Constraining Remote River Discharge Estimation Using Reach-Scale Geomorphology

C.B. Brinkerhoff1\*, C.J. Gleason1, D. Feng1, P. Lin2

1 Department of Civil & Environmental Engineering, University of Massachusetts, Amherst MA

2 Department of Civil & Environmental Engineering, Princeton University, Princeton NJ

\*Corresponding author: Craig Brinkerhoff (cbrinkerhoff@umass.edu)

Key Points:

* Discharge prediction accuracy is substantially increased by improving the quality of prior information, tested on both SWOT-simulated data and Landsat imagery

**Keywords:** remote sensing, river discharge, fluvial geomorphology, open-channel flow, McFLI

Abstract

Remote sensing of river discharge is presently possible in any river on Earth, but the algorithms used to do rely upon prior understandings of river geomorphology and hydraulics. Our current methods for estimating prior knowledge on rivers is not necessarily reflective of the physics experienced in a river. To address this problem, we studied the sensitivity of algorithm performance due to the quality of the prior information received by two algorithms capable of estimating discharge in uncalibrated settings: Bayesian-AMHG-Manning’s (BAM- Hagemann et al. 2017). We trained new models for estimating prior river knowledge on a dataset of over 370,000 field hydraulics measurements and statistically clustered geomorphology in order to prescribe prior river knowledge curated specifically for the river at hand, using only remotely-sensed observations. These interventions were tested by running BAM on the entire Mackenzie River basin, where median relative improvement in Nash-Sutcliffe Efficiency (NSE) was 72%; we also ran BAM and MetroMan on NASA Surface Water and Ocean Topography (SWOT)-simulated rivers and median NSE across rivers (for BAM) improved by 9%. For BAM, almost all improvement came in high flow events, as opposed to baseflow. These findings are implementable in any related algorithm and will play a vital role in both current attempts at global remotely-sensed river discharge as well as future discharge estimates from SWOT, where prior knowledge is essential to estimating flow in ungauged rivers where nothing is known *a priori*.

1 Introduction

In recent decades, remote sensing (RS) of rivers has flourished as a sub-field within fluvial geomorphology and hydrology. At the global scale, RS of rivers is changing current perceptions of rivers and their role in the earth system: globally modeled hydrography at fine-spatial scales (Yamazaki et al. 2019; Lehner et al. 2008), daily runoff routed through almost 3 million river reaches over 30 years (Lin et al. 2019), assessments of rivers and climate (Yang et al. 2020), water quality (Ross et al. 2019), surface area (Allen & Pavelsky, 2018), and hydrological connectivity (Grill et al. 2019) have all debuted recently. These examples, along with similar recent work quantifying global fluvial geomorphic patterns (Chen et al. 2019; Frasson et al. 2019), suggest that RS is coming of age in its ability to provide global scale data that honors local differences in rivers. These ideas will be further explored with the launch of the Surface Water and Ocean Topography (SWOT) satellite in 2021, which is expected to provide measurements of water surface elevation and extent at unprecedented spatial and temporal resolutions (Biancamaria et al. 2016).

A particular subset of this literature is showing that global RS of river discharge (RSQ) is presently possible with some gauging information in hand and should be globally possible in ungauged basins in the near future (Gleason and Durand, 2020). In basins with stream gauges or extensive field-measurements, RSQ approaches calibrate RS to local channel hydraulics (e.g. Brackenridge et al. 2007; LeFavour & Alsdorf 2005; Pavelsky, 2014; Pavelsky & Smith, 2009; Tarpanelli et al. 2013) or introduce RS data into hydrologic or hydraulic models (e.g. Bjerklie et al. 2005; Chandanpurker et al., 2017; King et al. 2018; Lin et al. 2019; Neal et al. 2009; Silvestro et al., 2015; Siquera et al., 2018; Zhang et al., 2016). These approaches (i.e., merging *in situ* and RS data) yield good predictive accuracy and can extend existing gauge records in space and time. In ungauged settings however, there are no gauge records to extend. Ground-based knowledge would improve RSQ accuracy in these scenarios, but in lieu of such information these methods must produce reasonably accurate results without relying on *in situ* knowledge (Gleason & Durand, 2020). In ungauged settings, standard practice is again to introduce RS data into hydrologic models (e.g. Emery et al 2018; Sun et al. 2015) or hydraulic models (e.g. Andreadis et al. 2007; Biancamaria et al. 2011; Durand et al. 2008; Yoon et al. 2012). The most recent and sophisticated methods for assimilating RS into hydraulic models (Larnier et al. 2019; Oubanas et al. 2018a; b) are highly accurate in ungauged settings but computationally burdensome for global application (Gleason & Durand, 2020).

A recent branch of RSQ has emerged with global application, SWOT, and ungauged basins in mind. This approach is termed Mass Conserved Flow-Law Inversion or McFLI (Gleason et al., 2017). McFLIs assume a river reach is mass conserved and then inversely solve for the unknown parameters in a flow law given some set of RS observations. This means that no hydrologic or hydraulic model is necessary and dischargeis exclusively estimated from RS by inverting basic geomorphic theories. McFLIs are therefore defined by their flow laws. To date, all published McFLIs have used either Manning’s equation (Bjerklie et al. 2018; Durand et al. 2014; Garambois & Monnier 2015; Hagemann et al. 2017; Sichangi et al. 2018) or at-many-stations hydraulic geometry (AMHG; Gleason et al. 2014; Hagemann et al. 2017; Feng et al., 2019) as a flow law, where AMHG reflects relationships between at-a-station hydraulic geometry (AHG) parameters along a river’s course (Gleason & Smith, 2014; Brinkerhoff et al., 2019). Regardless of the geomorphology driving McFLI, all McFLIs suffer from equifinality, as multiple sets of flow law variables can solve the inversion in this ill-posed estimation problem (Garambois & Monnier, 2015).

At the core of McFLI inversion is a reliance on initial guesses for parameters not observable from RS, termed ‘priors’ in Bayesian parlance. For example, to invert Manning’s equation, priors are generally needed for discharge, channel roughness, and channel cross-sectional area. These priors have previously been estimated from global hydrologic model output (Durand et al. 2016; Bonnema et al., 2016; Feng et al., 2019) and/or from external training data of geomorphic and hydraulic variables (e.g. Canova et al. 2016; Hagemann et al., 2017). Priors take the form of a probability distribution of these RS-unobservable parameters. If *in situ* data are available, then priors have extremely low variance: for example we might know channel roughness a priori. The less certain we are about a parameter a priori, the wider the distribution. Durand et al. (2016) found that McFLIs are sensitive to their priors in a test of five McFLIs on simulated SWOT observations (as SWOT has not launched, McFLIs are tested on ‘SWOT-like’ simulated data). In a similar comparison of algorithms using simulated rivers, Bonnema et al. (2016) found that AMHG inversion is particularly sensitive to its priors, and Tuozzolo et al. (2019) found that McFLI estimation bias is sensitive to the discharge prior in the first test of McFLIs from airborne Ka-band interferometry measurements of rivers. Finally, Andreadis et al. (*in review*) found that an expert classification of river planform morphology used to define channel shapes a priori yielded improved discharge prediction. This is logical- the more we know about a river, the better we can invert discharge.

These findings indicate that priors play a pivotal role in McFLI discharge accuracy, yet despite the geomorphic foundations of the McFLI paradigm, present McFLIs have used the same set of geomorphic priors for every river on earth, regardless of differences in planform geometry, hydraulics, and river size. This means that McFLIs use the same expectations for, e.g., Manning’s *n*, width/depth ratios, and AHG exponents in a braided river and a canal. No study to date has explicitly explored the sensitivity of McFLIs to the *quality* of their priors, where ‘quality’ refers to the hydraulic and geomorphic representativeness of a prior for a given river. High-quality priors would be accurate, river-specific knowledge that closely approximates field measurements and contextualizes discharge inversion to the specific hydraulics of the river. While Andreadis et al. (*in review*) did briefly address prior quality, their method was only applied to the channel shape prior and, more importantly, was not scalable globally. Further, Lin et al. (2019) xxx this seems an ideal spot to mention GRADES, as this is why she developed it! xxx

There are also troves of existing in situ data that can be mapped onto these global products to inform McFLI. The United States Geological Survey (USGS) makes periodic field measurements of discharge and other hydraulics to calibrate the rating curves for their stream gauges, and all of these are meticulously archived and freely available. These measurements are easily joined to existing hydrographic datasets, thereby providing reach-scale geomorphic attributes for each hydraulic measurement.. For example, Brinkerhoff et al. (2019) joined over 730,000 USGS in situ measurements to the National Hydrography Dataset (NHD,USGS, 2019), building on earlier work (e.g. HYDRoSWOT- Bjerklie et al. 2020; Canova, et al. 2016).

In order to be useful for McFLI, datasets of *in situ* measurements must be mapped to specific reaches. Hagemann et al (2017) provide one such approach, where they trained simple regression models on HYDRoSWOT to predict priors from just river width. This is a ‘global’ approach, where all rivers in the dataset are used to predict priors without attempting to differentiate between rivers, following Bjerklie et al (2003, 2005).. A potential alternative s that does address geomorphic differences is statistical classification. Distinct river classes, or types, should exhibit fundamentally different hydraulics that could be assigned to rivers using RS observations as predictors. By assigning representative values from the *in situ* datasets to each river type, global hydraulic priors are reduced to look-up tables of ballpark hydraulic estimates for each river type. However, scaling these river types to every river in a network poses a significant challenge (e.g. Guillon et al. 2020).

Given this context, m we hypothesize that McFLI performance will be improved by acknowledging geomorphic differences between rivers and assigning different priors to different rivers, building on recent global RS of rivers and decades of detailed *in situ* work mapped to global rivers via RS Further, we hypothesize that this intervention alone should be sufficient to improve accuracies, and no new RS-observations or updates to McFLI algorithms are needed to make better predictions of discharge. We use the Bayesian-AMHG-Manning’s (BAM) algorithm (Hagemann et al. 2017) as a case study for McFLIs. We provide BAM with improved prior river knowledge by 1) obtaining priors from the largest known repository of *in situ* data joined to hydrography (Brinkerhoff et al., 2019) and 2) constructing a river classification framework to reduce hydraulic variation to geomorphically distinct river types. We test these interventions to produce river discharge from Landsat observations on 7,522 river reaches in the Mackenzie River basin (validated at 95 gauges) and from SWOT-simulated data representing 17 rivers from Durand et al (2016). Ultimately, we provide a method for improving discharge estimation that is globally scalable using only RS observations and could theoretically be applied to any river on Earth large enough to see from space.

2 Data

The goal of this study was to improve McFLI accuracy by improving the quality of its priors. This required creating three distinct datasets: 1) in situ measured training data for generating new prior river knowledge (Section 2.1), 2) RS observations for RSQ in the Mackenzie River basin (Section 2.2), and 3) modelled RS observations for RSQ using SWOT-simulated rivers (Section 2.3).

**2.1 Hydraulic dataset used for prior river knowledge**

Our first task was to gather a comprehensive dataset of measured river hydraulics to generate training data. We started with Brinkerhoff et al. (2019)’s dataset. This dataset merges USGS surface water measurements of channel discharge, xxx, xxx, xx and xxx with the NHD and filters the data to include only those rivers that have at least six stations of 20 measurements each to derive that river’s AMHG. Brinkerhoff et al.’s dataset ultimately contains 730,072 unique measurements of hydraulics. For this study, we further filtered their data to exclude out impossible measurements (i.e. *Q* < 0) and measurements identified by the USGS as ‘poor’, yielding 372,109 unique in-situ hydraulic measurements across 190 rivers in the continental United States. We added to these data by calculating river and landscape geomorphic variables from the NHD and xxx for each observation in the training data. Table 1 gives the variables available to our ultimate training dataset used to differentiate rivers.

**Table 1.** The 15 geomorphic variables used to define river types in this study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Symbol** | **Description** | **Source** |
| Channel Width | W | Observed channel width | measured- NWIS\* |
| Channel Velocity | V | Observed mean channel velocity | measured- NWIS |
| Channel Depth | D | Observed mean channel depth | measured- NWIS |
| Slope | S | Observed slope | measured-NHD\*\* |
| Manning’s n | *n* |  | calculated- NWIS |
| Stream Order | SO | Strahler stream order |  |
| Distance Downstream | Dd |  |  |
| Hydraulic Residence Time | HRT |  |  |
| Minimum Grain Size Entrained | De |  |  |
| USGS Waterbody Type | WB |  |  |
| Unit Power | omega |  |  |
| Channel Shape | r |  |  |
| Drainage Area | DA |  |  |
| Froude Number | Fr |  |  |
| Shear Stress | tau |  |  |
| \* link |  |  |  |
| \*\*link |  |  |  |

This dataset is built exclusively for the United States, and we acknowledge that the continental US is not reflective of all global landscapes. Further, we have also limited our dataset to those rivers where there are six or more stations to parameterize AMHG. With that said, this is to our knowledge the largest possible freely available fluvial geomorphology dataset that covers a wide range of geographies from temperate and semi-arid climates to deserts and sub-tropical regions, and these data represent a best-case scenario for our analysis. Notably, our training data are missing observations in equatorial and Arctic/subarctic regions. With the aim of improving global RSQ at the center of this study, we chose to use one of these poorly represented regions as a case study: the Mackenzie River basin in the Canadian Arctic.

**2.2 RS data for the Mackenzie River basin**

The Mackenzie River basin is representative of Arctic hydrology and a good test for applying our new prior river knowledge at a large network scale. We use the Mackenzie as a test basin using real satellite data (as opposed to modelled SWOT data, Section 2.3). To do so, we need RS observations of river widths to drive the BAM algorithm. We extracted multi-temporal widths for 228,659 cross-sections in the basin following Feng et al. (2019, Text S1). In short, we (1) used MERIT hydrography (Lin et al. 2019) to define river centerlines and generate with measurement stations at varying intervals along the centerlines based on river widths (Yamazaki et al. 2019- Table Sx). (2) We constructed orthogonal cross section for every station following Yang et al. (2019), and (3) these orthogonal lines were used as inputs to RivWidthCloud, an automated algorithm for river width extraction using the Google Earth Engine (Yang et al. 2019). After filtering for clear-sky images, RivWidthCloud classifies a pixel as water using an algorithm detailed in Yang et al. (2019) and Text S1. The classified water mask is intersected with the orthogonal lines to estimate wetted width at each cross-section. We measured widths this way at 228,659 cross-sections from 7,522 Landsat-visible reaches in the Mackenzie River Basin using Landsat imagery from 1984-2013.

For validation, we collected in situ daily discharge data for the Mackenzie River basin for 1984-2013 from Water Survey of Canada (WSC, (). These gauge data are analogous to USGS gauging data provided by the NWIS. WSC gauge stations were linked to our hydrography based on their geospatial locations (distance within 500 m of a centerline) and drainage areas (i.e., difference within +/-10% of a hydrography reach), resulting in 327 validation gauges.

**2.2 SWOT-simulated rivers**

Once SWOT launches, Manning’s-based McFLIs can be run globally from SWOT’s novel simultaneous observations of river width, height, and slope. However, existing SWOT-like data are limited to an airborne Ka-band InSAR, which is currently available for less than five rivers globally with only a handful of observations each (Tuozollo et al. 2019a), or painstaking data fusion of altimetry and imagery (Bjerklie et al., 2019). Therefore authors typically use simulated SWOT data to test Manning’s-based McFLIs. A SWOT simulator has been built by the Jet Propulsion Laboratory (JPL) to simulate the errors that are expected to come with actual SWOT data (namely, radar layover errors and random noise), and has been used to benchmark McFLIs before (e.g. Oubanas et al. 2018). However, for this study we are interested purely in algorithm performance and so seek a test scenario that assumes perfect measurement conditions with no introduced errors. Satisfying data is to achieve this are simple reach-averaged hydraulic model outputs with water surface slopes and river widths labelled as ‘RS observations’. For this test, we used 17 test rivers from Frasson et al. (2019) which were developed for benchmarking McFLIs by Durand et al. (2016) and outlined in Table Sx. These rivers cover the United States, Canada, Great Britain, France, and Italy. Median discharge ranged from 62-14,199 m3/s and ‘observation’ windows ranged from 22-365 days over 3-16 reaches. Simulated rivers mimic perfect measurement conditions and represent the best-case scenarios of what SWOT will provide to hydrologists. These model rivers provide ‘observations’ of xxx, xxx, xxx needed to run both a Manning’s and AMHG based McFLI.

**3 Methods**

Our experimental design is as follows. For each of our two datasets, we ran a default McFLI as published by Hagemann et al., (2017): the ‘Bayesian-AMHG-Manning’s algorithm’ (BAM). This establishes the status quo of McFLI retrievals for one algorithm. We then ran the algorithm using new classified prior information, termed ‘geoBAM’. geoBAM’s new prior river knowledge refers to the two interventions previously outlined: 1) using a larger and more geomorphically-varied dataset to obtain prior knowledge (Section 2.1), and 2) classifying these prior data into discrete river types (Sections 3.1, 3.2).

**3.1 River Classification**

Recall our hypothesis that defining priors for specific river types will improve McFLI performance. To test this we constrain prior river knowledge using expert and unsupervised statistical classifications to extract geomorphically distinct river types.

Statistical learning is generally binned into unsupervised and supervised approaches (James et al. 2013; Hastie et al. 2009). Both use a suite of variables extracted from a training dataset to define a feature space and then identify patterns in that space. Unsupervised learning identifies these patterns and then clusters observations solely from the training data given user guidance only on algorithm parameters, while supervised learning uses user-defined predictors to model ‘target variables’ that are known a priori. The familiar concepts of regression (if the target variable is continuous), or classification (if it is discrete) are forms of supervised learning. In this study, we lack a priori river types in the training data to take a supervised approach to learning, and so instead we implement both unsupervised learning and an expert classification framework that forces specific knowledge on river geomorphology to guide statistical methods of classification.

As a representative unsupervised clustering approach, we used the ‘density-based spatial clustering of applications with noise’ (DBSCAN: Ester et al. 1996) algorithm. DBSCAN is a density-based clustering algorithm that groups observations in the multi-dimensional feature space using proximity. Distance between points is determined using Euclidean distance. Unlike simpler unsupervised clustering algorithms, DBSCAN does not assume all clusters have a convex shape in the feature space and instead uses density to group observations. This means clusters can be arbitrarily shaped or completely surround other clusters. This also permits DBSCAN to identify ‘noise’ points which are outside of the dense areas of the feature space, differing in practice from other simple unsupervised learning methods (e.g. K-Means clustering will assign every observation to a cluster). The user must provide a minimum number of points for a cluster and a maximum cluster radius, and DBSCAN determines the number of clusters (unlike simpler unsupervised algorithms). After testing, we settled on a cluster radius of 0.2 as the best balance between number of clusters, within-cluster variance, and computational efficiency. Initial testing also suggested a minimum cluster size of 1,000 hydraulic measurements. We ran DBSCAN on non-dimensional forms of the fifteen geomorphic variables in Table 1, yielding 9 clusters in approximately 85% of the data and 15% classified as ‘noise’.

We also developed a bespoke expert classification framework for extracting river types, built specifically such that river width is a predictor of these types. By using principal component analysis (PCA) as a guiding tool, the approach follows similar methods used to extract global hydro-climatic river types (Dallaire et al. 2018) and hydrologic flow regimes (Olden et al. 2012). Here, a PCA was used to dimensionally reduce our dataset and create multivariate, non-dimensional principal components (PCs) responsible for some amount of geomorphic variation across the feature space (and ultimately the United States). We ran a PCA on non-dimensional forms of the fifteen geomorphic variables in Table 1 and selected the five most influential PCs (cumulatively responsible for 73% of the variance in the feature space). In order of most variance explained, these PCs qualitatively represented 1) stream competency, 2) longitudinal location along the channel, 3) velocity/Froude number, 4) waterbody type, and 5) channel shape (Table S3).

There are two parts to the expert classification. In the first, we used PC scores to assist in classifying hydraulic observations. PC scores were calculated for every observation in the training data and are simply linear combinations of the standardized feature values multiplied by their loadings and calculated for each PC. Xxx definitely need a dumb reader plain language follow up sentence xxx Because the loading vectors, and thus the scores, for each PC are scaled the same (James et al. 2013; Hastie et al. 2009), we simply summed the five PC scores for each observation. This provides a single value per observation, where observations with similar values have similar geomorphology. We then use this metric to classify all hydraulic observations into river types by segmenting into 8 classes using quantiles of this metric as class thresholds. We subjectively chose eight classes to explicitly maintain river width as a predictor of river types.

The second part of the expert classification framework was designed to parse out unique river geomorphology from the previously defined 8 classes, namely ‘big’ rivers and ‘highly width-variable’ rivers. For some very large rivers (e.g. the Mississippi or St. Lawrence rivers), the training data had very few measurements in rivers of similar size and thus hydraulics are ill defined. We defined ‘big’ rivers as those with a mean width greater than . We then defined ‘highly width-variable’ rivers as those with a channel shape parameter *r* < 1, which guarantees significant variability in river width for both single channel and multi-channel rivers (Text S3).

Both classification frameworks yield fundamentally different hydraulics across river types. Figure 1 plots truncated, parametric distributions of six hydraulic terms calculated from the training data needed to run BAM for both classifications. *A0* is median cross-sectional area, *Wb* is bankfull width,*Db* is bankfull depth*, n* is Manning’s roughness term, *r* is a channel shape parameter, and *b* is an AHG exponent (Text S2 details calculating these terms). For both classifications, *A0*, *Wb*, *Db,* and *b* fundamentally vary by river type, while *n* and *r* less so. Per the definition of *r* and *b* in the expert system, their ‘highly width-variable’ distributions (Type 9) are significantly different from the other types’ distributions. For the unsupervised distributions, river types 8, 6, and 4 exhibit fundamentally distinct distributions on *r* and *b*, with river type 4 largely emulating our ‘highly-width-variable’ class (Type 9) in the expert classification*.*  While these plots visually justify both classification frameworks as ways to extract differentiable hydraulic distributions, we confirmed this by running two One-Way ANOVA tests on the medians of the six hydraulic variables’ distributions in Figure 1, grouped by river type (Table Sx). For the expert framework, all distributions except for *n* were significantly different. For the unsupervised framework, *A0*, *Wb*, *Db* all had p-values of approximately 0.10 while *b* and *n* were not significantly different. This is likely due to distribution overlap of river types 6 and 3. Overall, Figure 1 and Table Sx confirm that both classification frameworks yield differentiable river types across these six hydraulic terms.

A bunch of different types of map

Description automatically generated

**Figure 1:** Truncated, lognormal distributions of hydraulic terms as defined using new training data and the expert classification (left column) or the unsupervised classification (right column). Right column y-axes are truncated for visibility, though river type 5’s distribution has very little variance and thus rises above the axis limits.

**3.2 Mapping river type from remote sensing**

The goal of this study is to improve ungauged discharge prediction at the global scale. Thus, river types must be assignable to rivers using only RS observations. However, our classifications above are defined using *in situ* measurements that are not remotely sensible. This presents a unique supervised classification challenge necessitating width and/or slope (i.e., current synoptically available fluvial parameters) as the sole predictors of the river types. Here, we use ‘river type mapping’ to refer to a supervised classification where the target variables are our river types and the predictors are river width and/or slope.

For the DBSCAN framework, neither width nor slope were strong predictors of river types. Therefore, we needed a way to map RS observations at the scale of the Mackenzie basin (228,659 cross sections) to the DBSCAN-generated classes (which use no RS data). To do so, we turned to basic supervised machine learning to assign river types, which has seen some success at the regional scale (Guillon et al., 2020). After testing simpler regression-based approaches with no predictive success, we settled on a random forest classifier. Random forests are ensemble methods for classification (or regression) that use a series of decision trees, each estimated using a random subset of predictors, to assign classes. Individual decision trees are prone to overfitting and so an ensemble approach is used to curtail that tendency (Hastie et al. 2009). Using a classic validation-set approach to model training, we trained a Random Forest classifier on 80% of the training data using mean and standard deviation of cross-sectional width as predictors. When tested on the remaining 20% of the data, the classification accuracy was 90%. Thus, we are able to reproduce our classification, which carries rich in-situ prior information, solely from RS.

As previously stressed, the expert framework was developed explicitly such that river width was a strong predictor of river class as produced by the PCA. This is confirmed visually in Figure Sx. Thus, we assigned river class by proximity of MERIT hydro widths to the characteristic width (i.e., median of expert class river width) in the expert classes. We mapped the ‘highly width-variable’ type to rivers using a standard deviation of at-a-station log-transformed width > 0.45 meters as the threshold.

**3.3 Discharge estimation**

McFLI RSQ was performed using the BAM algorithm (Hagemann et al. 2017). BAM probabilistically estimates discharge via Bayesian inference and a Hamiltonian Monte Carlo sampling scheme. The user chooses to invert Manning’s equation, AMHG, or a combination of both as its flow law. The user also chooses whether to run BAM at every cross-section or to use reach-averaged observations in line with how SWOT will observe rivers. For this study, we ran AMHG on every cross-section in the Mackenzie River basin and ran a ‘switch’ flow law on the reach-averaged SWOT-simulated data. This ‘switch’ always inverts Manning’s equation, and inverts AMHG when it is deemed suitably strong (Text Sx). We also updated AMHG’s flow law to reflect new findings on the physical basis of AMHG (Brinkerhoff et al. 2019) and, following recent work on temporally defined roughness terms in McFLIs (Tuozollo et al. 2019b), tested space-varying and space-and-time-varying roughness terms in BAM. We ultimately implemented a space-varying roughness term as a compromise between computational efficiency and increased predictive accuracy (Text Sx).

To run BAM, the user provides width and/or slope RS observations, as well as prior distributions of 34 parameters. These priors, which are the six terms in Figure 2??, flow law errors, and AMHG’s ‘congruent width’ term, are formalized within BAM as truncated, lognormal distributions where ) for , using mean (), standard deviation (σ), and upper () and lower bounds () as parameters. For our tests (see Figure 2 for a flowchart), we used geomorphic river types to redefine the prior river knowledge on the six hydraulic terms in Figure 2 using the river mapping in Section 3.2 to assign a river type to each BAM reach. The prior estimate on discharge was defined differently for our two tests: for the Mackenzie River basin, we used a stream gauge when available and otherwise used the mean daily discharge estimate from GRADES (Lin et al 2019). For the SWOT-simulated rivers, output from a water-balance model (Wisser et al. 2010) was used in line with Durand et al. (2016).

A picture containing text, map

Description automatically generated

**Figure 2.** Flowchart detailing our two classifications and the tests we ultimately ran. In the final row of discharge products: ‘Q’ is discharge, ‘m’ is Mackenzie River basin, and ‘s’ is SWOT-simulated rivers.

**3.4 Validation**

Error metrics (Table 3) to quantify RSQaccuracy followed Hagemann et al. (2017). rBIAS and RRMSE define the range and central tendency of prediction errors, respectively. NRMSE is a normalized variant of RRMSE to account for RRMSE’s high sensitivity to errors in low flow estimation, and NSE represents the amount of variance in the observed data that the model explains. An NSE greater than zero indicates that our model estimates better than guessing mean flow every time.

**Table 3:** Error metrics used in this study. Also included is our definition of percent change.

|  |  |  |
| --- | --- | --- |
| **Description** | **Abbreviation** | **Definition** |
| Relative root-mean-square error | RRMSE |  |
| Normalized root-mean-square error | NRMSE |  |
| Relative bias | rBIAS |  |
| Nash-Sutcliffe efficiency | NSE |  |
| Percent Change | % Change |  |

1. Results

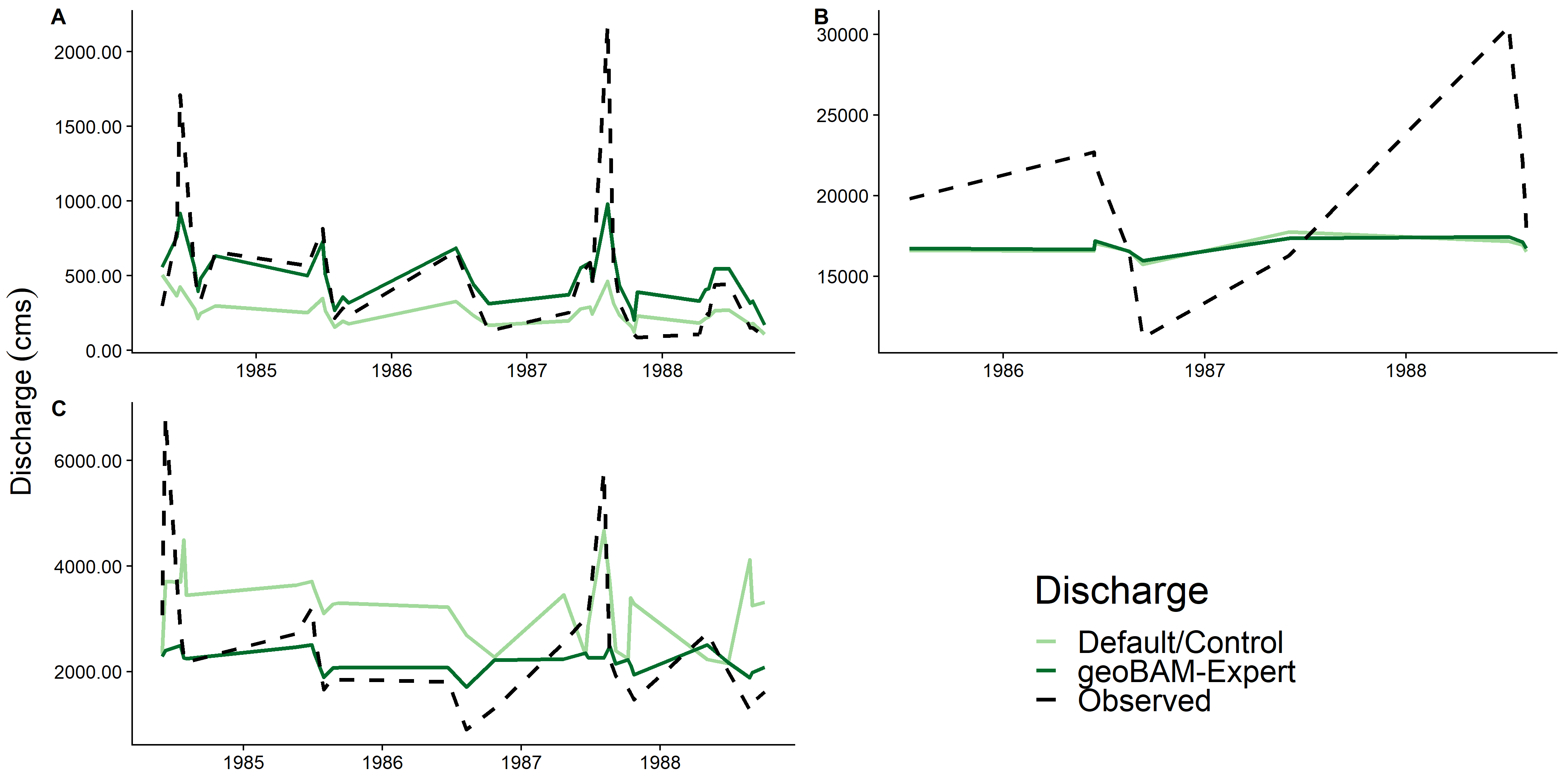
We present geoBAM-Expert results for the Mackenzie River basin at the cross-section scale (Section 5.1) and SWOT-simulated rivers at the reach scale (Section 5.2). We also compared RSQ for both classifications across the suite of error metrics in Table 3 (Section 5.3) and mapped the expert classification onto all rivers in the Mackenzie River basin (Section 5.4).

**5.1 Mackenzie River basin**

geoBAM-Expert predicted discharge is significantly improved over default/control BAM (Figure 3). Despite no new RS observations and no changes to BAM, we have yielded a median percent increase in NSE (Figure 3c) of 72% across the xxx validation gauges of the Mackenzie. Most of the improvement distribution is greater than 0, and improvement that is often greater than 100%. The largest improvement occurs in reaches that were poorly estimated (NSE < 0) by default/control BAM (Figure 3a). Default/control BAM had positive NSE less than half of the reaches (42%). geoBAM-Expert, however, has a positive NSE in most reaches (62%). In Figure 3b, median NSE across all reaches improved from -0.11 to 0.15 with geoBAM-Expert. Finally, the entire distribution of NSE scores is higher and more consistent: the middle 50% of scores increased and the inter-quartile range (IQR) narrowed substnatially (from 1.12 to 0.53) . In sum, geoBAM-Expert substantially outperforms BAM in both the magnitude and consistency of NSE. Similar analysis for other error metrics can be found in SIxxx



**Figure 3:** NSE improvement for 95 validation cross-sections in the Mackenzie River basin: A) Empirical cumulative density functions (CDFs) of the NSE scores for every reach; B) boxplots of the same results; c) probability density function (PDF) of the % change in NSE. Axes are truncated for visualization’s sake. We have provided arrows and number of rivers not plotted when necessary.



**Figure 4.** Representative hydrographs from the Mackenzie River basin: A) is an example of significant improvement in BAM-estimated discharge when using geoBAM-Expert. B) is an example of little to no change in BAM-estimated discharge when using geoBAM-Expert. C) is an example of worsened performance.

In some rivers, hydrographs (Figure 4) of discharge estimated using geoBAM-Expert significantly improved from those estimated using default/control BAM. FINISH WITH FULL RESULTS.

X

X

X

X

X

X

X

X

**5.2 SWOT-simulated rivers**

A close up of a map

Description automatically generated

**Figure 5:** NSE improvement for 17 SWOT-simulated rivers: A) Empirical cumulative density functions (CDFs) of the NSE scores across rivers; B) boxplots of the same results; C) probability density function (PDF) of the % change in NSE. Axes are truncated for visualization’s sake. We have provided arrows and number of rivers not plotted when necessary.

There is continued substantial improvement in NSE on the SWOT-simulated rivers (Figure 5, Table S5), though there is less overall improvement than in the Mackenzie River test (Figure 3). Upon implementing geoBAM-Expert, there is a median percent increase in NSE for all rivers of 11% (Figure 5c), as well as consistently better performance: the IQR shrinks from 1.53 to 0.87, pushing up against the upper limit of 1.0 for NSE (Figure 5b). While this percent improvement is lower than in the Mackenzie River test, the median NSE across all 17 rivers significantly improved from 0.03 to 0.49 (Figure 5b). Again, most of the significant improvement occurred in rivers which performed poorly using default/control BAM, with relatively less improvement observed in rivers with already good prediction accuracy.

Finally, significant variation in performance within river types is noted. For example, NSE scores for geoBAM-Expert’s river type 8 range from -21.7 for the Tanana to 0.94 for the Cumberland (Table Sx).



**Figure 6.** Normalized Hydrographs for the SWOT-simulated rivers: observed discharge (dashed black) is plotted alongside default BAM (light blue) and geoBAM-Expert (dark blue).

Moving beyond summary statistics, we now analyze hydrographs. In some rivers, geoBAM-Expert hydrographs (Figure 6- dark blue lines) significantly improved from those estimated using default/control BAM (light blue lines- e.g. St. Lawrence Upstream, Garonne Upstream, Cumberland, and Severn). In other rivers, very little changed (e.g. Wabash, Mississippi Downstream, and Tanana). Both implementations of BAM adequately modeled observed flow dynamics, but geoBAM-Expert has ‘filled in’ many errors in predicting the magnitude of peak events with varying degrees of success. For example, the Ohio River was mostly modeled correctly except for errors in the three peak events (Figure 6- ‘Ohio’). geoBAM-Expert has minimized the error in these peak events, while continuing to accurately model the rest of the hydrograph. Some rivers with little to no change in NSE scores (Figure 5) still appear to have partially ‘filled in’ these discrepancies (e.g. Seine, Severn, Cumberland). geoBAM-Expert results for both the Sacramento Upstream and St. Lawrence Downstream show visibly worsened reproduction of observed discharge than default/control BAM.

**5.3 Classifications compared**

*Figure 7) six sets of boxplots, 4 error metrics (both Mackenzie and swot)*

Compare classifications here across four error metrics. Compare differences in scores across the four error metrics.

X

X

X

X

X

X

X

X

X

X

Significant performance improvement is limited to NSE and not the other metrics (Figure 7). **MORE.** RRMSE and rBIAS generally track together and are easily inflated due to errors in baseflow prediction. Our interventions alter baseflow predictions very little (Figures 4 and 6) and because of this it is possible that these two metrics are relatively insensitive to geoBAM’s interventions.

**5.4 Mackenzie River basin classification**



**Figure 8.** River classification using geoBAM-Expert for Mackenzie River basin. The Canadian Shield and major lakes are overlaid upon the classified hydrography (Lin et al. 2019; Yamazaki et al. 2019). ‘River Type’ labels align with those in the left column of Figure 1.

Rivers mostly fall into two regimes composed of Type 1 and 2 rivers in the western half of the watershed, and Type 8 and 9 in the eastern. These two regimes align nearly perfectly with the boundary of the Canadian Shield, an xxx geology language here saying its old and hard with thin soils xxx which underlies large lakes and boreal forests in the eastern portion of the basin while the western portion drains the Rocky Mountains- a fundamentally different landscape. Thus, geoBAM-Expert qualitatively reflects observed geomorphology patterns in the landscape of the Mackenzie River basin through its different river types.

Xx Can you add a paragraph on what the classes actually are or are not? Motivated by discussion xxx

1. **Discussion**

**6.1 RSQ something something**

The quality of prior knowledge significantly influences RSQaccuracy, as evidenced through large performance improvements in BAM (Figures 4-7). While algorithm developments also continue to evolve our understandings of McFLIs and dischargeinversion, we have found that our relatively simple interventions to prior estimation are easy to implement and did not require new geomorphology within the algorithm. We also achieved skill improvement without relying on new and computationally intensive data assimilation schemes or hydraulic models (e.g. solving full Saint-Venant equations). Xxx we need to throw a bone to those folks, who will be reviewers. Sentence here saying that we think those are good ideas xxx

Improvements occur from real-world RS observations in the Mackenzie River basin over a massive spatial scale, which is noteworthy for two reasons. First, all reaches used in this study are in the Arctic/Subarctic, where our training data are unrepresentative (i.e. none of the field hydraulic measurements were made there). The success of these interventions in a blind case study like this, with training data only from the continental United States, suggests that this approach is implementable globally. Second, we have satisfyingly replicated Feng et al. (2019)’s results that were run only on eleven Alaskan rivers. We have extrapolated their RSQ workflow to tens of thousands of reaches, relying on GRADES (Lin et al. 2019) for reach-explicit prior knowledge on discharge. These results corroborate the aggregate use of Feng et al. (2019)’s method, geoBAM, and GRADES for big-data RSQ and functionally open the door for uncalibrated RSQ across global-scale river networks.

We are aware of one other McFLI that has explicitly tested the influence of prior quality. Andreadis et al.’s (2020) SWOT Assimilated Discharge (SAD) algorithm was developed to address parameter equifinality issues in McFLIs by constraining their parameter space for discharge inversion. To test this, they introduced a suite of interventions to SAD, one of which was an expert geomorphic classification that defined their prior on channel shape. While it is difficult to directly compare results across algorithms and with different interventions, when they isolated this intervention it yielded a median NSE increase of 0.35 on SWOT-simulated rivers. Using geoBAM-Expert, median NSE increased on the SWOT-simulated rivers by 0.46. Such similar results suggest that this degree of improvement occurs across McFLIs, assuming the algorithm is flexible enough to ingest different definitions of prior river knowledge (like BAM and SAD are). Despite the similarities, the SAD geomorphic classification is not globally scalable because xxx, and classification is limited to exclusively channel shape. The method developed here is globally scalable, can be used to extract distributions for any desired hydraulic term, and yields greater improvement in discharge prediction.

It is unclear if this approach in improving prior quality will work with all McFLIs. Hagemann et al. (2017) purposefully designed BAM to be flexible in both the priors it ingests and the flow laws it inverts, allowing us to run this study in two disparate settings: 1) using simulated altimetry data to invert Manning’s equation and 2) using river widths to invert AMHG. Our similar results in both settings (Sections 5.1-5.2) suggest that the McFLI paradigm would broadly benefit from improved prior quality, regardless of the flow law or RS observations used. However, we exclusively tested this within BAM, meaning these interventions may not co-exist with the mathematical and/or computational setups in other McFLIs. In particular, other McFLI algorithms may have priors that do not correspond to our in situ data or classes. Not all McFLIs are the same even though many use Manning’s equation: each algorithm has its own unique prior set. Whether or not our classifications are useful to all McFLIs or just to BAM is left for future work.

For RSQ, classifying rivers to provide better priors will only yield improvements if classes are assigned correctly. Because we reduce priors to look-up tables with a set of proirs assigned to each river type, if the wrong class assigned to a river then RSQ will considerably worsen. Andreadis et al. (2020) found a similar result when their expert classification misclassified two rivers and yielded poor performance. When they artificially assigned the correct river type, they improved RSQ accuracy. Similar behavior was identified in this study: the Sacramento Upstream’s NSE score improved by 5.4% using geoBAM-Unsupervised but worsened by 10.9% using geoBAM-Expert**.** Clearly, there is room for improvement, and it appears geoBAM-Expert is assigning an incorrect class, or the Expert classes are unrepresentative of this river. This suggests that RSQ accuracy hinges on how river types are defined and ultimately mapped to rivers. While these errors occurred in a few rivers (xxx list them xxx), both of our classification frameworks are globally scalable with good predictive accuracy and use just time-varying river widths as predictors.

6.2 Classification something something

It is useful to orient our classification workflow in the context of other global-scale river classification frameworks. Interestingly, geoBAM-unsupervised largely replicated the results from the more complicated geoBAM-Expert, signifying that unsupervised algorithms can nearly replicate the different types of rivers experienced globally and that future implementations might equal or outperform a more manual approach. The success of such simple unsupervised classification suggests that a generalized, global river typology framework is possible. There are currently very few of these frameworks (e.g. Fernandex & Sayama, 2015; Puckridge et al. 1998; Haines et al. 1988), only one of which is globally consistent in coverage (Dallaire et al. 2019). Dallaire et al. (2019) clustered features explicitly associated with every observation/river reach within HydroSHEDS (Lehner et al. 2008), however fluvial geomorphology data does not generally exist in this form at the global-scale and so their analysis was largely limited to hydro-climatic river types. Conversely, Guillon et al. (2020) successfully used machine learning models to upscale a priori geomorphic river types to over 100,00 reaches. Our study presents a novel amalgamation of these two methods, first using automated clustering of field data to define a geomorphic typology framework, and then using supervised learning to upscale river types to anywhere on Earth. Section 5.4 shows that this approach is viable in the Mackenzie River basin, where we accurately represent landscape geomorphology. However, more sophisticated methods will likely be needed to produce a generalized, global river typology framework that moves beyond RSQ. A novel amalgamation of existing RS and hydrographic data (e.g. Chen et al. 2019; Lin et al. 2019), machine learning, and field-measured hydraulics is a proposed way forward.

Finally, Durand et al. (2016) noted the potential homogeneity of test rivers used for benchmarking McFLIs, but until now there has not been a geomorphically explicit way to quantify this homogeneity. Using the geoBAM-Expert river types, we found that the SWOT-simulated rivers are quite homogenous and large (4/17 were classified ‘big’ rivers and 7/17 were assigned river type 8, the widest in the classification). These are not reflective of the global variation in hydraulics and geomorphology SWOT will encounter. What is thus needed is a wider range of river types for validating and testing McFLIs before SWOT launches, such that we can parse out specific river types that McFLIs model well, and those that McFLIs model poorly.

The fact that Durand et al. (2016)’s test rivers are almost all very large rivers might explain why Manning’s-based McFLIs generally perform well in these tests. Manning’s equation simulates discharge in deeper, larger rivers quite well, but not in shallower, smaller ones (Ferguson, 2010). It is well-established that the equation’s fixed velocity~depth relation exponent and roughness term oversimplify flow (Ferguson 2010; Bjerklie et al. 2005; Katul et al. 2002; Ferguson 1986) If the goal is network-scale RSQ, like performed in the Mackenzie River basin, a generalized flow resistance relation might be used to improve prediction accuracy in smaller streams.

Xx good discussion. Can you add a paragraph on what the classes actually are or are not? xxx

7 Conclusions

This study presents a first attempt at quantifying the sensitivity of the BAM McFLI priors for global-scale RSQ in two distinct settings: on thousands of Arctic river reaches using Landsat imagery and on simulated rivers representing NASA SWOT satellite outputs prior to launch in 2021. We found significant improvement in the accuracy of discharge predictions for both test cases, with the median NSE improvement the Mackenzie rivers to be 72% and 11% in the SWOT-simulated rivers. These findings are significant and highlight the importance of prior knowledge in a Bayesian mathematical setting, where we have shown that starting from a more informed understanding of the river yields more accurate results. These priors are implementable in any McFLI and will play a pivotal role in both current RSQ efforts in global river networks, as well as future global RSQ from SWOT.

Acknowledgments

C. B. Brinkerhoff was supported by a grant to C. J. Gleason through the NSF RoL FELS RAISE program (PI Peter Raymond, Grant 1840243). Code to run our analysis is available at <https://github.com/craigbrinkerhoff/2019_GeoClass>. geoBAM-Expert is available in the R programming language at <https://github.com/craigbrinkerhoff/geoBAM>. The authors would like to thank Mark Hagemann for building the original BAM algorithm.

References