Author Contributions:

Conceptualization: Renato Prata de Moraes Frasson, Michael T. Durand, Kevin Larnier, Colin Gleason, Mark Hagemann, Robert Dudley, David Bjerklie, Hind Oubanas, Pierre-André Garambois, Pierre-Olivier Malaterre, Tamlin M. Pavelsky
Data curation: Renato Prata de Moraes Frasson
Formal analysis: Renato Prata de Moraes Frasson, Michael T. Durand, Kevin Larnier, Colin Gleason, Hind Oubanas, Pierre-André Garambois,

Exploring the Factors Controlling the Error Characteristics of the Surface Water and Ocean Topography Mission Discharge Estimates

Renato Prata de Moraes Frasson[1,2] , Michael T. Durand[2,3] , Kevin Larnier[4,5], Colin Gleason[6] , Konstantinos M. Andreadis[6] , Mark Hagemann[7] , Robert Dudley[8] , David Bjerklie[8] , Hind Oubanas[9] , Pierre-André Garambois[10] , Pierre-Olivier Malaterre[9] , Peirong Lin[11] , Tamlin M. Pavelsky[12] , Jérôme Monnier[5], Craig B. Brinkerhoff[6] , and Cédric H. David[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [2]Byrd Polar and Climate Research Center, The Ohio State University, Columbus, OH, USA, [3]School of Earth Sciences, The Ohio State University, Columbus, OH, USA, [4]Space Department, CS Corporation, Toulouse, France, [5]INSA Toulouse - Math. Institute of Toulouse (IMT), Toulouse, France, [6]Department of Civil and Environmental Engineering, University of Massachusetts Amherst, Amherst, MA, USA, [7]EAB: Education Technology, Services, and Research, Richmond, VA, USA, [8]New England Water Science Center, U.S. Geological Survey, Augusta, ME, USA, [9]Irstea, Montpellier, France, [10]Irstea, Aix Marseille Université, RECOVER, Aix-en-Provence, France, [11]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [12]Department of Geological Sciences, University of North Carolina, Chapel Hill, NC, USA

Abstract The Surface Water and Ocean Topography (SWOT) satellite mission will measure river width, water surface elevation, and slope for rivers wider than 50–100 m. SWOT observations will enable estimation of river discharge by using simple flow laws such as the Manning-Strickler equation, complementing in situ streamgages. Several discharge inversion algorithms designed to compute unobserved flow law parameters (e.g., friction coefficient and bathymetry) have been proposed, but to date, a systematic assessment of factors controlling algorithm performance has not been conducted. Here, we assess the performance of the five algorithms that are expected to be used in the construction of the SWOT product. To perform this assessment, we used synthetic SWOT observations created with hydraulic model output corrupted with SWOT-like error. Prior information provided to the algorithms was purposefully limited to an estimate of mean annual flow (MAF), designed to produce a "worst case" benchmark. Prior MAF error was an important control on algorithm performance, but discharge estimates produced by the algorithms are less biased than the MAF; thus, the discharge algorithms improve on the prior. We show for the first time that accuracy and frequency of remote sensing observations are less important than prior bias, hydraulic variability among reaches, and flow law accuracy in governing discharge algorithm performance. The discharge errors and error sensitivities reported herein are a bounding benchmark, representing worst possible expected errors and error sensitivities. This study lays the groundwork to develop predictive power of algorithm performance, and thus map the global distribution of worst-case SWOT discharge accuracy.

Plain Language Summary Measurements of river flow are essential for the allocation of water resources, flood and drought forecast and mitigation efforts, and others. Access to local measurements is not ubiquitous and is particularly difficult for rivers flowing in remote locations or across country borders. Measurements taken by satellites such as the upcoming Surface Water and Ocean Topography (SWOT) mission will offer freely available global data and methods to estimate discharge using such data have been in development. We conducted a comprehensive assessment of the accuracy and precision of five SWOT discharge inversion algorithms under three conditions: (a) ideal, that is if the measurements were available once a day and contained no error; (b) with no measurement error but changing how frequently the measurements were taken, and (c) under different levels of measurement error. We found that the methods consistently improved over the initial estimates of discharge and we identified river hydraulic properties that increase the chances of successful parameter estimation. We also found that despite the use of very similar discharge equations, the subtle differences in equations

Pierre-Olivier Malaterre, Tamlin M. Pavelsky

Funding acquisition: Michael T. Durand

Methodology: Renato Prata de Moraes Frasson, Michael T. Durand, Kevin Larnier, Colin Gleason, Konstantinos M. Andreadis, Mark Hagemann, Robert Dudley, David Bjerklie, Hind Oubanas, Pierre-André Garambois, Pierre-Olivier Malaterre, Peirong Lin, Tamlin M. Pavelsky, Jérôme Monnier, Craig B. Brinkerhoff, Cédric H. David

Project Administration: Renato Prata de Moraes Frasson, Michael T. Durand

Software: Michael T. Durand, Kevin Larnier, Colin Gleason, Konstantinos M. Andreadis, Mark Hagemann, Robert Dudley, David Bjerklie, Pierre-André Garambois, Jérôme Monnier, Craig B. Brinkerhoff

Visualization: Renato Prata de Moraes Frasson, Michael T. Durand, Mark Hagemann, Craig B. Brinkerhoff

Writing – original draft: Renato Prata de Moraes Frasson, Michael T. Durand, Kevin Larnier, Colin Gleason, Konstantinos M. Andreadis, Mark Hagemann, Robert Dudley, David Bjerklie, Hind Oubanas, Pierre-André Garambois, Pierre-Olivier Malaterre, Peirong Lin, Tamlin M. Pavelsky, Jérôme Monnier, Craig B. Brinkerhoff, Cédric H. David

among the methods can be important. Finally, we found that at least two methods can work well with the expected amount of measurement error and frequency.

1. Introduction

As most global rivers are ungaged (Pavelsky et al., 2014), methods to complement the sparse network of in situ observations that rely on remotely sensed data have been developed (a review of such methods is presented by Gleason & Durand, 2020). The upcoming Surface Water and Ocean Topography (SWOT) satellite mission, with an expected launch in 2022, will augment the capabilities of the current constellation of satellites by providing an observation-based estimate of global discharge for rivers wider than 100 m and possibly as narrow as 50 m (e.g., Biancamaria et al., 2010, 2016; Pavelsky et al., 2014). The scientific impact of SWOT discharge will be determined, in large part, by its accuracy. In general, discharge accuracy will be governed by three factors (Yoon et al., 2016): accuracy of flow laws used to translate SWOT measurements into discharge, the accuracy of the unobservable flow law parameters estimated by the discharge inversion algorithms, and the accuracy of the SWOT observations of height, width, and slope.

Because parameter accuracy tends to dominate the discharge error budget (Yoon et al., 2016), understanding the factors that determine inversion algorithm accuracy is vital to the SWOT mission. Past studies (e.g., Bonnema et al., 2016; Durand et al., 2016; Oubanas, Gejadze, Malaterre, Durand, et al., 2018) have explored SWOT discharge inversion algorithm accuracy in ungaged basins but have not gone far enough in probing the controls on algorithm performance. Durand et al. (2016) tested five algorithms on a suite of 19 rivers using daily synthetic data created with hydraulic models, but without realistic levels of observation error, and found that at least one algorithm retrieved discharge with relative root mean square errors equal to or less than 35% on 14 out of the 16 non-braided rivers. While the authors identified specific algorithm deficiencies in some cases, they did not perform a systematic assessment of the controls on algorithm performance.

Tuozzolo, Lind, et al. (2019) showed that discharge algorithms are sensitive to the accuracy of prior information. Indeed, all discharge algorithms leverage at least some prior information, typically mean annual flow (Durand et al., 2016). Most algorithms treat this prior information in the Bayesian sense; its uncertainty is traded off against the information contained in the remote sensing observations, rather than simply accepting it as a "known" input (Hagemann et al., 2017). Some algorithms are also capable of incorporating additional prior information, such as expected discharge temporal variability and reach geomorphology. The exact configuration of which prior information will be used by which SWOT discharge inversion algorithms during the SWOT mission is still to be determined and is under examination by the science team. In the meantime, this prior information is being incorporated into algorithms (e.g., Andreadis et al., 2020; Brinkerhoff et al., 2020), and as a result, discharge accuracies are improving. Additionally, we expect that as more prior information is included, the sensitivity to any individual piece of prior information will likely decrease. Thus, studies such as Durand et al. (2016) as well as the current study, which use only a single prior (mean annual flow) represent an upper bound on both the benchmark error metrics and the sensitivity of discharge error to priors. In other words, the discharge errors and error sensitivities reported herein represent a bounding benchmark, representing worst possible expected errors and error sensitivities.

This study explores the controls on algorithm performance defined in terms of retrieved discharge error, contrasting these error metrics with quantifiable reach properties, flow law errors, and quality of prior information. Further, we test the sensitivity of SWOT discharge algorithms to temporal sampling and to observation error, with the latter being estimated with a parameterized model for SWOT observation uncertainty that takes into account the most important sources of error in SWOT observations of river height, width, and slope, such as the effects of layover, thermal noise, instrument resolution, and others. The benchmarked discharge inversion algorithms are summarized in Section 2, Section 3 describes the experimental protocol, the construction of the synthetic test cases for each of the three phases of our study: Phase 1: idealized daily measurements (Section 3.3), Phase 2: varying temporal sampling settings (Section 3.4), and Phase 3: varying degrees of measurement uncertainty (Section 3.5). We present the results of each phase in Section 4, which ends with a discussion of future and ongoing algorithm development (4.4), followed up by our conclusions in Section 5.
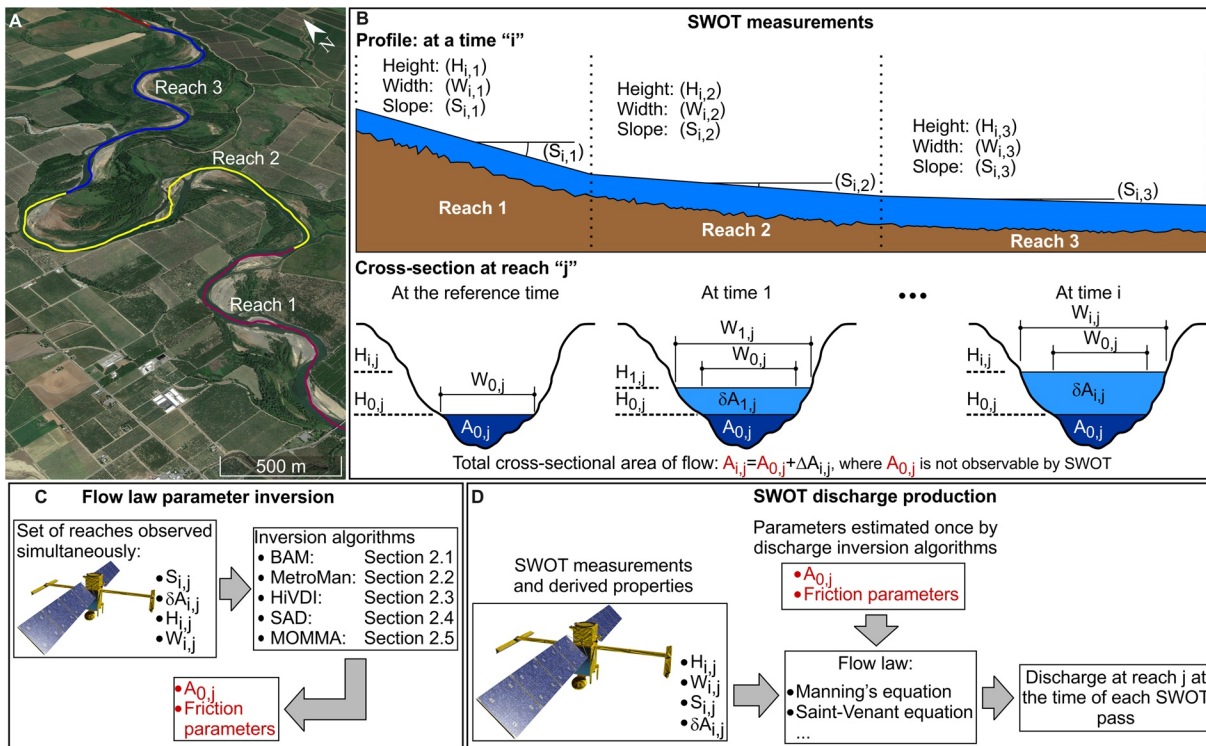
Figure 1. Panel a: Example of a set of contiguous reaches along the Sacramento River containing no visible tributaries, thus suitable for discharge parameter inversion based on the Mass conserved Flow Law Inversion paradigm. Panel b: Visualization of the Surface Water and Ocean Topography (SWOT) observations over a generic set of three reaches at a given time "i" (top) and bottom showing a cross-sectional cut made at reach "j" and the decomposition of the cross-sectional area into the unobservable fraction ($A_{0,j}$) and an observable fraction $\delta A_{i,j}$, derived from multitemporal observations of reach height and width. Panel c illustrates the flow law inversion step, when the unobservable fraction of the cross-sectional area and the friction parameters are estimated. Panel d: SWOT discharge production step which uses simple flow laws with parameters estimated by the different algorithms to produce discharge every time a new set of SWOT observations is collected. Base map data from Google, 2020.

2. SWOT Discharge Inversion Algorithms

SWOT will make high precision measurements of river surface slope (Desai, 2018; Esteban-Fernandez, 2013) averaged to nominal 10 km reaches (Frasson et al., 2017; Montazem et al., 2019) as well as water surface elevation and river width, thus measuring most of the terms in the Manning-Strickler flow law for discharge estimation (Bjerklie et al., 2003). However, SWOT will not be able to directly measure river bathymetry or quantify channel friction, which need to be indirectly estimated for the calculation of river discharge. Figure 1a shows a mass-conserved set of reaches, that is, a set of contiguous reaches containing no known tributaries. Figure 1b provides a conceptual list of SWOT measurements and the quantities that need to be estimated by discharge inversion algorithms. Figure 1d illustrates the discharge production process, which occurs every time a new set of SWOT observations is collected. As opposed to the flow law parameter inversion step (Figure 1c), which is only repeated at occasional reprocessing steps.

Several discharge inversion algorithms to estimate bathymetry and channel friction based on time series of SWOT measurements have been proposed. Here, we evaluate the five algorithms that are expected to run globally during the SWOT mission. The benchmarked algorithms are described in Sections 2.1–2.5 and summarized in Table 1. The evaluated algorithms can be grouped into three families: (a) Mass conserved Flow Law Inversion (McFLI) algorithms (Durand et al., 2014; Garambois & Monnier, 2015; Gleason & Durand, 2020; Gleason et al., 2017; Hagemann et al., 2017), (b) data assimilation algorithms (e.g., Andreadis et al., 2007; Durand et al., 2008; Garambois et al., 2020; Larnier et al., 2020; Monnier et al., 2018; Oubanas, Gejadze, Malaterre, Durand, et al., 2018; Oubanas, Gejadze, Malaterre, & Mercier, 2018); and (c) "big data" methods.

Table 1
Summary of the Discharge Inversion Algorithms

Algorithm	Theoretical basis	Applied variables from observation	Estimated variables	Variable estimation method
BAM	Manning flow resistance equation with constant n (Equation 1) and At-Many-stations-Hydraulic-Geometry (AMHG), Equation 2	Water surface width (W), slope (S), cross-sectional area anomaly (dA)	n, A_0, AMGH: b, Wc, Qc	Parameters optimized to preserve continuity between a set of reaches (assuming dQ/dx = 0) using Hamiltonian Bayesian inference.
MetroMan	Manning flow resistance with stage-dependent friction parameterization (Equation 3)	Water surface height (H), W, S, δA	n_a, b, A_0	Parameters optimized to preserve continuity between a set of reaches as defined by Equation 4, using the Metropolis algorithm.
HiVDI	1D Saint-Venant and Manning flow resistance with depth dependent flow resistance (Equation 5)	H, W, S, δA	a, b, A_0	Calibration of Manning-Strickler equation using discharge obtained by assimilation of SWOT measurements into the 1-D Saint Venant model.
SAD	Gradually varied flow equations (Equation 6) assuming friction slope that follows Manning-Strickler equation (Equation 7)	H, W, S	Q	Assimilation of SWOT observations into a steady-state, non-uniform hydraulic model using the Local Ensemble Transform Kalman Filter algorithm to estimate river discharge.
MOMMA	Manning flow resistance with stage-dependent friction parameterization (Equation 9)	H, W, S	n_b, H_b, B	Bankful height (Hb) is estimated by identifying breakpoints in the observed H versus W relationship. Height at 0 flow (B) is estimated by extrapolating the H versus W relationship to W = 0. Bankful flow friction is calibrated using the prior estimate of mean annual flow.

Notes. In this table, SWOT stands for the Surface Water and Ocean Topography mission. Description of the meaning the estimated variables can be found in each algorithm description: BAM in Section 2.1, MetroMan (2.2), HiVDI (2.3), SAD (2.4), MOMMA (2.5).
Abbreviations: BAM, Bayesian At-many-stations-hydraulic-geometry Manning; HiVDI, Hierarchical Variational Discharge Inference; MOMMA, Modified Optimized Manning Method Algorithm; SAD, SWOT Assimilated Discharge.

The McFLI methods evaluated here are the Metropolis-Manning (MetroMan) and the Bayesian At-many-stations-hydraulic-geometry Manning (BAM) algorithms. Both methods rely on invoking conservation of mass between a set of consecutive reaches to estimate channel friction parameters and a reference cross-sectional area of flow. However, the methods differ on the discretization of the continuity equation, the parameterization of friction, and the derivation of prior distributions for parameters.

Currently, two algorithms that rely on assimilating SWOT data into hydraulic models are expected to be used for discharge estimation at worldwide scales: the Hierarchical Variational Discharge Inference (HiV-DI) and the SWOT Assimilated Discharge (SAD). The HiVDI algorithm operates at a higher resolution scale (node scale), with height and width measurements aggregated into equally spaced points located every 200 m along the river centerline, which are assimilated into a 1D Saint-Venant model. In contrast to the SAD algorithm, which models flow at the SWOT reach scale using the gradually varied flow equations.

The final method that is currently expected to be operated by the SWOT mission is the Modified Optimized Manning Method Algorithm (MOMMA). MOMMA relies on extrapolation of geometric and hydraulic characteristics derived from large datasets in conjunction with SWOT data to estimate discharge. Figure 2 shows the general flow of information and contrasts the assumptions made by the five examined discharge inversion algorithms. The inversion algorithms evaluated in our study are described in Sections 2.1–2.5.

2.1. BAM: Bayesian At-Many-Stations-Hydraulic-Geometry Manning Algorithm

BAM is an evolution of the At-Many-stations-Hydraulic-Geometry (AMHG) algorithm initially proposed by Gleason and Smith (2014) and further explored by Gleason and Wang (2015) and by Brinkerhoff et al. (2019).
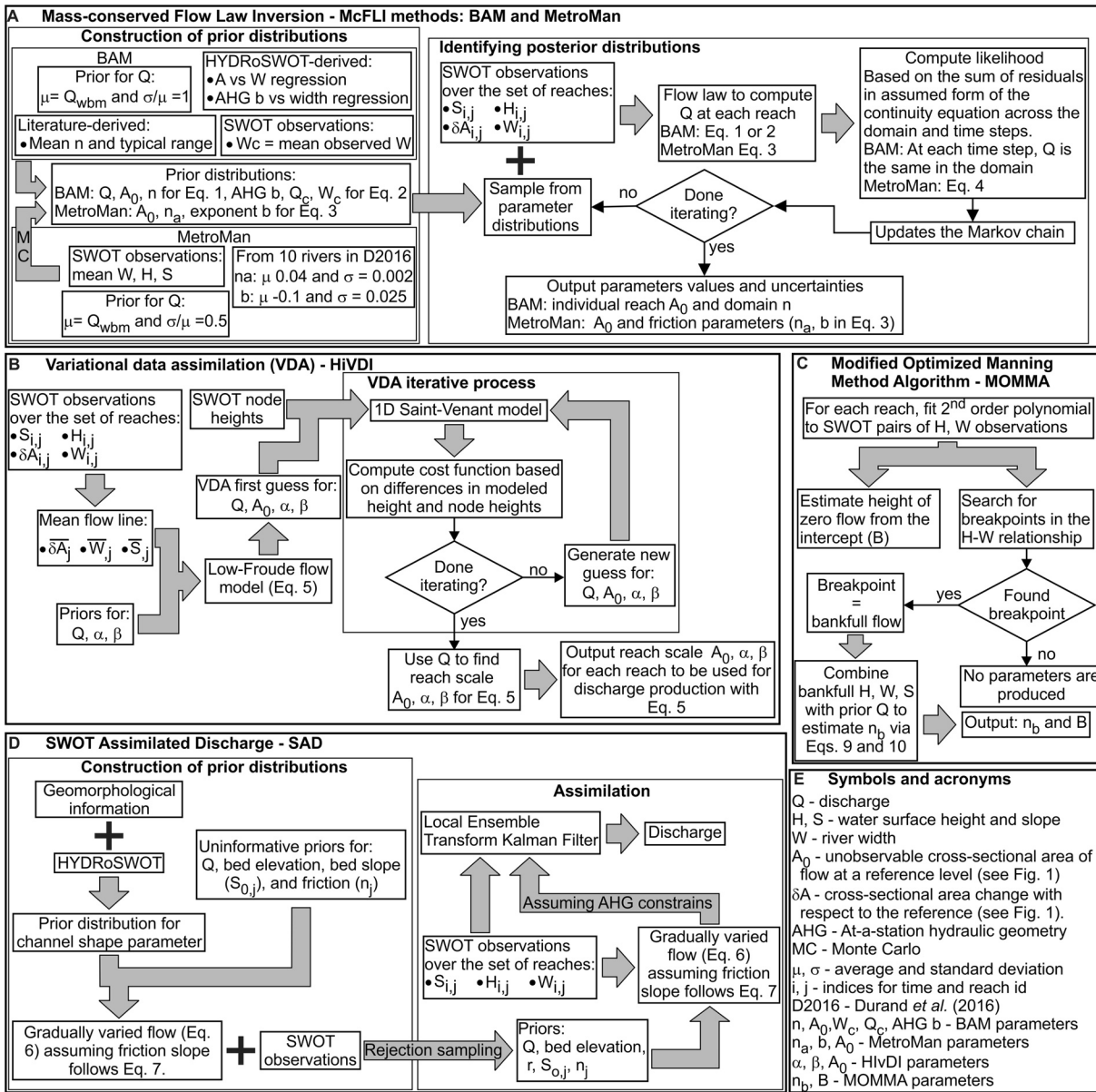
Figure 2. Summary of the inversion algorithms evaluated in the present study. Panel a: general flow of information used by the two McFLI methods: BAM (described in Section 2.1) and MetroMan (Section 2.2). Panel b: illustration of the Hierarchical Variational Discharge Inference (HiVDI described in Section 2.3). Panel c: The Modified optimized Manning Method Algorithm (MOMMA described in Section 2.5). Panel d: Surface Water and Ocean Topography (SWOT) Assimilated Discharge (SAD, described in Section 2.4). In this panel, Q stands for discharge, Q_{wbm} stands for the prior estimate of mean annual flow, $W_{i,j}$ for the river width at time i, reach j, $\delta A_{i,j}$ represents the cross-sectional area anomaly with respect to an arbitrary reference level (see Figure 1b), and $H_{i,j}$ and $S_{i,j}$ represent the water surface elevation and slope respectively. Panel e: list of symbols and previously non defined acronyms. Observations are at reach scale, which for the SWOT mission is approximately 10 km long with the exception of the VDA interactive process in panel B, which assimilates heights at the node scale, that is, equally spaced points located every 200 m along the river centerline.

AMHG is designed to estimate river discharge based on measurements of river width using hydraulic geometry relationships, which were originally observed by Leopold and Maddock (1953). Since SWOT also measures water surface elevation and slopes, Hagemann et al. (2017) added the capability to estimate discharge using the Manning-Strickler equation. The algorithm chooses between AMHG and Manning based on a width variability criterion (see Hagemann et al., 2017).

The version of BAM evaluated here uses the low Froude approximation (Garambois & Monnier, 2015; Larnier et al., 2020) for the estimation of the friction slope. This approximation, states that under low Froude

numbers, that is, less than 0.3, the friction slope can be approximated by the water surface slope. The implementation of the Manning-Strickler equation in BAM also assumes that the river width is much larger than the depth, which causes the wetted perimeter to reduce to the river width resulting in the following form of the Manning-Strickler and the AMHG equations:

$$Q_i = \frac{1}{n}\left(A_{0,j} + \delta A_{i,j}\right)^{5/3} W_{i,j}^{-2/3} \sqrt{S_{i,j}} \tag{1}$$

$$Q_i = \frac{W_{i,j}^{\frac{1}{b_j}}}{W_c} Q_c \tag{2}$$

where n is the friction coefficient, assumed to be constant for all reaches within the inversion set and invariable in time, $A_{0,j}$ represents a reference cross-sectional area, which is unique for each reach, $\delta A_{i,j}$ represents the cross-sectional area anomaly with respect to the selected reference for a reach j at time i, $W_{i,j}$ represents the top width of the reach j at time i, and $S_{i,j}$ represents the water surface slope. Additionally, W_c, Q_c, b_j, in Equation 2 represent the two AMHG parameters (Gleason & Wang, 2015) and the at-a-station hydraulic geometry coefficient, respectively. In Equations 1 and 2, the terms $W_{i,j}$ and $S_{i,j}$ are observed by SWOT and $\delta A_{i,j}$ can be calculated from the time series of SWOT-observed water surface elevation and width. However, n, Q_i, $A_{0,j}$, W_c, and Q_c need to be estimated by BAM.

Flow law parameter estimation is based on a Bayesian algorithm, which starts from prior distributions for discharge, Manning's n, reference cross-sectional area of flow, and the AMHG parameters and attempts to identify posterior distribution for these parameters given the SWOT observations. The acceptance or rejection of possible parameter values is guided by a likelihood function. This likelihood function is created by imposing the continuity equation to a set of mass-conserved reaches (e.g., Figure 1a), which is defined as a set of contiguous reaches with no known tributaries. The form of the continuity equation assumed by BAM neglects the cross-sectional area change with respect to time, essentially assuming that at a given time, the discharge in all reaches within a set is the same.

The prior distribution for A_0 and for the AMHG and hydraulic geometry parameters are constructed using the HYDRoacoustic data set in support of the Surface Water and Ocean Topography satellite mission (HYDRoSWOT; Bjerklie et al., 2020; Canova et al., 2016) following Hagemann et al. (2017). BAM also requires truncation bounds on Q_i, Q_c, n, and $A_{0,j}$, and the lognormal standard deviation of Q and b_j in space and time. These were set as follows: For rivers with a maximum observed $\delta A_{i,j}$ of less than 2,000 m^2, the lower and upper bounds on $A_{0,j}$ were set to 1 and 1,000 m^2, respectively. The lower bound on Q was set assuming a minimum depth of 0.5 m and a minimum velocity of 0.5 m/s, reasonable assumptions for these large rivers. b_{sd} was set to 0.2 (unitless), and upper bound on n was set to 0.05 following Chow (1959). Q_{sd} was set based on Q_{wbm} (prior estimate of mean annual flow) assuming a coefficient of variation of 1.

2.2. MetroMan: Metropolis-Manning Algorithm

Similarly to BAM, the Metropolis-Manning (MetroMan) algorithm (Durand et al., 2014) uses a Bayesian inference framework to estimate the unobservable flow law parameters needed by the Manning-Strickler equation to estimate discharge. However, unlike BAM, the version of MetroMan used in the present work allows n to vary in both time and space, in an attempt to account for the additional energy dissipation caused by spatial variability of hydraulic properties within a reach (Durand et al., 2016; Rodríguez et al., 2020; Tuozzolo, Langhorst, et al., 2019). Given the parameterization of friction, the assumption that width is much larger than depth, and that the friction slope can be approximated by the water surface slope (low Froude approximation), the Manning-Strickler equation used by MetroMan becomes:

$$Q_i = \frac{1}{n_{a,j}\left(H_{i,j} - H_{\min,j}\right)^{b_j}}\left(A_{0,j} + \delta A_{i,j}\right)^{5/3} W_{i,j}^{-2/3} \sqrt{S_{i,j}} \tag{3}$$

where $n_{a,j}$ and b_j are time invariant friction parameters estimated by MetroMan for each reach, $A_{0,j}$ is the reference cross-sectional area estimated by MetroMan for each reach in the domain, $H_{i,j}$, $W_{i,j}$, and $S_{i,j}$ are

the observations of water surface elevation, width, and slope for each reach j at each time i, and $\delta A_{i,j}$ the cross-sectional area anomaly for a reach j at time i computed from the time series of heights and widths. Additionally, when MetroMan applies a discretized version of the continuity equation for unsteady flow to the inversion domain, the algorithm retains both the derivative of discharge with respect to space and the time derivative of the cross-sectional area. This is in contrast to BAM that neglects the change in cross-sectional area in time, leading to equal discharge for all reaches. The continuity equation at a reach j at a time i is discretized as:

$$\frac{Q_{\text{dn}} - Q_{\text{up}}}{\Delta x} + \frac{A_{i+1,j} - A_{i,j}}{\Delta t} = 0 \tag{4}$$

where Q_{dn} represents the downstream discharge interpolated to the boundary between reaches $j-1$ and j, averaged between time steps i and $i + 1$, Q_{up} represents the upstream discharge interpolated to the boundary between reaches $j + 1$ and j, averaged between time steps i and $i + 1$, Δx represents the reach length, $A_{i+1,j}$ and $A_{i,j}$ stand for the cross-sectional area for reach j at time steps $i + 1$ and i respectively, and Δt represents the time interval between time steps $i + 1$ and i. MetroMan uses the sum of the residues in Equation 4 computed over the reaches within the inversion domain to construct a likelihood function, that guides the acceptance or rejection of parameter values for the construction of their posterior distributions (Equations 10 and 12 in Durand et al., 2014).

The likelihood function used by MetroMan for parameter acceptance is inversely proportional to the sum of the left-hand side of Equation 4 applied to each of the reaches contained in the inversion domain.

2.3. HiVDI: Hierarchical Variational Discharge Inference

HiVDI algorithm is presented by Larnier et al. (2020). HiVDI estimates flow law parameters by assimilating SWOT node heights and widths into a 1D Saint-Venant model using the variational data assimilation (VDA) framework described by Brisset et al. (2018) and Garambois et al. (2020). The VDA output includes optimal estimates of river discharge and reach averaged cross-sectional area. These outputs are used to calibrate a version of the Manning-Strickler equation:

$$Q_i = \alpha_j d_{i,j}^{\beta_j} \left(A_{0,j} + \delta A_{i,j} \right)^{5/3} W_{i,j}^{-2/3} \sqrt{S_{i,j}} \tag{5}$$

where α_j and β_j are fit parameters describing how the Manning-Strickler friction coefficient varies with river depth, $d_{i,j}$ is the cross-sectional average river depth, computed by $d_{i,j} = \left(A_{0,j} + \delta A_{i,j} \right) W_{i,j}^{-1}$, and all other terms are as defined in Equations 1 and 3.

2.4. SAD: SWOT Assimilated Discharge

The SAD algorithm (Andreadis et al., 2020) assumes that the river channel profile observed by a SWOT overpass can be modeled by the gradually varied flow equations as presented by Chow (1959) with a general form of:

$$\frac{\partial d_{i,j}}{\partial x_{i,j}} = \frac{S_{0,j} - S_{f,i,j}}{1 - Fr_{i,j}^2} \tag{6}$$

where $S_{o,j}$ is the bed slope for a reach j, $S_{fi,j}$ the friction slope at time i and reach j, d_{ij} is the average depth, x the longitudinal distance, and $Fr_{i,j}$ is the Froude number. SAD models the friction slope using the Manning-Strickler equation:

$$S_{f,i,j} = \frac{n_j^2 Q_i^2}{A_{i,j}^2 R_{i,j}^{4/3}} \tag{7}$$

where $R_{i,j}$ is the hydraulic radius, and other terms are as defined previously.

The gradually varied flow equations are combined with prior distributions for bed elevation, discharge, friction coefficient, and channel geometry (via the channel shape parameter) to form the forward model in a data assimilation scheme that is the core of the discharge estimation. The prior distributions are estimated with a data-driven approach based on rejection sampling and geomorphological classification. The SWOT observations can then be assimilated into this prior model in order to derive the posterior distribution of discharge for the observed river reaches. The assimilation algorithm used by SAD is the Local Ensemble Transform Kalman Filter (LETKF) that can be implemented efficiently over large areas and has been shown to be superior to other Kalman Filter algorithms partly due to its explicit localization (Hunt et al., 2007). In its current implementation, SAD is an instantaneous data assimilation algorithm, where updates to the current state of the model do not affect future discharge estimates.

Within this framework, the posterior estimate of discharge at a time i (Q_i^a) for each reach is estimated as:

$$Q_i^a = Q_i^b + X_i^b P_i^a M_i^{bT} Er_i^{-1} \left(m_i^o - m_i^b \right) \tag{8}$$

where Q_i^b is the prior estimate of discharge at a time i, X_i^b are the prior ensemble perturbations of the state variable, M_i^b is similar to X_i^b but for model-predicted measurements, m_i^o are the SWOT measurements, m_i^b are the model-predicted measurements, and Er is the measurement covariance matrix. The LETKF assimilation can include an additional regularization step that incorporates at-a-station hydraulic geometry relationships to better constrain the discharge estimates.

2.5. MOMMA: Modified Optimized Manning Method Algorithm

The MOMMA algorithm is an updated version of the Mean Flow and Geomorphology (MFG) algorithm examined in previous SWOT discharge algorithm benchmarking experiments (Bonnema et al., 2016; Durand et al., 2016). The MOMMA algorithm is designed to distinguish between in-bank and out-of-bank flow and therefore is dependent on identifying the bankfull stage and flow resistance (friction). MOMMA uses a modified version of the Manning-Strickler equation:

$$Q_{i,j} = \frac{1}{n_{i,j}} \left[\left(H_{i,j} - B_j \right) \left(\frac{r}{1+r} \right) \right]^{5/3} W_{i,j} \sqrt{S_{i,j}} \tag{9}$$

where r is the channel shape parameter also used by SAD. In MOMMA, the parameter r assumes the value of 2 based on the work of Bjerklie et al. (2020), and $n_{i,j}$ is computed based on an assumed logarithmic scaling function that is conceptually based on relative roughness (Lang et al., 2004; Limerinos, 1970):

$$n_{i,j} = n_{b,j} \left[1 + log \left(\frac{H_{b,j} - B_j}{H_{i,j} - B_j} \right) \right] \tag{10}$$

B_j used in both Equations 9 and 10 represents the water surface elevation of zero flow, which is assumed to be equal to the intercept of second order polynomials fitted to observed pairs of water surface elevation and widths obtained for each reach. The term $n_{b,j}$ in Equation 10 is the friction coefficient at bankfull flow and $H_{b,j}$ is the bankfull water surface elevation, estimated by identifying breakpoints in observed water surface elevation-width relationships. The bankfull flow friction is estimated using the empirical relation derived by Dudley (2004), and a prior estimate of mean annual flow.

3. Methodology

3.1. Hydraulic Models and Definition of Test Cases

In order to construct synthetic SWOT observations, we used output from 7 hydraulic models: the Ganges-Brahmaputra-Meghna model (Maswood & Hossain, 2016; Sikder & Hossain, 2015), the combined Sacramento River System model (Rogers, 2014), the Ohio River model (Adams et al., 2010), Iowa River model (Gilles et al., 2012), Missouri River model (USACE, 2015), Seine River model (Even et al., 1998, 2007), and Olentangy River model (Tuozzolo, Langhorst, et al., 2019). From each model, we isolated individual rivers,

Table 2
Summary of Test Case Properties, Including Average Domain Width, Number of Cross-Sections, Modeled Length, Number of Days for Which the Model Ran, Water Balance Model Estimates (WBM) of Mean Annual Flow and True Mean Flow, and Which Phases Used Each Test Case

	Test case name	Hydraulic model	Width (m)	Number of cross-sections	Length (km)	Simulation length (days)	Mean annual flow WBM (m^3/s)	True (m^3/s)	Study phases 1	2	3	Reference
1	Ash Slough	HEC-RAS	56	53	7.7	12	36	38	X			Rogers (2014)
2	Berenda Slough	HEC-RAS	43	62	11.4	12	36	16	X			Rogers (2014)
3	Brahmaputra	HEC-RAS	2,275	39	250.0	107	5,162	2,039	X		X	Maswood and Hossain (2016)
4	Chowchilla Canal	HEC-RAS	78	74	26.1	12	36	21	X			Rogers (2014)
5	Fresno River	HEC-RAS	52	61	16.2	12	46	52	X			Rogers (2014)
6	Grant Line Canal	HEC-RAS	200	34	12.5	12	244	212	X			Rogers (2014)
7	Iowa River	HEC-RAS	157	201	32.3	366	147	158	X	X	X	Gilles et al. (2012)
8	Jamuna	HEC-RAS	5,066	39	224.7	250	16,623	20,974	X	X	X	Maswood and Hossain (2016)
9	Kushiyara	HEC-RAS	318	48	264.9	366	2,133	1,171	X	X	X	Maswood and Hossain (2016)
10	Mariposa Bypass	HEC-RAS	165	34	6.3	12	76	108	X			Rogers (2014)
11	Merced River	HEC-RAS	87	153	43.7	12	76	95	X			Rogers (2014)
12	Middle River	HEC-RAS	45	55	16.2	12	172	22	X			Rogers (2014)
13	Mississippi	HEC-RAS	746	133	95.1	162	2,459	5,496	X			Adams et al. (2010)
14	Missouri upstream	HEC-RAS	274	58	33.0	595	498	922	X	X	X	USACE (2016)
15	Missouri midsection	HEC-RAS	241	36	22.2	595	498	1,017	X	X		USACE (2016)
16	Missouri downstream	HEC-RAS	247	58	33.0	595	498	1,033	X	X	X	USACE (2016)
17	Ohio section 1	HEC-RAS	377	240	96.2	220	2,459	1,636	X	X	X	Adams et al. (2010)
18	Ohio section 2	HEC-RAS	422	216	88.1	220	2,459	2,450	X	X	X	Adams et al. (2010)
19	Ohio section 3	HEC-RAS	506	362	140.8	220	2,459	2,945	X	X	X	Adams et al. (2010)
20	Ohio section 4	HEC-RAS	564	37	51.5	220	2,459	3,454	X	X	X	Adams et al. (2010)
21	Ohio section 5	HEC-RAS	678	95	86.5	220	2,459	3,831	X	X	X	Adams et al. (2010)
22	Ohio section 7	HEC-RAS	976	33	71.8	220	2,459	5,376	X	X	X	Adams et al. (2010)
23	Ohio section 8	HEC-RAS	1,162	50	71.2	220	2,459	7,601	X	X	X	Adams et al. (2010)
24	Olentangy	HEC-RAS	36	115	6.4	13	26	3	X			Tuozzolo et al. (2019)
25	Padma	HEC-RAS	8,831	32	105.1	366	2,133	30,017	X	X	X	Maswood and Hossain (2016)
26	San Joaquin	HEC-RAS	272	403	57.1	12	155	208	X			Rogers (2014)
27	San Joaquin 2	HEC-RAS	73	253	59.4	12	155	26	X			Rogers (2014)
28	Seine Downstream	ProSe	224	240	128.1	365	206	304	X	X	X	Even et al. (2007)
29	Seine Upstream	ProSe	162	219	84.6	365	206	197	X	X	X	Even et al. (2007)
30	Stanislaus River	HEC-RAS	114	55	24.3	12	155	101	X			Rogers (2014)
31	Tuolumne River	HEC-RAS	386	32	10.0	12	155	228	X			Rogers (2014)

extracting time series of cross-sectional top width, water surface elevation, area, and discharge, as well as the flow distance associated with the cross-sections. Each river was broken into mass-conserved test cases, hereafter referred to as cases, which are defined as a stretch of a river that contains no visible tributaries, allowing the application of either forms of the continuity equation assumed by BAM and MetroMan. An example of such segmentation is shown in Figure 4a for the Ohio River model. In all, 31 cases were constructed, covering a wide range of discharges, from 1 m^3/s to 80,000 m^3/s. The list of test cases and a summary of model properties are given in Table 2 and Figure 3 shows the approximate location of the test cases.
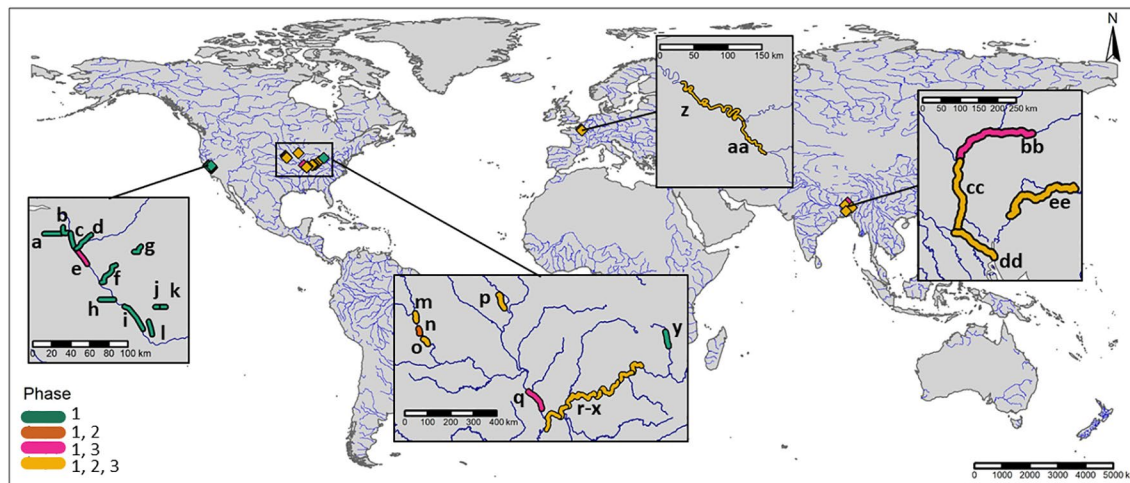
Figure 3. Location of the test cases used for algorithm benchmarking color coded by the phase or phases of the study for which each model is used. The test case names are: a: Grant Line Canal, b: Middle River, c: San Joaquin 2, d: Stanislaus, e: San Joaquin, f: Merced River, g: Tuolumne, h: Mariposa Bypass, i: Fresno River, j: Ash Slough, k: Berenda Slough, l: Chowchilla Canal, m: Missouri Upstream, n: Missouri Midsection, o: Missouri Downstream, p: Iowa River, q: Mississippi Intermediate, r–x: Ohio 8, Ohio 7, Ohio 5, Ohio4, Ohio3, Ohio2, Ohio 1 respectively, y: Olentangy, z: Seine Downstream, aa: Seine Upstream, bb: Brahmaputra, cc: Jamuna, dd: Padma, ee: Kushiyara. The locations are approximate due to the use of non-georeferenced hydraulic models. Test case characteristics in model references are included in Table 2.

Each case is further segmented into reaches, which is the scale where flow law parameters estimated by the discharge algorithms in conjunction with the synthetic SWOT measurements are used to calculate discharge. For this study, we use functional reaches following a segmentation criterion similar to the hydraulic controls reach definition method presented by Frasson et al. (2017). Here, we manually selected reach boundaries defined with the intent of minimizing slope heterogeneity within a reach, while including at least four cross-sections per reach. Additional boundaries were placed around visible hydraulic structures to isolate such structures as illustrated by Figure 4b.

Reach averaged height, width, cross-sectional area, and discharge were calculated by taking the arithmetic average of the corresponding values belonging to cross-sections within each reach. In the absence of measurement error, the reach averaged slope was computed as the elevation loss throughout the reach divided by the reach length (Phases 1 and 2 described in Sections 3.3 and 3.4). When elevation errors are taken into consideration (Phase 3 described in Section 3.5), the reach averaged slopes are computed by fitting a linear polynomial to the cross-section heights. The resulting data set can be retrieved from Frasson et al. (2020).

3.2. Experimental Protocol and Design

We subjected the algorithms to a strict experimental control in which the only information the algorithms received was a prior estimate of mean annual flow in addition to the SWOT synthetic measurements. This decision is in keeping with previous experimental design (i.e., Durand et al., 2016), and in order to control the number of experimental parameters. We acknowledge that many and much richer sources of prior information exist, for example, flow quantiles other than mean annual flow are available globally as part of the Global Reach-Level A Priori Discharge Estimates for SWOT (GRADES) database created by Lin et al. (2019), or, where available, in situ measurements of discharge and bathymetry, both of which can be used by the algorithms to provide better constraints for the unobservable flow law parameters. However, this experimental design allows us to test algorithm capabilities under the most stringent scenario to achieve the following goals: (a) to identify controls of algorithm performance in ungaged basins (the worst case scenario), (b) to explore the effects of temporal sampling on the ability to retrieve discharge, and (c) to assess to what extent discharge algorithms are impacted by SWOT measurement error.

The prior estimates of mean annual flow (Q_{wbm}), for each test case needed by the five evaluated discharge algorithms were obtained from a 50-year run of the global hydrological water balance model WBMsed (Cohen et al., 2014). No additional information was given to the discharge algorithms and no changes to the
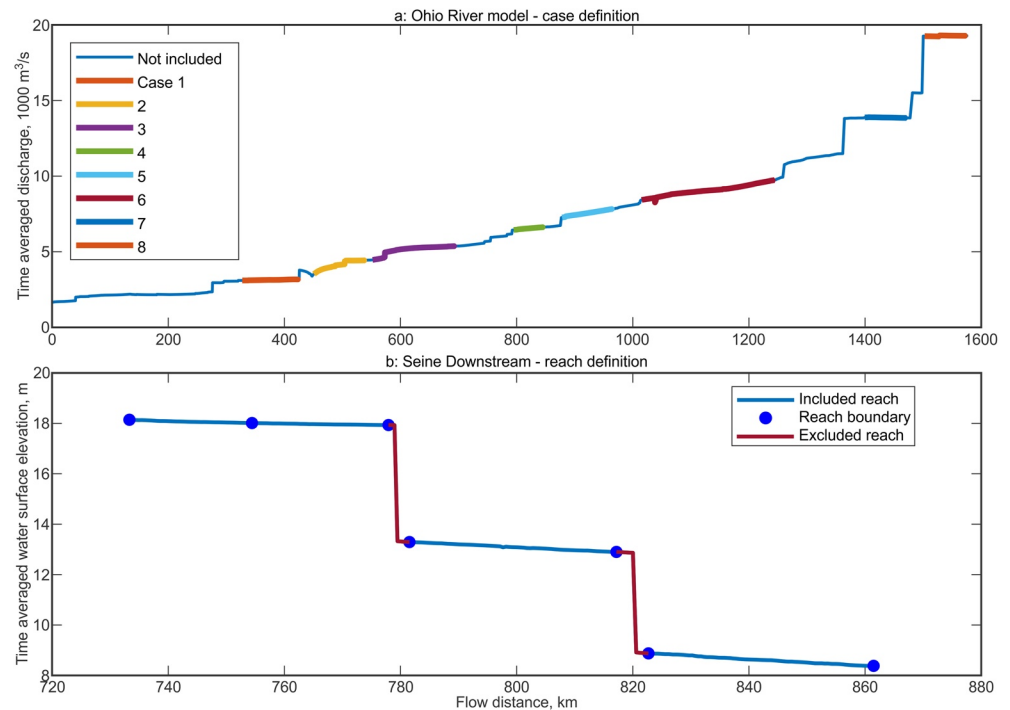
Figure 4. Panel a: the division of the Ohio River model into mass conserved cases. Line colors identify the different cases; note that reach 6 was used by Durand et al. (2016), therefore excluded from the present analysis. Each case contains multiple reaches. Panel shows the reach definition for the downstream Seine case. Short reaches shown in red were created to identify dams.

algorithms were allowed once testing began. We evaluated the discharge algorithms in three phases, which differ on temporal sampling strategy and the presence or absence of measurement error. We averaged all discharge estimates across all reaches within each case, following Durand et al. (2016). We judged algorithm performance based on a set of metrics described in Section 3.6. Inversion results obtained with the tested algorithms for each of the three study phases can be downloaded from Frasson et al. (2020).

3.3. Phase 1: Daily Sampling and No Measurement Error

In this phase, we ran the discharge algorithms using a set of synthetic observations that mirror SWOT measurements of reach averaged heights, widths, and slopes, however, without added measurement uncertainty and assuming daily sampling. In addition to providing the best case scenario for our study by applying the algorithms in the absence of noise and with high frequency sampling, this phase has three objectives: to investigate how algorithm performance is impacted by the quality of the prior estimate of mean annual flow (Q_{wbm}), assess how errors associated with the choice of the flow law used by each method affect algorithm performance, and to identify a predictor of inversion success.

3.3.1. Effect of Flow Law Errors on Discharge Inversion

Two different forms of the Manning-Strickler equation are used by BAM, MetroMan, and HiVDI. While the version of BAM evaluated here assumes a constant value of n in both space and time, MetroMan and HiVDI implement a parameterization of the friction coefficient that allows it to vary from reach to reach and with depth. We computed the error associated with the flow law used by each algorithm by calibrating the friction coefficient parameters directly to the true discharge and cross-sectional area. For this exercise, we identified the values of the friction parameters that minimized the squared differences between the true discharge for each reach and a discharge estimated with the true values of reach averaged cross-sectional area, width, and water surface slope.

After finding optimal friction coefficient parameters for each of the flow law formulations, that is, Equations 1, 3, and 5, we estimated discharge for each reach and time step using the optimal friction parameters and the true reach averaged values of cross-sectional area, width, and slope. Since optimal friction parameters and true reach averaged cross-sectional area were used in conjunction with error-free synthetic observations of width and slope, the remaining differences between the true discharge and the estimated discharge are caused by the assumed form of the flow law. Thus, this exercise results in an estimate of the error due to the structure of the flow law itself, measured in terms of normalized root mean squared error defined in Section 3.6, separate from the error in the parameters and measurement error. The true cross-sectional areas and discharges are used solely for the identification of optimal parameters and the estimation of the error associated with the use of either parameterizations of channel friction. The values of true cross-sectional areas and discharges were not used to inform any of the parameter inversions conducted in Phases 1, 2, and 3.

3.3.2. Predicting Inversion Success—The Importance of Hydraulic Spatial Variability

In the presentation of early McFLI discharge inversion methods, Durand et al. (2010) and Garambois and Monnier (2015) stated that sufficient slope variability is required for the joint estimation of bathymetry and discharge. We expanded on this concept by evaluating the spatial variability of two additional properties: top width at median flow, and cross-sectional area at the lowest positive flow rate, quantified by their coefficients of variation computed between reaches within a test case.

We hypothesized that cases with high spatial variability of at least one of the three hydraulic parameters should yield successful inversions, whereas those with low spatial variability should not be successful. We defined a successful discharge inversion as an inversion that results in a lower absolute value of the normalized bias (see Section 3.6) as compared with the prior estimate of mean annual flow (Q_{wbm}). We define high and low spatial variability thresholds in terms of the statistical percentiles of the coefficient of variation of each property across all cases considered.

The selection of the variability threshold was conducted based on the inspection of Receiver Operating Characteristic curves (ROC; Fawcett, 2006) constructed for each algorithm. The ROC was created by changing the high variability threshold from the 0th percentile to the 100th percentile, computing the number of true positives, that is, cases with high spatial variability of at least one hydraulic property that showed normalized discharge bias reductions, and the number of false positives, that is, cases with high spatial variability of at least one hydraulic property that showed increased normalized discharge biases. A curve for each algorithm was created by plotting the ratio between the false positives and the total number of unsuccessful inversions versus the ratio between the true positives and the total number of successful inversions.

When the high variability threshold is set to the 0th percentile, all cases are classified as having high variability, leading to a ratio between true positives and the total number of positives equal to 1. However, all unsuccessful inversions are also classified as high variability, leading to a ratio between false positives and the number of unsuccessful inversions equal to 1. As the high variability threshold increases, the true positive and false positive ratios tend to decrease. Curves that track the 1:1 line indicate that the predictor is no better than guessing, whereas curves that fall above and to the left of the 1:1 line indicate that the predictor has more skill than random selections. The optimal threshold depends on the user tolerance for false positives and the need for detection of true positives. For our study, we selected the same high variability threshold for all hydraulic properties and algorithms, with a value that aimed to place all algorithms as far (in terms of Euclidian distance) away from the 1:1 line in the ROC curve as possible.

3.4. Phase 2: Gradual Degradation in Sampling Frequency Without Measurement Uncertainty

In order to explore the effect of satellite revisit time, we created noise-free datasets at seven different sampling frequencies: 2, 3, 4, 5, 7, 10, and 21 days. The latter four sampling strategies cover the typical range of SWOT sampling frequencies (Biancamaria et al., 2016). The resampling was done by assuming that the first day was observed and from there, a new observation would occur at the desired time sampling in regular time steps. Figure S1, included in the supporting information, shows an example of the temporal sampling for the Iowa River case. While SWOT sampling of a particular location in time will in fact be irregular (e.g., Frasson et al., 2019), this irregularity in and of itself should not affect algorithm performance; thus, the
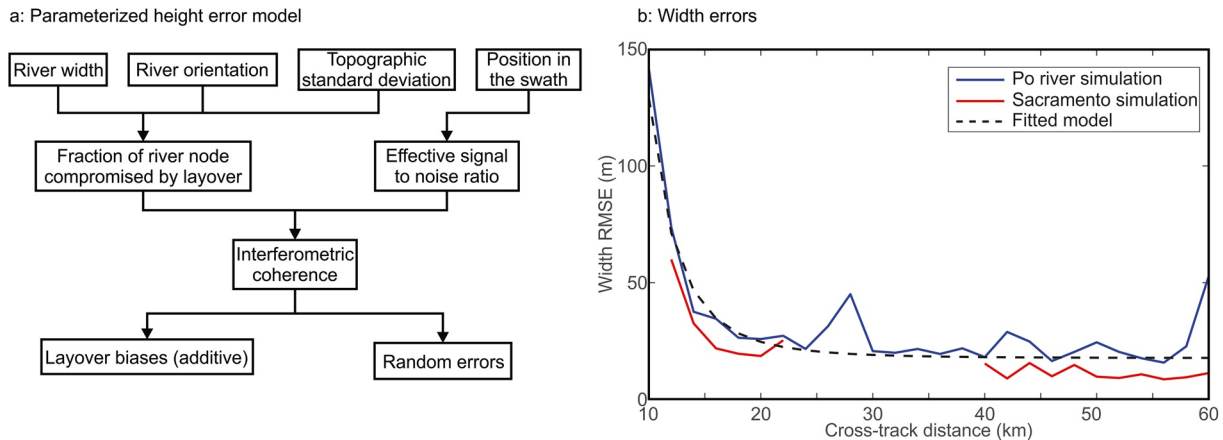
Figure 5. Cross-section height and width error model. Panel (a) schematics illustrating the height error model (Durand et al., 2020). Panel (b) dependence of the width root mean square error (RMSE) on cross-track position based on simulated Surface Water and Ocean Topography (SWOT) returns described by Frasson et al. (2017). In addition to the modeled errors, systematic biases accounting for atmospheric conditions, spacecraft roll, and others are taken from Esteban-Fernandez (2013).

regular sampling frequencies used in this study should provide a reasonable measure of the effect of sampling frequency on algorithm performance.

Test cases extracted from the combined Sacramento River System, the Olentangy, and the Brahmaputra models had short durations, which when downsampled do not result in at least six observations, the minimum number of observations for which BAM, HiVDI, and MetroMan are known to work (e.g., Tuozzolo, Lind, et al., 2019). Therefore, cases extracted from these models were not included in the set of 16 models used in Phase 2 as shown by Table 2.

3.5. Phase 3: Increasing Measurement Error

The purpose of this phase is to evaluate how measurement uncertainty impacts discharge algorithm performance. To address this question, we applied the error model presented by Durand et al. (2020) to generate height and width errors initially at the cross-section scale, increasing the error levels from half to full and to 1.5 times the estimated SWOT errors. Noisy cross-sectional heights are used to compute reach slope and height and noisy cross-sectional widths are used to compute the reach averaged width. Width errors due to misclassification of surfaces (see Grippa et al., 2019) are ignored in this study.

The cross-section height errors account for three main elements: layover error, thermal noise, and systematic errors. We compute the first two elements using the parameterized error model described by Durand et al. (2020) which we illustrate in Figure 5a, whereas the systematic errors accounting for media delays, spacecraft roll, and others are taken from the SWOT error budget document (Esteban-Fernandez, 2013). The position of a cross-section within the swath determines the cross-track and along track resolutions, which combined with the time-variable width and the distance between cross-sections, allow us to compute how many pixels are associated with a cross-section at a given point in time. Next, we computed the fraction of the cross-section that is affected by layover. The important factors for this calculation are the river orientation with respect to the swath, the topographic roughness (measured as described by Durand et al., 2020), and the number of pixels associated with a cross-section. The effective signal-to-noise ratio varies significantly across the swath, and is combined with the fraction of the node compromised by layover to compute the complex-valued interferometric coherence, which in turn exerts a direct control on error levels in interferometry as described by for example, Rosen et al. (2000).

The random height error Eh for a cross-section j at time i is drawn from a normal distribution with zero mean and standard deviation equal to $\sigma_{hi,j}$; $\sigma_{hi,j}$ is computed from the interferometric coherence. Layover bias $h_{Li,j}$ is also computed from interferometric coherence, as described by Durand et al. (2020). The systematic height errors due to media delays, spacecraft roll, and other sources are assumed to be constant for

all cross-sections but variable in time. Therefore, it assumes the form of Es_i, which is normally distributed with 0 mean and standard deviation of 8.897 cm, in accordance to the SWOT error budget document (Esteban-Fernandez, 2013). The synthetic SWOT observation of cross-sectional height $\hat{H}_{i,j}$ is calculated by adding error terms to the true height ($H_{i,j}$):

$$\hat{H}_{i,j} = H_{i,j} + h_{Li,j} + Er_{i,j} + Es_i \tag{11}$$

For the derivation of realistic width errors, we analyzed two high fidelity SWOT simulations documented by Frasson et al. (2017) and Domeneghetti et al. (2018): one over the Sacramento River, which runs mostly parallel to the ground track, and the Po River, which runs almost perpendicularly to the relevant SWOT ground tracks. We computed the root mean square errors (RMSE) at the node scale available from the simulations, which is a spatial scale similar to our cross-sections. Figure 5b shows the width RMSE compared to the cross-track distance for both models, to which we fitted a model with the following functional form:

$$\sigma_{Wi,j} = \frac{a}{s_{i,j}^4} + b \tag{12}$$

where σ_{wij} is the width standard deviation at time i for the cross-section j, s is the cross-track distance in kilometers, and a and b are fit parameters found to be equal to $1.115 \cdot 10^6$ m km^4 and 17.6 m respectively.

The cross-track distance controls the size of the pixels, which along with classification errors dominate width errors. Classification errors are controlled by the contrast between water and land, as well as the smoothness of the water surface, which under especially calm wind and flow conditions can cause specular reflection, a situation referred to as dark water. The contrast between land and water for the two simulations was assumed to be described by a water-land scattering brightness ratio of 15, which is not uncommon at the SWOT incidence angle according to ranges of values presented by Ulaby and Dobson (1989) and Fjørtoft et al. (2014).

Finally, the reach slopes were estimated using the slope of the linear fit over the noisy cross-section heights and reach averaged heights were computed using the same linear fit evaluated at the center of the reach. Reach widths were simply computed as the arithmetic average of the widths of cross-sections belonging to each reach. An example of the error statistics for height, width, and slope for the Iowa River case is shown in Figure S2 (supporting information) with the time-varying cross-sectional height error statistics shown in the Movie S3. Noisy reach-averaged heights, widths, and slopes alongside the prior estimate of mean annual flow were given to each discharge algorithm. As in previous phases, the estimated discharge is averaged in space for the entire case, which is compared with the true discharge.

3.6. Error Metrics

The Nash-Sutcliff Efficiency (NSE) was computed as:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{T} \left(Q_i - Q_{Ti} \right)^2}{\sum_{i=1}^{T} \left(Q_{Ti} - \bar{Q}_T \right)^2} \tag{13}$$

where Q_i stands for the estimated case-averaged discharge at time i, which is calculated as:

$$Q_i = \frac{\sum_{j=1}^{Nr} Q_{i,j}}{Nr} \tag{14}$$

In Equation 14, Nr represents the number of reaches used by the discharge algorithm, and $Q_{i,j}$ is the estimated discharge for reach j at time i; Q_{Ti} represents the case-averaged true discharge computed as:

$$Q_{Ti} = \frac{\sum_{j=1}^{Nr} Q_{Ti,j}}{Nr} \tag{15}$$

where $Q_{Ti,j}$ is the true reach averaged discharge for reach j at time i; and $\bar{Q}_T$ is the time and case averaged true discharge, computed by averaging Q_{Ti} along the time steps.

Similarly, the normalized root mean square error (NRMSE) is computed as:

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{Nt}\left(Q_i - Q_{Ti}\right)^2}{Nt}}}{\bar{Q}_T} \tag{16}$$

where Nt stands for the number of time steps.

The normalized bias (nbias) is computed as:

$$\text{nbias} = \frac{\frac{\sum_{i=1}^{Nt}\left(Q_i - Q_{Ti}\right)}{Nt}}{\bar{Q}_T} \tag{17}$$

And the prior normalized bias (nbias$_{\text{WBM}}$) is calculated as:

$$\text{nbias}_{\text{WBM}} = \frac{Q_{\text{WBM}} - \bar{Q}_T}{\bar{Q}_T} \tag{18}$$

Finally, the normalized standard deviation of residuals ($n\sigma_e$, also known as standard error) is represented by:

$$n\sigma_e = \frac{\sqrt{\frac{\sum_{i=1}^{Nt}\left(E_i - \bar{E}\right)^2}{Nr - 1}}}{\bar{Q}_T} \tag{19}$$

where E_i represents the discharge error at time i computed as $Q_i - Q_{Ti}$, and $\bar{E}$ represents the time-averaged discharge error.

4. Results and Discussion

4.1. Phase 1: Daily Sampling and No Measurement Error

4.1.1. Evaluation of the Retrieved Hydrographs and Summary of Algorithm Performance

Figure 6 illustrates the algorithm capabilities for daily, error free observations by showing six example hydrographs produced by the McFLI and data assimilation algorithms: MetroMan, BAM, HiVDI, and SAD, alongside the prior estimate of discharge, and the discharge extracted from the hydraulic model (referred as Truth in the plot). Generally, the algorithms track the flow patterns visible in the truth, matching the timing of increases and decreases of the true discharge, even in cases where the remaining bias in the discharge estimated by the algorithms is high, such as Figure 6d.

The normalized root mean square errors and the Nash-Sutcliffe efficiency for each of the 31 cases are shown in Figure 7, allowing a comparison of the performance for each case for the five algorithms. The two error metrics alongside the normalized biases, the normalized standard deviation of residuals and the correlation coefficient for the McFLI and data assimilation algorithms and for MOMMA are summarized in Figure 8. For completeness, we include a similar figure showing the Kling-Gupta Efficiency and its three elements in the supporting information, in Figure S5.

HiVDI and MetroMan show similar performance across most error metrics, a factor that may be related to the fact that both calculate discharge with a flow law that allows the effective friction coefficient to vary in space and in time (Equations 5 and 3, respectively). In contrast, the version of BAM benchmarked here assumes that all reaches have the same friction coefficient and that this coefficient is static in time. While median values of some error metrics for BAM, HiVDI, and MetroMan are comparable, the interquartile ranges
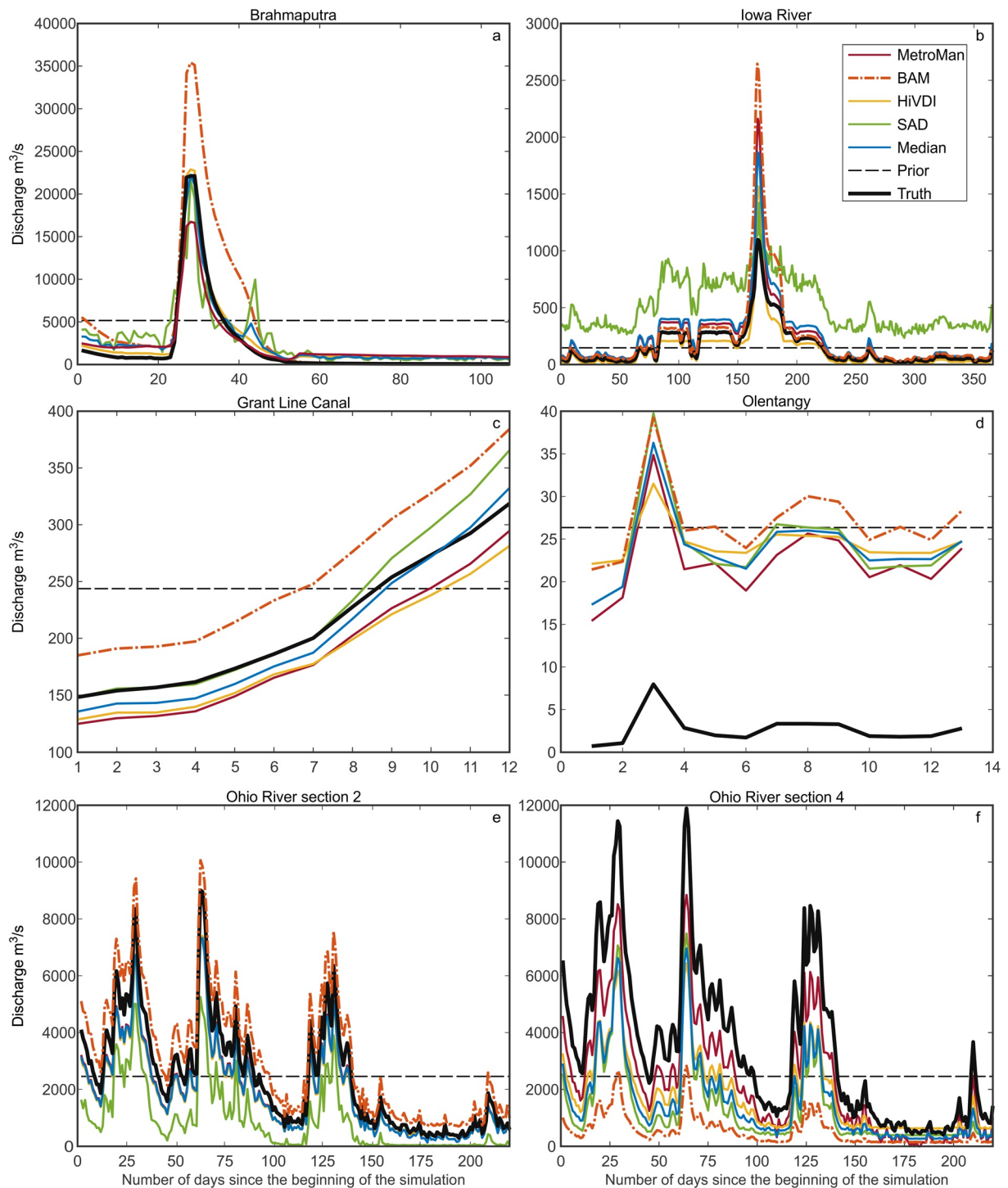
Figure 6. Examples of retrieved hydrographs. The line colors indicate the discharge estimated by four Mass conserved Flow Law Inversion and data assimilation algorithms alongside the algorithm ensemble median, the initial estimate of mean annual flow (prior), and the truth. MetroMan, Hierarchical Variational Discharge Inference, and the median are nearly overlapping in panel e.

for BAM are often considerably wider than those for HiVDI and MetroMan. SAD error metrics for Phase 1 are somewhat better than BAM, and somewhat worse than HiVDI and MetroMan. MOMMA performance is significantly worse than other algorithms when judged on most metrics. Indeed, MOMMA significantly increased median normalized absolute bias (0.67) compared with the median bias of the WBM prior (0.44),
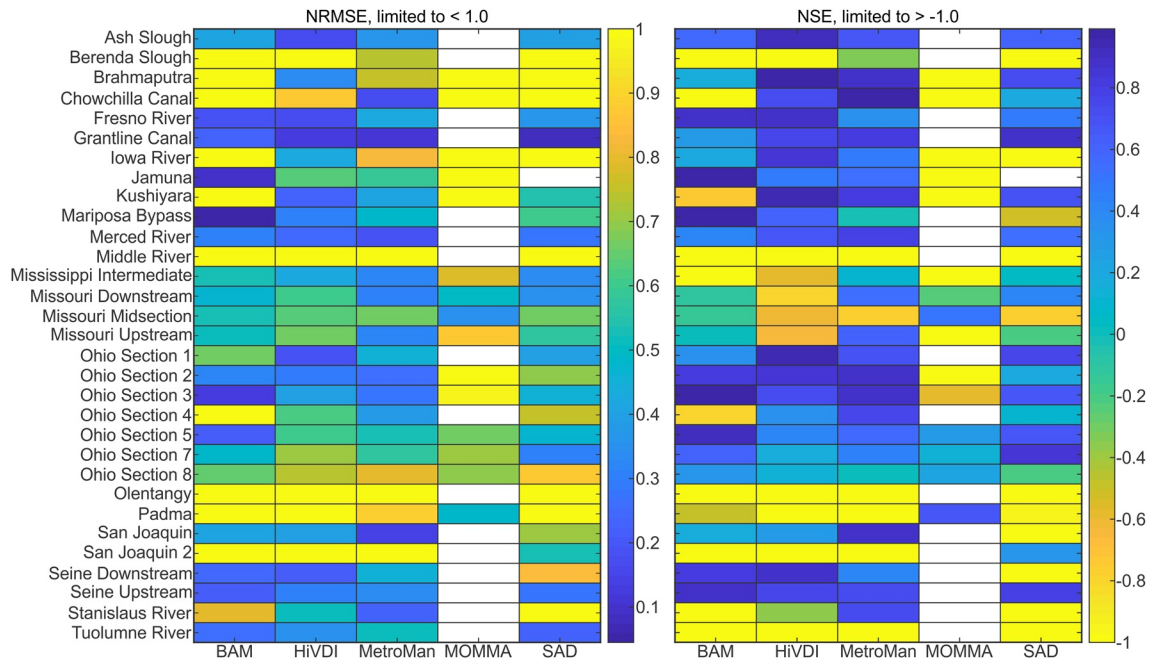
Figure 7. Comparison of the discharge Normalized Root Mean Square Errors (NRMSE) and Nash-Sutcliffe Efficiency (NSE) for each of the 31 idealized cases (organized in rows) as estimated by each of the discharge algorithms (columns). Cases shown in white for the Modified Optimized Manning Method Algorithm algorithm are cases where bankfull flow was not successfully identified, leading to no output. The missing case for SAD is the Jamuna river, where SAD did not run successfully. NRMSE was capped at 1 and NSE at −1 to allow for better visualization of differences between model performance. Outliers can be seen in Figure 8.

while BAM (0.43), HiVDI (0.42), and MetroMan (0.34) reduced the median bias and SAD caused a slight increase (0.47).

MOMMA relies on two key parameters that are derived from the observational width and stage data. These are the water surface elevation of zero flow (B used in both Equations 9 and 10) and the bankfull water surface elevation (H_b, also in Equations 9 and 10). The initial value of B is estimated using the intercept from a linear model with the dependent variable stage and independent variable width squared (due to the assumption that the channel has a parabolic shape) derived from all stage and width observations equal to or less than the median observed stage. The value of H_b is determined by an automated routine that finds the point in the stage-width curve where the width-to-depth ratio is at its minimum. If no bankfull stage is identified, which is the case in many of the river reaches, no discharge values are estimated. Because of this limitation, fewer river reaches were assessed with MOMMA than the other algorithms. If a bankfull stage is identified, the bottom of the channel is recomputed as described above using all observed width and stage data below the bankfull stage rather than below the median stage. Because H_b and B are both dependent on the width-stage relation, the linearity of that relation is critical to the success of MOMMA. Any nonlinearity in that relation at a given river reach is a source of error in MOMMA discharge estimates.

An additional source of error for MOMMA comes from the estimation of the friction coefficient parameter (n_b) (used in Equation 10). For this study, n_b is assumed to be the Manning n value at the bankfull discharge. One of the priors given for this study is an estimate of the mean annual discharge, but not the bankfull discharge. In lieu of this, the bankfull discharge is estimated from an empirical relation between mean annual discharge (Q_{wbm}) and bankfull discharge by Dudley (2004) that indicates that the bankfull discharge is roughly three times the magnitude of the mean annual discharge for a set of rivers in Maine. Using the estimate of bankfull discharge, n_b is calibrated to yield that bankfull discharge given the observed bankfull width, stage, and slope. As the empirical relation between mean annual flow and bankfull discharge was derived for a specific region in the northeastern United States, it is not expected to be true for other climatic and hydrophysiographic regions around the globe. This is therefore a likely source of error with the
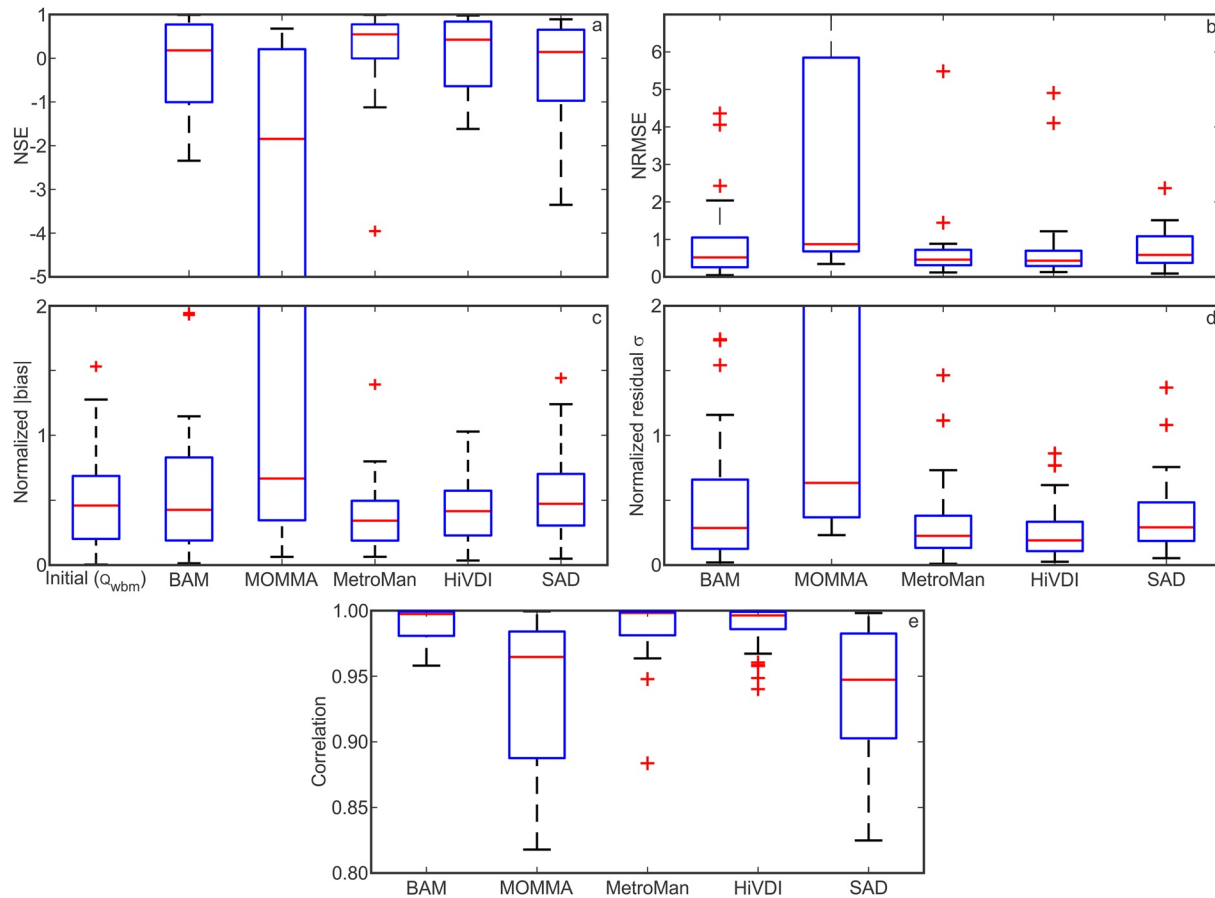
Figure 8. Summary statistics for case-averaged discharge for the five algorithms over the 31 examined cases. Panel a: box plot of the discharge Nash-Sutcliffe Efficiency for each of the algorithms. Panel b: discharge root mean square error normalized by the true mean discharge. Panel c: estimated discharge biases normalized by the true mean annual flow. Panel d: residual standard deviation normalized by the true mean discharge. Panel e: correlation coefficient. The vertical axes in panels a, c, and d are cut to allow comparisons between Bayesian At-many-stations-hydraulic-geometry Manning, MetroMan, Hierarchical Variational Discharge Inference, and SWOT Assimilated Discharge. Panels showing the full range of vales can be found in the Figure S4.

application of MOMMA to many of the rivers in this study indicating that the application of MOMMA to river systems in other settings would require the input of appropriate empirical relations.

These results point to the need for quality control on the width-stage relation and an independent method for estimating n_b for the successful use of MOMMA, which led us to exclude it from further analysis in this study. We note, however, that MOMMA tracks flow dynamics as evidenced by a correlation between estimated discharge and true discharge larger than 0.9 for more than 75% of the examined cases (Figure 8e). Moreover, MOMMA does not rely on enforcing mass conservation between neighboring reaches, thus does not require reaches to be simultaneously observed and is not affected by the presence of undetectable confluences, allowing discharge to be estimated in situations where the other algorithms cannot be applied. As such, future iterations of the MOMMA algorithm may be useful for estimating discharge from SWOT.

4.1.2. Impact of Flow Law

To investigate if flow law differences are driving discharge algorithm outcomes, we computed the discharge errors caused by the chosen form of the Manning-Strickler equation (see Section 3.3.1) and divided the cases into two classes: the 15 cases with the lowest errors and the 16 cases with the highest errors. Figure 9 panels i, j, and k compare discharge NRMSE for low versus high flow law errors. For BAM, the median NRMSE increases from 0.32 to 0.72 which is considerably larger than the increase in the NRMSE due to the flow law itself, which shows a median of 0.04 in the 15 best cases and 0.22 in the 16 worst cases. The two algorithms that use a spatiotemporally variable friction parameterization showed more modest increases:
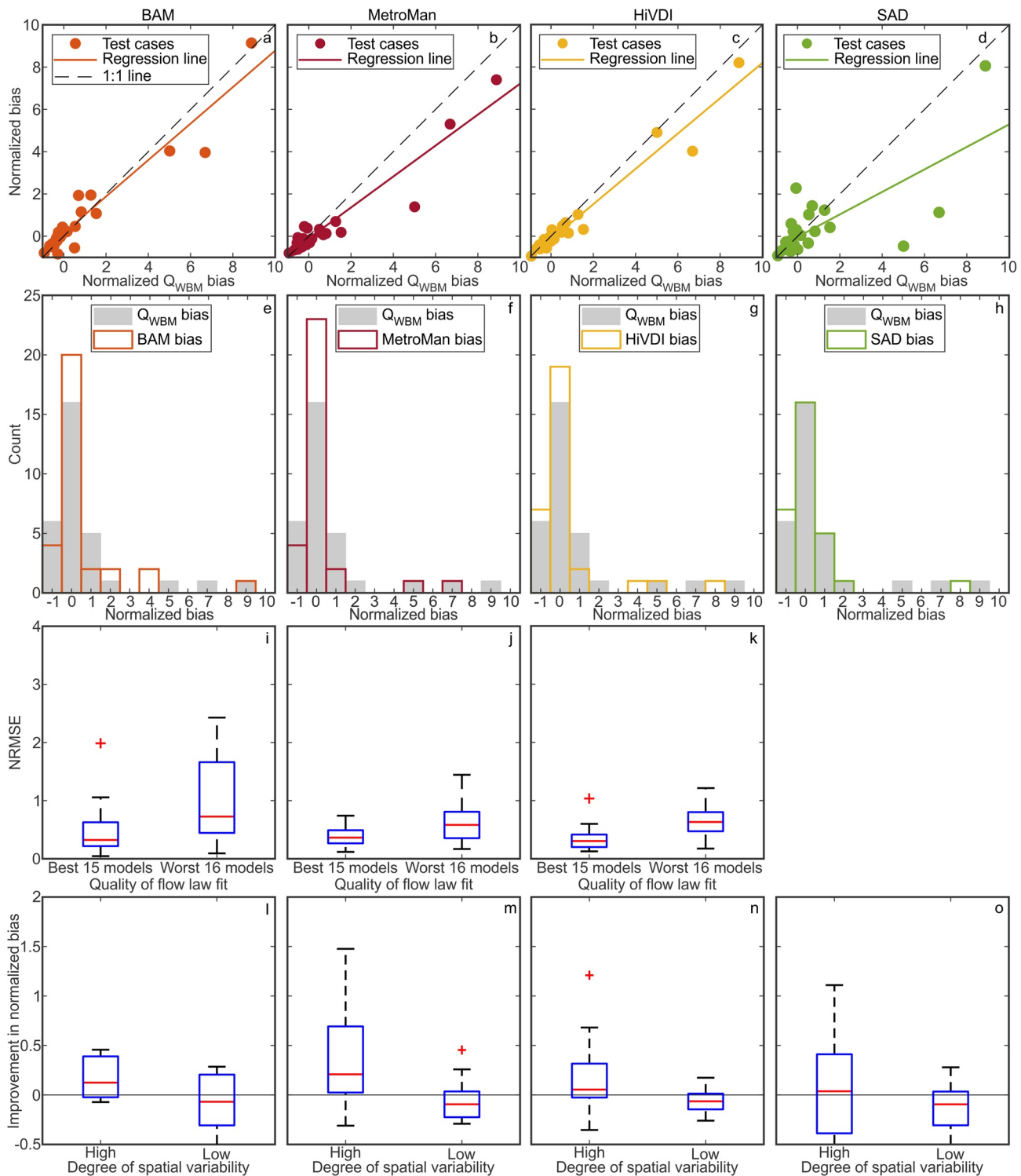
Figure 9. Panels a–d show how biases in the prior estimate of mean annual discharge (Q_{WBM}) affect the remaining biases after running the inversion algorithms, panels e–h show the histogram of remaining normalized bias after running each inversion method superimposed on the prior normalized biases (gray), panels i–k show how the quality of the Manning parameterization fit impacts the normalized root mean square errors, panels l–o show boxplots of the difference in magnitude between prior and final normalized discharge biases, distinguishing cases containing high spatial variability of hydraulic properties from those containing low spatial variability. In panels l–o, positive numbers represent decreases in the magnitude of the bias. The histogram for BAM represents only the 16 best fit cases shown in panel i.

for MetroMan, the median NRMSE increased from 0.36 to 0.58, whereas for HiVDI the median discharge NRMSE increased from 0.30 to 0.63, with the errors due to the spatiotemporally variable friction flow law itself taking median values of 0.02 for the 15 best cases and 0.08 for the 16 worst cases. This result suggests that the higher the errors caused by the parameterization of friction are, the harder it is for discharge algorithm to retrieve optimal flow law parameters, thus causing non-linear increases in discharge errors. This finding has not been reported previously, and bolsters the need to evaluate the performance of classical flow laws when applied to natural rivers at scales that are compatible with remote sensing methods as well as to improve flow law formulation to better suit such scales as done by for example, Tuozzolo, Langhorst, et al. (2019) and Rodríguez et al. (2020).

4.1.3. Impact of Bias in the Prior Estimate of Mean Annual Flow

Similarly to the SWOT algorithm discharge benchmarking study by Durand et al. (2016), we found that a considerable fraction of the observed discharge NRMSE is caused by biases. Contrasting Figures 8c and 8d, we see that bias is larger than error in flow variations: the median ratios between the magnitude of the normalized biases and the normalized residual standard deviation for BAM, MetroMan, HiVDI, and SAD were: 1.5, 1.8, 2.2, and 1.7, respectively. Therefore, we searched for the factors that control the biases in the discharge estimated by the algorithms.

Similarly to Tuozzolo, Lind, et al. (2019), we found that discharge algorithm biases are sensitive to biases in the prior discharge estimates, as shown in Figure 9 panels a–d, which relate prior (Q_{WBM}) normalized biases with normalized biases in discharge estimated by BAM, MetroMan, HiVDI, and SAD. The correlation between biases in Q_{WBM} and in discharge estimated by BAM, MetroMan, HiVDI, and SAD were 0.95, 0.95, 0.98, 0.72, respectively, and the slopes of the fitted lines are 0.85, 0.73, 0.84, and 0.53, respectively. The high correlation between prior and posterior biases in three of the algorithms shows that one should expect the biases in SWOT discharge to be closely related to the prior biases, however, the regression slopes being less than one suggests that the McFLI and the data assimilation algorithms tend to reduce bias. Although this observation is based on regressions that contain influential points, they are corroborated by Figure 9 panels e–h, which superimpose histograms of the estimated biases by BAM, MetroMan, HiVDI, and SAD on the prior biases. The increased number of cases falling into the class centered around zero and the smaller range observed in all four histograms support the observation that the algorithms generally improve over the prior discharge. We next sought to identify factors that control whether discharge algorithms decrease prior biases.

4.1.4. Effect of Hydraulic Spatial Variability

Following the framework described in Section 3.3.2, we classified cases into high and low spatial variability, based on a threshold of the 66th percentile (sensitivity to choice of this threshold is explored using receiver operating characteristic curves, described subsequently). We quantified the improvement in normalized bias as the difference between the absolute values of the normalized biases for the prior estimate of discharge (Q_{wbm}) and the normalized biases in the discharge estimated by each algorithm. Positive values of improvement denote cases where the algorithm improves bias. Error boxplots for high and low variability shown in Figure 9 panels l–o, for BAM, MetroMan, HiVDI, and SAD, respectively. The comparison shown in Figure 9 panel l for BAM only represents the 15 cases with the smaller NRMSE errors caused by the static friction parameterization in BAM. Figure 9 shows that bias generally improves in the cases with high spatial variability, especially for MetroMan and HiVDI: the p-value for two sample t test on whether the high variability results have lower bias was 0.01 for MetroMan and 0.04 for HiVDI. Neither BAM nor SAD showed different performance at the 0.05 level ($p = 0.07$ and 0.79, respectively); while BAM could show significant differences if tested for additional cases, we hypothesize that other issues dominate performance for SAD.

We explored sensitivity to the choice of threshold for high versus low variability using receiver operating curves, described in Section 3.3.2. Figure 10a shows curves built for MetroMan, HiVDI, SAD and for BAM, the latter showing two curves, one for all the 31 cases and a second curve named "BAM filtered" considering only the 16 cases with the lowest level of flow law errors. The selection of an excessively stringent minimum variability threshold, for example, the 100th percentile, would fail to identify any cases as being highly variable. Such threshold would yield a true positive rate of 0 and a false positive rate of 0. As the threshold is relaxed, more cases are flagged as highly variable, leading to an increase in the true positive rate (tpr) and
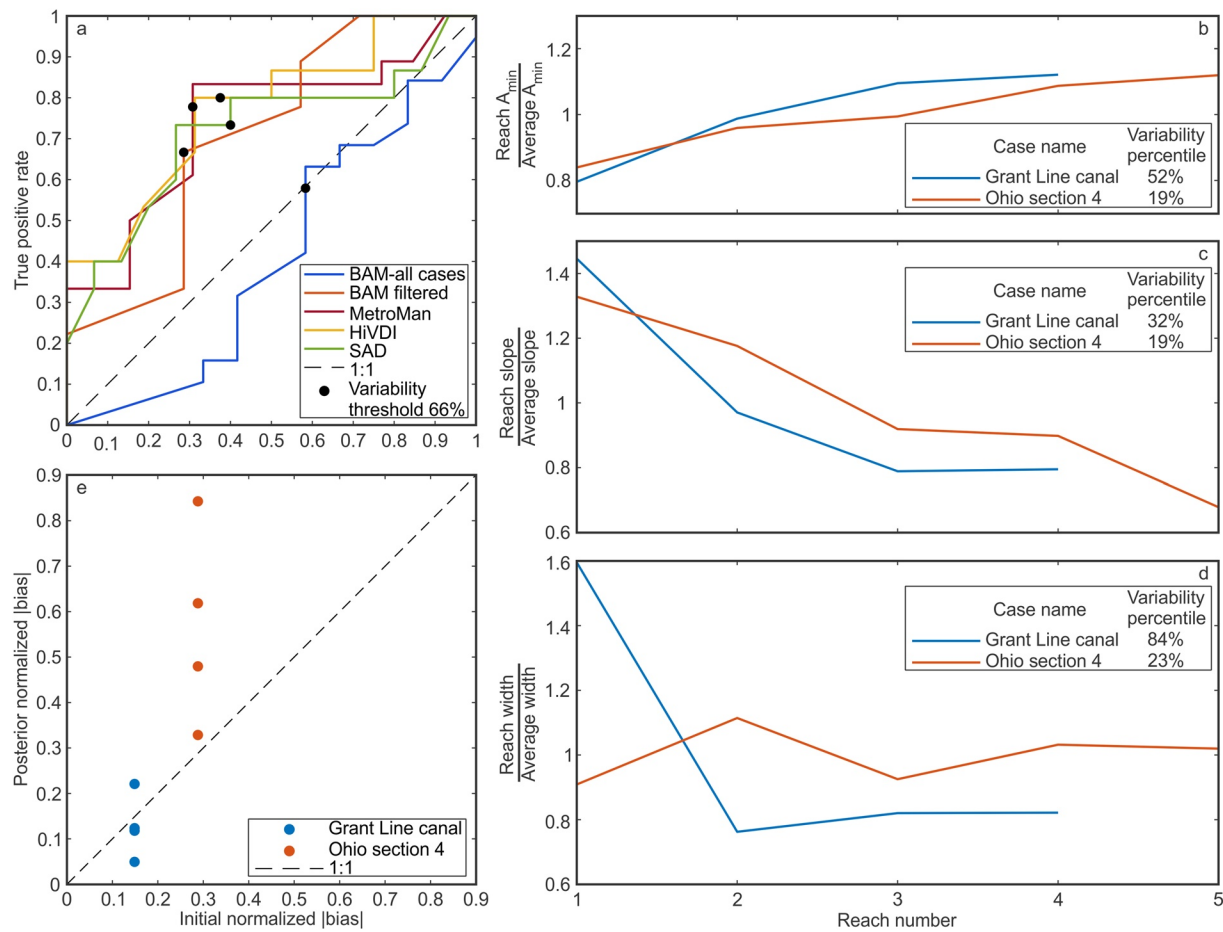
Figure 10. Panel a: Receiver operating characteristic curves for the four algorithms considering the degree of spatial variability as a predictor of inversion success. Lowering the minimum variability threshold will predict a higher number of successes, thus increasing both the number of true positives and false positives. Predictors following the 1:1 line are not better than guesses, therefore the further the predictor is from the 1:1 line, the better is the predictor. BAM is represented by two curves, one including all 31 cases and a second curve that only includes models where the normalized root mean square error due to Manning's equation considering a constant n for the set of reaches is less than 0.08, that is, the median value. Panels b–e illustrate how the predictor works. Both cases have hydraulic variability below the selected variability threshold for cross-sectional area at low flow (panel b) and slope (panel c), however, the coefficient of variation of reach widths for the Grant Line canal is at the 84th percentile (panel d), thus this case is classified as likely to succeed. The Ohio section 4 case does not show high variability of any of the three parameters; thus, it is classified as likely to worsen biases. Panel e: Grant Line canal case shows reduction in biases after running three of the four algorithms (represented by the four blue circles) whereas the Ohio section 4 case shows increased biases for all four algorithms. The Grant Line canal hydrograph is shown in Figure 6 panel c and Ohio River section 4 hydrograph in Figure 6 panel f.

often an increase in the false positive rate (fpr) as well. Under ideal circumstances, the number of false positives would remain at 0 until the threshold is overly relaxed. The curve for an ideal predictor begins at 0 tpr and 0 fpr, immediately rising to 1 tpr, 0 frp where it remains until the threshold is overly relaxed, when the curve moves toward 1 tpr, 1 fpr. For real curves, good predictors would stay above and as far from the 1:1 line (in terms of Euclidian distance) as possible, while curves tracking the 1:1 line or falling below would be as good as uninformed guesses.

The optimal threshold depends on the balance between an application's tolerance for false positives and the need to detect as many true positives as possible. In our study, we selected the threshold that would place the most algorithms as far away from the 1:1 line as possible, which happens at the 66th percentile. The use of the ROC for the selection of the optimal threshold has the purpose of decreasing the subjectivity involved in using an arbitrary threshold, thus allowing us to better examine how spatial variability affects the inversion process. Figures 10b–10d illustrate the process for two cases: Grant Line canal, a case that is identified as likely to reduce discharge biases (high spatial variability of hydraulic properties), and Section 4 of the Ohio River, which is classified as likely to increase discharge biases due to having reaches with low spatial

variability of hydraulic properties. Figure 10b shows the ratio of the reach averaged cross-sectional area at low flow (A_{min}) and the domain averaged A_{min} for both cases. When the coefficient of variation for A_{min} is contrasted to the other cases, we see that Grant Line canal ranks near the median for that criterion, while Ohio section 4 is at the 19th percentile. Therefore, both cases fail to reach the threshold for high variability in terms of cross-sectional area. The same happens for reach slope, however, when this exercise is repeated for width, we see that the coefficient of variation of width for the Grant Line canal is above the 84th percentile. Therefore, this case is classified as having high spatial variability of at least one of the criteria and is likely to decrease biases. Since section 4 of the Ohio River fails all three criteria, it is classified as likely to worsen discharge biases. This is illustrated in Figure 10e, where Grant Line canal shows smaller normalized posterior biases than those computed with the prior estimates of discharge for three of the algorithms, the exception being BAM, which increased the discharge bias from 0.15 to 0.22, whereas the Ohio section 4 shows larger posterior biases for all four algorithms.

A possible interpretation for the opposite behavior observed in the two curves made for BAM, one with and one without filtering the cases where the flow law had provided worse fits to the data, is that the spatial variability of the hydraulic properties provides the information used by the inversion algorithms to find the flow law parameters, while errors in the flow law add confounding effects that can overwhelm the amount of available information. This interpretation is also supported by the difference in NRMSE between the best and worst flow law fit groups shown in Figure 9i. The filtering was neither necessary nor beneficial to HiVDI and MetroMan, likely because the level of error associated with the version of the Manning-Strickler flow law modified to incorporate depth dependency (see Equations 3 and 5) was not sufficient to obscure the signal present in the spatial variability of hydraulic properties.

While a predictor that simultaneously considers the spatial variability of the cross-sectional area at low flow, the water surface slope, and width provides a more complete picture than a predictor that only considers one of the three properties, the need to know the cross-sectional area precludes the use of our framework in large scales, as information on channel bathymetry is often unavailable. Future work should also search for quantitative relationships between hydraulic variability and the magnitude of bias improvement or worsening, which would allow the prediction of posterior biases in discharge, as such biases constitute a large fraction of overall discharge errors and their knowledge is essential to fulfilling the goal of improving the current state of knowledge of global runoff. The development of quantitative relationships between hydraulic variability and bias improvement over prior estimates requires a specific experimental protocol featuring several synthetic cases covering a large range in the degree of hydraulic variability, far beyond the 31 cases used in our study. Ultimately, our predictor is useful for diagnosing inversions and for bringing insight into the factors that lead to inversion success, with an important potential use for the identification of optimal reach definition strategies to maximize the chance of inversion success.

4.2. Phase 2: Gradual Degradation in Sampling Frequency Without Measurement Uncertainty

This phase allowed us to examine the robustness of the algorithms to decreasing time sampling without the confounding effect of measurement uncertainty. Figure 11 shows the different error metrics organized in rows with one column for each algorithm. For the Kling-Gupta Efficiency and its three associated elements see Figure S7 in the supplemental information. BAM had the most consistent performance of all algorithms at all sampling frequencies, showing no degradation as data became sparser in time. Despite the apparent increase in NSE, and decrease in normalized biases and NRMSE, one-way analysis of variance (ANOVA) test across the different time sampling settings for the error metrics failed to show any significant differences, with p-values of 0.37, 0.94, 0.98 for NSE, NRMSE, and normalized biases, respectively. The robustness to sampling frequency could be due to the way BAM simplifies the continuity equation, ignoring the time derivative of the cross-sectional area, leading to the assumption that at a given point in time, discharge is identical in all reaches.

HiVDI and MetroMan keep the time derivative term of the continuity equation, approximating it as the difference in cross-sectional area between two consecutive passes divided by the time interval between the passes. This assumption introduces a dependency on sampling frequency and could explain the modest decrease in NSE and increase in normalized bias for MetroMan when the revisit period is longer than 7 days. Visual inspection suggests a degradation in performance was more pronounced for HiVDI, with
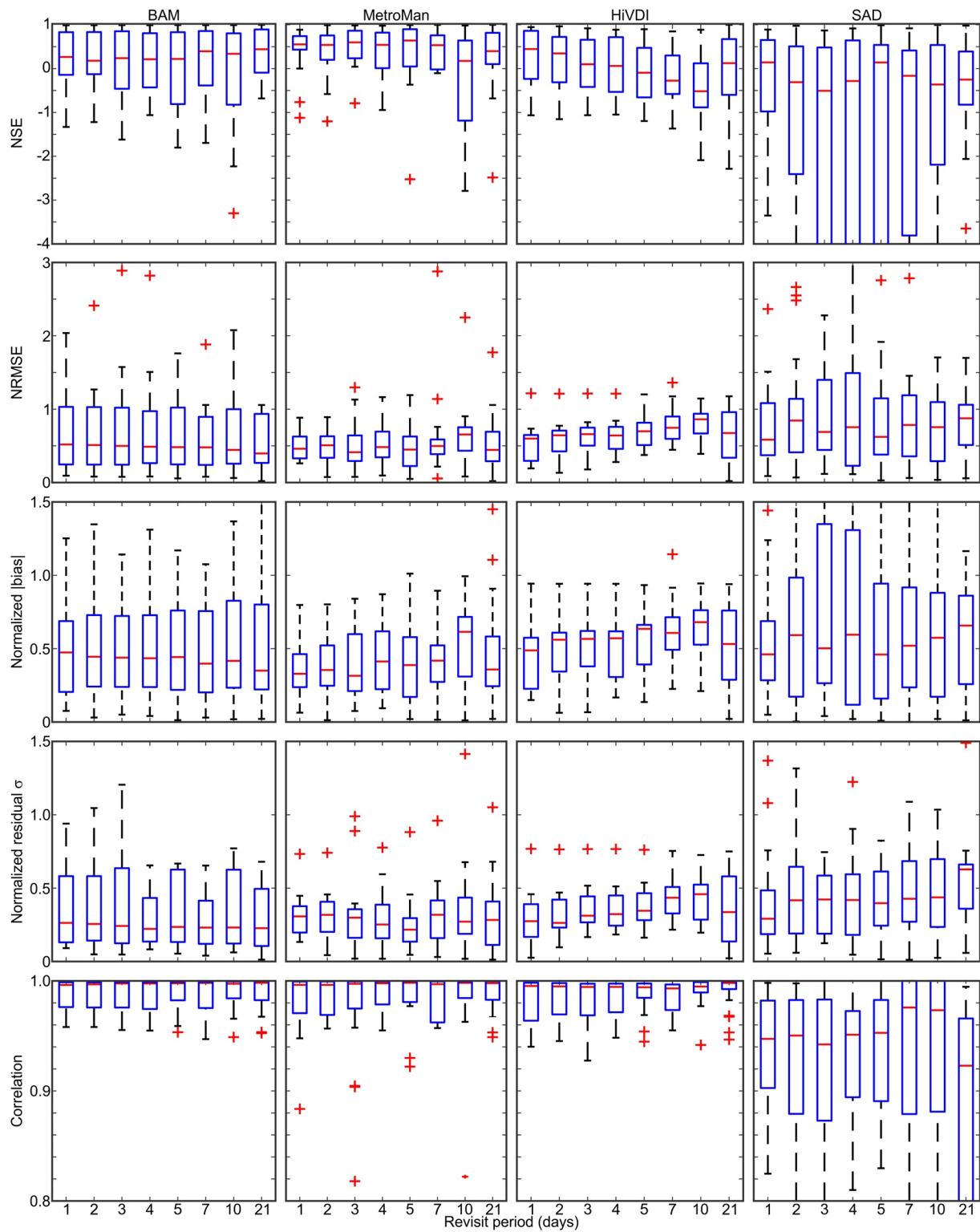
Figure 11. Effect of time sampling on inversion metrics. Each column represents an inversion algorithm and the rows represent the different metrics. First row: Nash-Sutcliff Efficiency, second row: normalized root mean square errors, third row: absolute value of the normalized bias, fourth row: normalized standard deviation of residuals, and fifth row: correlation.

error metrics worsening for revisit periods equal to or longer than 3 days. Based on one-way ANOVA tests performed across the different sampling frequencies, we observed that the changes in NSE and NRMSE were both statistically significant at the 0.05 level (*p*-values of 0.009 for both NSE and NRMSE). However similar tests were inconclusive for normalized biases and residual standard deviation and correlation, which resulted in p-values of 0.52, 0.14, 0.70, respectively. For SAD, the only metric that showed statistically significant changes at the 0.05 level was correlation (p-value of 0.02), the remaining metrics had p-values in excess of 0.3. As for MetroMan, none of the metrics showed statistically significant changes, with the lowest p-values corresponding to the NRMSE metric (0.31). We conclude that for two out of the four examined algorithms, temporal sampling does not play a major role in algorithm accuracy, with a third algorithm (SAD) only showing obvious degradation at the sparsest time sampling.

4.3. Phase 3: Increasing Measurement Error

Figure 12 summarizes the sensitivity of the error metrics for the four algorithms to four levels of observational error: no error, half the expected SWOT errors, the full amount of SWOT errors, and 1.5 times the expected SWOT errors, for the 16 cases identified in Table 2. The Kling-Gupta Efficiency and its three associated elements are included in Figure S8 in the supplemental information.

Measurement uncertainty can affect discharge estimation in two ways: (a) it will corrupt observations of reach properties, decreasing the amount of information available to the discharge algorithms and affecting their ability to retrieve flow law parameters and (b) it propagates directly through the chosen flow law into the estimated discharge. Both effects are visible for the SAD algorithm (fourth column in Figure 12), with increasing normalized biases (third row) and increasing normalized residual standard deviation (fourth row) with increasing observational uncertainty. Increasing magnitudes of bias and random errors lead to degrading NSE and increasing NRMSE (first and second rows respectively).

In terms of NSE, NRMSE, normalized biases, and residual standard deviations, BAM, MetroMan and HiVDI show little change in either median values or interquartile range with increasing measurement errors. BAM and MetroMan show a small degradation in the correlation between estimated discharges and true discharges, however, one-way ANOVA tests performed across the error metrics yield *p*-values larger than 0.3 for all error metrics for both algorithms. HiVDI shows slight degradation of NSE and NRMSE with the inclusion of observational uncertainties. These were mostly caused by increases in the normalized residual standard deviation, as the magnitude of normalized biases remained largely unchanged.

4.4. Future and Ongoing Work

It is important to remember that discharge error metrics and sensitivities presented in this study are a worst-case scenario benchmark, because the amount of prior information provided to each algorithm was limited. Thus, the performance of SAD and BAM may have been affected by the limited amount of ancillary information given to the algorithms. Because BAM expects the specification of different flow quantiles other than the mean, which were not available in this study, we had to make severely restrictive assumptions to compensate, for example, setting a coefficient of variation of 1 for the prior distribution of discharge for BAM. For SAD, the absence of information on geomorphology forced us to use a less informative prior for the shape parameter. Our experiment was designed to reflect the worst-case scenario for the algorithms when they operate at an extremely sparse landscape of prior information (just giving mean annual flow). Because HiVDI and BAM operate better when information on the expected dynamic range of discharge is available, we are working on extracting more flow quantiles from the WBMSed simulations. In operational settings, more prior information may be available and we expect that the lessons learned from this exercise will be put into practice to improve algorithms in the future.

Additionally, our experimental protocol required a strict version control: we did not rerun our algorithms to improve upon the results here even when algorithm improvements were suggested by poor results (e.g., in the case of the MOMMA algorithm results shown in Section 4.1.1). Thus, this analysis represents a snapshot in time of the evolution of SWOT discharge algorithms. Inspired by the results of this study, the BAM algorithm is undergoing improvements, including a new parameterization of the friction coefficient that allows it to vary in both space and time and the incorporation of geomorphological information as prior
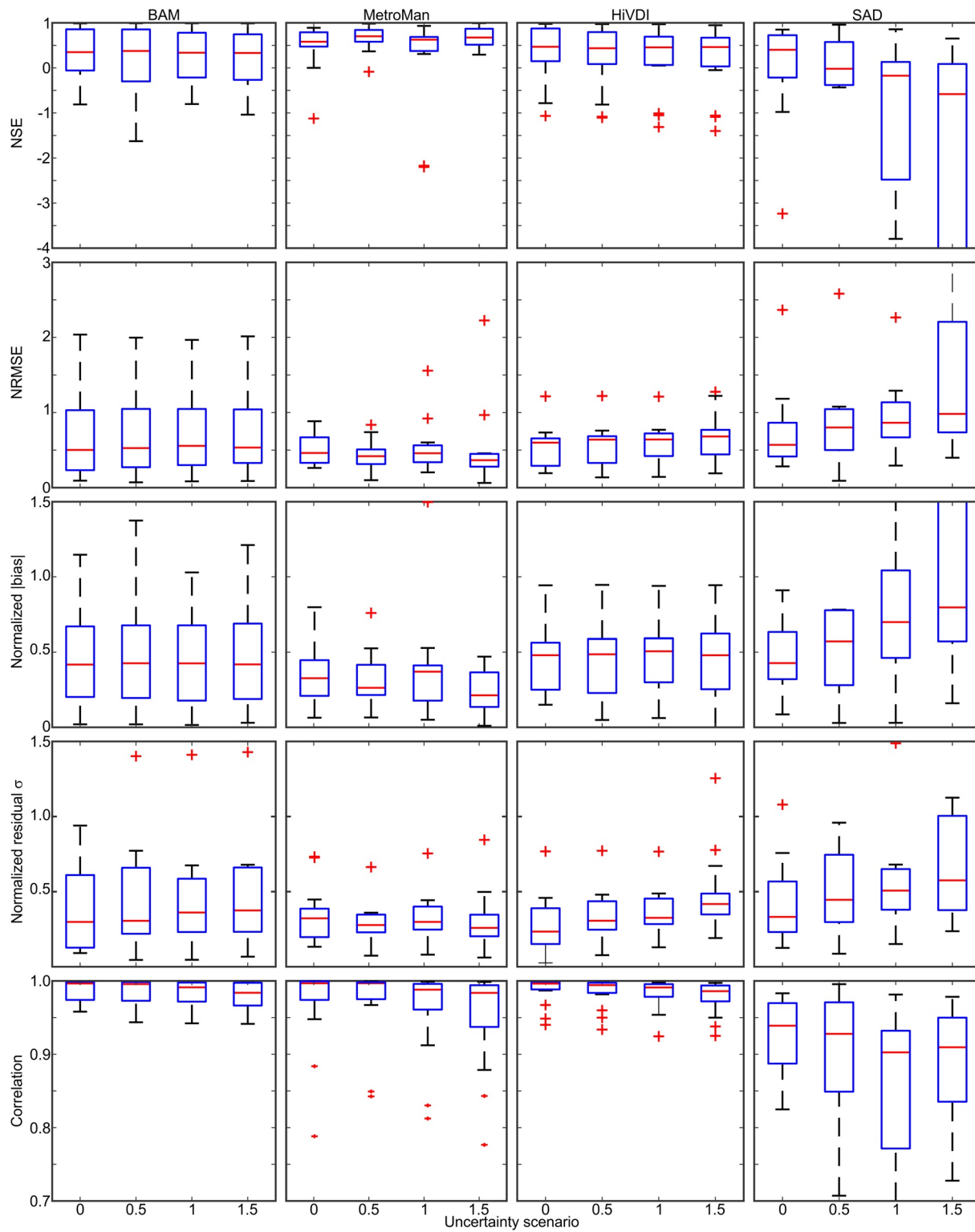
Figure 12. Effect of the Surface Water and Ocean Topography (SWOT) measurement error on inversion metrics. Each column represents an inversion algorithm and the rows represent the different metrics. First row: Nash-Sutcliff Efficiency, second row: normalized root mean square errors, third row: absolute value of the normalized bias, fourth row: normalized standard deviation of residuals, and fifth row: correlation.

information. Furthermore, other promising techniques such as machine learning, able to take advantage of a wider variety of measurements, including more extensive use of in situ observations, may be applied during or after the end of the SWOT mission by members of the community. However, due to time and computational constrains, such methods are not expected to be included in the early versions of the SWOT discharge product.

Synthetic studies such as Bonnema et al. (2016), Durand et al. (2016), as well as the present study require detailed hydraulic models to be executed. Despite the insightful results that are only possible in the presence of hydraulic river models, the extensive data requirements limit the number of test cases that can be realistically evaluated. SWOT observations of calibration sites with collocated in situ gages to be collected during the calibration and validation phase of the mission will be of utmost importance to increase the number of rivers where discharge inversion algorithms can be evaluated, thus producing a more complete picture of the SWOT discharge performance.

Furthermore, future work should also strive to evaluate discharge uncertainty estimates, as such knowledge is essential for the proper assimilation of SWOT discharge into hydrological and land surface models. Examples of assimilation of discharge estimated with data collected by altimeters can be found in the literature, for example, Emery et al. (2018) who assimilated ENVISAT-derived discharge into the ISBA-CTRIP land surface model (Decharme et al., 2019) through the use of ensemble Kalman filter. Despite discharge not being directly measured by ENVISAT, Emery and coauthors found assimilation of the ENVISAT discharge to be effective in reducing the overall discharge errors across the Amazon Basin. More recently, Wongchuig-Correa et al. (2020) evaluated the assimilation of synthetic SWOT observations of water surface elevation, width, and the estimated discharge into a continental scale hydrologic model which explicitly represents hydrodynamic routing through a basin. Wongchuig-Correa and coauthors not only observed that the assimilation of SWOT discharge was beneficial, but also that the assimilation of discharge, water surface elevation, and width together yielded better improvements than assimilating just water surface elevation, width or discharge alone. Such findings suggest that SWOT discharge products have potential for improving the modeling and prediction of continental scale runoff, particularly when assimilated into models that do not explicitly account for hydrodynamic routing, which precludes the assimilation of direct SWOT observations such as inundation extent and water surface elevations.

5. Conclusion

We have made important progress in understanding the underlying controls on SWOT discharge algorithm accuracy and have tested algorithms with realistic temporal sampling and observation error in a context mimicking algorithm deployment in ungaged basins. Knowledge of the factors that control algorithm accuracy guide further algorithm development and enable the global estimation of SWOT discharge uncertainty. Our results reveal that accuracy of the prior estimates of mean annual flow, flow law adequacy, and hydraulic variability are more important to the final discharge accuracy than either revisit time or observation error. This is encouraging as further studies are likely to identify improvements to flow laws, better and more complete sources of prior information, and the optimal selection of mass-conserved sets of reaches.

Consistent with previous studies, we show that discharge algorithms are better at tracking temporal variations in discharge than they are at estimating absolute discharge. Under ideal conditions, the ability to track flow dynamics described by the median normalized standard deviation of residuals computed across the 31 cases was benchmarked at between 0.19 (HiVDI) and 0.29 (SAD) in this study, excluding the MOMMA algorithm, whereas median absolute normalized biases varied from 0.34 (MetroMan) to 0.47 (SAD). Such results suggest that the current generation of algorithms designed for the SWOT mission will be invaluable for the characterization of discharge dynamics particularly in ungaged basins. However, the ability to provide accurate representation of climatology, in terms of mean discharge and the amplitude of its variation, will depend on the quality of the prior estimate of mean annual flow or the presence of ancillary information and its quality. Indeed, because only prior information on mean annual flow was provided to algorithms in this study, and because algorithms are evolving to use more prior information, the results presented in this study are essentially a worst-case benchmark.

Although we found discharge algorithm biases to be dependent on biases present in the prior estimate of mean annual flow, the algorithms consistently, even if moderately, improved upon prior biases. Algorithm performance was found to further depend on both the accuracy of the flow law parameterization used in each algorithm, and on hydraulic variability, that is, mass-conserved sets of reaches that contain higher spatial variability of slope, width, and cross-section area help inversions, whereas flow law errors hinder inversions.

In contrast, our experiments showed little sensitivity to measurement error or temporal revisit time in the discharge algorithm results. Nevertheless, the interaction of sparse temporal sampling with high observation errors could lead to unexpected poor algorithm performance, particularly in cases where the temporal sampling is not sufficient to resolve variations in flow conditions reflected in the observed heights, widths, and slopes, which is needed by the inversion algorithms. This unfavorable condition could be exacerbated by the measurement error associated with the SWOT observations by preventing us from resolving the temporal variations that remain in the sparsely sampled observations. Further experimentation can identify the existence and the magnitude of such interactions. However, our results suggest that the quality of the prior estimates of mean annual flow, the maximization of hydraulic variability in the inversion domain, and the accuracy of the flow law when applied to remote sensing-relevant scales are likely to be more important determinants of inversion performance.

This study has four important implications. First the optimal strategy for reach definition should strive to maximize hydraulic variability between reaches, since such variability improves the ability to estimate river discharge using McFLI and assimilation algorithms. Second, future work should explore retrieval studies in controlled contexts in order to better understand the inversion problem itself with the purpose of developing quantitative relationships between measurable reach properties and bias reduction. Third, given the importance of the prior estimates of mean annual flow, this study highlights the need to identify the best possible sources of prior information, which should preferably contain smaller biases and include all the elements needed by the discharge inversion algorithms. Lastly, our results pave the way to build the capability to predict SWOT discharge algorithm performance, and thus map expected algorithm performance globally.

Data Availability Statement

The full datasets needed for the reproduction of the work can be freely downloaded from https://zenodo.org/record/3817817 under a creative commons attribution 4.0 license.

References

Adams, T., Chen, S., David, R., Shade, T., & Lee, D. (2010). *The Ohio River Community HEC-RAS Model* (Techinical Report). American Society of Civil Engineering. https://doi.org/10.1061/41114(371)160

Andreadis, K. M., Brinkerhoff, C. B., & Gleason, C. J. (2020). Constraining the assimilation of SWOT observations with hydraulic geometry relations. *Water Resources Research*, *56*. 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*(10). https://doi.org/10.1029/2007gl029721

Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodríguez, E., Mognard, N. M., et al. (2010). Preliminary characterization of SWOT hydrology error budget and global capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 6–19. https://doi.org/10.1109/jstars.2009.2034614

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (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., Dingman, S. L., Vorosmarty, C. J., Bolster, C. H., & Congalton, R. G. (2003). Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, *278*(1–4), 17–38. https://doi.org/10.1016/s0022-1694(03)00129-x

Bjerklie, D. M., Fulton, J. W., Dingman, S. L., Canova, M. G., Minear, J. T., & Moramarco, T. (2020). Fundamental hydraulics of cross sections in natural rivers: Preliminary Analysis of a large data set of acoustic doppler flow measurements. *Water Resources Research*, *56*(3), e2019WR025986. https://doi.org/10.1029/2019WR025986

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*(4), 2439–2461. https://doi.org/10.1002/2015WR017296

Brinkerhoff, C. B., Gleason, C. J., Feng, D., & Lin, P. (2020). Constraining remote river discharge estimation using reach-scale geomorphology. *Water Resources Research*, *56*(11), e2020WR027949. https://doi.org/10.1029/2020WR027949

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*(16), 9637–9647. https://doi.org/10.1029/2019GL084529

Brisset, P., Monnier, J., Garambois, P.-A., & Roux, H. (2018). On the assimilation of altimetric data in 1D Saint-Venant river flow models. *Advances in Water Resources*, *119*, 41–59. https://doi.org/10.1016/j.advwatres.2018.06.004

Canova, M. G., Fulton, J., & Bjerklie, D. M. (2016). *USGS HYD Roacoustic dataset in support of the Surface Water Oceanographic Topography satellite mission (HYDRoSWOT): U.S. Geological Survey data release*. https://doi.org/10.5066/F7D798H6

Chow, V. (1959). *Open-channel hydraulics*. Mc Graw-Hill.

Cohen, S., Kettner, A. J., & Syvitski, J. P. M. (2014). Global suspended sediment and water discharge dynamics between 1960 and 2010: Continental trends and intra-basin sensitivity. *Global and Planetary Change*, *115*, 44–58. https://doi.org/10.1016/j.gloplacha.2014.01.011

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J. P., Alias, A., et al. (2019). Recent changes in the ISBA-CTRIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1207–1252. https://doi.org/10.1029/2018MS001545

Desai, S. (2018). *Surface Water and Ocean Topography mission (SWOT), science requirements document (JPL document D-61923 revision B)*. Retrieved from https://swot.jpl.nasa.gov/system/documents/files/2176_2176_D-61923_SRD_Rev_B_20181113.pdf

Domeneghetti, A., Schumann, G. J.-P., Frasson, R. P. M., Wei, R., Pavelsky, T. M., Castellarin, A., et al. (2018). Characterizing water surface elevation under different flow conditions for the upcoming SWOT mission. *Journal of Hydrology*, *561*, 848–861. https://doi.org/10.1016/j.jhydrol.2018.04.046

Dudley, R. W. (2004). *Hydraulic-geometry relations for rivers in coastal and central Maine*. US Department of the Interior, US Geological Survey. https://doi.org/10.3133/sir20045042

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*(20). https://doi.org/10.1029/2008gl034150

Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., et al. (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

Durand, M., Neal, J., Rodríguez, E., Andreadis, K. M., Smith, L. C., & Yoon, Y. (2014). Estimating reach-averaged discharge for the River Severn from measurements of river water surface elevation and slope. *Journal of Hydrology*, *511*, 92–104. https://doi.org/10.1016/j.jhydrol.2013.12.050

Durand, M., Rodríguez, E., Alsdorf, D. E., & Trigg, M. (2010). Estimating river depth from remote sensing swath interferometry measurements of river height, slope, and width. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 20–31. https://doi.org/10.1109/jstars.2009.2033453

Durand, M. T., Chen, C., Frasson, R. P. D. M., Pavelsky, T., Williams, B., Yang, X., & Fore, A. (2020). How will radar layover impact SWOT measurements of water surface elevation and slope, and estimates of river discharge? *Remote Sensing of Environment*. https://doi.org/10.1109/igarss39084.2020.9323156

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*(4), 2135–2162. https://doi.org/10.5194/hess-22-2135-2018

Esteban-Fernandez, D. (2013). *SWOT mission performance and error budget (JPL D-79084 Revision A)*. Retrieved from http://swot.jpl.nasa.gov/files/SWOT_D-79084_v5h6_SDT.pdf

Even, S., Mouchel, J.-M., Servais, P., Flipo, N., Poulin, M., Blanc, S., et al. (2007). Modeling the impacts of combined sewer overflows on the river Seine water quality. *The Science of the Total Environment*, *375*(1), 140–151. https://doi.org/10.1016/j.scitotenv.2006.12.007

Even, S., Poulin, M., Garnier, J., Billen, G., Servais, P., Chesterikoff, A., & Coste, M. (1998). River ecosystem modeling: Application of the PROSE model to the Seine river (France). *Hydrobiologia*, *373*(0), 27–45. https://doi.org/10.1023/A:101704552233610.1007/978-94-011-5266-2_3

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fjortoft, R., Gaudin, J.-M., Pourthie, N., Lalaurie, J.-C., Mallet, A., Nouvel, J.-F., et al. (2014). KaRIn on SWOT: Characteristics of near-Nadir Ka-band interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(4), 2172–2185. https://doi.org/10.1109/tgrs.2013.2258402

Frasson, R. P. D. M., Durand, M. T., Larnier, K., Gleason, C. J., Andreadis, K. M., Hagemann, M., et al. (2020). *Synthetic river datasets built for testing and development of the Surface Water and Ocean Topography mission discharge algorithms*. Retrieved from https://zenodo.org/record/3817817

Frasson, R. P. D. M., Schumann, G. J. P., Kettner, A. J., Brakenridge, G. R., & Krajewski, W. F. (2019). Will the Surface Water and Ocean Topography (SWOT) satellite mission observe floods? *Geophysical Research Letters*, *46*, 10435–10445. https://doi.org/10.1029/2019gl084686

Frasson, R. P. D. M., Wei, R., Durand, M., Minear, J. T., Domeneghetti, A., Schumann, G., et al. (2017). Automated river reach definition strategies: Applications for the Surface Water and Ocean Topography Mission. *Water Resources Research*, *53*(10), 8164–8186. https://doi.org/10.1002/2017wr020887

Garambois, P.-A., Larnier, K., Monnier, J., Finaud-Guyot, P., Verley, J., Montazem, A.-S., & Calmant, S. (2020). Variational estimation of effective channel and ungauged anabranching river discharge from multi-satellite water heights of different spatial sparsity. *Journal of Hydrology*, *581*, 124409. https://doi.org/10.1016/j.jhydrol.2019.124409

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

Gilles, D., Young, N., Schroeder, H., Piotrowski, J., & Chang, Y.-J. (2012). Inundation mapping initiatives of the Iowa Flood Center: State-wide coverage and detailed urban flooding analysis. *Water*, *4*(1), 85–106. https://doi.org/10.3390/w4010085

Gleason, C. J., & Durand, M. T. (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., & 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 of the United States of America*, *111*(13), 4788–4791. https://doi.org/10.1073/pnas.1317606111

Gleason, C. J., & Wang, J. (2015). Theoretical basis for at-many-stations hydraulic geometry. *Geophysical Research Letters*, *42*(17), 7107–7114. https://doi.org/10.1002/2015GL064935

Grippa, M., Rouzies, C., Biancamaria, S., Blumstein, D., Cretaux, J.-F., Gal, L., et al. (2019). Potential of SWOT for monitoring water volumes in Sahelian ponds and lakes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*, 2541–2549. https://doi.org/10.1109/JSTARS.2019.2901434

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*(11), 9692–9707. https://doi.org/10.1002/2017WR021626

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

Lang, S., Ladson, T., & Anderson, B. (2004). A review of empirical equations for estimating stream roughness and their application to four streams in Victoria. *Australian Journal of Water Resources*, *8*(1), 69–82. https://doi.org/10.1080/13241583.2004.11465245

Larnier, K., Monnier, J., Garambois, P.-A., & Verley, J. (2020). River discharge and bathymetry estimation from SWOT altimetry measurements. In *Inverse problems in science and engineering* (pp. 1–31). https://doi.org/10.1080/17415977.2020.1803858

Leopold, L. B., & Maddock, T. (1953). The hydraulic geometry of stream channels and some physiographic implications. *Geological Survey Professional Paper*, *252*, 64.

Limerinos, J. T. (1970). *Determination of the Manning coefficient from measured bed roughness in natural channels*. Retrieved from https://pubs.usgs.gov/wsp/1898b/report.pdf

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Maswood, M., & Hossain, F. (2016). Advancing river modeling in ungauged basins using satellite remote sensing: The case of the Ganges-Brahmaputra-Meghna basin. *International Journal of River Basin Management*, *14*(1), 103–117. https://doi.org/10.1080/15715124.2015.1089250

Monnier, J., Larnier, K., & Garambois, P. A. (2018). Discharge and bathymetry estimations of rivers from SWOT like data: The hierarchical variational discharge inference (HiVDI) algorithm. In *Paper presented at the A.G.U. Fall Meeting*.

Montazem, A. S., Garambois, P. A., Calmant, S., Finaud-Guyot, P., Monnier, J., Moreira, D. M., et al. (2019). Wavelet-based river segmentation using hydraulic control-preserving water surface elevation profile properties. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL082986

Oubanas, H., Gejadze, I., Malaterre, P. O., 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

Oubanas, H., Gejadze, I., Malaterre, P.-O., & Mercier, F. (2018). River discharge estimation from synthetic SWOT-type observations using variational data assimilation and the full Saint-Venant hydraulic model. *Journal of Hydrology*, *559*, 638–647. https://doi.org/10.1016/j.jhydrol.2018.02.004

Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H., & Miller, Z. F. (2014). Assessing the potential global extent of SWOT river discharge observations. *Journal of Hydrology*, *519*, 1516–1525. https://doi.org/10.1016/j.jhydrol.2014.08.044

Rodríguez, E., Durand, M., & Frasson, R. P. D. M. (2020). Observing rivers with varying spatial scales. *Water Resources Research*, *56*. https://doi.org/10.1029/2019WR026476

Rogers, W. (2014). *Central valley floodplain evaluation and delineation, subtask 5: Combined Sacramento River system model*.

Rosen, P. A., Hensley, S., Joughin, I. R., Li, F. K., Madsen, S. N., Rodriguez, E., & Goldstein, R. M. (2000). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, *88*(3), 333–382. https://doi.org/10.1109/5.838084

Sikder, M. S., & Hossain, F. (2015). Understanding the Geophysical Sources of Uncertainty for Satellite Interferometric (SRTM)-based discharge estimation in river deltas: The case for Bangladesh. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(2), 523–538. https://doi.org/10.1109/JSTARS.2014.2326893

Tuozzolo, S., Langhorst, T., de Moraes Frasson, R. P., Pavelsky, T., Durand, M., & Schobelock, J. J. (2019). The impact of reach averaging Manning's equation for an in-situ dataset of water surface elevation, width, and slope. *Journal of Hydrology*, *578*, 123866. https://doi.org/10.1016/j.jhydrol.2019.06.038

Tuozzolo, S., Lind, G., Overstreet, B., Mangano, J., Fonstad, M., Hagemann, M., et al. (2019). Estimating river discharge with swath altimetry: A proof of concept using AirSWOT observations. *Geophysical Research Letters*, *46*(3), 1459–1466. https://doi.org/10.1029/2018GL080771

Ulaby, F. T., & Dobson, M. C. (1989). *Radar Scattering statistics for terrain*. Artech House Inc.

USACE. (2015). *Missouri river recovery management plan and environmental impact statement: Existing conditions, unsteady HEC-RAS model calibration report*. Retrieved from https://usace.contentdm.oclc.org/digital/collection/p16021coll7/id/7886

Wongchuig-Correa, S., Cauduro Dias de Paiva, R., Biancamaria, S., & Collischonn, W. (2020). Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models. *Journal of Hydrology*, *590*, 125473. https://doi.org/10.1016/j.jhydrol.2020.125473

Yoon, Y., Garambois, P. A., Paiva, R. C. D., Durand, M., Roux, H., & Beighley, E. (2016). Improved error estimates of a discharge algorithm for remotely sensed river measurements: Test cases on Sacramento and Garonne rivers. *Water Resources Research*, *52*(1), 278–294. https://doi.org/10.1002/2015wr017319