g., Andreadis et al., 2020). The NRMSE assesses the extent of errors in model predictions normalized by mean observed flow to compare reaches. The NSE represents whether our model predictions are better than using gauged mean flow at all times and exhibits a positive value when predictions perform better than mean flow (Table 1). The KGE, introduced by Gupta et al. (2009), merges correlation, variation, and volumetric bias terms. We validate discharge daily to give as honest a picture as possible of our results, and our discussions include the utility of each metric.

3. Results

Figure 2a shows NSE scores validated through 403 USGS gauge data sets. Overall, out of 403 gauge stations, we had large improvement (>1 NSE change) at 78 reaches, fair improvement ($0.1 <$ NSE change ≤ 1) at 123 reaches, almost no improvement or impairment ($-0.1 <$ NSE change ≤ 0.1) at 174 reaches, and degradation (≤ -0.1 NSE change) at 28 reaches. Thus 50% (201/403) reaches experienced improvement by more than 0.1 NSE after assimilation, but we assert that performance was even better than this. The relationship is roughly linear ($r^2 = 0.71$) with a slope of -1.68: for every additional unit of initial NSE error in the baseline, we improved NSE by 1.68. Reaches which originally had poor NSE experienced large improvements, whereas reaches which already had good NSE in the baseline had less improvement. Figure 2b shows a histogram of NSE change for reaches whose baseline NSE is less than 0 (225/403 reaches), and Figure 2d shows reaches with baseline NSE greater than or equal to 0 (178/403 reaches). We see major improvements in Figure 2b (NSE $<$ 0), and importantly, we see we have "done no harm" in Figure 2d (NSE $\geq$ 0). The median of improvement in Figure 2b was 0.50 NSE, the mode of the histogram was 0.00, and 207 out of these 225 reaches were improved ($\geq$0 NSE change) by the addition of McFLI, which is a substantial improvement. For those initially well-estimated reaches (Figure 2d), we see essentially no change in NSE (mode = 0.00, median = -0.01), aligning precisely with the theory of data assimilation when assimilating uncertain observations. Figure 2c shows the cumulative distribution function (CDF) of NSE at all 403 gauge stations, and Figure 2e shows the CDF at 225 stations where initial performance in baseline was poor. Again, we see a shift between the CDFs in Figures 2c and 2e toward improvement, especially in the reaches whose original NSE was less than 0. Generally positive NSE indicates that model prediction is better than the gauged mean annual flow, and 49 reaches (out of a possible 225%, 22%) moved from negative NSE in the baseline to positive NSE after assimilation. Figure 2f shows a correlation between drainage area and improvement for 225 reaches where baselines NSE was negative. The correlation coefficient was 0.44, suggesting that larger subcatchments (i.e., larger downstream reaches) were more likely to improve among the ill-modeled baseline rivers. The rest of Figure 2 establishes that the dominant control on improvement, however, is still the accuracy of the baseline.

We also investigate these results spatially. Figure 3 shows the spatial distribution of baseline NSE and NSE improvements over the Missouri basin, plotting NSE improvements from Figure 2 at each gauge. Reaches with BAM discharges are marked as thick black lines. We saw major improvements (dark blue) around BAM reaches, especially upstream. Following from Figure 2a, these results are also an inverse way of interrogating baseline accuracy: there is less improvement at reaches that are already well modeled in the baseline (Figure 3a). We note also that many reaches are improved despite not having BAM observations nearby (especially around 44°N, −104°E). These reaches improved via our use of the local patch, which transfers information spatially throughout connected reaches.

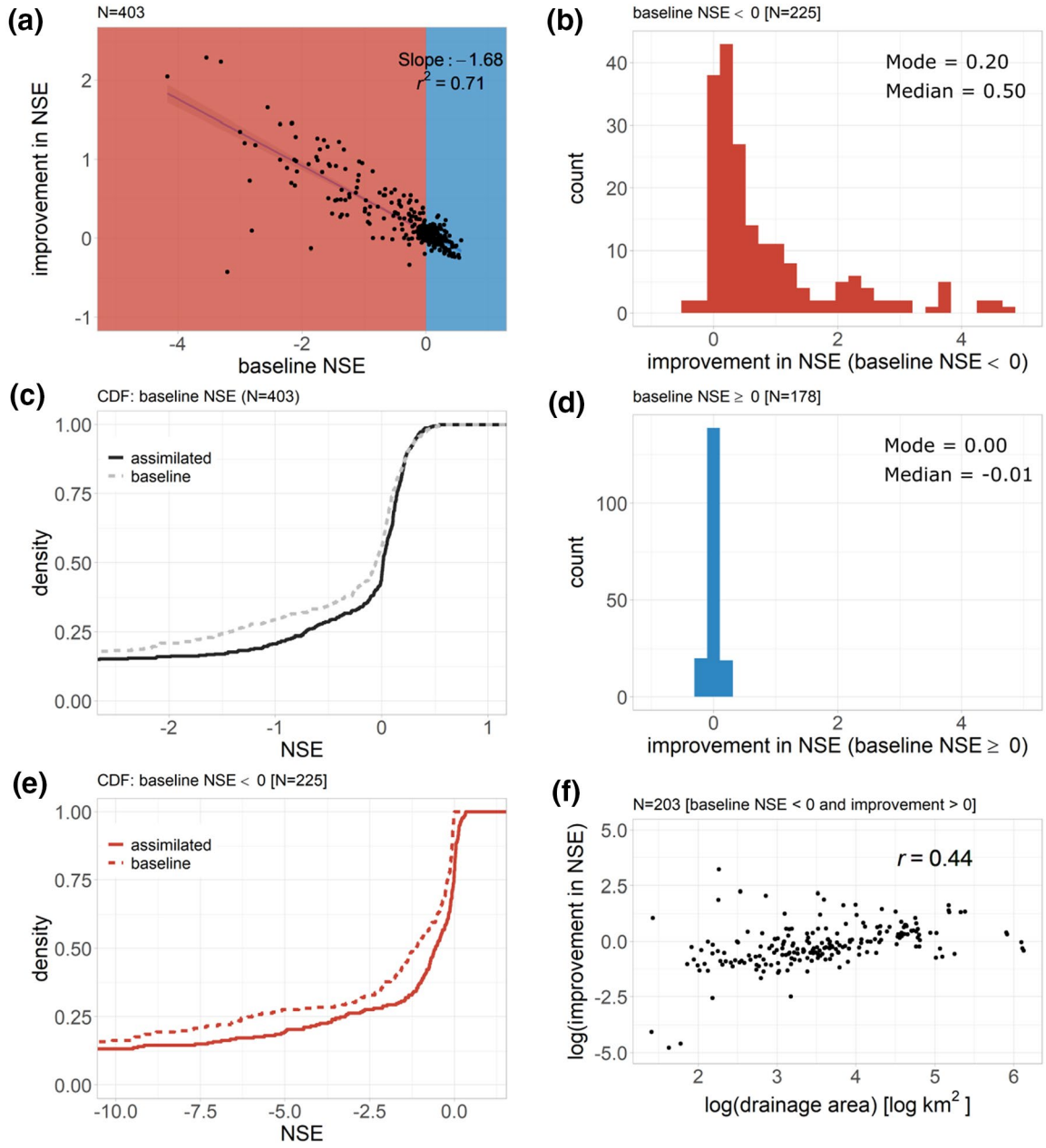

Figure 2. Improvement in NSE, where positive numbers indicate improvement. (a) Negative relationship between baseline NSE and improvement in NSE (assimilated – baseline). (b) Histogram of improvement in NSE for the reaches whose NSE was less than 0 in the baseline (n = 225/403 reaches). (c) CDF of assimilated and baseline NSE for all reaches (403). (d) Same as (b), but for the reaches whose NSE was greater than or equal to 0 in baseline (n = 178/403 reaches). The range of x axis is the same as (b). (e) Same as (c) but for reaches whose baseline NSE was less than 0. (f) Scatter plot of drainage area and improvement in NSE in $\log_{10}$–$\log_{10}$ space for reaches where original NSE was negative. Note that negative improvement was masked out. NSE, Nash-Sutcliffe efficiency; CDF, cumulative distribution function.

The trends in Figure 2 were similar for our other validation metrics. We report here only on the cases of baseline NSE < 0 for conciseness, although we show all other results in Figure 4. Figures 4c and 4d are histograms of improvement for KGE and NRMSE, respectively. We again see the major improvement in NRMSE: the mode of this histogram was 20% improvement, and its median was 28%, meaning that we improved baseline by at least 28% in normalized error in more than half of the reaches. Overall, we had positive improvement in 204/225 stations using NRMSE. For KGE, the improvement was more moderate, and we saw 54/225 reaches impaired by our data assimilation. However, while KGE improvement

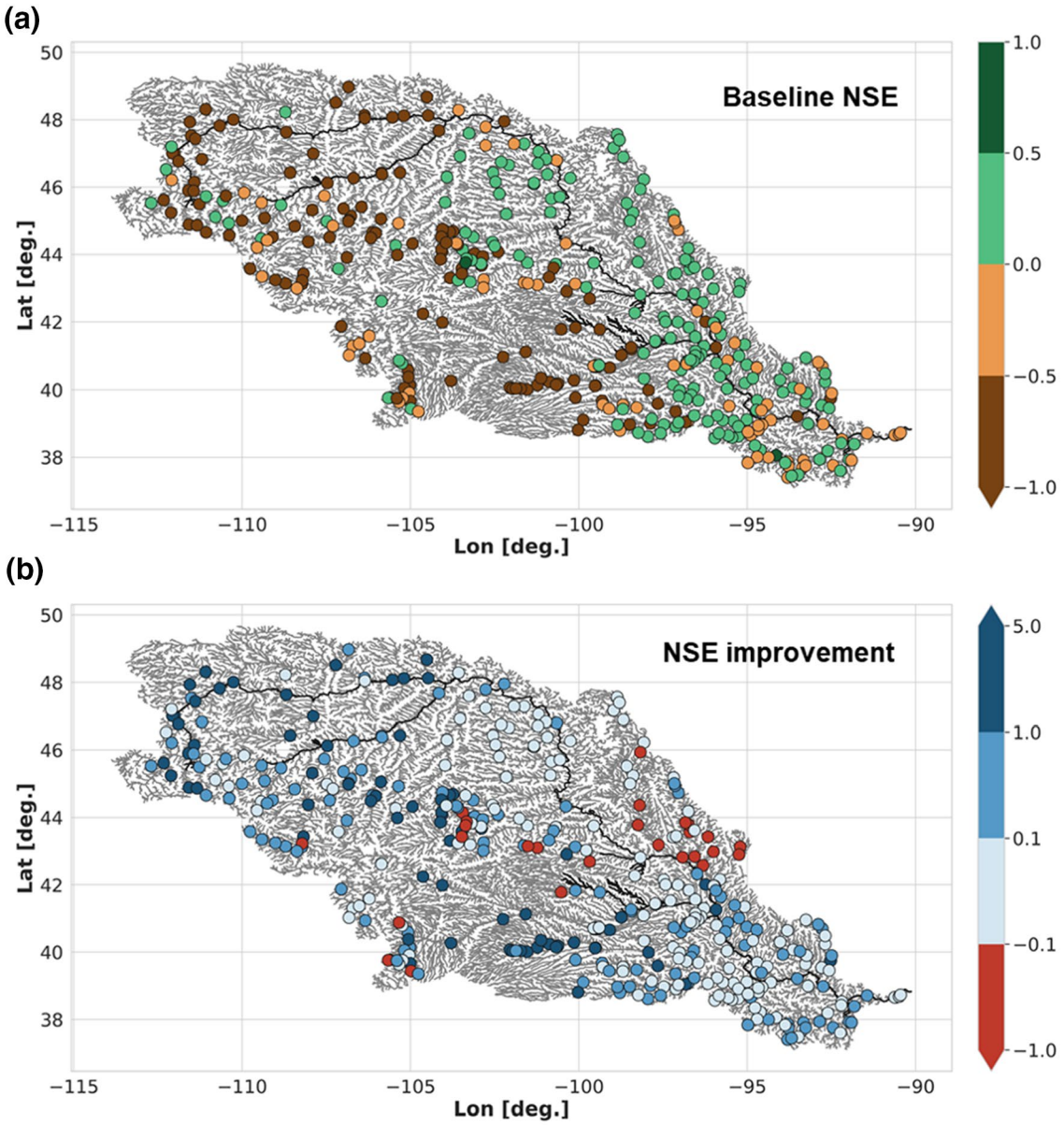

Figure 3. Spatial distribution of the (a) baseline NSE and (b) NSE improvements over Missouri basin. Thick black lines show BAM reaches. We see most large improvement (dark blue) upstream, where original NSEs were poor. The downstream reaches were well modeled in the baseline already, and we made almost no improvement but did no harm. Improvement in the southwest of the basin is a result of our large local patch (5,000 km^2) that propagated BAM information upstream. BAM, Bayesian At Many-stations Hydraulic Geometry Manning; NSE, Nash–Sutcliffe efficiency.

was closer to a zero-mean normal distribution compared to the other two metrics (which were skewed toward improvement) and the mode was 0 KGE, 171/225 reaches had positive improvement, and median improvement was 0.23 KGE. Figures 4a and 4b summarize those improvements with CDFs. Ultimately, using KGE as a metric, we improved (≥0 KGE change) baseline discharge at 54% of 403 validation gauge stations without compromising the performance at other reaches. Using NSE and NRMSE, we improved at 71% (≥0 NSE/NRMSE change) of the total gauge stations without impairing original performance at rest of the reaches.

Figure 5 shows hydrographs at several reaches for 2010. The columns of Figure 5 are divided by the baseline accuracy (negative/positive), whereas the rows correspond to small/medium/large drainage areas. Note that

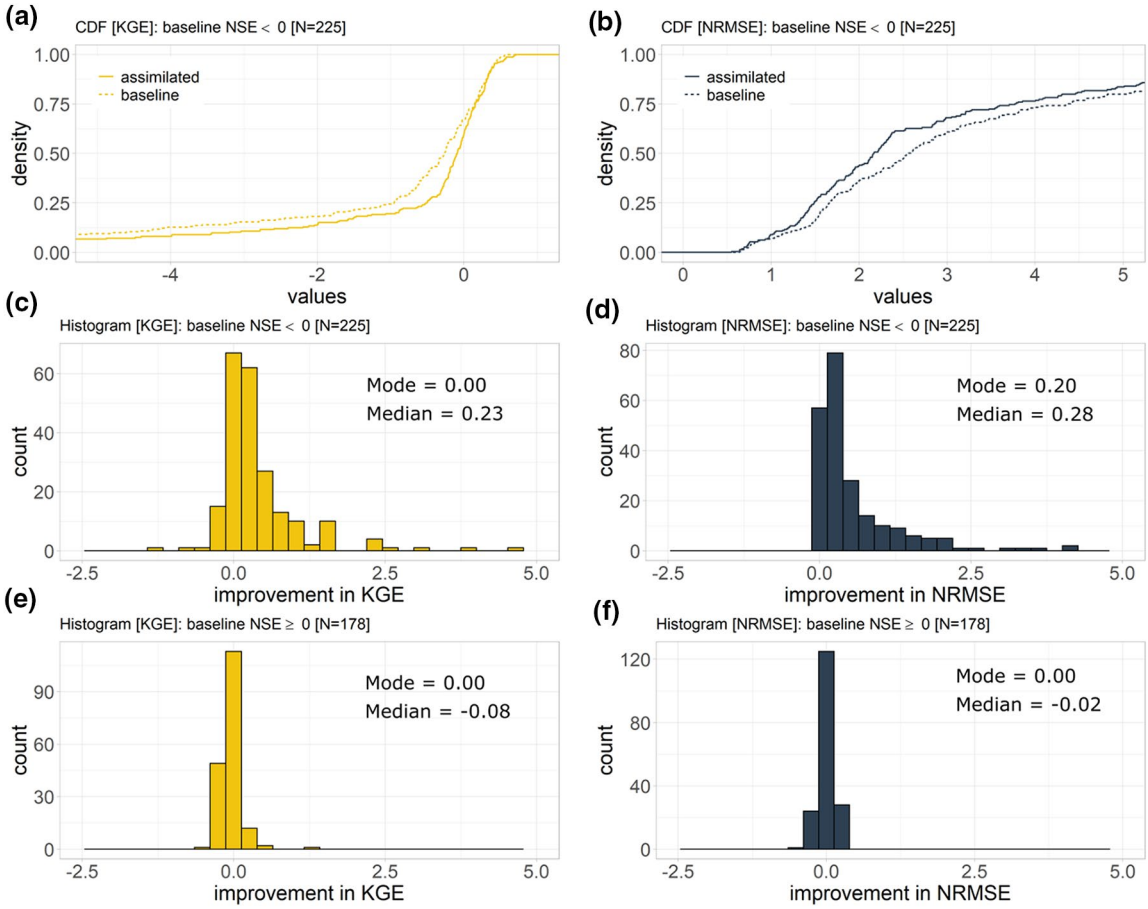

Figure 4. Improvement performance in NSE, KGE, and NRMSE. (a) CDF of KGE for reaches whose baseline NSE was less than 0 (225/403). (b) Same as (a) but for NRMSE. (c) Histogram of improvement in KGE for the reaches whose NSE was less than 0 in the baseline. (d) Same as (c), but for improvement in NRMSE. (e) Same as (c), but for the reaches whose NSE was greater than or equal to 0 in baseline. The range of *x* axis is the same as (b). (f) Same as (e), but for improvement in NRMSE. NSE, Nash-Sutcliffe efficiency; KGE, Kling–Gupta efficiency; NRMSE, normalized root mean square error; CDF, cumulative distribution function.

we did not have a positive baseline NSE in the large drainage area case. Panel (a) (negative baseline NSE/small drainage area) shows improved discharges even without BAM observations: information is passed to this reach via the local patch and manifests as sudden peaks without BAM observations present. These peaks are pulling the hydrograph toward the truth (remembering that the assimilation algorithm does not know the truth), which represents a best case scenario for our setup and justifies use of the local patch. These corrections, however, quickly attenuated as there are few BAM corrected reaches near enough in time or space to sustain them. In panel (c) (negative baseline NSE/medium drainage area), we see uncertain BAM values (which were ultimately inaccurate) were rejected by data assimilation, whereas in July data assimilation trusted the BAM value, which was certain (and ultimately accurate). This represents another excellent outcome where the dynamic acceptance of BAM uncertainty within the same reach is apparent. In panel (e) (negative baseline NSE/large drainage area), we see a significantly different hydrograph between the baseline and the assimilated case. Improvements again come from the local patch/accumulated routing corrections, but unlike panel (a), improvements resulted in longer duration of corrections especially in July, October, and November and explain why discharge improvements improve with drainage area in Figure 2f (correlation coefficient *r* between improvement and drainage area is 0.4). However, panels (b) and (d) show when our methods have produced no change or made flows worse, consistent with Figure 2. Panel (b) represents a well-modeled hydrograph that was largely unchanged by data assimilation, and there continues to be a missing peak flow in July. Panel (d) shows how inaccurate BAM data can pull the hydrograph away

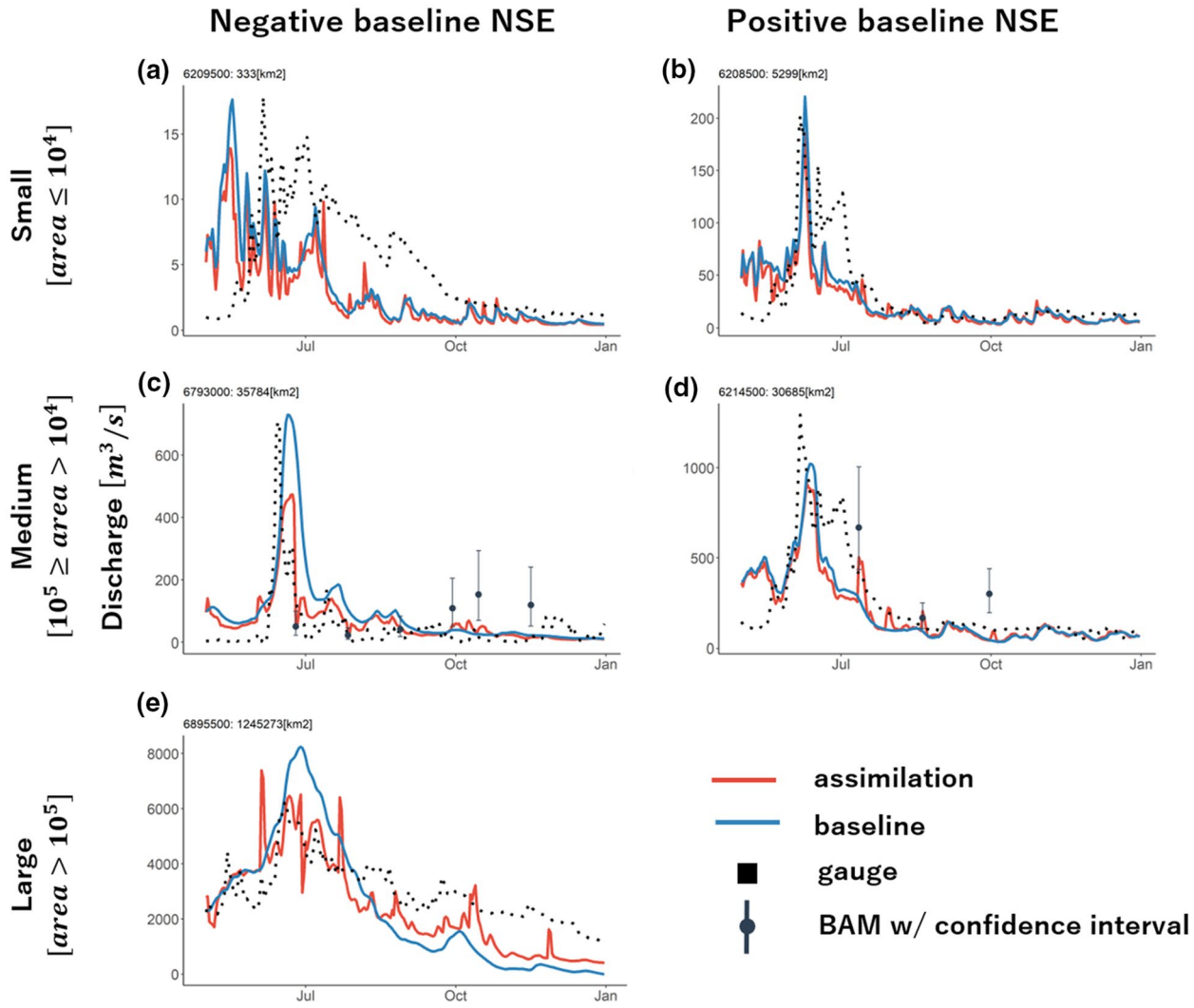

Figure 5. Hydrographs in 2010, chosen to exemplify BAM improvements via BAM, improvements via local patch, degradation due to BAM, and lack of improvement. Columns correspond to baseline NSE score (positive/negative), and rows are divided by drainage area size. We defined small, medium, an large drainage area by area in km^2, where $10{,}000 \geq$ small $> 0, 100{,}000 \geq$ medium $> 10{,}000$, and large $> 100{,}000$, respectively. Note that whiskers in BAM values represent confidence intervals. We see the data assimilation and local patch working as intended: panels (a), (c), and (e) either move toward BAM flows for a more accurate hydrograph or are moved toward a more accurate hydrograph by the local patch. Uncertain and inaccurate BAM flows (e.g., panel (c)) are rejected, but certain and inaccurate flows (panel (d)) are largely retained. BAM, Bayesian At Many-stations Hydraulic Geometry Manning; NSE, Nash–Sutcliffe efficiency.

from the truth (again, remembering that the gauge is imposed only for validation). In this case, a short duration correction prevented an even worse performance.

4. Discussion

4.1. McFLI, GHMs, and Data Assimilation

Existing satellite data sets like Landsat, AMSR-E, and MODIS have provided global water extent observations for at least a decade (and multiple decades for Landsat). Other data sets, like Sentinel-2 and cubesats are available more recently with higher revisit time and spatial resolution compared to Landsat. This study uses McFLI to estimate discharge directly from these satellite data and assimilates the derived discharge into a GHM in a computationally efficient manner suitable for global use. We assimilated BAM discharges

to produce daily flows across nearly 29,000 reaches, marking the first time that McFLI-derived discharges have been applied beyond the single-reach level. This assimilation improved (>0 NSE change) on the best available global discharge estimates at 71% of 403 gauge stations, and at 92% of stations (207/225) whose initial modeled discharge was poor. The median improvement for these 225 reaches was 0.50 improvement in NSE, 28% in NRMSE, and 0.23 in KGE. We thus assert that we have achieved our objectives, and that McFLI and GHMs as merged here are more than the sum of their parts by improving poor initial hydrographs (NSE < 0) and not degrading initially good hydrographs.

Global discharge data sets (e.g., Döll et al., 2003; Lin et al., 2019; Oki & Sud 1998; Qian et al., 2006; Sperna Weiland et al., 2010) provide historical insight into terrestrial hydrology and priors for future predictions. More specifically, these data sets can be used for water resource management (Busker et al., 2019), deriving climatological flood thresholds and providing initial conditions for flood forecasts (Alfieri et al., 2013), or priors for remote-sensing algorithms, including McFLI. However, global gauge data sets used to produce these global discharge products are still spatiotemporally limited and in decline, posing challenges in basins where gauge observations are sparse or unavailable. In this study, we demonstrated that we could indeed improve GHM discharge baselines by the addition of Landsat information through McFLI. We emphasize that these improvements were achieved using existing data sets and state-of-the-art GRADES discharge without gauge calibration of BAM flows or our routing model, and that our results are globally scalable.

Previous studies (Feng et al., 2019; Hagemann et al., 2017) have shown that BAM discharge estimates can contain large errors. In particular, if a reach does not have sufficient width variations observable from Landsat, BAM cannot correctly simulate discharge dynamics. While this problem is sometimes unavoidable, BAM does provide an explicit uncertainty estimate for its flows, estimated from Bayesian inference. However, previous BAM studies tended to only focus on the mean of BAM's posterior distribution, leaving valuable information underused. The data assimilation here was designed to leverage these distributions and reject those BAM discharge estimates with high uncertainty. Indeed, Figures 2d, 4e, and 4f clearly show a large peak of improvement at "0," suggesting that our method correctly retained the prior estimates that were accurate by explicitly accounting for observational uncertainty. That is, those data correspond to BAM estimates with large uncertainty.

Although we improved our baseline by incorporating BAM discharge, as a first incorporation of McFLI and GHMs we acknowledge that the approach we took here is still preliminary and further methodological improvement should be performed. While we did see improvements by considering uncertainty in BAM, this process is not always foolproof. There were many points in individual hydrographs where BAM was confident (i.e., had a low uncertainty) in an incorrect discharge. Because gauges are not invoked, the assimilation algorithm gravitates with uncertainty, not error. Therefore, most of our postassimilation error can be attributed to this false-confidence problem. This problem could be addressed with techniques such as adaptive covariance inflation (Anderson, 2007), which changes inflation of covariance matrix in time via hierarchical Bayesian model. The relationship between BAM accuracies and uncertainties is also not well studied, and further quantification and improvement of BAM behavior would improve the results presented here by providing better uncertainty. While we performed data assimilation with discharge estimated via BAM, we could also use Landsat and Sentinel-2 river widths directly in the assimilation, following the logic of Oubanas et al. (2018) in treating satellite observations as assimilation variables explicitly by including them in the state vector. Doing so would require a more complex routing model capable of calculating time-varying width with a more physically explicit treatment of channels and floodplains. Some of these models are already available at the global scale with a feasible computational time (e.g., Yamazaki et al., 2011). Using widths directly would omit McFLI, which on one hand would preserve the primary data and not add uncertainty from McFLI physics, but on the other hand would lose the rich prior information on discharge available in McFLI. Comparison between width and McFLI assimilation should be performed in a future study. We should also note that upcoming SWOT observations will provide river slopes and water surface elevations in addition to widths. Using these additional dimensions should constrain BAM better and result in further improvement via data assimilation (Durand et al., 2016).

Improvements in NSE and NRMSE (Figures 2b and 4c) were substantial and skewed heavily toward improved hydrographs after assimilating BAM. Improvements to KGE were rather moderate, a likely result of our data assimilation slightly degrading the correlation coefficient of the hydrograph (median −0.01 change

out of 403 stations). The correlation coefficient, however, does not consider volumetric bias but rather only the similarity of dynamics between two time series. BAM observations only cover a small fraction of total time series; therefore, hydrographs were corrected discontinuously, further worsening this correlation, even though absolute error or bias was improved. NSE and NRMSE are insensitive to this discontinuity, as they assess volumetric difference between prediction and observation, and no intertemporal variations are considered. This impairment of correlation coefficient may be eased after increasing spatiotemporal frequency of BAM, as more observations should make the hydrograph more consistently constrained, which warrants future investigations.

McFLI, especially AMHG-based McFLI used in this study (via BAM), poses a different way to inform GHMs of an actual river state from satellite observations. We were thus interested in whether this McFLI, which estimates discharge as a completely standalone process apart from GHMs, could inform our best available model estimates of discharge using existing data sets. While we do assert that our study is a first of its kind, it is hardly the first satellite/model fusion study. As stated, enormous amount of effort has been already poured into the integration of remote-sensing data sets and GHMs via data assimilation (e.g., Andreadis & Lettenmaier, 2006; Pan & Wood, 2006; Durand et al., 2008; Neal et al., 2009; R. E. Beighley et al., 2011; Biancamaria et al., 2011; Munier et al., 2015; Revilla-Romero et al., 2016; Emery et al., 2018, 2020; Maxwell et al., 2018; Oubanas et al., 2018; Häfliger et al., 2019; Oaida et al., 2019; Stampoulis et al., 2019; Yoon et al., 2012; Yoon & Beighley, 2015). The upcoming SWOT mission has accelerated this integration even further. SWOT examples include data assimilation for reservoir operations (Munier et al., 2015), estimating river bathymetry (Almeida et al., 2018; Durand et al., 2008; Hostache et al., 2015; Yoon et al., 2012), and acquiring better discharge/water-level estimation (Emery et al., 2020; Häfliger et al., 2019; Oubanas et al., 2018). Specific to discharge estimation, mock SWOT observations have been used to correct river water depths and widths in models requiring us a priori knowledge of river bed elevation and river bathymetry (e.g., Andreadis et al., 2007), to match anomalies of observation (e.g., Emery et al., 2020), and as assimilation variables directly (Oubanas et al., 2018). Our results thus add to this pantheon of work and suggest an alternate route for eventual SWOT merging of satellite observations and GHM globally.

4.2. Spatiotemporal Frequency of BAM

In this study, we processed more than 1 M width observations from Landsat satellites from 1984 to 2010 to estimate discharge estimates from BAM. Although this is a huge amount of data, after the filtering of clouds and other quality controls we only had about 40 days of width measurements for each reach, or approximately five images per year, per reach. As stated, there were 784/3,015 days (26%) when a BAM observation was available somewhere across the 40 BAM reaches, yet a given BAM reach only had observations at 1.4% of total 2002–2010 time series on average. Furthermore, the BAM reaches only cover 1.5% of the total 28,998 reaches. This quality control was necessary, as errors in width measurement propagate exponentially to errors in discharge in BAM (Gleason & Smith, 2014). Recent developments in GEE processing (e.g., Feng et al., 2019; Yang et al., 2020) have automated these procedures, making our study feasible in a global scale, although we have not tested whether or not our manual editing of GEE widths would be necessary following e.g., Yang et al. (2020). Despite this sparse coverage in space and time, we still saw substantial improvements in discharge simulations. We attribute this performance to the use of a routing model and the fact that the BAM reaches are located in the largest streams in their respective subwatersheds. By routing flows, improved discharges are propagated downstream and can correct poorly modeled baseline flows using upstream BAM observations. The other major factor in overcoming the sparseness of BAM data was data assimilation, specifically spreading the correction information via background covariance matrix.

Increasing the spatiotemporal frequency of BAM could improve results. Feng et al. (2019) showed that a fusion of existing satellites (Landsat/Sentinel-2/Planet) resulted in higher spatiotemporal frequency of observations, and SWOT in particular should improve this aspect, as SWOT's cloud-penetrating Ka-band radar observations and higher temporal repeat is designed to maximize these surface water observations. However, even after the multisatellite fusion or the launch of SWOT, we still may not have sufficient number of observations enough to constrain hydrographs from observations alone. Current data assimilation schemes only consider a spatial local patch. We tested the 4D-LETKF (Hunt et al., 2004; Kalnay & Yang, 2010) to

increase the number of observations used in an assimilation window, yet this was not substantially better than our presented results and further increased the computation burden. More sophisticated treatment of localization, spatiotemporal errors, or variance inflation (Emery et al., 2020) should also be tested. Recent studies (e.g., David et al., 2019; Y. Yang et al., 2019) have shown an effective way to propagate errors in discharge to runoff fields by inverting routing schemes. This results in longer retention of assimilation corrections (i.e., corrected discharge rapidly fades away after assimilation, whereas runoff tends to affect river discharge for a longer time). In conjunction with multisatellite fusion, these methodological interventions should be addressed in future McFLI assimilation studies.

4.3. Remote Sensing and Ungauged Basins

Despite the high quality of the GHM we used as prior information, there were many reaches with negative NSE values for baseline discharge. This lack of fidelity is partially explained by the fact that we validated against discharge daily, which is very difficult to simulate accurately at this scale, and especially because the effect of reservoirs is not directly included in our routing. Further, runoff was generated at 0.25° (~25 km), which may not represent small catchments in the basin, while errors can also propagate from atmospheric forcings. We also did not calibrate our routing consistent with ungauged basin information and imposed one binned parameter (Manning's n) and one global parameter (Hillslope friction) which certainly contributed to the errors in Figures 2 and 3 for the baseline. Despite these acknowledged sources of error, our baseline is representative of current global knowledge and methods for ungauged basins, and we have substantially improved upon the baseline with McFLI, and so we consider the methods and results presented here useful. Consequently, we have focused our discussion on improvements made via remote sensing, rather than on the original or final validation metric values (i.e., we argue that the final NSE, NRMSE, or KGE values here are less important than the remote-sensing improvements). Our absolute skill scores would undoubtedly improve if we allowed gauge data to calibrate both the routing and BAM.

Note that the GRADES-forced baseline represents a globally available estimate of daily river discharge through thousands of river reaches. While GRADES runoff itself is calibrated hydrologic GHM output, these estimates are uniformly available at the global scale and represent the current state of knowledge for many global rivers. We thus consider this baseline applicable in any ungauged basin, as we do not introduce any gauge data or local context to improve upon GRADES. We acknowledge that repeating this study in other basins would result in baselines of differing accuracies, largely driven by available calibration data to generate the forcing. However, the Missouri is extraordinarily well gauged by global standards, and thus the fact that our baseline runoff sometimes produces negative NSE values at the validation gauges when routed in this manner is indicative of the scope of the challenge for global ungauged basin hydrology. We remind readers that all of our results proceed from the baseline, and no gauge data were used to calibrate any facet of our results: our improvement is "over the top" of GRADES and applicable in any basin on earth. We chose the Missouri in part because of the 403 available validation gauges, as demonstrating this study in a truly ungauged basin would prevent us from validating it.

5. Conclusion

In this study, we assimilated McFLI discharges into an uncalibrated routing model forced with state-of-the-art global runoff estimates. We tested our assimilation scheme in the Missouri basin and routed flows daily over 28,998 reaches from 2002 to 2010. As a result, 286, 283, and 215 out of total 403 validation reaches had positive improvements in NSE, NRMSE, and KGE, respectively. Where initial modeled discharges were poor (NSE < 0, 225/403 reaches), 207/225 of these reaches improved substantially after assimilation, and median improvements in NSE, NRMSE, and KGE were 0.50%, 28%, and 0.23, respectively. Our results are even more promising when considering that the baseline discharge already performed well in 178/403 reaches (NSE greater than or equal to 0), and for those reaches we had either modest improvement or at worst maintained original accuracy. These results were achieved despite a small set of input remote-sensing observations: these only cover 1.5% of total reaches and 26% of days in the validation period.

It had not been previously shown whether McFLI discharge could add new hydrologic understanding to the current best available global discharge estimates. To our knowledge, this study is a first attempt to use McFLI discharges, which were estimated solely from satellite images, to inform our best understanding of hydrology at the continental scale using actual satellite data rather than an identical-twin experiment or synthetic remote-sensing data. While we solely used Landsat data here, multisatellite fusion with higher spatiotemporal frequency may achieve further improvements. Further, we have used only widths to estimate discharge from McFLI, yet previous research (e.g., Durand et al., 2016) clearly indicates that future SWOT McFLI retrieval using river width, height, and slope should improve these results even more. By coupling McFLI with the routing model via data assimilation, we overcame McFLI issues of downstream discontinuity and temporal/spatial sparseness and simultaneously overcame GHM's lack of quality primary input data. We argue that while future studies may further improve the approach pioneered here (including testing different McFLI algorithms and testing ingesting the original signal of width directly into a model without McFLI), we have provided a template for integrating McFLI and hydrologic models, and we have used real observations rather than synthetic remote sensing or an identical-twin experiment. This methodology could be deployed at a global scale in the near future and could potentially change our understanding of the water cycle in data-poor environments, especially after the launch of SWOT.

Data Availability Statement

All data in this study are publicly available and were accessed at the links given above in the text (USGS Gauge data set, http://usgs-r.github.io/dataRetrieval/#citing-dataretrieval; BAM, https://github.com/mark-wh/bamr), from the literature (GRADES, Lin et al., 2019; MERIT HYDRO, Yamazaki et al., 2019), or on the Google Earth Engine platform (Landsat images, https://earthengine.google.com/).

Acknowledgments
NSF CAREER grant 1748653, NASA New Investigator Program grant 80NSSC18K0741, and NASA SWOT Science Team grant NNX13AD96G awarded to C. J. Gleason supported Y. Ishitsuka, D. Feng, and C. J. Gleason, respectively. T. M. Pavelsky and G. H. Allen were supported by NASA Terrestrial Hydrology Program grant NNX13AD05G, and E. Beighley was supported by NASA grant NNX16AQ39G. G. H. Allen was partially supported by an NC Space Grant.

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., & Famiglietti, J. S. (2018). Global estimates of river flow wave travel times and implications for low-latency satellite data. *Geophysical Research Letters*, *45*, 7551–7560. https://doi.org/10.1029/2018GL077914

Allen, G. H., & Pavelsky, T. M. (2015). Patterns of river width and surface area revealed by the satellite-derived North American River Width data set. *Geophysical Research Letters*, *42*, 395–402. https://doi.org/10.1002/2014GL062764

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, *361*(6402), 585–588. https://doi.org/10.1126/science.aat0636

Allen, G. H., Pavelsky, T. M., Barefoot, E. A., Lamb, M. P., Butman, D., Tashie, A., & Gleason, C. J. (2018). Similarity of stream width distributions across headwater systems. *Nature Communications*, *9*, 610. https://doi.org/10.1038/s41467-018-02991-w

Almeida, T. G., Walker, D. T., & Warnock, A. M. (2018). Estimating river bathymetry from surface velocity observations using variational inverse modeling. *Journal of Atmospheric and Oceanic Technology*, *35*(1), 21–34. https://doi.org/10.1175/jtech-d-17-0075.1

Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, *59*, 210–224. https://doi.org/10.1111/j.1600-0870.2006.00216.x

Andreadis, K. M., Brinkerhoff, C. B., & Gleason, C. J. (2020). Constraining the assimilation of SWOT observations with hydraulic geometry relations. *Water Resources Research*, *56*, e2019WR026611. https://doi.org/10.1029/2019WR026611

Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., & Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, *34*, L10403. https://doi.org/10.1029/2007GL029721

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*, 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*(1), 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2015). Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology*, *16*(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., et al. (2019). MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

Beighley, R. E., Eggert, K. G., Dunne, T., He, Y., Gummadi, V., & Verdin, K. L. (2009). Simulating hydrologic and hydraulic processes throughout the Amazon river basin. *Hydrological Processes*, *23*, 1221–1235. https://doi.org/10.1002/hyp.7252

Beighley, R. E., Ray, R. L., He, Y., Lee, H., Schaller, L., Andreadis, K. M., & Shum, C. K. (2011). Comparing satellite derived precipitation datasets using the hillslope river routing (HRR) model in the Congo river basin. *Hydrological Processes*, *25*(20), 3216–3229. https://doi.org/10.1002/hyp.8045

Biancamaria, S., Hossain, F., & Lettenmaier, D. P. (2011). Forecasting transboundary river water elevations from space. *Geophysical Research Letters*, *38*, L11401. https://doi.org/10.1029/2011GL047290

Biancamaria, S., Lettenmaier, D., & Pavelsky, T. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, *37*(2), 307–337. https://doi.org/10.1007/s10712-015-9346-y

Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., & Garambois, P. (2018). Satellite remote sensing estimation of river discharge: Application to the Yukon river Alaska. *Journal of Hydrology*, *561*, 1000–1018. https://doi.org/10.1016/j.jhydrol.2018.04.005

Bjerklie, D. M., Lawrence Dingman, S., Vorosmarty, C. J., Bolster, C. H., & Congalton, R. G. (2003). Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, *278*, 17–38. https://doi.org/10.1016/S0022-1694(03)00129-X

Bjerklie, D. M., Moller, D., Smith, L. C., & Dingman, S. L. (2005). Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology*, *309*, 191–209. https://doi.org/10.1016/j.jhydrol.2004.11.022

Bonnema, M. G., Sikder, S., Hossain, F., Durand, M., Gleason, C. J., & Bjerklie, D. M. (2016). Benchmarking wide swath altimetry-based river discharge estimation algorithms for the Ganges river system. *Water Resources Research*, *52*, 2439–2461. https://doi.org/10.1002/2015WR017296

Brakenridge, G. R., Nghiem, S. V., Anderson, E., & Mic, R. (2007). Orbital microwave measurement of river discharge and ice status. *Water Resources Research 43*, W04405. https://doi.org/10.1029/2006WR005238

Brinkerhoff, C. B., Gleason, C. J., & Ostendorf, D. W. (2019). Reconciling at-a-station and at-many-stations hydraulic geometry through river-wide geomorphology. *Geophysical Research Letters*, *46*, 9637–9647. https://doi.org/10.1029/2019GL084529

Brown, C. M., Lund, J. R., Cai, X., Reed, P. M., Zagona, E. A., Ostfeld, A., et al. (2015). The future of water resources systems analysis: Toward a scientific framework for sustainable water management. *Water Resources Research*, *51*, 6110–6124. https://doi.org/10.1002/2015WR017114

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., & Cottam, A. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669–690. https://doi.org/10.5194/hess-23-669-2019

Cloke, H., Weisheimer, A., & Pappenberger, F. (2011). *Representing uncertainty in land surface hydrology: Fully coupled simulations with the ECMWF land surface scheme. Paper presented at ECMWF Workshop on Model Uncertainty, 20–24 June 2011*. Shinfield Park, Reading: ECMWF: https://www.ecmwf.int/node/8740

David, C. H., Hobbs, J. M., Turmon, M. J., Emery, C. M., Reager, J. T., & Famiglietti, J. S. (2019). Analytical propagation of runoff uncertainty into discharge uncertainty through a large river network. *Geophysical Research Letters*, *46*, 8102–8113. https://doi.org/10.1029/2019GL083342

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology*, *270*(1), 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4

Durand, M., Andreadis, K. M., Alsdorf, D. E., Lettenmaier, D. P., Moller, D., & Wilson, M. (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters*, *35*, L20401. https://doi.org/10.1029/2008GL034150

Durand, M., Fu, L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., & Esteban-Fernandez, D. (2010). The Surface Water and Ocean Topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, *98*(5), 766–779. https://doi.org/10.1109/JPROC.2010.2043031

Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., & Vilmin, L. (2016). An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, *52*, 4527–4549. https://doi.org/10.1002/2015WR018434

Ehalt Macedo, H., Beighley, R. E., David, C. H., & Reager, J. T. (2019). Using GRACE in a streamflow recession to determine drainable water storage in the Mississippi river basin. *Hydrology and Earth System Sciences*, *23*(8), 3269–3277. https://doi.org/10.5194/hess-23-3269-2019

Emery, C. M., Biancamaria, S., Boone, A., Ricci, S., Rochoux, M. C., Pedinotti, V., & David, C. H. (2020). Assimilation of wide-swath altimetry water elevation anomalies to correct large-scale river routing model parameters. *Hydrology and Earth System Sciences*, *24*(5), 2207–2233. http://dx.doi.org/10.5194/hess-24-2207-2020

Emery, C. M., Paris, A., Biancamaria, S., Boone, A., Calmant, S., Garambois, P.-A., & Santos da Silva, J. (2018). Large-scale hydrological model river storage and discharge correction using a satellite altimetry-based discharge product. *Hydrology and Earth System Sciences*, *22*, 2135–2162. https://doi.org/10.5194/hess-22-2135-2018

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., & Yamazaki, D. (2019). Hillslope hydrology in global change research and earth system modeling. *Water Resources Research*, *55*, 1737–1772. https://doi.org/10.1029/2018WR023903

Feng, D., Gleason, C. J., Yang, X., & Pavelsky, T. M. (2019). Comparing discharge estimates made via the BAM algorithm in high-order arctic rivers derived solely from optical CubeSat, Landsat, and Sentinel-2 data. *Water Resources Research*, *55*, 7753–7771. https://doi.org/10.1029/2019WR025599

Garambois, P.-A., & Monnier, J. (2015). Inference of effective river properties from remotely sensed observations of water surface. *Advances in Water Resources*, *79*, 103–120. https://doi.org/10.1016/j.advwatres.2015.02.007

Gleason, C., & Durand, M. (2020). Remote sensing of river discharge: A review and a framing for the discipline. *Remote Sensing*, *12*(7), 1107. https://doi.org/10.3390/rs12071107

Gleason, C. J., Garambois, P. A., & Durand, M. (2017). Tracking river flows from space. *Eos*, *98*. https://doi.org/10.1029/2017EO078085.

Gleason, C. J., & Hamdan, A. N. (2017). Crossing the (watershed) divide: Satellite data and the changing politics of international river basins. *The Geographical Journal*, *183*, 2–15. https://doi.org/10.1111/geoj.12155

Gleason, C. J., & Smith, L. C. (2014). Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, *111*, 4788–4791. https://doi.org/10.1073/pnas.1317606111

Gleason, C. J., Smith, L. C., & Lee, J. (2014). Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: Sensitivity to river form and optimization parameters. *Water Resources Research*, *50*, 9604–9619. https://doi.org/10.1002/2014WR016109

Gleason, C. J., Wada, Y., & Wang, J. (2018). A hybrid of optical remote sensing and hydrological modeling improves water balance estimation. *Journal of Advances in Modeling Earth Systems*, *10*, 2–17. https://doi.org/10.1002/2017MS000986

Gleason, C. J., & Wang, J. (2015). Theoretical basis for at-many-stations hydraulic geometry. *Geophysical Research Letters*, *42*, 7107–7114. https://doi.org/10.1002/2015GL064935

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Häfliger, V., Martin, E., Boone, A., Ricci, S., & Biancamaria, S. (2019). Assimilation of synthetic SWOT river depths in a regional hydrometeorological model. *Water*, *11*(1), 78. https://doi.org/10.3390/w11010078

Hagemann, M. W., Gleason, C. J., & Durand, M. T. (2017). BAM: Bayesian AMHG-Manning inference of discharge using remotely sensed stream width, slope, and height. *Water Resources Research*, *53*, 9692–9707. https://doi.org/10.1002/2017WR021626
Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*, 1191–1200. https://doi.org/10.1002/hyp.7794
Her, Y., Yoo, S., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: Multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports*, *9*(1), 4974. https://doi.org/10.1038/s41598-019-41334-7
Hostache, R., Matgen, P., Giustarini, L., Teferle, F. N., Tailliez, C., Iffly, J.-F., & Corato, G. (2015). A drifting GPS buoy for retrieving effective riverbed bathymetry. *Journal of Hydrology (Amsterdam)*, *520*, 397–406. https://doi.org/10.1016/j.jhydrol.2014.11.018
Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., & Zimin, A. V. (2004). Four-dimensional ensemble Kalman filtering. *Tellus*, *56*(4), 273–277. https://doi.org/10.1111/j.1600-0870.2004.00066.x
Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1), 112–126. https://doi.org/10.1016/j.physd.2006.11.008
Kalnay, E., & Yang, S. (2010). Accelerating the spin-up of ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *136*(651), 1644–1651. https://doi.org/10.1002/qj.652
Lee, H., Beighley, R. E., Alsdorf, D., Jung, H. C., Shum, C. K., Duan, J., et al. (2011). Characterization of terrestrial water dynamics in the Congo Basin using GRACE and satellite radar altimetry. *Remote Sensing of Environment*, *115*, 3530–3538. https://doi.org/10.1016/j.rse.2011.08.015
Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14415–14428. https://doi.org/10.1029/94JD00483
Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*, 6499–6516. https://doi.org/10.1029/2019WR025287
Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, L. R. (2013). A physically based runoff routing model for land surface and earth system models. *Journal of Hydrometeorology*, *14*(3), 808–828. https://doi.org/10.1175/JHM-D-12-015.1
Manning, R. (1891). On the flow of water in open channels and pipes. *Transactions of the Institution of Civil Engineers of Ireland XX*, 161–207.
Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. *Journal of Hydrology*, *560*, 127–140. https://doi.org/10.1016/j.jhydrol.2018.03.015
Munier, S., Polebistki, A., Brown, C., Belaud, G., & Lettenmaier, D. P. (2015). SWOT data assimilation for operational reservoir management on the upper Niger river basin. *Water Resources Research*, *51*, 554–575. https://doi.org/10.1002/2014WR016157
Neal, J., Schumann, G., Bates, P., Buytaert, W., Matgen, P., & Pappenberger, F. (2009). A data assimilation approach to discharge estimation from space. *Hydrological Processes*, *23*, 3641–3649. https://doi.org/10.1002/hyp.7518
Oaida, C. M., Reager, J. T., Andreadis, K. M., David, C. H., Levoe, S. R., Painter, T. H., & Famiglietti, J. S. (2019). A high-resolution data assimilation framework for snow water equivalent estimation across the western United States and validation with the airborne snow observatory. *Journal of Hydrometeorology*, *20*(3), 357–378. https://doi.org/10.1175/JHM-D-18-0009.1
Oki, T., & Sud, Y. C. (1998). Design of total runoff integrating pathways (TRIP)—A global river channel network. *Earth Interactions*, *2*(1), 1. https://doi.org/10.1175/1087-3562(1998)002<0001:DoTRIP>2.0.CO;2
Oubanas, H., Gejadze, I., Malaterre, P., Durand, M., Wei, R., Frasson, R. P. M., & Domeneghetti, A. (2018). Discharge estimation in ungauged basins through variational data assimilation: The potential of the SWOT mission. *Water Resources Research*, *54*, 2405–2423. https://doi.org/10.1002/2017WR021735
Pan, M., & Wood, E. F. (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *Journal of Hydrometeorology*, *7*, 534–547. https://doi.org/10.1175/JHM495.1
Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P., & Seyler, F. (2016). Stage-discharge rating curves based on satellite altimetry and modeled discharge in the amazon basin. *Water Resources Research*, *52*, 3787–3814. https://doi.org/10.1002/2014WR016618
Pavelsky, T. M. (2014). Using width-based rating curves from spatially discontinuous satellite imagery to monitor river discharge. *Hydrological Processes*, *28*, 3035–3040. https://doi.org/10.1002/hyp.10157
Qian, T., Dai, A., Trenberth, K. E., & Oleson, K. W. (2006). Simulation of global land surface conditions from 1948 to 2004. *Journal of Hydrometeorology*, *7*(5), 953–975. https://doi.org/10.1175/JHM540.1
Reichle, R. H., De Lannoy, G. J. M., Forman, B. A., Draper, C. S., & Liu, Q. (2014). Connecting satellite observations with water cycle variables through land data assimilation: Examples using the NASA GEOS-5 LDAS. *Surveys in Geophysics*, *35*, 577–606. https://doi.org/10.1007/s10712-013-9220-8
Revel, M., Ikeshima, D., Yamazaki, D., & Kanae, S. (2019). A physically based empirical localization method for assimilating synthetic SWOT observations of a continental-scale river: A case study in the congo basin. *Water*, *11*(4), 829. https://doi.org/10.3390/w11040829
Revilla-Romero, B., Wanders, N., Burek, P., Salamon, P., & de Roo, A. (2016). Integrating remotely sensed surface water extent into continental scale hydrology. *Journal of Hydrology*, *543*(Pt B), 659–670. https://doi.org/10.1016/j.jhydrol.2016.10.041
Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*, 651–659. https://doi.org/10.1038/s41586-018-0123-1
Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*, 999–1002. https://doi.org/10.1038/nature08238
Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., et al. (2016). Estimating continental river basin discharges using multiple remote sensing data sets. *Remote Sensing of Environment*, *179*, 36–53. https://doi.org/10.1016/j.rse.2016.03.019
Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., & Bierkens, M. F. P. (2010). The ability of a GCM-forced hydrological model to reproduce global discharge variability. *Hydrology and Earth System Sciences*, *14*(8), 1595–1621. https://doi.org/10.5194/hess-14-1595-2010
Stampoulis, D., Reager, J. T., David, C. H., Andreadis, K. M., Famiglietti, J. S., Farr, T. G., & Liu, Z. (2019). Model-data fusion of hydrologic simulations and GRACE terrestrial water storage observations to estimate changes in water table depth. *Advances in Water Resources*, *128*, 13–27. https://doi.org/10.1016/j.advwatres.2019.04.004
Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., et al. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018
Syed, T. H., Famiglietti, J. S., & Chambers, D. P. (2009). GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *Journal of Hydrometeorology*, *10*, 22–40. https://doi.org/10.1175/2008jhm993.1
Tarpanelli, A., Brocca, L., Lacava, T., Melone, F., Moramarco, T., Faruolo, M., et al. (2013). Toward the estimation of river discharge variations using MODIS data in ungauged basins. *Remote Sensing of Environment*, *136*, 47–55. https://doi.org/10.1016/j.rse.2013.04.010

Tarpanelli, A., Santi, E., Tourian, M. J., Filippucci, P., Amarnath, G., & Brocca, L. (2019). Daily river discharge estimates by merging satellite optical sensors and radar altimetry through artificial Neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 329–341. https://doi.org/10.1109/TGRS.2018.2854625

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*, 555–561. https://doi.org/10.1038/nature09440

Wada, Y., van Beek, L. P. H., & Bierkens, M. F. P. (2011). Modeling global water stress of the recent past: On the relative importance of trends in water demand and climate variability. *Hydrology and Earth System Sciences*, *15*(12), 3785–3808. https://doi.org/10.5194/hess-15-3785-2011

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., & Coulston, J. (2014). The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data. *Remote Sensing*, *6*(8), 7424–7441. https://doi.org/10.3390/rs6087424

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*, 3025–3033. https://doi.org/10.1080/01431160600589179

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*, 5053–5073. https://doi.org/10.1029/2019WR024873

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*, W04501. https://doi.org/10.1029/2010WR009726

Yang, Y., Lin, P., Fisher, C. K., Turmon, M., Hobbs, J., Emery, C. M., & Pan, M. (2019). Enhancing SWOT discharge assimilation through spatiotemporal correlations. *Remote Sensing of Environment*, *234*, 111450. https://doi.org/10.1016/j.rse.2019.111450

Yang, X., Pavelsky, T. M., Allen, G. H., & Donchyts, G. (2020). RivWidthCloud: An automated Google Earth Engine algorithm for river width extraction from remotely sensed imagery. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), 217–221. https://doi.org/10.1109/LGRS.2019.2920225

Yoon, Y., & Beighley, E. (2015). Simulating streamflow on regulated rivers using characteristic reservoir storage patterns derived from synthetic remote sensing data. *Hydrological Processes*, *29*(8), 2014–2026. https://doi.org/10.1002/hyp.10342

Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., & Alsdorf, D. E. (2012). Estimating river bathymetry from data assimilation of synthetic SWOT measurements. *Journal of Hydrology*, *464–465*, 363–375. https://doi.org/10.1016/j.jhydrol.2012.07.028

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, *548*, 552–568. https://doi.org/10.1016/j.jhydrol.2017.03.022

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028