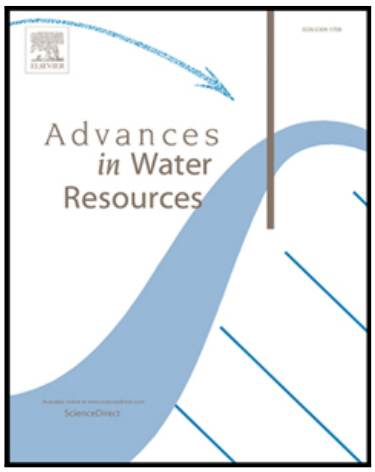
Journal Pre-proof



Exploring the Effect of River Spatial Variability on SWOT Discharge Accuracy

Yushan Zhang , Michael Durand , Steve P. Coss ,
Renato Prata de Moraes Frasson , Elizabeth M. Prior ,
Colin J. Gleason

PII: S0309-1708(26)00179-X
DOI: https://doi.org/10.1016/j.advwatres.2026.105384
Reference: ADWR 105384

To appear in: *Advances in Water Resources*

Received date: 6 February 2026
Revised date: 11 June 2026
Accepted date: 11 June 2026

Please cite this article as: Yushan Zhang , Michael Durand , Steve P. Coss , Renato Prata de Moraes Frasson , Elizabeth M. Prior , Colin J. Gleason , Exploring the Effect of River Spatial Variability on SWOT Discharge Accuracy, *Advances in Water Resources* (2026), doi: https://doi.org/10.1016/j.advwatres.2026.105384

Highlights

- High within reach spatial variability reduces median discharge NSE from 0.78 to 0.69.
- Hypsometry improves calibration and narrows the NSE gap to 0.05.
- Effective resistance is more sensitive in high width variability reaches.
- The specific slope variability parameterized flow law tested here yields no consistent calibration improvement.

Exploring the Effect of River Spatial Variability on SWOT Discharge Accuracy

Yushan Zhang[1], Michael Durand[1], Steve P. Coss[1], Renato Prata de Moraes Frasson[2], Elizabeth M. Prior[3], Colin J. Gleason[4]

[1]School of Earth Sciences, The Ohio State University, USA

[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

[3]Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, USA

[4]Civil and Environmental Engineering, University of Massachusetts Amherst, Amherst, MA, USA

Corresponding author: Michael Durand (durand.8@osu.edu)

Abstract

The Surface Water and Ocean Topography (SWOT) mission measures river width, water surface elevation (WSE), and slope for discharge estimation. In this study, we explore the impact on remotely sensed discharge of within-reach river spatial variability, defined as the coefficient of variation ($\varepsilon_p = \sigma/\mu$). After quality control, we analyze 484 representative reaches across the United States with paired SWOT observations and United States Geological Survey (USGS) discharge records from January 2023 to October 2024. Discharge is calibrated using modified Manning equations and evaluated using the Nash-Sutcliffe efficiency (NSE), absolute normalized residuals and correlation coefficient. Reaches with high spatial variability exhibit substantially lower calibration skill than reaches with low variability, with the median NSE reduced by 0.09 (0.78 vs. 0.69). Two possible reasons for this phenomenon are examined: SWOT observation errors and structural limitations of modified Manning's equations. Regarding observation errors, we apply a river hypsometry

constraint to correct width and WSE. Overall calibration accuracy improved, and the median NSE gap between low- and high-variability reaches reduced to 0.05 (0.86 vs. 0.81), emphasizing the significance of observation errors in this degradation. Conversely, evidence linking the remaining performance degradation to structural flow-law error is inconclusive. Although effective resistance changes more for reaches with higher spatial variability, adding the specific slope-variability term tested here does not lead to consistent improvement. This research has important implications for better estimating discharge from SWOT data.

Keywords

River discharge, SWOT mission, Spatial variability, Remote sensing

1. Introduction

Monitoring river discharge is critical for characterizing the global water cycle, optimizing water resource management, and assessing hydrological hazards (Depetris, 2021; Yin et al., 1997). In regions where ground-based gauging networks are sparse, remote sensing presents a vital alternative, offering broad spatial coverage and frequent revisit intervals (Casagli et al., 2023; Gholizadeh et al., 2016). Spaceborne platforms are capable of retrieving key hydraulic variables across extensive spatial domains, including river width, water surface elevation (WSE), and slope (Bjerklie et al., 2003; Bjerklie et al., 2020; Durand et al., 2009; Eggleston et al., 2024; C. J. Gleason & Durand, 2020). When synthesized through established hydraulic relationships, these variables enable discharge estimation independent of continuous *in-situ* measurements (Andreadis et al., 2020; Durand et al., 2014; Frasson et al., 2021; C. Gleason et al.,

2017; Hagemann et al., 2017; Oubanas et al., 2018). Among existing sensing modalities, optical imagery proves effective for mapping river planforms and measuring width, whereas interferometric synthetic aperture radar (InSAR), radar altimetry, and laser altimetry yield high-precision measurements of WSE (Allen & Pavelsky, 2018; Bjerklie et al., 2018; Durand et al., 2008; Larnier et al., 2021).

Building upon previous advancements in satellite remote sensing, the NASA–CNES Surface Water and Ocean Topography (SWOT) mission, launched in December 2022, represents the first satellite mission dedicated to producing high-resolution, two-dimensional maps of surface water elevation across rivers, lakes, and oceans (Andreadis et al., 2025; Biancamaria et al., 2016; Fu et al., 2024). Its Ka-band Radar Interferometer (KaRIn), an interferometric SAR system, facilitates global mapping of rivers typically wider than 100 m at spatial resolutions of tens of meters, with a 21-day repeat cycle (Fjørtoft et al., 2013; Peral et al., 2024). For rivers, SWOT provides reach and node level profiles of WSE, width, slope, and corresponding quality flags to support discharge estimation (Altenau et al., 2021; Paiva et al., 2015).

The estimation of discharge from SWOT data generally entails a two-stage procedure: first, establishing a functional relationship between SWOT observations and river discharge (hereafter referred to as flow laws); and second, applying these flow laws to concurrent SWOT observations to compute discharge for each satellite overpass. Typically, these flow laws are based on a modified form of the Gauckler-Manning-Strickler equation (hereafter referred to as the "modified Manning's equation"). Due to the nonlinearity of these flow laws, reliable discharge inference from reach-averaged width, depth, and slope is compromised when these hydraulic parameters exhibit substantial spatial variability (Beven, 1989; Tuozzolo et al., 2019). Furthermore, scale discrepancies in remote sensing observations, when compounded by measurement noise, can result in significant retrieval errors (Durand et al., 2023; Durand et al., 2014; Frasson et al., 2021). Additionally, as SWOT cannot resolve all terms within the flow laws, the unobserved components—termed flow law parameters (FLPs)—require

estimation to facilitate discharge production. To address this, the SWOT science team has developed six reach-scale Flow Law Parameter Estimation algorithms designed to infer these parameters from SWOT time series, utilizing prior or ancillary information (Durand et al., 2023). Given the absence of gauging stations along the majority of SWOT-observed rivers, discharge must be inferred directly from SWOT-derived geometry via flow laws, independent of site-specific calibration (Durand et al., 2020). Traditional regionalization methods for ungaged basins range from simple scaling techniques, such as the drainage area ratio method (Archfield & Vogel, 2010), to data assimilation frameworks that integrate observations with hydrologic models (Reichle, 2008; Verma et al., 2026). Within the context of SWOT, purely data-driven approaches are beginning to be explored as alternatives to traditional flow laws. Machine learning models are increasingly being tested to map SWOT-derived geometric variables directly to discharge (Tebaldi et al., 2026; Zou et al., 2026).

A first look shows that SWOT discharge estimates track discharge variations as expected but have a higher bias than predicted in pre-launch studies, with median of the absolute value of the bias being 56% (Andreadis et al., 2025). Strategies to improve SWOT discharge estimation require characterizing the three primary sources of error: observation errors, structural errors in flow laws, and FLPs' errors (Durand et al., 2023).

Previous research indicates that pixel misassignments near channel boundaries result in significant width errors, which are amplified in morphologically complex reaches (Allen & Pavelsky, 2018; Legleiter & Roberts, 2005). Since morphological variables often covary (Moody & Troutman, 2002), width serves as a practical proxy for within reach spatial variability in this study, defined as the coefficient of variation ($\varepsilon_p = \sigma/\mu$), especially when SWOT lacks direct depth measurements and slope retrievals are comparatively noisy. Although reach averaging enhances precision by mitigating measurement noise, it may concurrently increase calibration error in spatially and temporally heterogeneous reaches. This arises because, under nonlinear flow laws,

discharge calculated from reach-averaged geometry differs systematically from the mean discharge derived from locally varying geometry.

Li et al. (1992) demonstrated that spatial variability can induce nonlinear effects on hydraulic resistance parameters, which can be modeled as "effective resistance ñ" (Tuozzolo et al., 2019). Rodríguez et al. (2020) further established that the effective resistance ñ is inversely proportional to the depth variance and increases as the river discharge decreases. Fundamentally, as discharge increases, the relative contribution of skin friction diminishes; consequently, the friction coefficient typically decreases with rising discharge (Gust, 1988; Mahfoze & Laizet, 2017). As an example of this dynamic, consider the cartoons in Figure 1, which contrast two idealized reaches. Figure 1a depicts a reach characterized by a riffle-pool structure; this morphology exhibits substantial within-reach spatial variability in water-surface slope under low-flow conditions, but the slope becomes more spatially uniform as discharge increases. Figure 1c shows a contrasting reach in which the water-surface slope is idealized as spatially uniform across flow regimes. Figure 1b explicitly illustrates the corresponding relationship between within-reach slope variability and discharge for both reaches. Finally, Figure 1d illustrates the theoretical implication of this variability for effective resistance ñ. For Reach 1, high spatial variability at low flows produces a dynamic, flow-dependent effective resistance that decreases as discharge increases. Conversely, for Reach 2, because the water-surface slope is idealized as spatially uniform and nearly invariant across flow conditions, the contribution of within-reach spatial variability to effective resistance is assumed to be negligible, and ñ is shown as approximately constant with discharge. This does not imply that roughness is strictly discharge-independent in natural rivers; rather, Reach 2 serves as a limiting reference case against which the effect of flow-dependent spatial variability in Reach 1 can be contrasted.

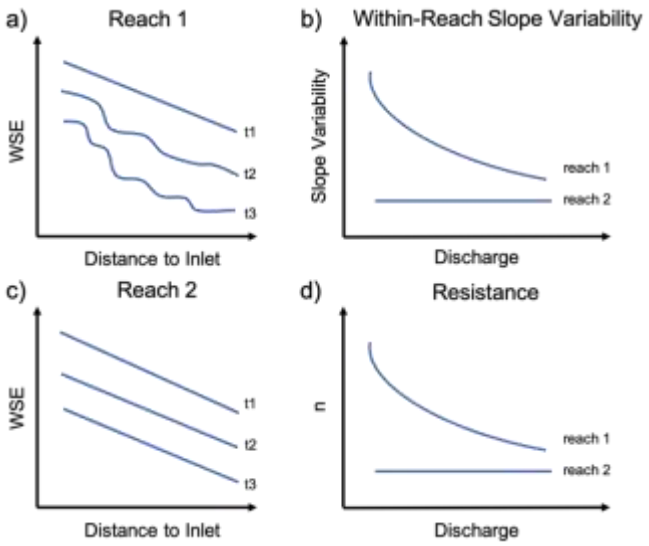


Figure 1. Conceptual diagram illustrating within-reach spatial variability. (a) WSE profiles for Reach 1 showing spatial variation with distance to outlet across multiple times. (b) WSE profiles for Reach 2. (c) Discharge (Q) versus slope variability with Reach 1 (blue) and Reach 2 (red). (d) Q versus effective resistance ñ.

The "spatial averaging theory" advanced by Rodríguez et al. (2020) explicitly links effective resistance to within-reach spatial variability, providing a mathematical framework to predict the reduction in effective resistance as a function of this variability. Quantifying the spatial variability of hydraulic parameters, such as water surface slope, is therefore critical for characterizing the temporal dynamics of effective resistance. Moreover, Durand et al. (2023) introduced novel flow laws wherein discharge depends not only on reach-averaged hydraulic properties but also explicitly on within-reach spatial variability, specifically riverbed elevation variability. Consequently, it is hypothesized that flow law accuracy is maximized by explicitly incorporating terms that account for spatial variability (Durand et al., 2023).

Therefore, this paper addresses three primary research questions:

1. How large is the discharge calibration performance gap between reaches with low and high within-reach spatial variability in paired SWOT–USGS observations?

2. How much of this performance gap can be reduced by applying a hypsometric constraint to mitigate SWOT width and WSE observation errors?
3. Is the remaining performance gap consistent with model structural error in reach-averaged Manning-type flow laws? Specifically, we investigate whether predictions of spatial averaging theory are substantiated by actual SWOT data and whether parameterizing flow laws with slope spatial variability improves calibration accuracy.

To answer these questions, we compare the discharge calibration performance of multiple flow laws between reaches with relatively low and high within-reach variability. By quantifying the impact of spatial variability, we reveal current limitations in SWOT observations and flow laws, providing guidance for improving SWOT data processing and refining flow-law structures to enhance discharge calibration accuracy.

2. Study Area and Datasets

2.1 Study Area and Study Period

Following rigorous quality control and filtering (detailed filter procedures in Appendix A), this analysis encompasses 484 representative river reaches where SWOT observations coincide with United States Geological Survey (USGS) discharge gage records. These reaches capture a broad spectrum of hydroclimatic and geomorphic settings, facilitating a robust assessment of spatial variability effects across diverse river systems. The study period spans from January 2023 to October 2024, corresponding to the first two years of SWOT operations. Figure 2 shows the empirical cumulative distribution function (CDF) of width variability for all 484 gage-paired U.S. river reaches and their spatial distribution. In Figure 2a, the red dashed lines and labeled point P (0.5, 0.90) mark the 90th percentile of width spatial variability, with 434 reaches

exhibiting width variability less than 0.5. Because the distribution of width variability is strongly skewed, reaches with very high spatial variability are relatively rare. We therefore used a two-group classification to test the primary hypothesis that high-variability reaches exhibit reduced discharge-calibration performance relative to lower-variability reaches, rather than dividing the dataset into multiple bins with small sample sizes in the high-variability range. We acknowledge that this choice produces imbalanced group sizes. A threshold of 0.5 is adopted to stratify the dataset: reaches are classified as either the "low variability group" ($\varepsilon_w < 0.5$) or the "high variability group" ($\varepsilon_w \geq 0.5$). Figure 2b maps the geographic distribution of these reaches, distinguishing between the low-variability (blue) and high-variability (orange) cohorts.

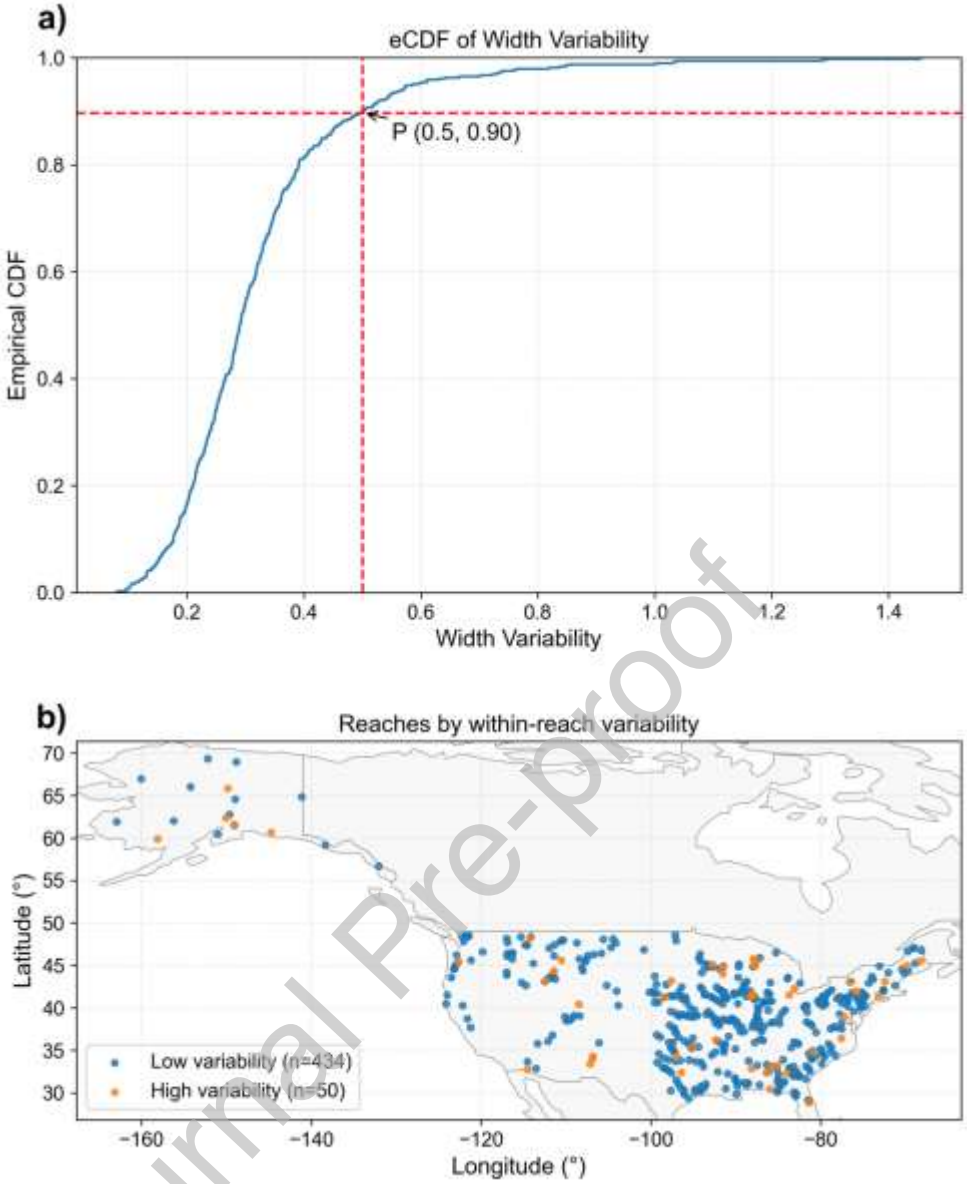


Figure 2. (a) Empirical cumulative distribution function (CDF) of width variability for all 484 gage-paired U.S. river reaches. The red dashed lines and labeled point P (0.5, 0.90) indicate that 90% of reaches (434/484) exhibit width variability less than 0.5. (b) Spatial distribution of gage-paired reaches, where blue points represent reaches with width variability below the threshold of 0.5, and orange points represent those with width variability above the threshold.

2.2 SWORD Database

To provide consistent river geometry and metadata, we used the Surface Water and Ocean Topography River Database (SWORD), a topologically organized global river

network database developed for SWOT river vector products (Altenau et al., 2021). Specifically, this study used SWORD version 17b, the official SWORD version underpinning the SWOT Version D RiverSP Vector Products, as archived in the SWORD v17b data release (Altenau et al., 2025). Unless otherwise noted, the width spatial variability reported in this study was derived from the reach-level mean width and width variance provided in SWORD v17b. Relying on the static SWORD prior rather than dynamic SWOT node observations ensures a consistent geomorphic baseline that is independent of transient satellite measurement uncertainties.

2.3 SWOT Level 2 River Data Product

We use the SWOT Level 2 River Single-Pass Vector Data Product Version D (SWOT_L2_HR_RiverSP_D) at reach and node levels. Node WSEs are used to quantify time-varying within reach slope variability, and reach discharge is calibrated against coincident USGS gage records. We access the SWOT time series by Hydrocron (Greguska et al., 2025). Detailed SWOT time series observation preprocessing procedures are presented in Appendix A.

Figure 3 provides an example of SWOT node time series data for one reach with high within-reach width spatial variability, the Cedar River (Reach ID: 74294400121). Note that while this figure illustrates spatial variability using dynamic SWOT node observations, the systematic classification of reach-level spatial variability utilized in our subsequent analysis is derived from the static SWORD prior database. Relying on SWORD for classification ensures a consistent geomorphic baseline that is independent of transient SWOT measurement uncertainties.

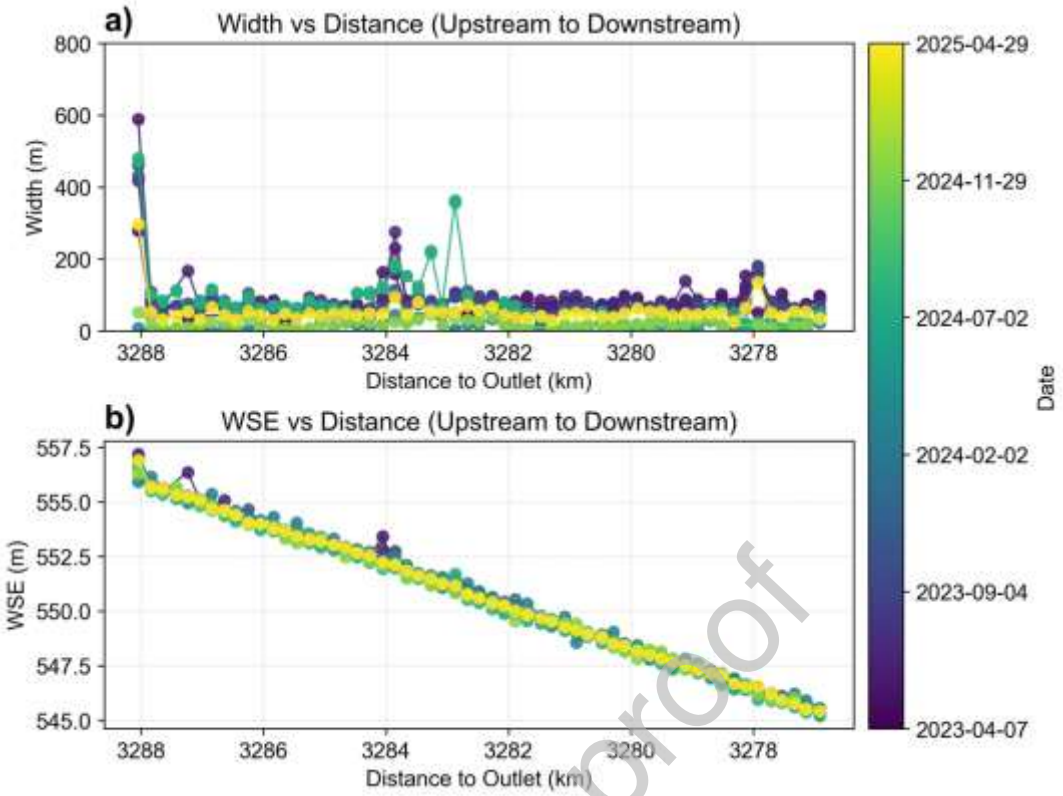


Figure 3. Spatial profiles of (a) width and (b) WSE along the Cedar River (Reach ID: 74294400121).

2.4 USGS Gage Data

Discharge records from the USGS National Water Information System are used as ground truth for calibration (Turnipseed & Sauer, 2010). Gage stations provide continuous discharge time series that overlap with the SWOT observation period. By pairing gage data with SWOT observations at the reach level, we assess calibration performance and quantify performance gaps between reaches with low and high spatial variability. The spatial associations identifying the specific USGS gage paired to each SWORD reach were obtained directly from the SWOT Prior River Database (SoS) (Altenau et al., 2021; Swot, 2025). To temporally align the datasets, the continuous USGS discharge time series were linearly interpolated to the exact timestamps of the SWOT satellite overpasses. Specifically, both datasets were aligned using standard numerical timestamps, allowing us to interpolate the continuous gage records to the

discrete satellite observation times. Finally, the interpolated USGS discharge was converted from English units to metric units (m^3/s).

3. Methods

3.1 Defining Within-Reach Spatial Variability

Following prior work that introduced variability indices to represent the relative fluctuation of hydraulic parameters (Harman et al., 2008; Moody & Troutman, 2002; Rodríguez et al., 2020), we quantify spatial variability as the coefficient of variation (ratio of standard deviation to mean). This measure provides a simple, dimensionless indicator of within-reach variability that can be derived from the SWORD dataset directly:

$$\varepsilon_p = \frac{\sigma_p}{\bar{p}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \bar{p})^2}}{\bar{p}} \quad (1)$$

where ε_p is the within-reach spatial variability of a hydraulic parameter p, $\bar{p}$ is the arithmetic mean of p and σ_p is the standard deviation of p. This formulation is consistent with earlier concepts of variability indices but offers a direct and widely used metric for analyzing spatial heterogeneity in river hydraulics.

3.2 Flow Laws

SWOT discharge estimation adheres to a gage-analogous two-step workflow; however, given the global scarcity of in-situ discharge data, the process relies on satellite-derived "primary data" rather than traditional rating-curve calibration. Typically, this estimation employs a modified Gauckler–Manning–Strickler (Manning's) equation to compute discharge for each satellite overpass (Durand et al., 2023). In this study, we evaluate

three distinct flow laws, all grounded in the modified Manning's equation, but distinguished by three different parameterizations for the hydraulic resistance term.

3.2.1 Modified Manning's Equation

Under a diffusive-wave approximation of gradually varied flow, Manning's equation is expressed as (Durand et al., 2024; Frasson et al., 2021; Tuozzolo et al., 2019).

$$Q_t = \frac{1}{n_t}(A_0 + \delta A_t)^{5/3} W_t^{-2/3} S_t^{1/2} \tag{2}$$

where n_t is Manning's coefficient, W_t is the river width, S_t is the water surface slope, A_0 is the reference cross-sectional area, and δA_t is the time-varying change in area. The validity of the wide-river approximation for our dataset is mathematically and empirically justified in Section S1 of the Supplementary Material.

3.2.2 Manning's Equation with Power Law

Several parameterizations of Manning's coefficient have been proposed. We assess the "Manning's equation for area with power law n" as MWAPN hereafter (Durand et al., 2024):

$$n_t = n_0 \left(\frac{A_0 + \delta A_t}{W_t}\right)^p \tag{3}$$

where n_0, p and A_0 are flow law parameters estimated by calibration.

3.2.3 Manning's Equation with Riverbed Spatial Variability

Following Rodríguez et al. (2020)'s idea, Durand et al. (2024) introduced within-reach spatial variability of riverbed elevation into Manning's coefficient, creating "Manning's equation for area with riverbed elevation spatial variability n" as MWAVN hereafter:

$$n_t = n_\infty \left[1 + \frac{5}{6}\left(\frac{\sigma_z W_t}{A_0 + \delta A_t}\right)^2\right] \tag{4}$$

where n_∞ is the friction coefficient at high flow where spatial variations in parameters tend to have minimal effect, and σ_z is the spatial standard deviation in riverbed elevation. The flow law parameters estimated by calibration are n_∞, σ_z, and A_0.

3.2.4 Manning's Equation with Riverbed and Slope Spatial Variability

Based on MWAVN, we extend the formulation by explicitly incorporating slope variability. To our knowledge, this is the first flow law that accounts for slope spatial variability in Manning's coefficient, which we refer to as MWAVNS ("Manning's equation for area with riverbed elevation and slope spatial variability n") hereafter:

$$n_t = n_\infty \left[1 + \frac{5}{6}\left(\frac{\sigma_z W_t}{A_0 + \delta A_t}\right)^2 + \frac{1}{4}\left(\frac{\sigma_S}{\bar{S}}\right)^2\right] \quad (5)$$

where $\bar{S}$ is the arithmetic mean of the water surface slope S, and σ_S is the standard deviation of S. We calculate slope spatial variability from SWOT node data and use the slope variability as input of this flow law. The flow law parameters estimated by calibration are still n_∞, σ_z, and A_0.

3.3 Flow Law Calibration

Discharge calibration is implemented in the Flow Law Parameter Estimation libraries designed at Byrd Polar (FLaPE-Byrd) with least squares and a bound-constrained optimizer (L-BFGS-B), which is available at https://github.com/mikedurand/FLaPE-Byrd.

In terms of flow law MWAPN, we estimate the parameter vector $\theta = \{n_0 \quad A_0 \quad p\}$ by modeling discharge $Q_{hat}(\theta)$ to USGS gage discharge Q_{gage}, using all valid SWOT observations within the calibration dataset used in this study. We impose physically motivated bounds with $n_0 \in [0.001, 1]$, $A_0 \geq 1 - \min(\delta A_t)$, $p \in \mathbb{R}$, avoiding ill-conditioned evaluations. For flow law MWAVN and MWAVNS, we estimate the

parameter vector $\theta = \{n_\infty \quad A_0 \quad \sigma_z\}$ with $n_\infty \in [0.001, 1]$, $A_0 \geq 1 - \min(\delta A_t)$, $\sigma_z \geq 0.001$.

3.4 Slope Spatial Variability Estimation

To quantify within-reach slope spatial variability, we analyze WSE–distance profiles from SWOT node data for each reach and timestamp. SWOT node quality may have a value of *good*, *suspect*, *degraded* or *bad* (Jet Propulsion Laboratory, 2023). Only node measurements flagged as *good* or *suspect* are retained.

For a given profile, the WSE-distance relationship is approximated by a piecewise linear function. Breakpoints are first identified using a piecewise linear fitting (pwlf) approach, which determines optimal segmentation points by minimizing the overall residual sum of squares across the reach (Hastie et al., 2008; Muggeo, 2003). In this study, each profile is divided into five sub-reaches, which we found sufficient in preliminary tests to capture localized variability while keeping computation tractable.

Within each segment, the water surface slope is estimated using ordinary least squares (OLS) regression:

$$\hat{y}(x) = \alpha + \beta x \tag{6}$$

where $\hat{y}(x)$ is the fitted WSE, x is the distance to outlet, α is the intercept, and β is the estimated water surface slope of the sub-reach. The unbiased OLS estimator of β is given by:

$$\beta = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \tag{7}$$

To ensure that the slope profile forms a continuous representation of water surface slope, adjacent linear segments are constrained to be connected at their breakpoints, analogous to the treatment of hypsometric curves in prior studies (Durand et al., 2024; Markose & Jayappa, 2011).

$$\alpha_i + \beta_i \tau_i = \alpha_{i+1} + \beta_{i+1} \tau_i \quad (8)$$

where τ_i is the ith breakpoint.

At each timestamp, we compute five segment-wise slopes along the reach and summarize within-reach variability by the segment mean and the standard deviation across segments. Repeating this over all timestamps yields a time series of spatial slope variability for each reach, which we use to quantify its influence on discharge calibration.

We then use the gage discharge, Q_{gage}, to build an idealized relationship describing how within-reach slope variability, sv, changes with flow (i.e., $sv = f(Q_{gage})$). We first remove invalid or negative values. To eliminate the severe observational noise often present at the absolute tails of the data distribution, we filter out extreme outlier observations by simply discarding values that fall outside selected percentile ranges (0.5–99.5% for Q_{gage} and 1–99% for sv) (Tukey & McLaughlin, 1963). Next, we sort the remaining observations by $\log(Q_{gage})$ and group them into 10–20 quantile bins. Within each bin, we remove outliers using a median-absolute-deviation (MAD) rule and summarize the bin using the median sv and the interquartile range (IQR) (Helsel et al., 2020; Rousseeuw & Croux, 1993). We discard bins with too few samples and smooth the sequence of bin medians using a rolling median with a window size of 3 bins. Finally, we fit a simple monotone decreasing curve (e.g., exponential, power-law, logarithmic, or rational decay) to the binned medians using robust bounded least squares (El Ghaoui & Lebret, 1997). The slope variability estimation procedures are presented in Figure 11.

3.5 Fluvial Hypsometry Constraints

Durand et al. (2024) introduced a method to infer reach-level fluvial hypsometry by fitting a piecewise-linear relationship between river width and WSE from noisy satellite observations. Individual width and WSE measurements are then projected onto this

curve to produce hypsometrically constrained estimates. While this regularization improves overall precision, it introduces a positive correlation between width and WSE errors. Since the hypsometric curve dictates that width must increase with WSE, an overestimation in one variable mathematically forces an overestimation in the other. In our study, applying these constraints reduces the apparent impact of within-reach spatial variability in SWOT-derived WSE and width primarily by lowering observational noise. The actual effect of hypsometry constraint on SWOT data will be presented in Figure 5.

3.6 Evaluation Error Metrics

To evaluate the accuracy of discharge calibration, we use three complementary metrics: (i) the 1σ error, (ii) the Nash–Sutcliffe efficiency (NSE), and (iii) the Pearson correlation coefficient r. Together, these metrics assess absolute error magnitude (1σ error), overall model efficiency (NSE), and linear association between observed and estimated discharge (r).

The 1σ error is defined as the 67th percentile of the absolute normalized residuals between estimated discharge $\hat{Q}_t$ and gage discharge Q_t:

$$\epsilon_{1\sigma} = P_{67}\left(\frac{|\hat{Q}_t - Q_t|}{\bar{Q}}\right) \tag{9}$$

where $\bar{Q}$ is mean discharge from USGS gage observations and $P_{67}(\cdot)$ denotes the 67th percentile. This metric captures the typical magnitude of relative error, analogous to a one-standard-deviation bound for residuals.

NSE measures the predictive skill of discharge estimates relative to the mean of observations:

$$NSE = 1 - \frac{\sum_t (Q_t - \hat{Q}_t)^2}{\sum_t (Q_t - \bar{Q})^2} \tag{10}$$

NSE values close to 1 indicate high predictive performance, while values near or below 0 indicate poor performance.

The Pearson correlation coefficient r evaluates the linear association between observed and estimated discharge:

$$r = \frac{\sum_t (Q_t - \bar{Q})\left(\hat{Q}_t - \bar{\hat{Q}}\right)}{\sqrt{\sum_t (Q_t - \bar{Q})^2}\sqrt{\sum_t \left(\hat{Q}_t - \bar{\hat{Q}}\right)^2}} \tag{11}$$

where $\bar{\hat{Q}}$ are the means of estimated discharge.

4. Experiment Design

We hypothesize that within-reach spatial variability degrades the accuracy of SWOT discharge calibration. Our experimental design therefore proceeds in three steps to (i) detect any performance degradation, (ii) diagnose whether this degradation arises from observation error, and (iii) check whether the performance degradation can be linked to structural error in the flow laws.

Our first objective is to test whether higher spatial variability is associated with reduced calibration performance. We perform nominal discharge calibration using MWAPN (Eq. 3) and compare accuracy metrics between low- and high-variability reaches. If spatial variability degrades the performance of a spatially averaged flow law, we expect systematically higher errors in the high-variability group; we refer to these higher errors as a "performance gap" between low- and high-variability reaches.

Our second objective is to determine whether any observed performance gap arises from SWOT measurement noise. We repeat the calibration using a hypsometric constraint that reduces width and WSE errors. If the constrained case shows a reduction

in the performance gap, this will indicate that observation noise contributes to the observed performance differences.

Our third objective is to assess whether reduced performance in high-variability reaches reflects structural error in the flow law. Because structural error cannot be measured directly, we use three predictions from spatial-averaging theory as indirect tests. In MWAPN, the exponent p controls how effective resistance ñ changes with discharge, and $|p|$ measures its temporal sensitivity (governing equations detailed in the Appendix B). We therefore examine three indicators: (i) whether calibrated p values imply ñ decreasing with Q ($p < 0$, as predicted by spatial-averaging theory) or the opposite trend, (ii) whether $|p|$ is systematically larger in high-variability reaches than in low-variability reaches, as predicted by spatial-averaging theory, and (iii) whether a flow law that explicitly incorporates measured slope variability reduces errors in high-variability reaches relative to MWAPN. Failure to confirm predictions (i) and (ii) from spatial-averaging theory would undermine the hypothesis that structural flow-law error is driving the performance gap. Demonstration that explicitly incorporating measured river spatial variability into flow laws improves calibration performance (iii) would support the hypothesis that the performance gap is driven by structural error in the flow law. Thus, evidence from these tests will indicate whether spatial-averaging theory is consistent with SWOT discharge calibration results, how we should treat the spatial-variability information, and provide a new direction for improving SWOT discharge performance via improved flow laws.

5. Results

5.1 Calibration Performance Gap

Figure 4 compares the discharge calibration performance of high- and low-variability reaches for the baseline experiment using the MWAPN flow law. Empirical CDFs of the 1σ error, NSE, and correlation coefficient r illustrate the systematic performance divergence between the two groups.

As shown in Figure 4a, the low-variability group consistently exhibits lower 1σ error. Specifically, the median error is 0.25 for low-variability reaches compared to 0.33 for high-variability reaches, representing a performance gap of 0.08. This degradation persists across the distribution, with gaps of 0.05 observed at both the 20th and 80th percentiles.

Figure 4b demonstrates that calibration skill, measured by NSE, is substantially higher for the low-variability group (median NSE = 0.78) than for the high-variability group (median NSE = 0.69), yielding a median performance gap of 0.09. Notably, this disparity widens at the lower tail of the distribution (20th percentile), where the gap extends to 0.14.

Similarly, Figure 4c confirms this trend for r. While median performance remains robust in both cohorts (0.88 vs. 0.85), the low-variability group maintains a distinct advantage, particularly at the 20th percentile where the gap reaches 0.15. Collectively, these results substantiate our hypothesis that elevated within-reach spatial variability significantly compromises the accuracy of SWOT discharge calibration.

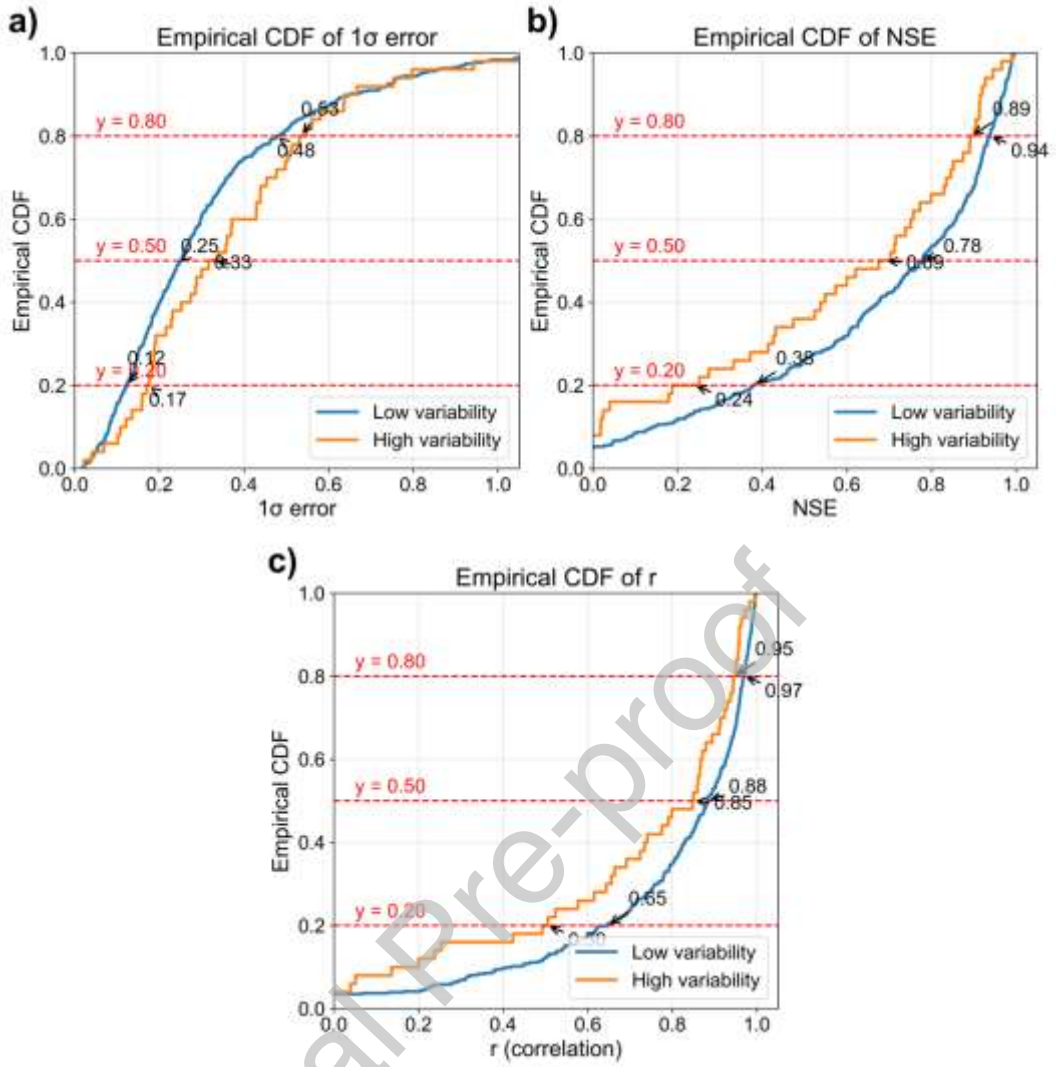


Figure 4. (a) Empirical CDF of the 1σ error for discharge calibration across U.S. paired reaches. The red dashed line marks the 80th percentile, median and 20th percentile. (b) Empirical CDF of NSE for discharge calibration. (c) Empirical CDF of r for discharge calibration.

5.2 SWOT Observation Error

To assess whether SWOT observation error is greater for high spatial variability rivers, thus driving the performance gap, we apply a hypsometry constraint to all the SWOT observations used in the nominal MWAPN baseline experiment. Figure 5 illustrates the hypsometry constraint applied to SWOT width and WSE observations for six example reaches. We apply a hypsometry constraint to all the SWOT observations used in the nominal MWAPN baseline experiment.

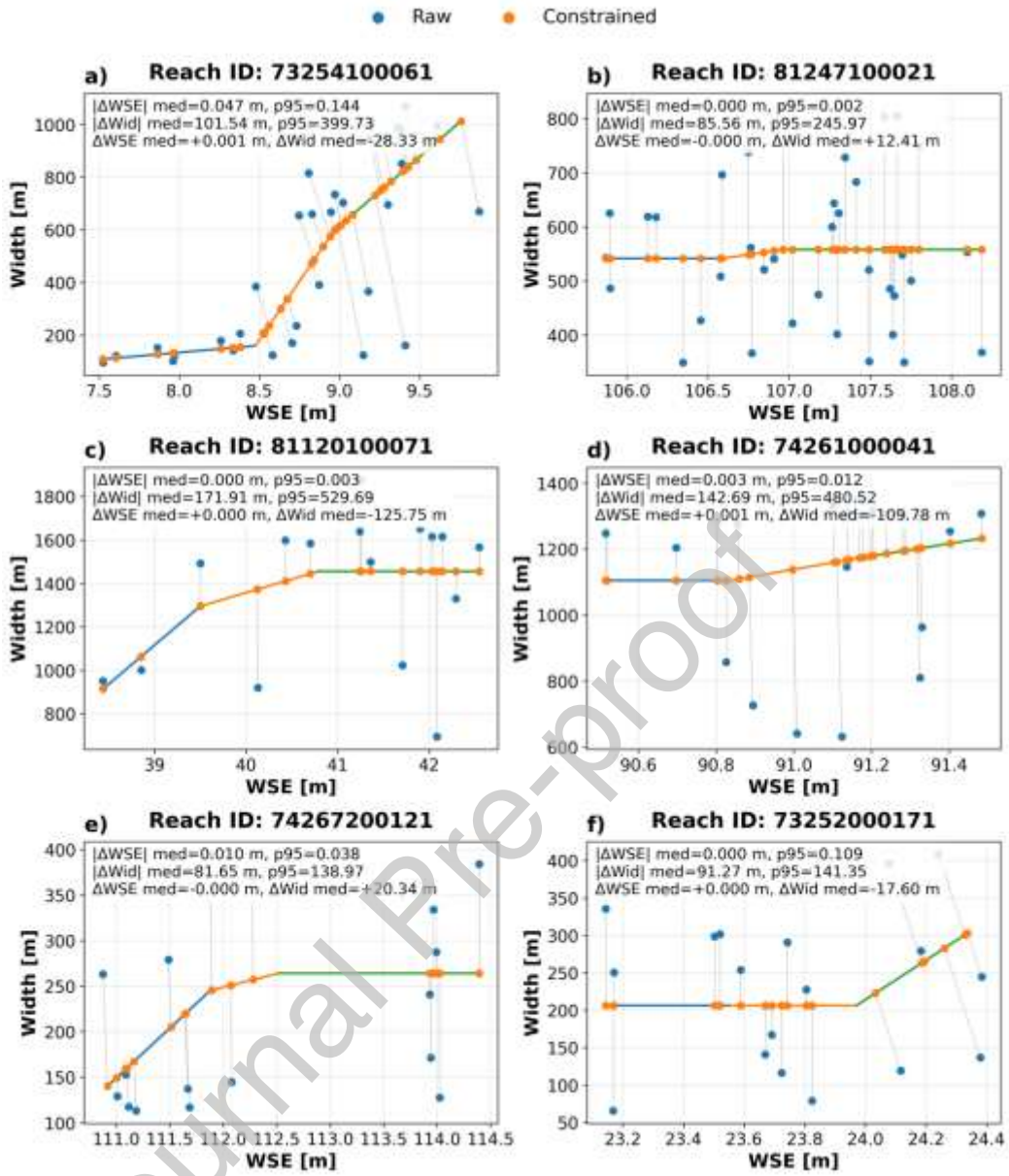


Figure 5. River hypsometric curves for six randomly chosen reaches, along with the unconstrained (blue) and the constrained (orange) observations of WSE and width. The constrained and unconstrained datapoints are connected with grey lines, to illustrate the effect of the hypsometric constraint.

Figure 6 presents the calibration performance following the application of the hypsometric constraint. Notably, this constraint yielded a substantial improvement in global calibration accuracy, exceeding initial expectations. For instance, calibration errors at the 20th percentile among low-variability reaches decreased from 0.14 to 0.09.

The broader implications of this accuracy gain for SWOT discharge estimation are detailed in Section 6.

Regarding the 1σ error (Figure 6a), the hypsometric constraint reduced the performance gap between the two groups. The median error gap narrowed to 0.07 (0.17 for low variability vs. 0.24 for high variability). At the tails, the gap was 0.06 at the 80th percentile and 0.02 at the 20th percentile.

This trend is mirrored in the NSE results (Figure 6b). The median NSE gap narrowed to 0.05 (0.86 vs. 0.81). Crucially, the performance gap effectively vanishes among high-performing reaches (80th percentile gap = 0.03) but persists among lower-performing reaches (20th percentile gap = 0.11).

Similarly, the r results (Figure 6c) show a reduced median gap of 0.03 (0.93 vs. 0.90). Consistent with NSE, the gap is minimal at the 80th percentile (0.02) but remains pronounced at the 20th percentile (0.09).

Figure 6d synthesizes these findings, illustrating that the calibration performance disparity is reduced when width errors are mitigated by the hypsometric constraint. Collectively, these results indicate that the hypsometric constraint not only improves overall calibration accuracy but also reduces the performance gap, particularly among poorly performing reaches. However, the performance gap remains, suggesting that other factors, such as flow-law structural error, may also contribute. We next test whether our data are consistent with this interpretation.

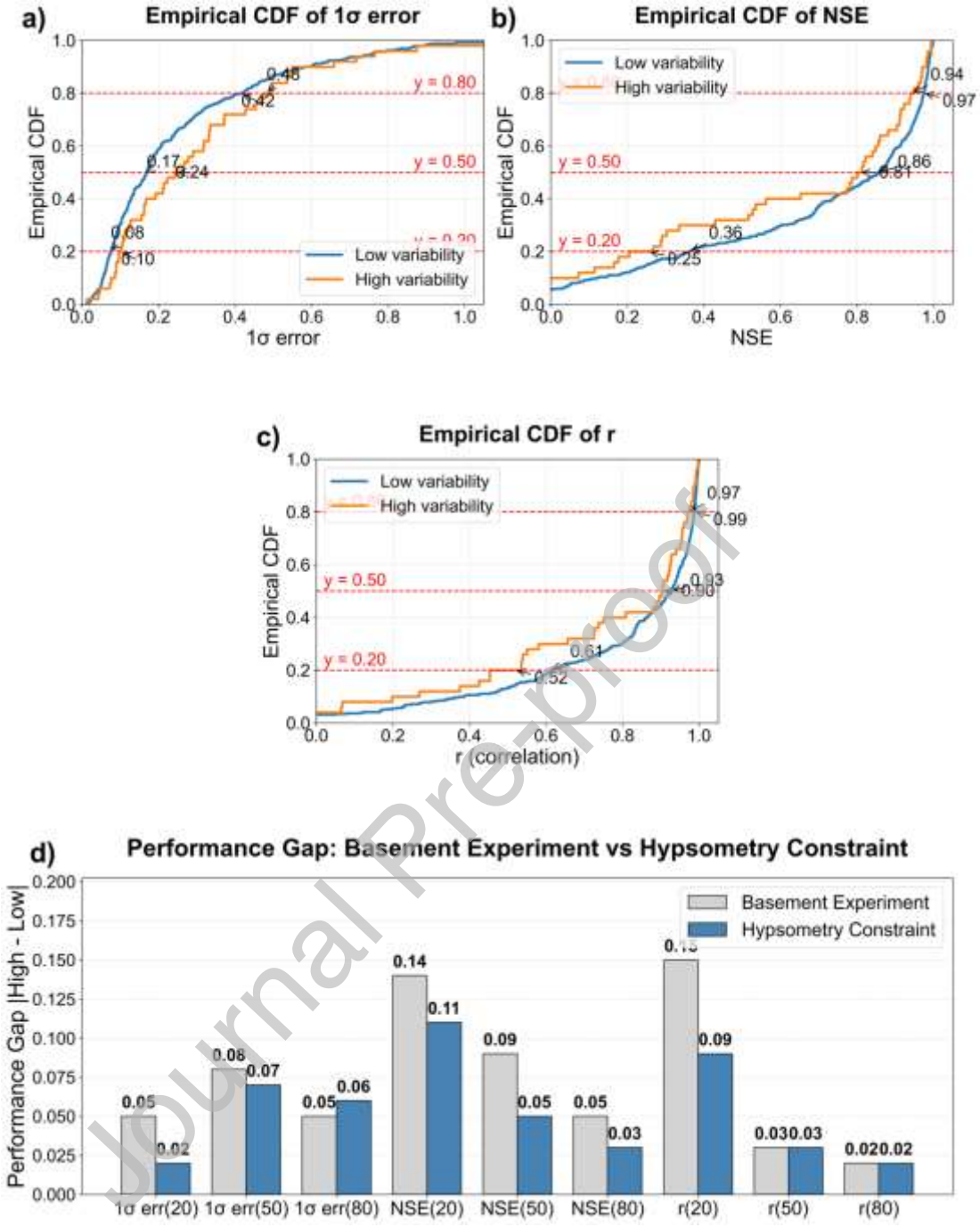


Figure 6. Empirical CDFs of (a) 1σ error, (b) NSE and (c) r under the hypsometric constraint, with dashed lines marking the 20th, 50th and 80th percentiles. (d) Performance gaps between low- and high-variability reaches at these percentiles for the baseline and hypsometric constraint experiments.

In addition to the reach-level summary metrics, we further examined how the hypsometric constraint affects discharge residuals across different flow regimes. For each reach, USGS gage observations were divided into low-, medium-, and high-flow

regimes using the 33rd and 66th percentiles of the reach-specific discharge record. We then computed the relative residual, defined as $(\hat{Q}_t - Q_t)/Q_t$, for both the baseline MWAPN experiment and the hypsometry-constrained experiment. As shown in Supplementary Figure S3, the effect of the hypsometric constraint is not uniform across flow regimes. The largest reduction in residual spread occurs under low- and medium-flow conditions, where the baseline experiment shows broader residual distributions and more frequent large relative errors. In contrast, the high-flow residual distribution is already narrower in the baseline experiment, and the additional improvement from the hypsometric constraint is smaller. Importantly, the constrained experiment does not show an inflation of high-flow residuals or a clear degradation relative to the baseline. This residual-level analysis suggests that the hypsometric constraint primarily improves calibration by reducing low- and medium-flow residual dispersion, while maintaining stable high-flow behavior. This helps explain why both the error metrics improve after applying the constraint, and why the overall low-/high-variability performance gap narrows but does not fully disappear.

5.3 Flow Laws Exploration

To test whether the remaining performance gap may be consistent with flow-law structural error, we hypothesize that systematic deviations may arise when the assumed relationship between flow resistance and discharge does not adequately capture spatial variability within a reach. Specifically, we test three related expectations: (1) that the relationship between resistance and discharge follows the theoretically expected trend; (2) that rivers exhibiting greater spatial variability demonstrate a stronger dependence of discharge on flow resistance; and (3) that the specific flow law parameterized by measured slope variability provides systematic improvement relative to a formulation based primarily on reach-averaged hydraulic properties. Building upon the findings of the previous section, which demonstrated the critical role of the hypsometric constraint

in enhancing SWOT discharge calibration, all analyses in this section are conducted using hypsometry constrained MWAPN results.

Theoretical predictions suggest that the effective resistance ñ should decrease with increasing discharge, implying a negative exponent in the MWAPN flow law. However, empirical calibrations in many reaches reveal the opposite apparent trend, where ñ appears to increase with discharge. To further investigate this discrepancy, we analyze the exponent p estimated from MWAPN to characterize the apparent ñ – Q relationship. Figure 7a shows the probability density function of p, which is approximately normal with the mean slightly skewed toward positive values; only 40.08% (194/484) of reaches exhibit $p \leq 0$, contrary to the theoretical expectation. Taken literally, this result appears to contradict the predictions of spatial-averaging theory. However, this unexpected finding may stem from parameter equifinality or weak identifiability of p, rather than a fundamental deviation from theory. Specifically, for reaches where calibration yields $p > 0$, an alternative parameter set satisfying $p \leq 0$ may still produce similar discharge estimates. Under such conditions, calibrated positive exponents should not be interpreted as direct physical evidence that effective resistance increases with discharge, nor do they necessarily constitute a refutation of spatial-averaging theory. Thus, we interpret this unexpected behavior as evidence of parameter equifinality and partial identifiability of p, rather than as a definitive deviation from spatial-averaging theory.

To test this equifinality interpretation, we imposed a negative-exponent constraint and compared the resulting calibration performance with the original hypsometry-constrained MWAPN calibration. Figure 7b and 7c summarize the performance of Run 1, using the hypsometry constraint only, and Run 2, using both the hypsometry constraint and the negative-exponent constraint. The combined-constraint case (Run 2) yields broadly similar performance to Run 1, indicating that discharge estimates are relatively robust to the sign constraint on p. Because this comparison does not by itself determine whether p is uniquely identifiable, we further evaluated the sensitivity of

calibration skill to fixed values of p in a supplementary profile-based identifiability analysis (Supplementary Figure S4).

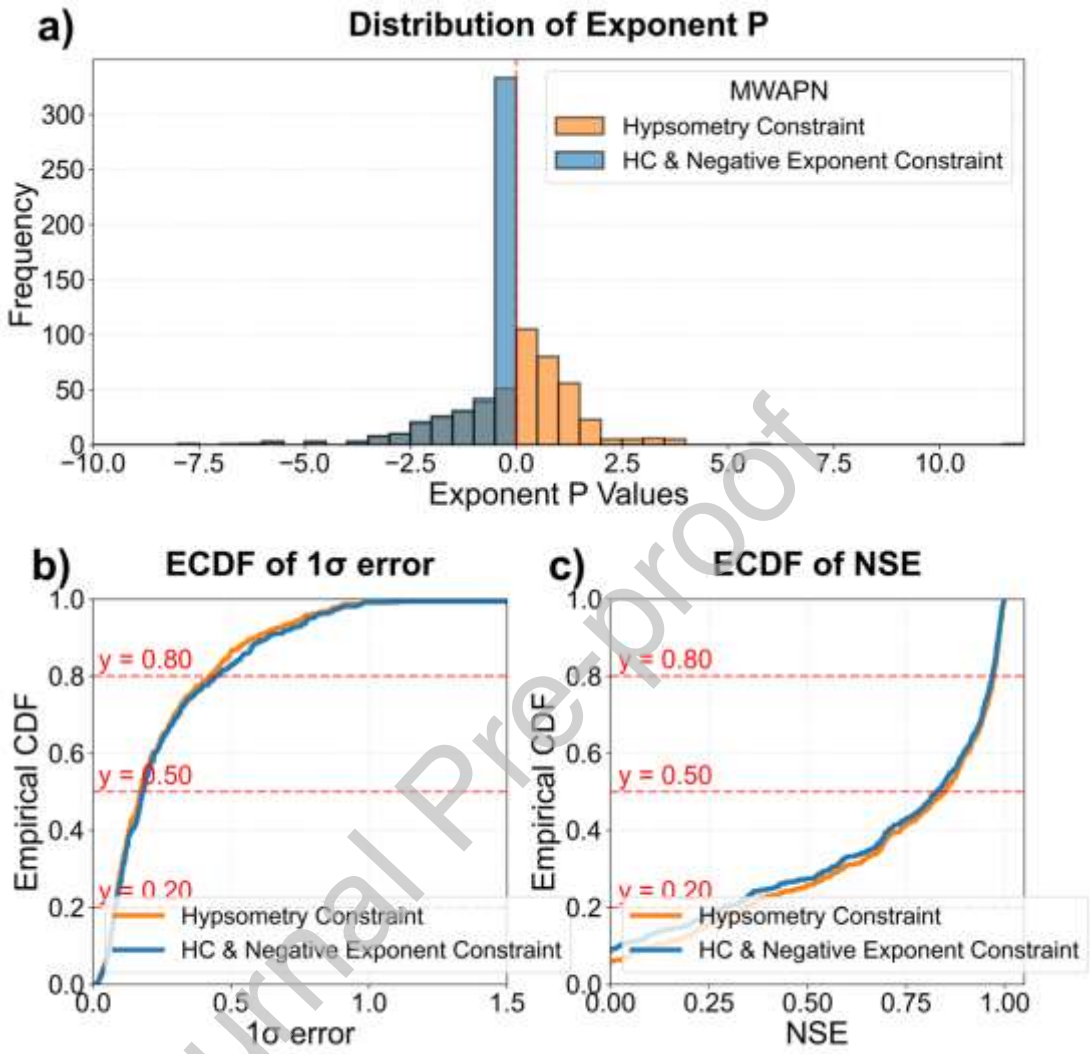


Figure 7. (a) Probability density functions of the exponent p from the MWAPN flow law, comparing the "Hypsometric Constraint" and "Hypsometric Constraint & Negative Exponent Constraint" experimental setups. Empirical CDFs of (b) 1σ error and (c) NSE comparing the "HC" and "HC & NEC" experimental setups

Figure 8 highlights a representative example (Reach 81250100101) of a reach where effective resistance ñ increases with discharge. Similar behavior is consistently observed in other locations (e.g., Reaches 74252600031, 78293000031, and 74287200101; see Figure S4 in the Supplementary Materials). Applying the negative-exponent constraint removes the positive resistance–discharge trend, generally

rendering ñ approximately constant or weakly decreasing over time. However, the calibrated discharge under the combined hypsometry and negative-exponent constraints remains similar to that obtained using the hypsometry constraint alone, illustrating that discharge prediction is more robust than the inferred temporal dynamics of ñ.

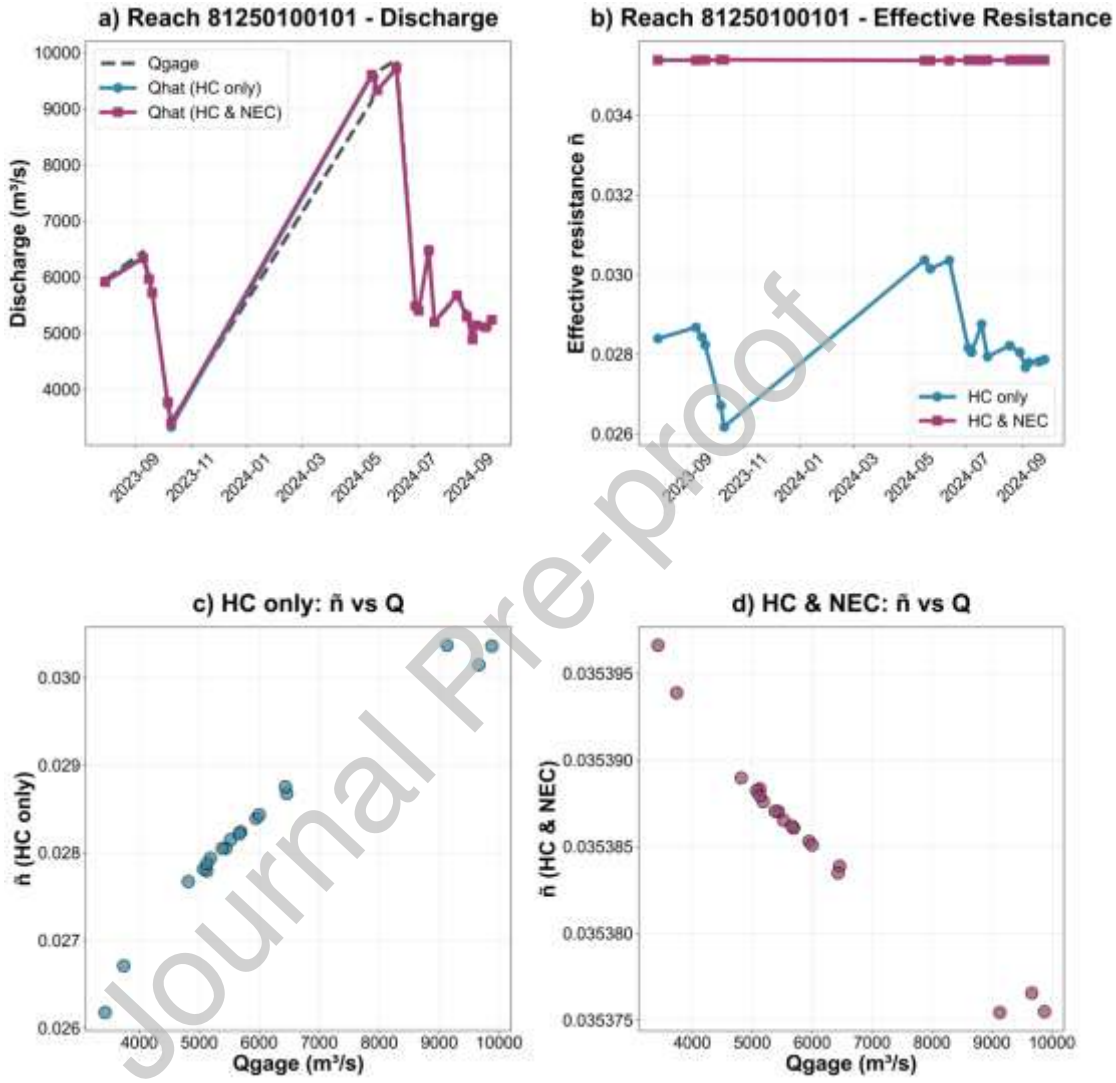


Figure 8. Time series of discharge and effective resistance ñ for Reach 81250100101. Panel (a) compares gage and calibrated discharge under the hypsometry constraint (HC) and the combined hypsometry and negative-exponent constraints (HC&NEC). Panel (b) shows the corresponding temporal variations in effective resistance ñ. Panels (c) and (d) illustrate the relationship between ñ and gage discharge for the HC only and the HC&NEC scenarios, respectively.

We next test the theory's prediction that reaches with greater spatial variability should exhibit larger temporal fluctuations in resistance. This hypothesis is tested by

comparing the magnitude of the resistance exponent ($|p|$)—a proxy for resistance sensitivity—between low- and high-variability groups, under the combined hypsometry and negative-exponent constraints.

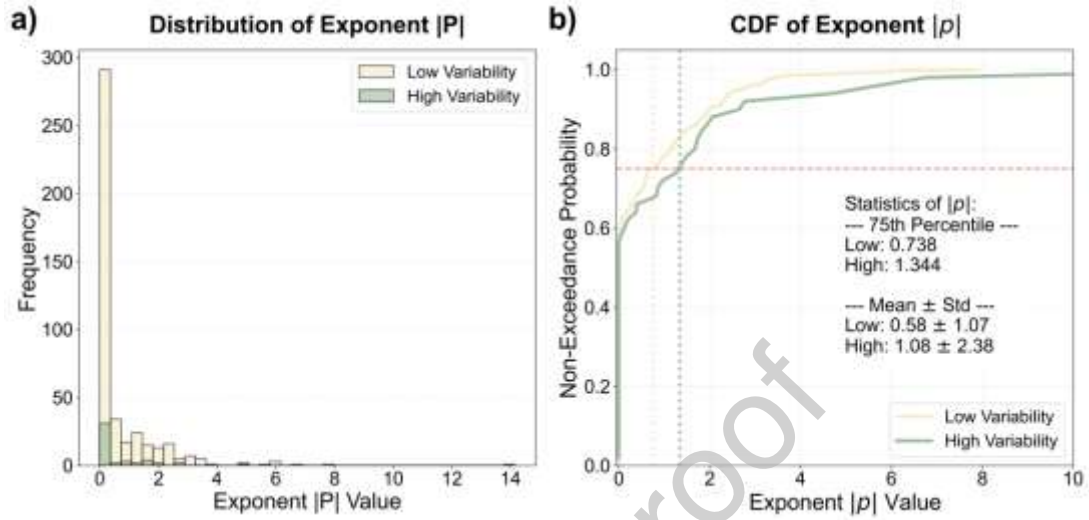


Figure 9. (a) Probability density function (PDF) of |p| for low- and high-variability reaches. (b) Non-exceedance probability distribution function (CDF) of |p|.

As shown in Figure 9a and b, the $|p|$ values for high-variability reaches are consistently larger than those for low-variability reaches, indicating that effective resistance in spatially variable river reaches is significantly more sensitive to changes in discharge. Statistical comparisons further support this pattern: the high-variability group exhibits a mean $|p|$ of 1.08 (standard deviation = 2.38), nearly double the mean of 0.58 (standard deviation = 1.07) observed in the low-variability group. Furthermore, this divergence is pronounced at the upper tail of the distribution, where the 75th percentile increases from 0.74 in the low-variability group to 1.34 in the high-variability group, representing a difference of 0.64.

Finally, we evaluate whether the specific slope-variability parameterization tested here provides improved calibration skill compared to formulations relying only on reach-averaged hydraulic properties. We quantify slope spatial variability through the following procedure, exemplified in Figure 10 using a reach on the Klamath River (Reach ID: 77480100021). Figure 10a and 10b contrast WSE–distance profiles under

low- and high-flow conditions, where slope fluctuations are more localized in the former and smoother in the latter. Figure 10c and 10d illustrate the application of piecewise OLS fitting (described in Section 3.4) for representative dates (2024-09-11 and 2023-04-24), highlighting how the segmentation adapts to different hydraulic regimes.

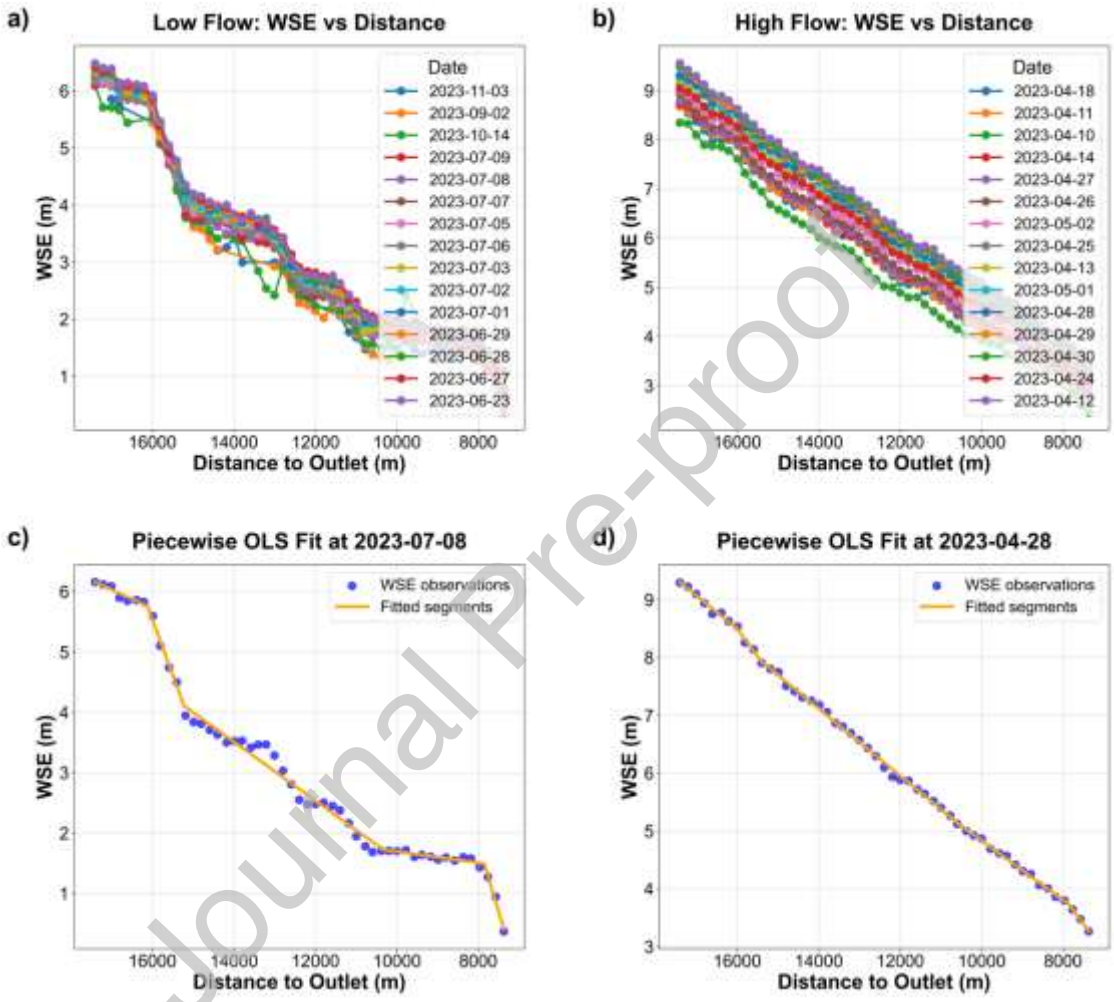


Figure 10. WSE vs. Distance to Outlet for Reach 77480100021: (a) Low flows, (b) High flows, (c) Fitted slope segments for 2023-07-08, (d) Fitted slope segments for 2023-04-28.

Figure 11 illustrates all the data processing procedures with an example of Saint John River (Reach ID: 72608900081).

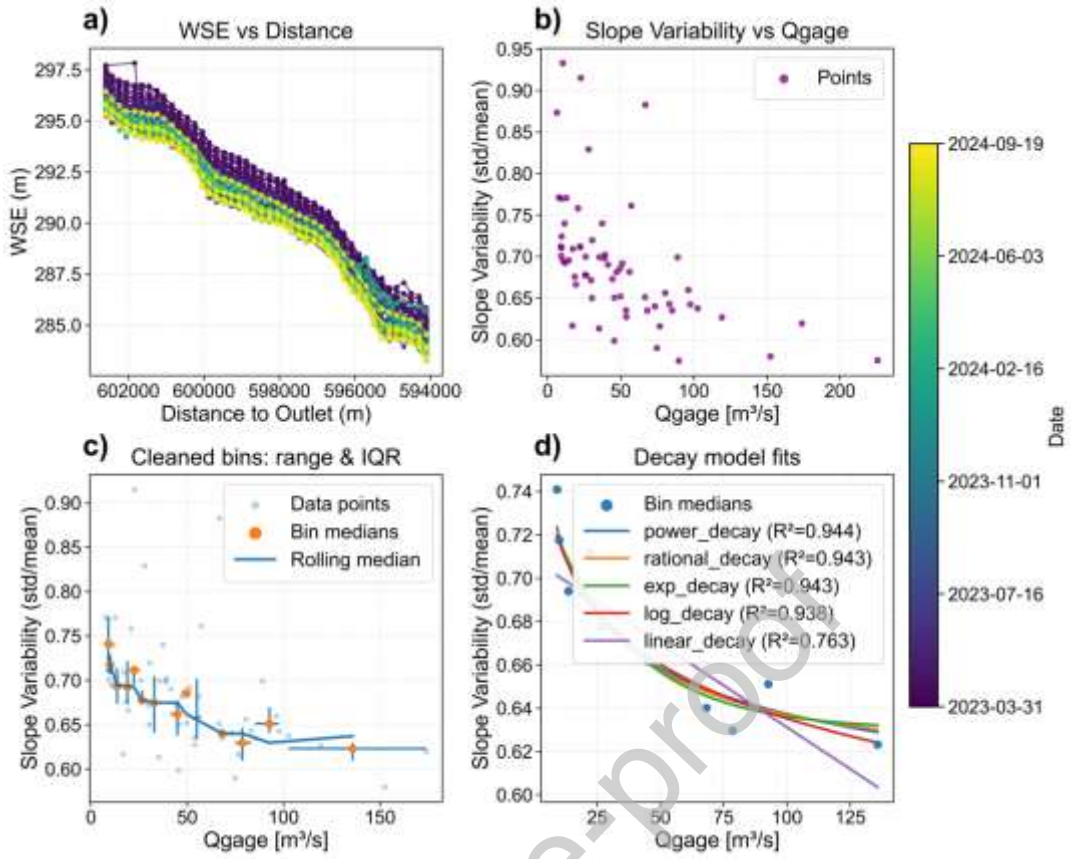


Figure 11. (a) Time series of WSE vs. Distance to Outlet for Reach 72608900081. (b) Scatter points of the estimated within-reach slope spatial variability vs. USGS discharge. (c) Data cleaning with median-absolute-deviation (MAD) rule. (d) Curve-fitting results for various decay functions describing the relationship between the slope variability and gage discharge.

We evaluate the calibration performance of MWAPN, MWAVN, and MWAVNS among a few reaches. Klamath River (Reach ID: 77480100021) shows an example of high calibration accuracy from three flow laws (Figure S6 in Supplementary Document). The overall NSE values of MWAVN and MWAVNS are 0.968 and 0.969, correlation coefficients r are 0.984 and 0.985, and the 1σ errors are 0.086 and 0.091, respectively. However, both MWAVN and MWAVNS perform slightly worse than MWAPN, which achieves an NSE of 0.990, an r of 0.995, and a 1σ error of 0.051.

For reaches with lower calibration accuracy, we introduce the hypsometry constraint as the control group. Umpqua River (Reach ID: 78100600061, width variability = 0.2313) presents a clear example that MWAVNS is not able to improve calibration accuracy

significantly. Calibration performance increases monotonically from MWAPN to MWAVN to MWAVNS in terms of NSE, both without constraints (0.7632 to 0.7946 to 0.8015) and with hypsometry constraints (0.9892 to 0.9894 to 0.9896).

Table 1. Calibration performance metrics for Umpqua River (Reach ID: 78100600061).

	NSE	**1σ error**
MWAPN	0.7632	0.2987
MWAVN	0.7946	0.2251
MWAVNS	0.8015	0.2280
MWAPN-HC	0.9892	0.0692
MWAVN-HC	0.9894	0.0687
MWAVNS-HC	0.9896	0.0676

6. Discussion

In this study, we examine the sensitivity of SWOT-derived river discharge estimates to within-reach spatial variability. While SWOT discharge data products produced thus far have been derived using methods applicable in ungaged basins, we here explore the impact of river-reach spatial variability in the context of calibrating in situ gage discharge to SWOT observations of WSE, width, and slope across river reaches. Because these calibrated accuracies are directly informed by ground truth, they represent an upper limit on achievable SWOT discharge accuracy. The insights we derive here shed light on two specific aspects of estimating SWOT river discharge even in ungaged basins, namely (i) the extent to which SWOT observation errors exacerbate performance degradation in morphologically complex reaches, and (ii) the sensitivity of flow laws to structural errors arising from within-reach spatial variability.

The results presented herein demonstrate that the accuracy of SWOT river discharge estimates is significantly compromised by within-reach spatial variability. In low

variability reaches, the median 1σ SWOT discharge error is 0.25, and in the high variability group the error is 0.33, representing an increase of approximately 32%. A similar performance gap is observed in the NSE: low variability reaches achieve a median NSE of 0.78, significantly outperforming the high variability group (0.69). While the existence of a performance gap was anticipated, its magnitude is notable. We therefore investigate two primary drivers of this degradation: errors inherent in the SWOT observations themselves, and structural limitations within the flow laws used to translate these observations into discharge estimates.

It would not be surprising for error levels of SWOT observations of river WSE, width, and slope to be higher in rivers exhibiting more spatial variability. Measurements of river width, in particular, are susceptible to increased noise in morphologically complex channels due to a higher prevalence of mixed water–land pixels along river boundaries. To address this, we employ the recently developed river hypsometric constraint described by (Durand et al., 2024), which forces SWOT width and WSE observations to conform to a monotonic increasing relationship, thereby mitigating variability-related width errors.

The results of these hypsometrically constrained analyses yield a finding of critical importance for SWOT discharge estimation, even beyond the immediate scope of spatial variability. Specifically, the performance of hypsometrically constrained calibrations is superior to the nominal case. For low-spatial-variability reaches, the median 1σ error decreases to 0.17 (compared to a baseline of 0.25), and the median NSE improves to 0.86 (baseline 0.78). These substantial gains strongly suggest that future operational SWOT discharge algorithms should incorporate hypsometric constraints on observations to optimize accuracy.

Furthermore, these results indicate that the hypsometric constraint not only enhances overall calibration accuracy but also narrows the performance gap between high- and low-variability reaches, particularly among lower-performing reaches. This outcome

supports our hypothesis that SWOT observation errors—especially width errors—are an important factor mediating the impact of spatial variability on discharge estimation performance, although the remaining performance gap indicates that other sources of error may also contribute. We note, however, that the hypsometric constraint should not be interpreted as a uniform reduction of all error sources. While the constraint reduces independent noise in width and WSE observations, it can also introduce covariance between width and WSE errors. The stratified residual analysis in Supplementary Figure S3 suggests that the net benefit is flow-regime dependent: residual spread is reduced most clearly under low- and medium-flow conditions, whereas the high-flow residual distribution changes less because the baseline residuals are already narrower. Importantly, the constrained case does not show an evident inflation of high-flow residuals, which helps explain why NSE improves despite its sensitivity to large discharge errors. These results suggest that the hypsometric constraint provides a net improvement in calibration diagnostics, but its benefit is not uniform across flow regimes.

The study also attempts to identify the signature of spatial variability in structural error. This is more difficult, for multiple reasons, especially because it is challenging to fully isolate structural and observational error. We achieved partial isolation of observational error by comparing identical calibrations using hypsometrically constrained versus unconstrained (baseline) SWOT data; since only the input data differed, the resulting performance shift can be attributed to observational quality. Building from the theoretical framework of (Rodríguez et al., 2020), we investigated whether temporal variability of the within-reach spatial variability leads to identifiable errors in flow laws. Mathematically, such variability is predicted to induce a monotonically decreasing dependence of effective flow resistance on river discharge.

When assessing calibration results using hypsometrically constrained data, we found that only 40.08% of the reaches exhibited the theoretically expected negative exponent ($p < 0$), implying a decreasing resistance-discharge relationship. However,

empirical contradictions—where resistance appears to increase with flow—do not necessarily refute the theory. As noted by Yoon et al. (2016), flow laws exhibit a high degree of equifinality, where multiple combinations of resistance and bathymetry parameters yield nearly identical discharge time series. Since imposing a negative-exponent constraint produces broadly similar calibration performance, the positive exponents observed in the unconstrained case should not be interpreted as direct physical evidence that effective resistance increases with discharge. Instead, they likely reflect parameter equifinality and partial identifiability of p, whereby compensating changes in n_0, A_0, and p can yield similar discharge time series.

To further evaluate this identifiability issue, we performed a supplementary profile-based analysis for the positive-p reaches with successful profile calibrations (n=289; Supplementary Figure S5). In this analysis, p was fixed over a range of values while n_0 and A_0 were re-optimized. The best negative-exponent solution produced a median ΔNSE of 0.01, with an interquartile range of 0.0014–0.0524. Approximately 48.4% of reaches had ΔNSE<0.01, and 57.1% had ΔNSE<0.02. These results indicate that p is weakly identifiable for a substantial subset of reaches, but not universally so. The best negative exponents were concentrated near $p = -0.001$, suggesting that the constrained model generally preserves discharge skill by selecting near-zero negative exponents rather than strongly negative ones. Therefore, the calibrated discharge estimates are relatively robust to the sign constraint on p, especially near the $p = 0$ boundary, whereas the inferred effective-resistance dynamics are less robust and should not be over-interpreted at the individual-reach level.

Importantly, this identifiability limitation does not rule out the influence of time-varying spatial variability in flow laws. Rather, it indicates that the present calibration framework cannot uniquely separate time-varying effective-resistance dynamics from compensating adjustments in other flow-law parameters. Because Rodríguez et al. (2020) predict that time-varying spatial variability can influence effective resistance,

our results should be interpreted as inconclusive regarding the detailed form of that relationship, rather than as evidence against the theory.

We also examine whether the magnitude of the exponent is systematically larger in high-variability reaches. The results corroborate this expectation: the low-variability group show a mean $|p|$ of 0.58 (standard deviation = 1.07), whereas the high-variability group displays a significantly elevated mean of 1.08 (standard deviation = 2.38). The high-variability group exhibits consistently larger $|p|$ values and a higher upper tail compared to the low-variability group, indicating a clear shift toward stronger magnitude exponents. We interpret this result as additional evidence that spatial averaging governs the observed discharge–resistance relationship.

We evaluated the efficacy of explicitly modeling time-varying spatial variability across several representative systems, including the Klamath, Umpqua, and Saint John Rivers. We found that the performance disparities among MWAPN, MWAVN, and MWAVNS were negligible. These results do not support the hypothesis that the specific explicit slope-variability term tested here yields consistent improvements in SWOT discharge calibration. However, they do not rule out structural flow-law error more generally. The lack of improvement may reflect uncertainty in SWOT node-derived slope variability, amplification of WSE errors when estimating local slopes, or parameter equifinality among resistance and bathymetry-related terms. Consequently, while spatial variability remains a theoretically important factor, further research is required to develop flow-law structures and uncertainty treatments capable of more effectively assimilating this morphological information.

We further examine the sensitivity of the observed performance gap to the specific classification threshold, as detailed in Figure 12. By varying the threshold from 0.20 to 0.60, we confirm the stability of our findings. Across this entire range, the difference in median 1σ error difference (blue line) remains positive, indicating consistently higher errors for the high-variability group. Conversely, the median NSE difference (orange

line) remains negative, confirming consistently lower predictive skill for high variability reaches. While the magnitude of the gap varies with the threshold, the adverse effect of high width variability on calibration performance is robust.

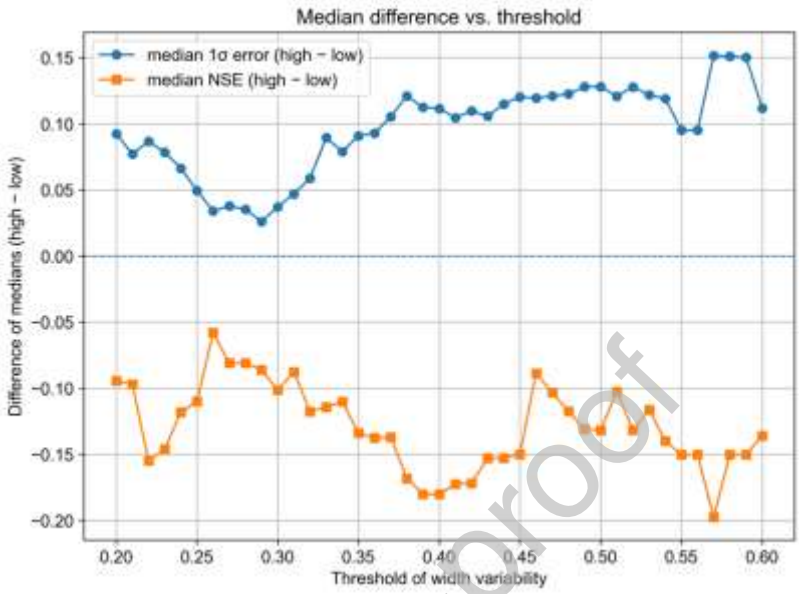


Figure 12. Relationship between the threshold used to separate low- and high-width variability groups and the resulting differences in calibration performance metrics. The blue line shows the difference in median 1σ error (high – low), and the orange line shows the difference in median NSE (high – low), as a function of the width variability threshold.

7. Conclusions

This study quantifies the impact of within-reach spatial variability on the accuracy of SWOT-derived discharge estimates. Reaches with higher spatial variability exhibit lower calibration performance: median 1σ error is roughly 32% higher and median NSE lower by about 0.09 in the high variability group relative to the low variability group. These results confirm that reach-scale spatial variability is a critical determinant of SWOT discharge calibration performance.

Applying a river hypsometric constraint to SWOT observations narrows the performance gap and improves overall calibration accuracy. This improvement suggests that the observed performance degradation in spatially variable reaches is related to observation noise—specifically, width errors arising from mixed water–land pixels along complex channel boundaries.

Analyses of the flow-law structure give mixed support for the theoretical prediction that effective resistance decreases with discharge. While many reaches exhibit anomalous positive exponents, imposing a negative-exponent constraint yields broadly similar calibration performance at the population level, especially when the constrained solution lies near $p = 0$. This suggests that positive exponents likely reflect parameter equifinality and partial identifiability of p, rather than a fundamental deviation from spatial-averaging theory. Thus, discharge estimates are more robust than the inferred effective resistance dynamics, which should be interpreted cautiously at the individual-reach level.

Furthermore, a comparative analysis reveals that high-variability reaches consistently exhibit larger absolute exponent values ($|p|$) than their low-variability counterparts. This systematic shift toward higher sensitivity aligns with theoretical expectations, providing corroborative evidence that spatial averaging mechanisms fundamentally shape the discharge–resistance relationship.

However, in the selected reaches examined, MWAVNS did not provide consistent improvement relative to MWAPN or MWAVN. This result does not rule out structural flow-law error; instead, it suggests that this particular formulation and the current SWOT-derived slope variability estimates are not sufficient to produce consistent calibration gains.

Overall, these findings suggest two actionable directions for improving SWOT discharge estimation. First, observation processing should explicitly account for morphologically complex and spatially variable reaches. In practice, hypsometric

constraints should be considered as a standard preprocessing or quality-control step for SWOT width and WSE observations, especially for reaches with high width variability. Width-variability metrics from SWORD could also be used to flag reaches so that these reaches receive stronger observation filtering, larger assigned uncertainty, or additional node-level quality control.

Second, our results show that the tested slope-variability term does not improve calibration, but this does not rule out structural flow law limitations. Future work should therefore test alternative representations of spatial variability, such as formulations based on node-level WSE profiles, reach-scale bathymetric variability, hydraulic regime classification, or uncertainty-aware calibration frameworks. These approaches should be evaluated not only by discharge accuracy, but also by whether the inferred resistance parameters are identifiable and physically robust.

Acknowledgements

A portion of this research was performed at the Jet Propulsions Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

Appendix

Appendix A: USGS-SWOT Paired Reaches Selection and SWOT Data Preprocessing

We preprocess SWOT reach-time observations of USGS gage paired reaches with a three-step screen:

1. Data Acquisition and Pre-processing

Data is ingested from the NASA Hydrocron API (Time Series endpoint) or local SWORD-formatted repositories. Upon ingestion, the system standardizes the dataset by identifying specific "No Data" markers and converting them to null values to ensure accurate statistical handling.

2. The Standard Filtering Pipeline

All river reaches undergo a sequential, four-stage quality control process to ensure data fidelity.

Stage 1: Slope Lower Bound Correction

To address non-physical slope values, a correction algorithm is applied to water surface slope and energy slope measurements.

Dynamic Limit Determination: A minimum physical slope threshold is established for each reach. If a prior SWORD slope is available and exceeds the default minimum (3.4e-5), the prior is used as the threshold; otherwise, the default minimum is applied.

Replacement: Any slope observations falling below this physical minimum (effectively zero or negative) are replaced with the determined threshold value.

Stage 2: Geometric and Environmental Thresholds

Data points are flagged as invalid if they fail to meet specific geometric or environmental criteria:

Reach Geometry: Reaches must meet minimum requirements for prior width (70 m) and reach length (7 km).

Satellite Geometry: Observations are discarded if the cross-track distance is too small (<10 km, near-nadir effects) or too large (>60 km, edge-of-swath effects).

Environmental Quality: Data is rejected if ice is detected, if "dark water" fractions are too high (indicating poor signal return), or if the fraction of valid observations within the reach is below 50%.

Calibration: Points with poor crossover calibration metrics are excluded.

Stage 3: Official Quality Flag Masking

The protocol applies the official SWOT quality bitmask to the discharge quality flag. A specific binary mask (value 507510784) is used to isolate critical system warnings. Any data point containing these specific error bits is immediately flagged as invalid.

Stage 4: Statistical Outlier Detection

A statistical filter is applied only to data that has successfully passed the Threshold and Quality Flag stages.

Method: A Tukey's Fences test (Interquartile Range method with a 1.5 multiplier) is utilized.

Application: The test checks for statistical outliers in three key variables: Water Surface Elevation (WSE), Width, and Slope.

Logic: A data point is rejected if it is a statistical outlier in any of these three variables.

3. Ground Truth Matching (USGS Integration)

To validate the satellite observations, SWOT data is paired with in-situ USGS gage records:

Temporal Alignment: USGS discharge records are linearly interpolated to the exact timestamp of the SWOT overpass.

Unit Standardization: USGS discharge values are converted from cubic feet per second to cubic meters per second to match SWOT native units.

Appendix B: Spatial Averaging Theory Mathematical Illustrations

At-a-station hydraulic geometry (AHG) relates river depth to discharge by:

$$Q_t = a(H_t - H_0)^b = aD_t{}^b \tag{12}$$

where a and b are the AHG coefficient and exponent, respectively. Depth D_t is approximated by the difference between measured water surface elevation H_t and riverbed elevation H_0.

Manning's equation involves the hydraulic radius, $R_{h,t}$. Strictly, $R_{h,t} = A_t/P_t$, where P_t is the wetted perimeter. Under the wide-river approximation used here, $P_t \approx W_t$, such that $R_{h,t} \approx A_t/W_t$. Since the mean flow depth is $D_t = H_t - H_0 \approx A_t/W_t$, D_t is used as a wide-river approximation to $R_{h,t}$, while the two quantities remain physically distinct.

$$D_t = H_t - H_0 \approx \frac{A_0 + \delta A_t}{W_t}, \qquad R_{h,t} \approx D_t \tag{13}$$

Combining the flow law MWAPN (Eq. 3) with Eq. 12 and Eq. 13 yields the time-varying resistance n_t:

$$n_t = n_0 \left[\left(\frac{Q_t}{a}\right)^{1/b}\right]^p = n_0 \left(\frac{Q_t}{a}\right)^{\frac{p}{b}} \tag{14}$$

Differentiating with respect to discharge:

$$\frac{dn_t}{dQ_t} = n_0 \cdot \left(\frac{p}{b}\right) \cdot (a)^{-\frac{p}{b}} \cdot (Q_t)^{\frac{p}{b}-1} = \left(\frac{p}{b}\right) \cdot \left(\frac{n_t}{Q_t}\right) \tag{15}$$

In Eq. 15, the terms n_0, a, b, and Q_t are strictly positive. Thus, the sign of the derivative depends entirely on p. For the effective resistance to decrease with discharge (i.e., $dn_t/dQ_t < 0$) as predicted by spatial averaging theory, p must be negative. We

therefore examine the distribution of estimated p to assess alignment with theoretical expectations.

To assess sensitivity, we linearize Eq. 15 for small perturbations:

$$\Delta n_t \approx \frac{dn_t}{dQ_t} \cdot \Delta Q_t = \left(\frac{p}{b}\right) \cdot \left(\frac{n_t}{Q_t}\right) \cdot \Delta Q_t \tag{16}$$

Dividing by n_t yields the fractional change in resistance. Taking the absolute value allows us to compare magnitudes:

$$\left|\frac{\Delta n_t}{n_t}\right| \approx \frac{|p|}{b} \cdot \left|\frac{\Delta Q_t}{Q_t}\right| \tag{17}$$

Eq. 17 demonstrates that for a given fractional change in discharge $|\Delta Q_t/Q_t|$, the fractional change in resistance $|\Delta n_t/n_t|$ scales linearly with $|p|$ (assuming positive b is fixed across reaches). Therefore, theory predicts that reaches with high spatial variability should exhibit larger $|p|$ values and consequently larger temporal fluctuations in resistance than low variability reaches.

Reference

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science, 361*(6402), 585-588.

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. *Water resources research, 57*(7), e2021WR030054.

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2025). *SWOT River Database (SWORD) (Version v17b)*. Retrieved from: https://doi.org/10.5281/zenodo.15299138

Andreadis, K., Brinkerhoff, C., & Gleason, C. (2020). Constraining the assimilation of SWOT observations with hydraulic geometry relations. *Water resources research, 56*(5), e2019WR026611.

Andreadis, K., Coss, S. P., Durand, M., Gleason, C. J., Simmons, T. T., Tebaldi, N., et al. (2025). A first look at river discharge estimation from SWOT satellite observations. *Geophysical Research Letters, 52*(9), e2024GL114185.

Archfield, S. A., & Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water resources research, 46*(10).

Beven, K. (1989). Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology, 105*(1-2), 157-172.

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. In *Remote sensing and water resources* (pp. 117-147): Springer.

Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., & Garambois, P.-A. (2018). Satellite remote sensing estimation of river discharge: Application to the Yukon River Alaska. *Journal of Hydrology, 561*, 1000-1018.

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.

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.

Casagli, N., Intrieri, E., Tofani, V., Gigli, G., & Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment, 4*(1), 51-64.

Depetris, P. (2021). The importance of monitoring river water discharge. *Frontiers in Water, 3*, 745912.

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).

Durand, M., Dai, C., Moortgat, J., Yadav, B., de Moraes Frasson, R. P., Li, Z., et al. (2024). Using river hypsometry to improve remote sensing of river discharge. *Remote Sensing of Environment, 315*, 114455.

Durand, M., Gleason, C., de Moraes Frasson, R. P., & Pavelsky, T. (2020). *The potential of SWOT river discharge estimates to constrain hydrological processes globally in Ungaged basins.* Paper presented at the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium.

Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., et al. (2023). A framework for estimating global river discharge from the surface water and ocean topography satellite mission. *Water resources research, 59*(4), e2021WR031614.

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.

Durand, M., Rodriguez, E., Alsdorf, D. E., & Trigg, M. (2009). 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.

Eggleston, J., Mason, C., Bjerklie, D., Durand, M., Dudley, R., & Harlan, M. (2024). Siting considerations for satellite observation of river discharge. *Water resources research, 60*(6), e2023WR034583.

El Ghaoui, L., & Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on matrix analysis and applications, 18*(4), 1035-1064.

Fjørtoft, R., Gaudin, J.-M., Pourthié, N., Lalaurie, J.-C., Mallet, A., Nouvel, J.-F., et al. (2013). KaRIn on SWOT: Characteristics of near-nadir Ka-band interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing, 52*(4), 2172-2185.

Frasson, R. P. d. M., Durand, M. T., Larnier, K., Gleason, C., Andreadis, K. M., Hagemann, M., et al. (2021). Exploring the factors controlling the error characteristics of the Surface Water and Ocean Topography mission discharge estimates. *Water resources research, 57*(6), e2020WR028519.

Fu, L. L., Pavelsky, T., Cretaux, J. F., Morrow, R., Farrar, J. T., Vaze, P., et al. (2024). The surface water and ocean topography mission: A breakthrough in radar remote sensing of the ocean and land surface water. *Geophysical Research Letters, 51*(4), e2023GL107652.

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors, 16*(8), 1298.

Gleason, C., Garambois, P.-A., & Durand, M. (2017). Tracking river flows from space. *EOS Earth & Space Science News*(98).

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.

Greguska, F., Tebaldi, N., McDonald, V., Vanesa, Nickles, C., & Wood, J. (2025). Hydrocron. Zenodo.

Gust, G. (1988). Skin friction probes for field applications. *Journal of Geophysical Research: Oceans, 93*(C11), 14121-14132.

Hagemann, M., Gleason, C., & Durand, M. (2017). BAM: Bayesian AMHG‐Manning inference of discharge using remotely sensed stream width, slope, and height. *Water resources research, 53*(11), 9692-9707.

Harman, C., Stewardson, M., & DeRose, R. (2008). Variability and uncertainty in reach bankfull hydraulic geometry. *Journal of Hydrology, 351*(1-2), 13-25.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). Linear methods for regression. In *The elements of statistical learning: Data mining, inference, and prediction* (pp. 43-99): Springer.

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., & Gilroy, E. J. (2020). *Statistical methods in water resources*. Reston, Virginia.

Jet Propulsion Laboratory. (2023). *SWOT Algorithm Theoretical Basis Document: Level 2 KaRIn High Rate River Single Pass (L2_HR_RiverSP) Science Algorithm Software* (D-105505). Retrieved from

Larnier, K., Monnier, J., Garambois, P.-A., & Verley, J. (2021). River discharge and bathymetry estimation from SWOT altimetry measurements. *Inverse problems in science and engineering, 29*(6), 759-789.

Legleiter, C. J., & Roberts, D. A. (2005). Effects of channel morphology and sensor spatial resolution on image-derived depth estimates. *Remote Sensing of Environment, 95*(2), 231-247.

Li, S.-G., Venkataraman, L., & McLaughlin, D. (1992). Stochastic theory for irregular stream modeling. Part I: flow resistance. *Journal of Hydraulic Engineering, 118*(8), 1079-1090.

Mahfoze, O., & Laizet, S. (2017). Skin-friction drag reduction in a channel flow with streamwise-aligned plasma actuators. *International Journal of Heat and Fluid Flow, 66*, 83-94.

Markose, V. J., & Jayappa, K. (2011). Hypsometric analysis of Kali River Basin, Karnataka, India, using geographic information system. *Geocarto International, 26*(7), 553-568.

Moody, J. A., & Troutman, B. M. (2002). Characterization of the spatial variability of channel morphology. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 27*(12), 1251-1266.

Muggeo, V. M. (2003). Estimating regression models with unknown break‐points. *Statistics in medicine, 22*(19), 3055-3071.

Oubanas, H., Gejadze, I., Malaterre, P. O., Durand, M., Wei, R., Frasson, R. P., & Domeneghetti, A. (2018). Discharge estimation in ungauged basins through variational data assimilation: The potential of the SWOT mission. *Water resources research, 54*(3), 2405-2423.

Paiva, R. C., Durand, M. T., & Hossain, F. (2015). Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. *Water resources research, 51*(1), 430-449.

Peral, E., Esteban-Fernández, D., Rodríguez, E., McWatters, D., De Bleser, J.-W., Ahmed, R., et al. (2024). KaRIn, the Ka-band radar interferometer of the SWOT mission: design and in-flight performance. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1-27.

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in water resources, 31*(11), 1411-1418.

Rodríguez, E., Durand, M., & Frasson, R. P. d. M. (2020). Observing rivers with varying spatial scales. *Water resources research, 56*(9), e2019WR026476.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association, 88*(424), 1273-1283.

Swot. (2025). SWOT Level 4 Sword of Science River Discharge Products, Version 3. Data set\. https://doi.org/10.5067/SWOT-SOS-RD3\

Tebaldi, N., Gleason, C. J., Durand, M. T., Coss, S., & Simmons, T. (2026). Lessons learned from building Confluence: Cloud-based global river discharge estimation software for the SWOT satellite. *Environmental Modelling & Software*, 106916.

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A*, 331-352.

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.

Turnipseed, D. P., & Sauer, V. B. (2010). *Discharge measurements at gaging stations* (2328-7055). Retrieved from

Verma, K., Munier, S., Boone, A., & Le Moigne, P. (2026). CTRIP‐HyDAS: A global‐scale data assimilation framework for SWOT‐derived discharge using synthetic observations at high resolution (1/12). *Water resources research, 62*(2), e2025WR040888.

Yin, K., Goldblatt, R. H., Harrison, P. J., John, M. A. S., Clifford, P. J., & Beamish, R. J. (1997). Importance of wind and river discharge in influencing nutrient dynamics and phytoplankton production in summer in the central Strait of Georgia. *Marine Ecology Progress Series, 161*, 173-183.

Yoon, Y., Garambois, P. A., Paiva, R. C., Durand, M., Roux, H., & Beighley, E. (2016). Improved error estimates of a discharge algorithm for remotely sensed river measurements: Test cases on S acramento and G aronne R ivers. *Water resources research, 52*(1), 278-294.

Zou, S., Manevski, K., Tian, J., & Tian, J. (2026). New Insights in Prediction of Daily River Flow Using SWOT Observations and Machine Learning. *Hydrology and Water Resources, 1*(2), 10.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: