UNIVERSITY OF CALIFORNIA

Los Angeles

Enabling remote-sensing based streamflow estimation at the continental scale

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Civil Engineering

by

Jacob Rubin Schaperow

2023

ABSTRACT OF THE DISSERTATION

Enabling remote-sensing based streamflow
estimation at the continental scale

by

Jacob Rubin Schaperow
Doctor of Philosophy in Civil Engineering
University of California, Los Angeles, 2023
Professor Steven A. Margulis, Chair

Forecasting streamflow on a continental scale is a challenge, especially in ungauged basins where in-situ data is rare. We can combine numerical modeling with information from remotely-sensed data to make best-estimates of discharge. The body of this dissertation is split into three chapters. Chapter 2 presents a method for characterizing river channels, a key input for hydraulic models, using information purely from remote sensing data. Chapter 3 describes a high resolution, near-global dataset of inputs to the Variable Infiltration Capacity model, a land surface model that has been widely used for hydrologic forecasting, trend analysis, and coupling with climate models. This dataset, "VICGlobal," represents a major improvement on past land surface parameterizations for VIC, with higher resolution than existing VIC inputs, and greatly simplifies the process for running VIC anywhere in the world. Lastly, Chapter 4 focuses on a procedure called "Inverse Streamflow Routing (ISR)," a method for runoff estimation and discharge interpolation, originally proposed by researchers at Princeton in 2013. We build upon their method and use an update borrowed from the Ensemble Kalman smoother to reduce uncertainty in the a priori runoff estimate.

The dissertation of Jacob Rubin Schaperow is approved.

William W-G Yeh

Timu Gallien

Dennis P. Lettenmaier

Steven A. Margulis, Committee Chair

University of California, Los Angeles

2023

To my parents . . .

to whose support and encouragement

I am eternally grateful

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my PhD advisor, Steve Margulis, for his guidance and insight and for helping me avoid rabbit-holes.

I would like to thank Dennis Lettenmaier for his support and eye for promising research directions.

I would like to thank my colleagues, fellow PhD students, and teachers for helping me grow and become the critical thinker I am today.

VITA

2016	B.S. (Civil and Environmental Engineering), University of Maryland, College Park, MD
2016	R&D Intern, Earth Networks. Carried out demand response optimization research for efficient energy use.
2017	Teaching Assistant, Civil and Environmental Engineering Department, UCLA. Taught C&EE 150 Introduction to Hydrology
2017	M.S. (Civil and Environmental Engineering), UCLA, Los Angeles, CA
2020-2023	Teaching Assistant, Civil and Environmental Engineering Department, UCLA. Taught C&EE 103 Applied Numerical Computing and Modeling

PUBLICATIONS

Cooper, M.G., Schaperow, J.R., Cooley, S.W., Alam, S., Smith, L.C., Lettenmaier, D.P. (2018). Climate Elasticity of Low Flows in the Maritime Western U.S. Mountains. *Water Resources Research*. `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018WR022816?af=R`

Schaperow, J.R., Gabriel, S.A., Siemann, M., Crawford, J. (2019). A Simulation-Based Model for Optimal Demand Response Load Shifting: A Case Study for the Texas Power Market. *Journal of Energy Markets*. `https://papers.ssrn.com/abstract=3507996`

Schaperow, J.R., Margulis, S.A., Li, D., Lettenmaier, D.P. (2019). A Curve-Fitting Method for Estimating Bathymetry From Water Surface Height and Width. *Water Resources Research*. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019WR024938

Schaperow, J.R., Margulis, S.A., Li, D., Lettenmaier, D.P. (2021). A near-global, high resolution land surface parameter dataset for the variable infiltration capacity model. *Nature Scientific Data*. https://www.nature.com/articles/s41597-021-00999-4

CHAPTER 1

Introduction

Streamflow is a key hydrologic variable. Accurate measurements of river flow are needed to make decisions about water resource management, such as how much water to allocate to different regional stakeholders and how to best manage reservoirs to optimize hydropower production, store water for dry spells, and provide storage capacity in case of floods. While stream gauge networks are relatively dense in the U.S., there are very few stream gauges worldwide, especially in developing countries in flood- and drought-prone areas such as Africa and south Asia.

Hydrological modeling is one way to estimate streamflow in ungauged basins. In a hydrological modeling approach, precipitation estimates and other meteorological variables are input into a computer model that uses conservation of mass, momentum, and energy to predict how precipitation is partitioned into runoff, infiltration, and evapotranspiration. The runoff, or the portion of water that runs off over the land surface, eventually winds up in rivers and streams and becomes streamflow. The runoff output from the hydrologic model can be input into a river routing model or a hydrodynamic model to estimate the discharge downstream. (Note that we use "streamflow" and "discharge" interchangeably throughout this dissertation.)

Satellite-based streamflow estimates are another way to fill in the gaps between stream gauges. The recently-launched Surface Water and Ocean Topography (SWOT) satellite will observe rivers worldwide and provide measurements of river width, surface slope, and water surface elevation. Both hydrologic modeling and remote-sensing based discharge retrieval

algorithms have associated challenges. Hydrologic models are prone to error caused by uncertain meteorological inputs. Satellite-based estimates are susceptible to errors due to retrieval algorithm assumptions. Both methods require knowledge of input parameters. Hydrological modeling requires knowledge of soil and vegetation parameters which control how much and how soon precipitation becomes runoff. Satellite-based estimates, while observation-based discharge estimates require knowledge of Manning's roughness coefficient and channel bathymetry. In this dissertion, we address these challenges related to streamflow estimation at the continental scale.

1.1 Objective of the Dissertation

The objective of this dissertation is to address three challenges associated with continental-scale hydrological modeling with the goal of estimating discharge in ungauged basins. First, we showcase a method for estimating river bathymetry from satellite observations of river width and water surface elevation. Bathymetry is an important parameter for hydrodynamic models such as HEC-RAS ([HEC-RAS, 2008]) and LISFLOOD-FP ([Bates and De Roo, 2000]). Bathymetry estimation via satellite could be a useful alternative to large-scale bathymetric surveys. Second, we present a global parameterization of the Variable Infiltration Capacity (VIC) hydrological model at higher resolution (1/16°) than any previous global soil and vegetation parameter dataset. The only other global VIC dataset is at 2°resolution ([Nijssen et al., 2001]). Hydrological models, such as VIC, solve the water and energy balance at the land surface to partition precipitation inputs into infiltration, evapotranspiration, and runoff, and are sensitive to soil and vegetation parameters ([Demaria et al., 2007]). Third, we develop a method for estimating runoff from streamflow measurements in order to estimate runoff at the grid cell scale and interpolate discharge between gauge locations. Our method builds on contributions by [Pan and Wood, 2013] and [Yang et al., 2020] while addressing key limitations of their methods in order to make improved estimates of runoff

in the face of error-prone discharge measurements.

1.2 Organization of the Dissertation

The following chapters are organized as follows. Chapter 2 describes a method for estimating river bathymetry from satellite measurements of river water surface elevation. In Chapter 3, we present the VICGlobal dataset, a 1/16°resolution soil and vegetation parameter dataset for the Variable Inlfiltration Capacity (VIC) hydrologic model. Chapter 4 is dedicated to the Inverse Streamflow Routing technique. In this chapter, we build on a previously-developed method for estimating runoff from streamflow measurements, addressing drawbacks in the original method, and focusing on producing runoff and discharge estimates with appropriate uncertainty bounds. Chapter 5 contains our concluding remarks and highlights directions for future research.

CHAPTER 2

A Curve-Fitting Method for Estimating Bathymetry From Water Surface Height and Width

River discharge estimation requires knowledge of bathymetry. However, aside from a few locations where surveys have been conducted, bathymetric data are unavailable, even for major rivers. It has been suggested that water surface elevation and flow width measurements from the upcoming Surface Water and Ocean Topography (SWOT) satellite mission (planned launch 2021) may be used to infer the submerged channel geometry; however, the full potential of these measurements for inferring bathymetry has yet to be explored. We apply four different techniques, with varying assumptions about height-width relationships, to predict unknown bathymetry. We call these "curve-fitting methods" the linear, slope break, nonlinear, and nonlinear slope break (NLSB) methods. The linear and slope break methods are based on a linear height-width relationship, while the nonlinear and NLSB methods are based on a height-width relationship derived from hydraulic geometry equations. We generate SWOT-like observations of height and width based on 5 m gridded Upper Mississippi River data, and evaluate the performance of each curve-fitting method given the SWOT-like observations. The NLSB method predicts bed elevation and low flow area with the least error, although the nonlinear method may be preferred in low data conditions. Additionally, we show that our method outperforms previously suggested methods, and we propose an NLSB-based bathymetry prior for Bayesian discharge estimation algorithms.

2.1 Introduction and Background

Estimating river discharge worldwide is a difficult problem because remote sensing platforms cannot directly observe discharge ([Lettenmaier et al., 2015]). Following its 2021 launch, the joint NASA/CNES Surface Water and Ocean Topography (SWOT) satellite mission will use a Ka-band radar interferometer to measure water surface height and water-covered area for rivers wider than about 100 m, with a repeat cycle of 21 days ([Biancamaria et al., 2016]). Although SWOT cannot measure river discharge directly, researchers have developed a variety of indirect methods to estimate streamflow from SWOT measurements. These methods fall into two categories: retrieval methods, which compute discharge at the time and location of the SWOT overpass; and data assimilation methods, which use measured data to constrain hydrodynamic model predictions of discharge at both measurement and non-measurement times. Recent papers about SWOT discharge retrieval algorithms include [Bjerklie et al., 2018], [Hagemann et al., 2017], [Durand et al., 2016], [Durand et al., 2014], [Bonnema et al., 2016], and [Garambois and Monnier, 2015]. Recent progress in estimating discharge through data assimilation includes [Brisset et al., 2018], [Oubanas et al., 2018b], [Oubanas et al., 2018c], and [Yoon et al., 2016]. With the exception of the at-many-stations hydraulic geometry method (AMHG; [Gleason and Smith, 2014]), which estimates discharge based solely on river width fluctuations, physically-based discharge algorithms all use Manning's equation or more general forms of conservation laws, and thus require information about the river channel bathymetry and roughness — essential parts of the governing equations. In the SWOT literature, these are known as "mass-conserved flow law inversion" (McFLI) algorithms ([Gleason et al., 2018]). Bathymetry and roughness factor into discharge algorithms through the shallow water equations (see e.g. [Chow et al., 1988]), which are the depth-averaged mass and momentum conservation equations. These equations govern the way streamflow area, depth, and discharge change in time and space due to upstream and lateral inflow boundary conditions.

Bathymetry and roughness are key parameters for physically-based river hydrodynamic modeling ([Neal et al., 2015]; [Aronica et al., 1998]), but they are impractical to measure in situ on a large scale. Acoustic Doppler current profilers are used to collect accurate, spatially dense measurements of river bathymetry, but they must be deployed from boats. Remote sensing techniques are limited by the ability of the sensor to penetrate water, especially turbid water ([Marcus and Fonstad, 2008];[Legleiter et al., 2009]; [Pan et al., 2015]). Compounding the problem, bathymetry gradually changes over time due to erosion, more quickly during major storm events ([Buffington, 2012]), so even if a global bathymetry survey were conducted, it would need to be redone periodically to account for changes in the river channel morphology. Roughness has been related to the river bed grain size ([Chow, 1959]), but it also depends on flow depth and channel geometry ([Moramarco and Singh, 2010]).

Parameter estimation methods have been used to estimate bathymetric and roughness parameters along with discharge; however, bathymetric and roughness parameters compensate for one another to produce the same discharge estimate — the "equifinality problem" ([Garambois and Monnier, 2015]). Prior information can be used to constrain the estimates. Several SWOT-based discharge algorithms are designed to incorporate prior information about these parameters. The Manning-Metropolis algorithm (MetroMan; [Durand et al., 2014]), the Manning component of the Bayesian AMHG-Manning algorithm (BAM; [Hagemann et al., 2017]), and variational data assimiliation algorithms ([Oubanas et al., 2018b]; [Oubanas et al., 2018c]) all postulate prior distributions for bathymetry and roughness. Below, we review how SWOT discharge algorithms handle unknown bathymetry.

MetroMan ([Durand et al., 2014]) is a McFLI algorithm that estimates bathymetric and roughness parameters using a Metropolis algorithm. It represents bathymetry using the unobserved cross-sectional area of flow below the minimum observed water surface elevation (A0). In the updated MetroMan method in [Durand et al., 2016], A_0 priors are postulated by solving Manning's equation for A0 given observations of width, water surface slope, cross sectional flow area change, and mean annual flow (QMAF) data from a global water balance

model. The accuracy of this method is limited by the QMAF estimate.

Bayesian-AMHG-Manning (BAM; [Hagemann et al., 2017]) is a more recent McFLI method that uses priors based on flow-equation parameters measured in-situ at 10,081 USGS gauge stations (HydroSWOT; [Canova et al., 2016]). They fit a multiple regression model to measurements of width vs $ln(A_0)$, where A_0 was taken as the smallest observed flow area. Because the residuals of $ln(A_0)$ were approximately normally-distributed, [Hagemann et al., 2017] used a lognormal prior for A_0, with mean and variance given by the regression model. This approach is limited by the USGS database's U.S.-only coverage, and it does not take height data into account, even though SWOT will measure water surface height.

[Oubanas et al., 2018b] and [Oubanas et al., 2018c] used variational data assimilation to estimate discharge, bathymetry, and roughness simultaneously using SWOT measurables. They found that under uncertainty in either bathymetry or roughness, but not both, the assimilation updates the parameter value correctly, but when both bathymetry and roughness are unknown, it does not. This is consistent with equifinality, or the idea that there are multiple parameterizations that result in the same model outputs ([Beven, 2006]).

[Bjerklie et al., 2018] estimated discharge using two different flow law equations parameterized using optical, laser, and radar remote sensing measurements. They assumed a parabolic cross section based on analysis of the HydroSWOT in-situ database to infer bed elevation from height-width measurements. They note that while the average cross section in the database was a parabola, it is not necessarily the correct shape for a particular cross section, and this contributes to error in discharge estimates — although their discharge estimate were still accurate to within 20% of measured values.

Most recently, [Brisset et al., 2018] inferred roughness and discharge simultaneously in a synthetic experiment and for a section of the Garonne River in France using variational data assimilation. They used the "low Froude" method to infer bathymetry given limited in-situ depth measurements ([Garambois and Monnier, 2015]; [Gessese et al., 2013]; [Gessese, 2011]). The method allows inversion of a simplified version of the shallow water

equations to solve to water depth along the river given knowledge of river depth, width, and surface slope at one location. However, the low-Froude method assumes steady flow and a small Froude number (typically less than 0.3), and these conditions may not hold true through the length of the river, thus the method still has considerable dependence on in-situ measurements.

Overall, accurate simultaneous inference of bathymetry, roughness, and discharge from satellite measurements remains a difficult problem. Yet, the current body of literature (e.g. [Oubanas et al., 2018c]; [Garambois and Monnier, 2015]) suggests that better knowledge of bathymetry could make inference of roughness and discharge significantly easier. While most existing work has focused on improving discharge estimates, only a few studies have focused on estimating bathymetry using SWOT-like observations ([Mersel et al., 2013]; [Yoon et al., 2012]; [Durand et al., 2010]). [Mersel et al., 2013] first introduced a curve-fitting approach for predicting unobserved bathymetry, using bathymetric survey data on the Upper Mississippi, Illinois, Rio Grande, and Ganges-Brahmaputra rivers as a substitute for SWOT measurements. This approach, from which we draw heavily in this paper, uses height-width observations collected over a range of flow regimes to fit a model and extrapolate to predict unobserved bathymetry. [Mersel et al., 2013] assumed measurements were available at different river stages, ranging from bankfull to a low flow stage occurring at some maximum channel exposure level, which they defined as the minimum observed fraction of flow area to bankfull flow area, $e = 1 - \frac{A}{A_{bf}}$. They tested a linear height-width model and a piecewise linear height-width model with one breakpoint, which they called the linear and slope break methods. Throughout this paper, we will refer to the methods of [Mersel et al., 2013] as M13. The slope break method predicted minimum bed elevation accurately even at relatively low channel exposure. Notwithstanding its success, the slope break method is lacking in that it assumes a piecewise linear height-width relationship — slope break predictions will be biased if this assumption is violated. M13 addressed this by restricting the application of the slope break method to a few "optimal" locations where

the slope break model is a good fit. They identified optimal locations using an arbitrary threshold determined by trial and error. The fraction of "optimal" cross sections out of the total number of cross sections depended on channel exposure: typically only 0-16% of cross sections were defined as optimal. While restricting the slope break method to optimal locations ensures accuracy, it limits the number of locations where the method may be applied. Furthermore, it may be possible to make more accurate predictions with less channel exposure using nonlinear models vs. linear models.

Here, we propose a new method to estimate bathymetry solely using SWOT-like water surface height and effective width measurements. An extension to M13's methods, we call our method a "curve-fitting approach" for predicting unobserved bathymetry. SWOT will measure water surface elevation h and effective river width w, i.e. water covered area per unit length of river. SWOT's 21-day repeat cycle means that it will observe every river between 78° N and 78° S approximately 100 times over its 3-year design life, more frequently for rivers at higher latitudes (Biancamaria et al., 2016). By fitting curves to height and width measurements at various river stages, we can predict the bathymetry below the lowest observed water surface elevation for rivers worldwide.

Compared with previous river bathymetry estimation methods, our approaches are novel in three respects:

1. Our curve-fitting methods are more general than those of M13; they can describe more types of height-width relationships. Breakpoints are selected using breakpoint analysis techniques, so there is no need for thresholds or trial and error.

2. We account for measurement uncertainty using Monte Carlo simulation.

3. Since our bathymetry estimates are not intended for use in simultaneous discharge estimation, they are unaffected by equifinality because they are made independently of roughness and discharge.

Parameterizing a height-width curve is much simpler than simultaneously estimating bathymetry, roughness, and discharge using a hydraulic model or direct application of hydrodynamic equations. The final product can be taken as a stand-alone bathymetry dataset, or it can be used as a prior for Bayesian methods such as MetroMan, BAM, or variational data assimilation. We test our method over 30 river reaches in the Upper Mississippi River system, and we postulate prior distributions for bathymetry parameters that could be used in SWOT discharge algorithms.

2.2 Methods, Data, and Study area

2.2.1 Theoretical justification

We investigate two unknown bathymetry parameters: the minimum bed elevation z0 and the low-flow wetted area A0, the area below the minimum observed water surface height. We also estimate the bed slope (s_0) by taking finite differences of z_0. We chose these parameters because hydrodynamic models incorporate bathymetry through flow area, wetted perimeter, and bed slope. SWOT will observe water surface height (h) and effective flow width (w) each time it passes over a sufficiently wide river — 100 m or more (Rodriguez et al., 2016). For such wide rivers, wetted perimeter is approximately equal to flow width (Dingman, 2009), so we focus on flow area and bed slope here. Both can be calculated given the full height-width relationship, from $h(w = 0) = z_0$ to $h(w = w_{bf}) = h_{bf}$, where w_{bf} and h_{bf} are bankfull width and height. Our method consists of two major steps. First, we fit a model $h(w)$ to observed height and width data, then we extrapolate the fitted $h(w)$ to predict $\hat{z_0} = h(w = 0)$, as well as the predicted submerged area $\hat{A_0}$.

Flow area is related to the integral of the height-width curve. Knowing the $h(w)$ relationship, we can write cross sectional area A for a given flow width w^*,

$$A(w^*) = \int_0^{w^*} [h^* - h(w)]dw \tag{2.1}$$

where $h^* = h(w^*)$. Bed slope s_0 can be estimated using finite differences of the bed elevations (Equation 2.2), where r is the coordinate along the river channel:

$$s_0 = \frac{z_0(r+1) - z_0(r)}{\Delta r} \tag{2.2}$$

Although a higher-order difference could be used, we chose a forward first difference because there were only a few adjacent river reaches in our study domain, and higher-order differences would have precluded us from inferring slope at a number of river reaches.

The question remains as to what model to use for h(w). There are a variety of models we might choose from: linear and nonlinear, piecewise or not. The hydraulic geometry equations ([Leopold and Maddock, 1953]) use power law relationships to relate width w, depth d, and mean velocity v to discharge:

$$w = aQ^b \tag{2.3}$$

$$d = cQ^f \tag{2.4}$$

$$v = kQ^m \tag{2.5}$$

Combining equations (2.3) and (2.4) yields:

$$d = \frac{c}{a^{f/b}} w^{f/b} \tag{2.6}$$

Substituting $d = h - z_0$ into Equation (2.6) yields:

$$h = z_0 + \frac{c}{a^{f/b}} w^{f/b} \tag{2.7}$$

Redefining the coefficient and exponent parameters yields:

$$h = z_0 + uw^s \tag{2.8}$$

The form of Equation (2.8) is equivalent to the cross section model derived by [Lawrence, Dingman, 2007] and which was used recently in a SWOT discharge context by [Bjerklie et al., 2018]. The exponent s is the shape parameter. When $s = 1$, the cross section is triangular; when $s = 2$, the cross section is parabolic; and when $s \to \infty$, the cross section is rectangular. The model may also be piecewise. The M13 slope break method assumes there is one height-width relationship at low flows and another at moderate-to-high flows. The slope break concept is supported in the fluvial geomorphology literature ([Lewis, 1966]; [Fahnestock, 1963]), so we consider the possibility of a piecewise $h(w)$ relationship with one breakpoint.

We use Equation (2.8) as the basis for our nonlinear (NL) height-width model, where we change the variable names here to be consistent with the methods that include breakpoints, and rewrite as:

$$h = z_0 + u_1 w^{s_1} \tag{2.9}$$

Incorporating a slope break, or breakpoint, in the nonlinear height-width model leads to the nonlinear slope break (NLSB) model:

$$h = \begin{cases} z_0 + u_1 w^{s_1} & \text{for } w \leq w_b \\ z_0 + u_1 w_b^{s_1} + u_2 \left(w^{s_2} - w_b^{s_2} \right) & \text{for } w \geq w_b \end{cases} \tag{2.10}$$

where w_b is the width at the breakpoint. We test how well each model predicts z_0 and A_0 from height-width observations. We also test the simpler linear (L) and slope break (SB)

models given respectively by:

$$h = z_0 + u_1 w \tag{2.11}$$

$$h = \begin{cases} z_0 + u_1 w & \text{for } w \leq w_b \\ z_0 + u_1 w_b + u_2 (w - w_b) & \text{for } w \geq w_b \end{cases} \tag{2.12}$$

The NL and NLSB models are based on hydraulic geometry equations, while the L and SB models are based on a linear height-width relationship. The SB and NLSB methods each allow one breakpoint in the height-width relationship, while the L and NL methods do not. Each model implies a theoretical channel shape, illustrated in Figure 2.1. For simplicity, we have assumed all cross sections are symmetric. Note that it is straightforward to implement the L and NL cross sections in the hydraulic model LISFLOOD-FP ([Neal et al., 2015]). This has been done, for example, by [Grimaldi et al., 2018], who parameterized LISFLOOD-FP with exponential channel shape models based on in-situ bathymetry measurements in the Clarence River, Australia. While our bathymetry predictions are intended to be used in conjunction with models (including but not limited to LISFLOOD-FP) we do not use any hydrodynamic modeling to derive our bathymetry predictions — they are derived from observed height-width relationships only.

2.2.2 Data and study area

SWOT measurements are not yet available, so we developed our method using SWOT-like measurements derived from a high-resolution bathymetry dataset corrupted with measurement error. The core was a 5 m resolution gridded bathymetry dataset maintained by the Upper Midwest Environmental Sciences Center (UMESC). The dataset covers 406 km of the Upper Mississippi River and 124 km of the Illinois River near its confluence with the Mississippi. Figure 2.2 shows the extent of the UMESC bathymetry data.

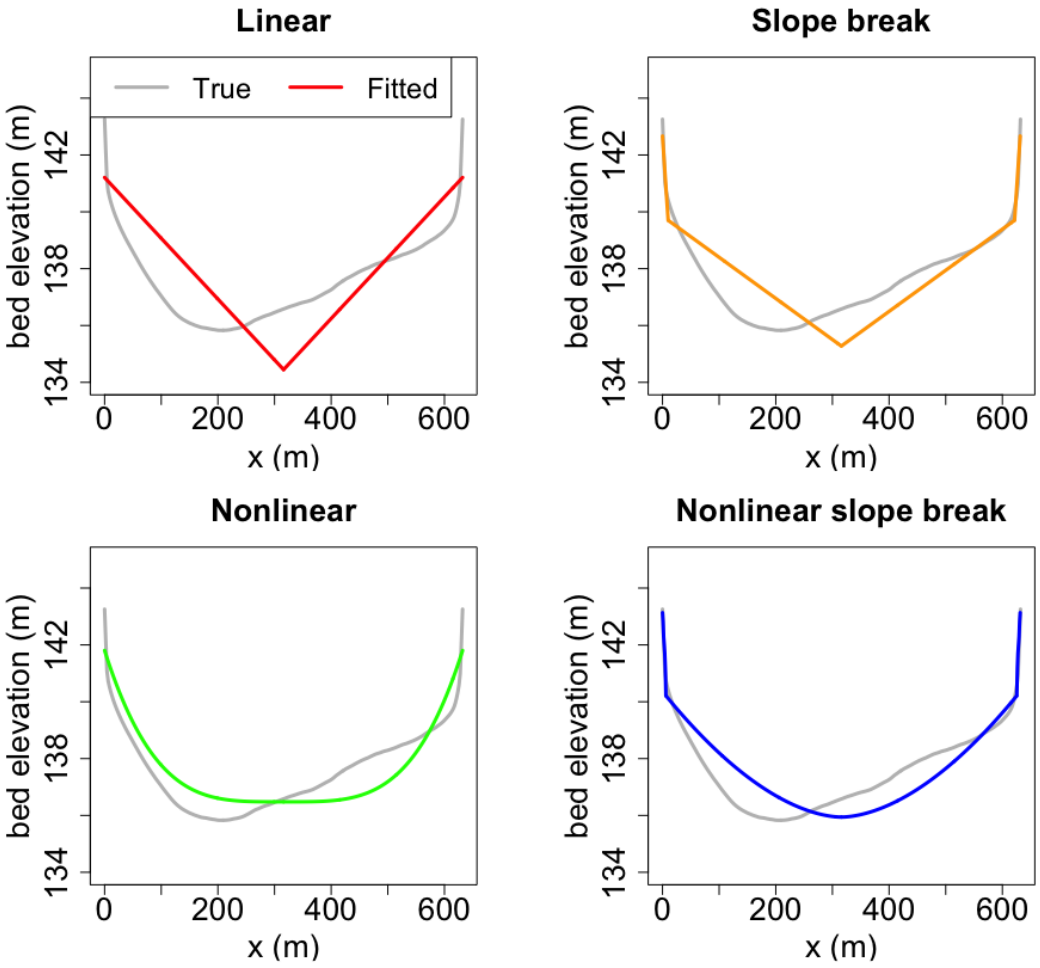

Figure 2.1: Theoretical cross section models for linear (L), slope break (SB), nonlinear (NL), and nonlinear slope break (NLSB) methods superimposed over the true reach geometry for reach 1 of Pool 21. The models were fit given full channel exposure and assumed no measurement error. The gray line shows the true, irregular reach geometry.

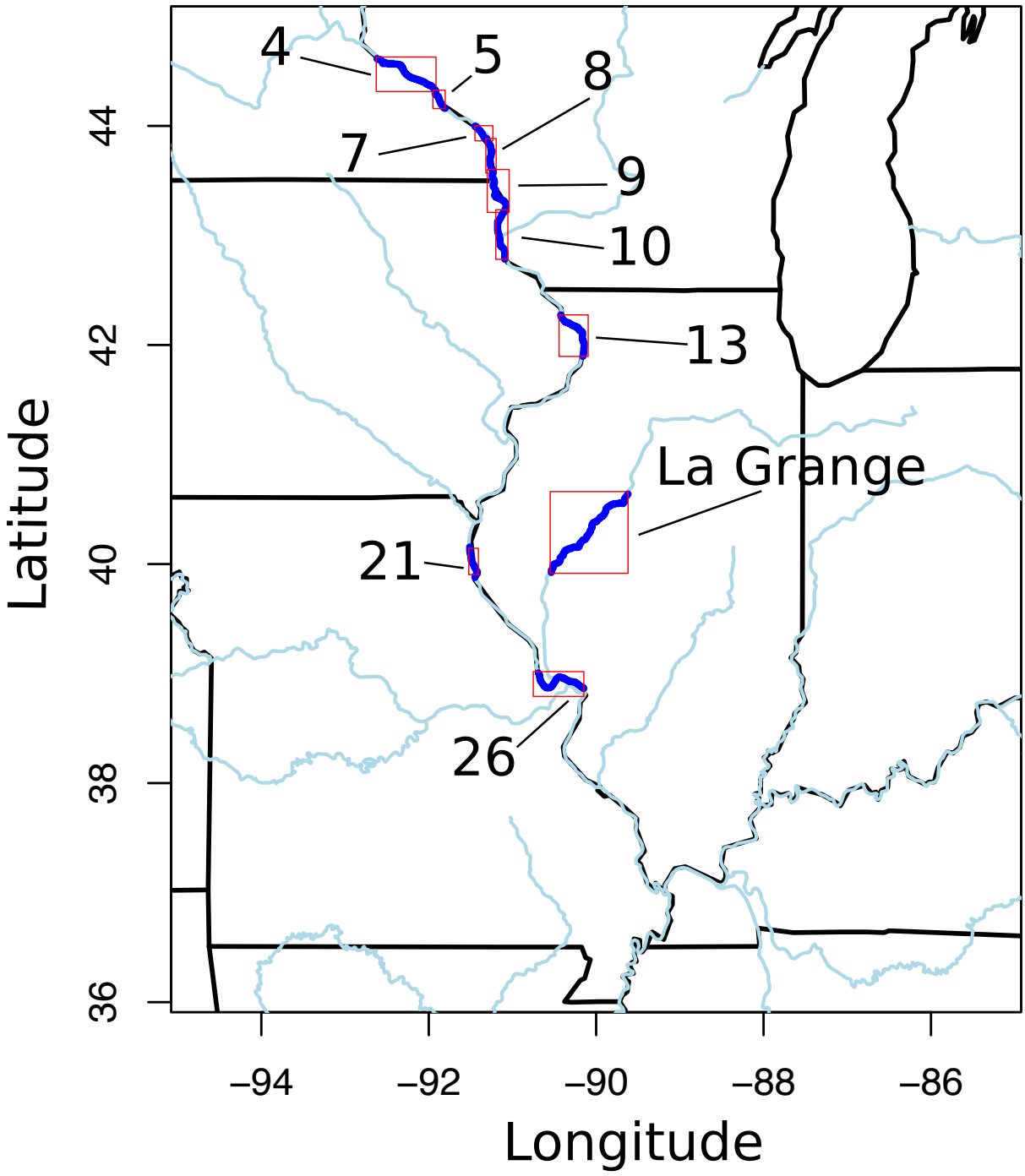

Figure 2.2: Study area showing locations of UMESC bathymetry data in the Upper Mississippi River system. Black lines show state boundaries, rivers are depicted in light blue. Pools with bathymetry data are boxed and colored dark blue.

We chose the Upper Mississippi River as the study area because it is one of a few rivers whose bathymetry is well-measured and for which the measurements are freely available. Locks and dams divide the Mississippi River into pools. UMESC provides access to bathymetry data for pools 4, 5, 7-10, 13, and 26 of the Upper Mississippi River, as well as the La Grange pool of its tributary, the Illinois River. The pools range from 19 to 124 km long. The bathymetry of each pool is summarized in the Supporting Information (SI; Table S6).

The UMESC bathymetry data (http://www.umesc.usgs.gov/aquatic/bathymetry/download.html) were collected in various years from 1992 to 2001 using a depth sounder calibrated with a sounding pole ([Rogala, 1999]). The raw sounding data were interpolated over a 5 m resolution raster grid for the published version. We used river centerlines from HydroSHEDS ([Lehner et al., 2008]). Since SWOT measurements will be averaged over river reaches to smooth out random errors ([Biancamaria et al., 2010]), for simplicity, we chose a reach length of 10 km for this study ([Frasson et al., 2017]). This resulted in 51, 10-km long study reaches. Shape parameters for these reaches varied from 0.6 to 4.5, with an average value of 1.9.

2.2.3 Generating SWOT-like observations

We derived SWOT-like observations using the UMESC data as "ground-truth." By sampling these bathymetry data under different scenarios and accounting for measurement errors, we produce SWOT-like height and width measurements for each river reach, to which we fit each of the four models described in Section 2.1. We pre-processed the gridded bathymetry data to obtain reach-average cross section geometry. Details are in the SI (Section S2), but the main features are as follows: bathymetry data were extracted along transects perpendicular to the river channel and averaged over 10-km reaches, over the non-braided portions of the river. We selected the 30 least braided reaches for our study. In order to define bankfull quantities, we imposed zero depth at the left and right river banks to make them even. While treating uneven banks in this way is a limitation of the methodology since it is a departure

from the true conditions existing in nature, our focus is on in-channel flow, and we found conducted sensitivity tests that showed that this modification did not strongly affect the performance of the curve-fitting methods (see SI, Section S3). Reach average geometry for each of the 30 study reaches is included in the SI (Section S6).

In order to study a range of possible observation scenarios, we assumed measurements were available from near bankfull elevation down to some low-flow elevation, in 5% channel exposure increments. Channel exposure is 0% flow is at bankfull stage, and it is 100% when the river channel is completely exposed. We defined channel exposure in terms of bankfull flow width as $e = 1 - w/w_{bf}$. Note that we chose to differ from M13's area-based definition because bankfull width is easier to measure than flow area. Figure 2.3 shows a cross section with horizontal lines drawn at bankfull height h_{bf} and two additional measured heights h_2 and h_1. The corresponding widths are w_{bf}, w_2, and w_1. Bankfull width wbf is not necessarily measured by SWOT, but it is known in the context of our study from the UMESC data. SWOT will make N measurements (w_n, h_n) over its lifetime. For a given river reach, there will be some low-flow observation (w_1, h_1). We say that this location has been observed down to the exposure level $e_1 = 1 - w_1/w_{bf}$.

We sampled the true height-width relationships at each river reach increments and added independent, identically Gaussian errors with standard deviations of 10 cm for height and 10 m for width to account for measurement error. While SWOT observations can theoretically be available at any exposure level, we used 5% increments to explore how accurately we can predict bathymetry given a particular maximum observed channel exposure level. We obtained 500 sets of SWOT-like observations for each river reach in order to represent measurement errors.

There is substantial reason to expect that SWOT height measurement error will be about 10 cm. This is the value given in the SWOT Science Requirements ([Rodriguez et al., 2016]), and it is close to SWOT Instrument Simulator height error estimates for the Sacramento and Platte rivers ([Frasson et al., 2017]). Width error is less well-quantified, but because

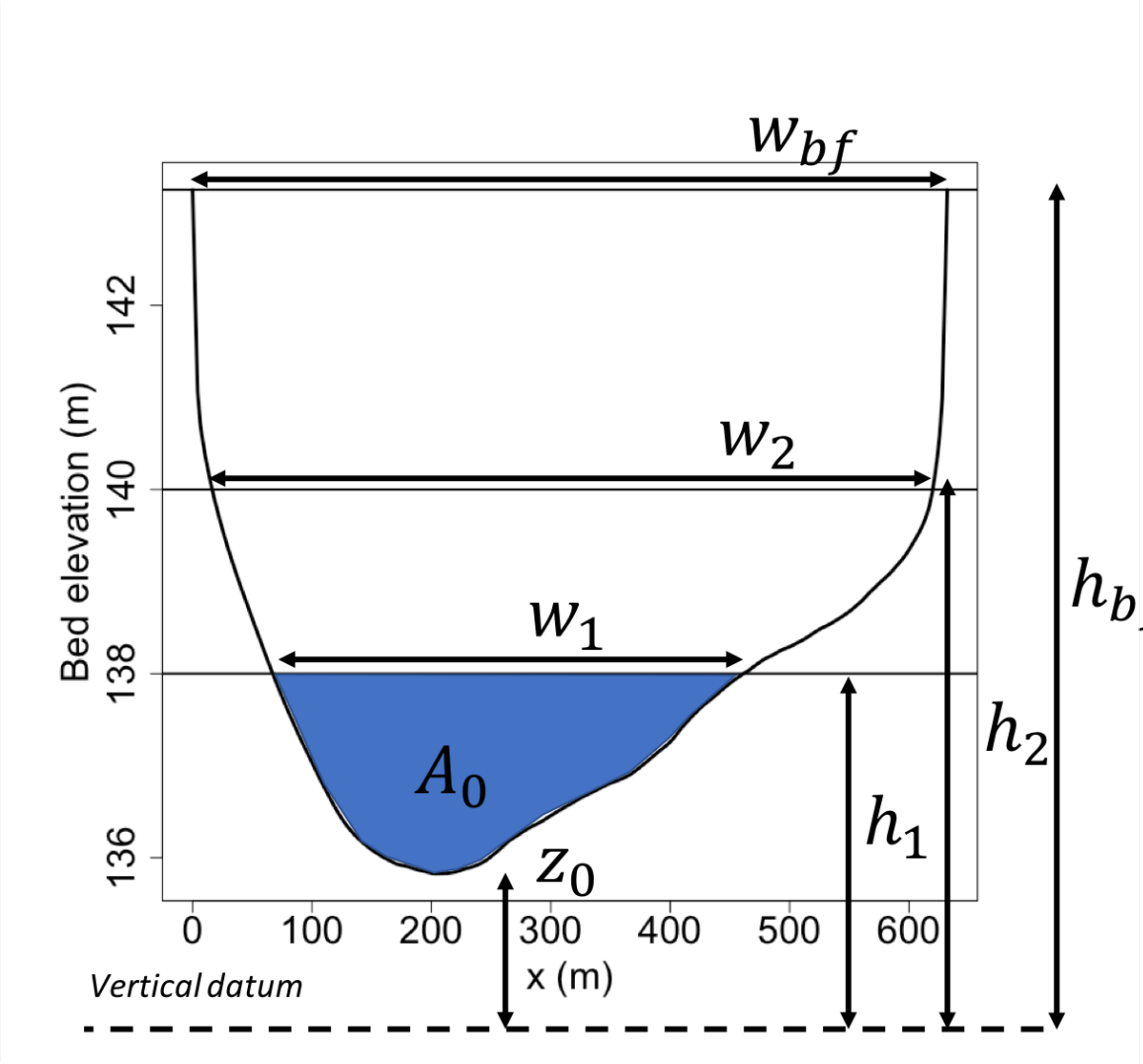

Figure 2.3: Cross section geometry for reach 1 of Pool 21, defining bankfull conditions and illustrating SWOT measurements of height and width. See Section 2.3 of the text.

SWOT widths depend on water-covered area classification accuracy, it will probably be at a similar magnitude to the pixel size, which ranges from 10 m in the far range to 60 m in the near range ([Biancamaria et al., 2016]). We used a Monte Carlo approach to propagate error, fitting the linear, SB, NL, and NLSB models to each of the 500 sets of observations at each river reach, and reporting statistics of the ensemble of estimated model parameters and predicted bathymetry parameters $\hat{z}_0$ and $\hat{A}_0$.

2.2.4 Curve Fitting

Given SWOT-like height and width measurements corrupted with Gaussian error and sampled at exposure levels from 5-95% as described in Section 2.3, we estimated parameters for each type of cross section model (L, SB, NL, and NLSB). We fit the models using measurements above some assumed maximum exposure level e_1, which we allowed to vary from 5-95%. Height-width data falling below e_1, where $w < w_1$, were excluded.

For the linear model, we estimated parameters using ordinary least squares regression. For the SB model, we constrained the piecewise linear fit to be continuous at the breakpoint, which we selected using a dynamic programming algorithm ([Zeileis et al., 2003]) that identifies the breakpoint such that the piecewise model has the minimum least squared error (LSE). We required at least 5 data points to perform a fit (10 for the piecewise SB and NLSB methods).

The hydraulic geometry power model in Equation (2.6) can be linearized by taking logs, but when using the form of Equation (2.8), with unknown z_0, nonlinear methods must be used, so we estimated parameters for the NL and NLSB methods using the Levenberg-Marquardt algorithm, which requires initial guesses for the model parameters. We used $s = 2$ as an initial guess for the shape parameter based on the finding in [Bjerklie et al., 2018] that cross section shapes across the U.S. are on average parabolic. We used the minimum observed water surface height as our initial guess for z_0. Finally, we used linear regression with the model $h = min(h_{obs}) + uw^2$ to get an initial guess for the coefficient u. Occasionally, this

initial guessing strategy led to a singular gradient at the initial guess, and the Levenberg-Marquadt algorithm failed to parameterize the height-width model. In those cases, we attempted the fit multiple times, with six sets of initial guesses starting with assumptions of $s = [1, 2, 3, 4, 5, 6]$, and we selected the fit with the lowest LSE; this was almost always sufficient to avoid singular gradients. For the NLSB model, we chose the breakpoint that gave the lowest LSE out of all possible breakpoints, using a brute force method to find the optimal breakpoint. We required a minimum of 4 points on either side of the breakpoint for the nonlinear fits.

Lastly, we tried to ensure physical realism in the fitted height-width relationships. Sometimes, particularly when the maximum channel exposure was low, the width error was large compared to the overall width fluctuations, and this led to negatively sloped height-width relationships, $\frac{dh}{dw} < 0$. As this cannot occur in natural river channels, we removed these fits from consideration. Also, natural river channels cannot be concave down at their lowest point. Otherwise, they would diverge to infinite depths. We also found that extrapolating concave down models could lead to wildly incorrect predictions. Therefore, whenever the estimated $\frac{d^2h}{dw^2} < 0$ for the NL or NLSB model, we used the L or SB model, instead.

Given fitted height-width models, it is possible to predict unobserved heights $h(w < w_1)$ and calculate the unobserved low flow wetted area (A_0) by substituting the h(w) relationship into Equation (2.1). For the linear and SB methods, $\hat{A}_0 = \frac{1}{2}\hat{u}w_1^2$. For the NL and NLSB methods, $\hat{A}_0 = \frac{\hat{u}}{\hat{s}+1}w_1^{\hat{s}+1}$. Under the above considerations for physical realism, A_0 is guaranteed to be non-negative.

2.2.5 Error metrics

We report relative depth error (ε_d), relative A_0 error (ε_{A_0}), and bed slope error (ε_{s_0}) as:

$$\varepsilon_d = (\hat{d} - d)/d * 100\% \tag{2.13}$$

$$\varepsilon_{A_0} = (\hat{A}_0 - A_0)/A_0 * 100\% \tag{2.14}$$

$$\varepsilon_{s_0} = (\hat{s_0} - s_0)/s_0 * 100\% \tag{2.15}$$

where ε_d is the percent by which the bankfull depth is overestimated, ε_{A_0} is the percent by which the low flow wetted area is overestimated, ε_{s_0} is the percent by which bed slope is overestimated, and $d = h - z_0$ is the flow depth.

2.3 Results

2.3.1 Validation of height-width model for a reach with no measurement error

We first present bathymetry prediction results under the deterministic case of no SWOT measurement errors to evaluate the applicability of each of the four types of $h(w)$ model to the Upper Mississippi reaches. This also gives us a bound for bathymetry prediction accuracy because prediction errors will be higher when there is measurement error than when there is not. We fit each model to $h - w$ observations at different exposure levels, and extrapolate to predict the unobserved heights and widths below the water surface. Figure 2.4 demonstrates how the $h - w$ models were fit to the synthetic observations and shows $h - w$ plots for Pool 21, reach 1, which has a shape parameter of about 4.5. At maximum channel exposure of $e_1 = 20\%$, the steep observed $h - w$ relationship is extrapolated downward, and the bottom depth is overestimated. The $h - w$ projections from each model become more accurate as more of the channel is exposed.

The choice of $h(w)$ model has significant implications. The NLSB method is able to predict unobserved height and width accurately even at 20% exposure, whereas the linear method continues to overpredict the depth, even at 80% exposure. Table 2.1 shows statistics of d and A_0 prediction errors for the deterministic case. Over the 30 nonbraided study

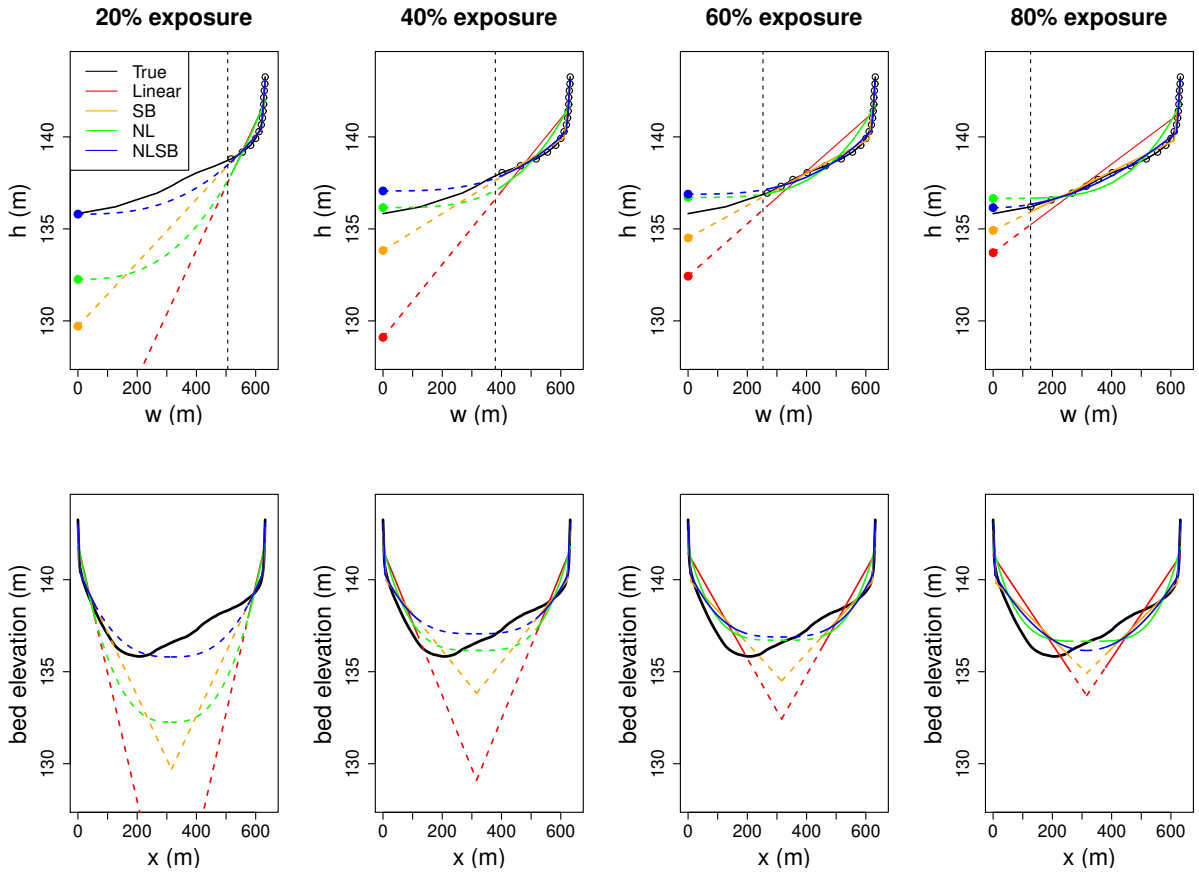

Figure 2.4: Cross section height-width models fit to water surface elevation and width data for Pool 21, reach 1, assuming perfect measurements are available at 5 percent increments of channel exposure. Dashed lines indicate extrapolation, while the solid gray line shows the true height-width relationship.

ε_d	L	SB	NL	NLSB	ε_{A_0}	L	SB	NL	NLSB
min	-33%	-39%	-49%	-32%	min	-69%	-99%	-99%	-91%
mean	40%	11%	-7.1%	-3.1%	mean	1100%	320%	95%	6.8%
max	88%	38%	15%	72%	max	2400%	840%	1300%	250%
25th percentile	27%	2.4%	-13%	-10%	25th percentile	340%	26%	-55%	-67%
median	45%	16%	-5.4%	-2.9%	median	1400%	370%	25%	-4.0%
75th percentile	61%	25%	1.1%	-0.8%	75th percentile	1800%	580%	130%	53%

Table 2.1: Relative depth error (ε_d) and A_0 prediction error (ε_{A_0}) statistics over the 30 nonbraided study reaches. Depth and A_0 predictions are for height-width data at 80% exposure and no measurement error.

reaches, at 80% channel exposure, the linear method overpredicts bankfull depth by 40% on average, while the NLSB method underpredicts by 3.1%. Low flow wetted area errors eA0 are larger, ranging from -69% to 2400% with a mean of 1100% for the linear method, and ranging from -91% to 250% with a mean of 6.8% for the NLSB method. The SB and NL methods perform better than the linear method but not as well as the NLSB method.

2.3.2 Bathymetry predictions over the study reaches: stochastic case

To assess how well bankfull depth, low flow wetted area, and bed slope are predicted when accounting for measurement errors, we applied each model at different levels of channel exposure, using Monte Carlo simulation to add random errors to synthetic width and depth observations.

As in the deterministic case, depth was overestimated at low exposure levels. The reach geometry considered in this study tended to have a steep height-width relationship at near-bankfull width. Extrapolating a height-width model fit to these data results in an excessively steep decline in height as width decreases. The linear and slope break methods were most susceptible to this extrapolation error, while the functional form of the NL and NLSB meth-

ods reduced the error. Figure 2.5 shows box and whisker plots of ε_d, as well as ε_{s_0} and ε_{A_0}, over the 30 study reaches. In this figure, we account for measurement error with an ensemble of 500 sets of error-corrupted observations per reach. There are 500 values of ε_d, ε_{s_0}, and ε_{A_0} for each reach, for each exposure level e1. Across reaches and exposure levels, the median ε_d was lowest for the NLSB method, somewhat higher for the NL method, higher for the SB method, and highest for the linear method (-7%, -8%, 27%, and 52% at 60% channel exposure, respectively). However, increasing channel exposure led to lower ε_d across all methods.

Bed slope prediction error is related to depth prediction error, with the distinction that it is possible for positive and negative depth errors to cancel. ε_{s_0} is fairly large across all models and decreases with increasing exposure levels. It is not until 80% channel exposure when the NL and NLSB methods show an advantage due to their superior depth-prediction accuracy. The NLSB method predicts s_0 with a median error of -14% of the true value at 60% channel exposure, but the 25th and 75th quartiles are -72% and 83%. Note that the box and whisker plots in Figure 2.5 show only the errors less than 200% for clarity. Summary statistics for the plots shown in Figure 2.5 can be found in the SI (Section S4).

ε_{A_0} is high at low exposure levels; this is a result of overestimating depth. The NLSB method is the least biased out of the four methods at all exposure levels; however, the median bias is actually lower at 40% exposure than at 80% exposure — NLSB tends to underpredict A_0 somewhat at high exposure levels.

2.4 Discussion

2.4.1 Evaluating performance in the context of breakpoint locations

We employed four curve-fitting methods to extrapolate height-width relationships fitted to error-corrupted measurements of flow width and water surface height to predict unknown bathymetry parameters. Next, we evaluate the different methods and recommend which is

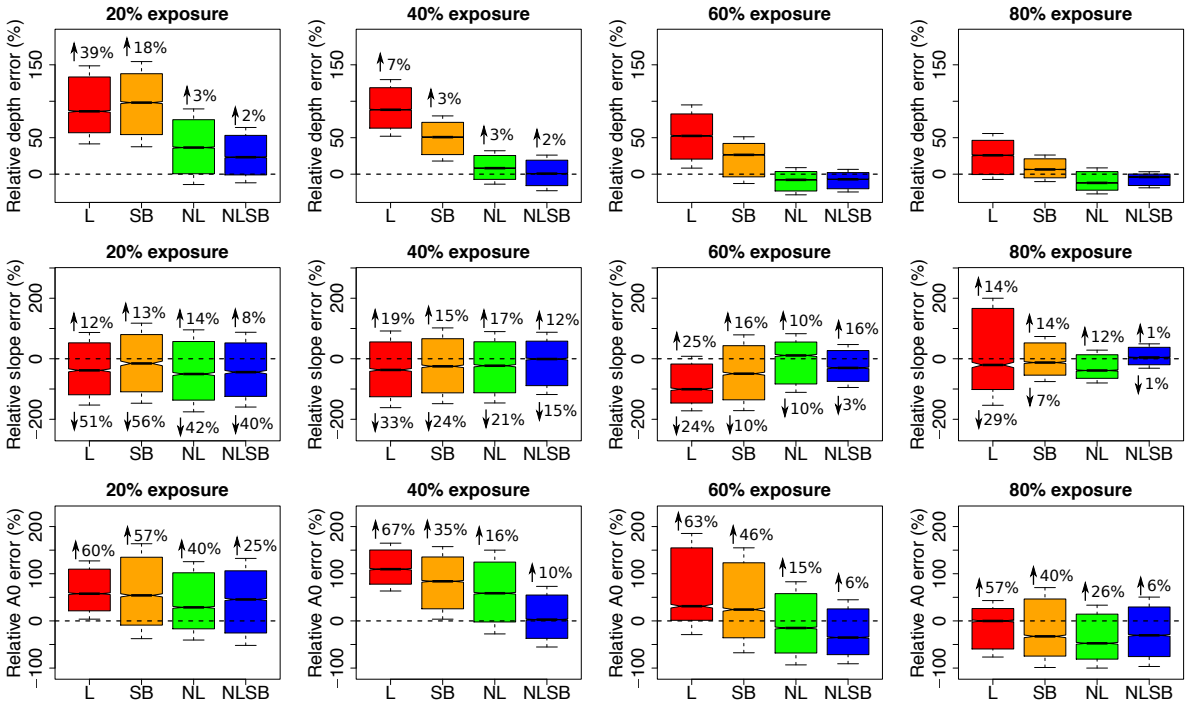

Figure 2.5: Relative depth, slope, and low flow area prediction error (ε_d,ε_{s_0} and ε_{A_0}) for the linear, slope break, nonlinear, and nonlinear slope break methods. The box and whisker plots show ε_d and ε_{A_0} for each method, lumped over all 30 study reaches. Slope errors (ε_{s_0}) are for 16 adjacent reaches where finite differences could be calculated. Errors greater than 200% have been excluded from the plots so the difference between models can be seen more clearly. Arrows and inset text show the percent of prediction errors higher (or lower) than the truncation threshold.

preferred. There are computational and accuracy tradeoffs between the methods. The linear method is the simplest and the fastest to apply but the least accurate, while the NLSB requires an iterative solution and good initial guesses for the parameters, but it is also more accurate. The utility of the method depends on how well the assumed $h(w)$ model matches reality. For the 30 nonbraided Upper Mississippi River reaches, the NLSB model is the best approximation of the true, fully-observed, height-width relationship, with an average RMSE of 0.12 m, compared to 0.65 m for the linear model, 0.23 m for the SB model, and 0.32 m for the NL model, where $RMSE = \sqrt{\frac{SSE}{N}}$ and SSE is the sum of squared error between the fitted $h(w)$ model and the measured data.

The NLSB method had the smallest relative errors (Eqs. 13-15), followed by the NL, SB, and linear methods. We recommend using the NLSB method, but note that there are also reasons to use the NL method. The NLSB method provides a somewhat tighter bound to the bathymetry parameters than does the NL method, but it requires more h-w data points to fit two models for the NLSB method than it does to fit one model for the NL method. The advantage of the NLSB method is that it can handle the case when there is a breakpoint, whereas the NL method cannot. For a given reach, we would expect observation of the breakpoint to reduce bathymetry prediction error. Specifically, our hypothesis was that errors would decrease with channel exposure, drop suddenly once the breakpoint was observed, and continue decreasing smoothly with channel exposure past the breakpoint. We plotted ε_d vs. maximum observed exposure level for each reach (Figure 2.6) to evaluate this hypothesis. Figure 2.6 shows ε_d for these 16 of the 30 study reaches. Recall from Section 2.3 that we modified the banks of the original UMESC cross sections in order to define bankfull geometry for uneven cross sections. Since this may introduce artificial breakpoints into the height-width data, we used only the 16 reaches where bank adjustments were less than 10% of bankfull depth for Figure 2.6. These were reaches 1, 3, 4, 5, 6 in Pool 4; reaches 1, 4, 5 in Pool 26; and reaches 1, 2, 3, 4, 7, 9, 11, 12 in La Grange Pool.

We calculated breakpoint locations for each reach assuming full channel exposure and

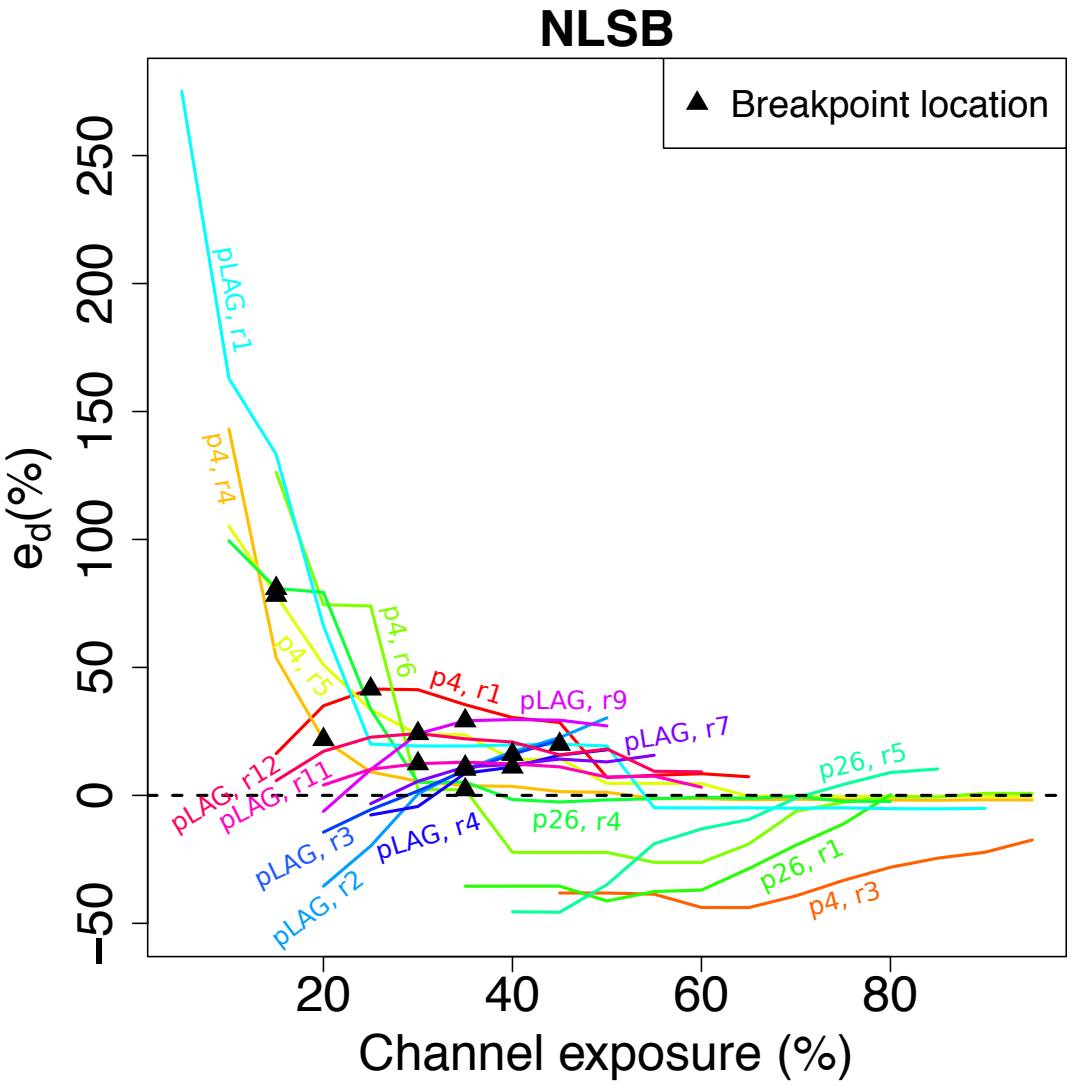

Figure 2.6: NLSB depth prediction errors vs. exposure level. Each line corresponds to one pool-reach combination (LAG=La Grange); markers show the breakpoint locations. Breakpoint locations are not shown if they occur at an exposure level before any NLSB predictions can be made at that reach, e.g. if there are fewer than 10 data points.

no measurement error. The average breakpoint occurred at 27% channel exposure, with a range of about 8-45% exposure. Breakpoints are indicated on Figure 2.6 — except they are omitted in the case where the breakpoint occurred at such a low exposure level that no NLSB predictions could be made, e.g. because there were fewer than 10 height-width observations available.

As might be expected, ε_d decreased with exposure level, but in contrast to our hypothesis, observing the breakpoint does not appear to be a prerequisite for good accuracy. Inspection of the reach-average cross section shapes reveals that while there are some cross sections with well-defined slope breaks, such as reach 5 in Pool 4, many cross sections (especially in La Grange Pool) are parabolic in shape and thus well-approximated even without a breakpoint. We found that spatial averaging from the cross-section scale to the reach scale caused the h(w) curves to become smoother and decreased the importance of the breakpoints.

Since the NLSB method had better accuracy than the NL method and since it is robust to the presence of a breakpoint in the reach-average height-width relationship, whether or not it actually exists, we advocate for the NLSB method. However, for applications with fewer than 10 observations the NL method may be the better choice. In the following sections, we report further applications of our bathymetry prediction methods and comparisons with previously developed methods.

2.4.2 Comparison to Mersel's slope break method

[Mersel et al., 2013] used height and width relationships to predict bed elevations z_0 for cross sections on the Upper Mississippi and several other rivers. They used cross section data from 1495 cross sections, but they only predicted bathymetry at a fairly small subset of those cross sections. They applied versions of the linear and SB methods at "optimal" locations only, classifying a cross section as an optimal location if the height and width observations were close to linear, possibly with one breakpoint. The number of optimal locations varied with exposure level and method; at best, the method predicted bathymetry

at 63% of cross sections using the linear model and 16% using the SB model. While they did not evaluate A_0, we can compare our z_0 estimation results to theirs. They reported z0 RMSE values relative to bankfull depth.

M13's results for the Upper Mississippi River were as follows: for the linear method, RMSE was as low as 11% at 95% channel exposure, and as high as 302% at 20% exposure. For the SB method, RMSE was as low as 3% at 95% channel exposure, and as high as 20% at 20% exposure. In general, the number of optimal locations where they were able to predict z_0 varied with exposure, as well, ranging from 63% of cross sections at 20% exposure to 11% of cross sections at 95% exposure for the linear method, and from 0.4% of cross sections at 20% exposure to 16% of cross sections at 95% exposure for the slope break method.

Our method makes predictions at all locations, not just "optimal" locations. Essentially, there is a tradeoff between the accuracy of the method and its domain of applicability. It is not a perfect comparison because M13 was applied only at optimal cross sections, whereas our method was applied at nonbraided reaches. We calculated standard errors (SE) of depth estimates $SE = \varepsilon_d^2$ for each reach and exposure level, using the mean z_0 prediction to calculate ε_d to compare with M13's deterministic approach. These values are plotted in Figure 2.7, along with the original M13 results. We also calculated $rRMSE = \sqrt{\frac{1}{N}\sum_{r=1}^{N} SE_r}$ to summarize the errors at each exposure level. Applying the linear and SB methods to all (not just optimal) reaches, reduces the accuracy of the depth prediction. Note that both the NL and NLSB methods outperform the linear and SB methods, with average depth SE as low as 15% for the NLSB method at 80% channel exposure.

2.4.3 Comparison to Hagemann's A0 regression model

[Hagemann et al., 2017] (hereafter H17) fit a regression model to flow width and area, taking A0 as the minimum observed flow area at a site and data from the HydroSWOT dataset ([Canova et al., 2016]) to formulate a prior for A_0 based on width measurements alone:

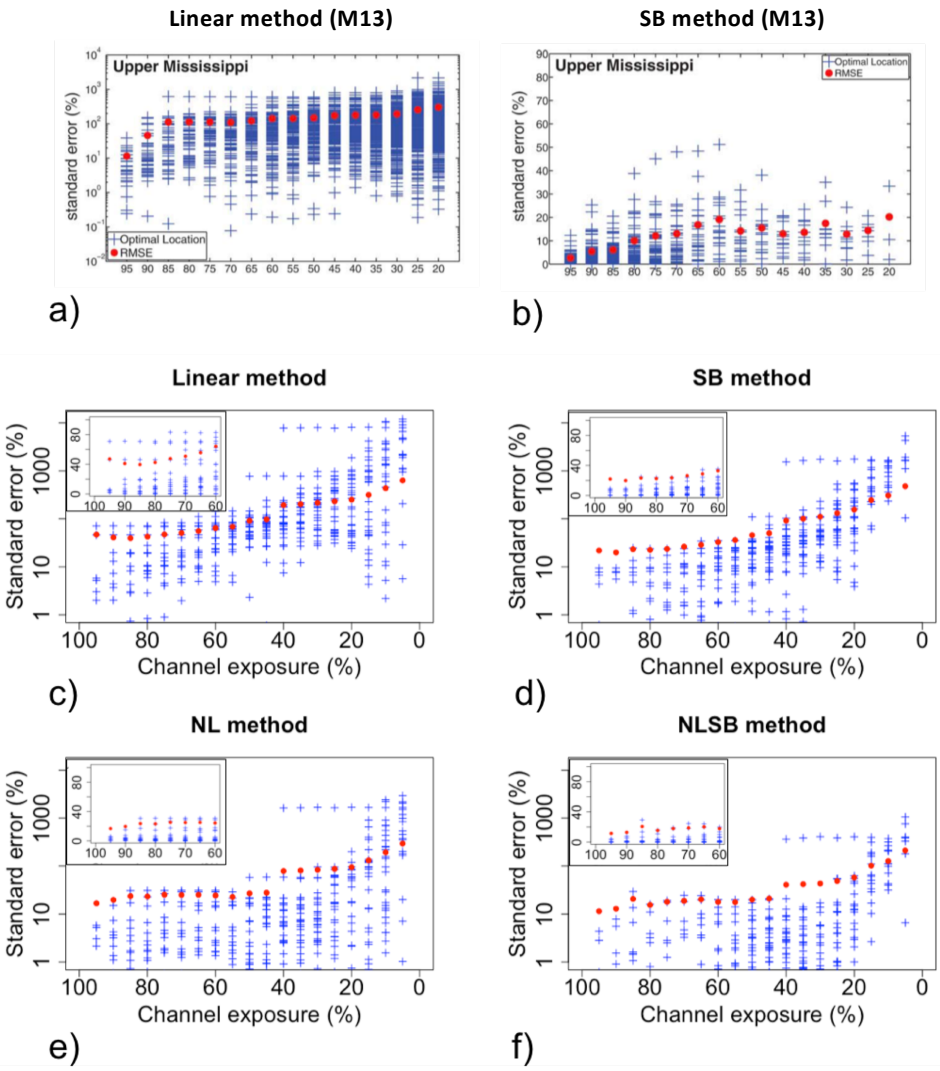

Figure 2.7: Panels (a) and (b) show depth prediction SE (blue plus signs) and relative RMSE (filled red circles) at optimal locations for the M13 linear and slope break methods (reprinted with permission from Mersel et al. (2013)). Panels c-f show the same statistics for the linear, slope break, nonlinear, and nonlinear slope break methods presented herein.

$$ln(A_0) = 0.855 + 1.393 * mean(ln(w_{obs}) - 1.7249043 * sd(ln(w_{obs}) + \varepsilon_{A_0} \quad (2.16)$$

where w_{obs} are the observed width values. R^2 for the model was 0.70. They found that ε_{A_0} was normally-distributed with standard deviation 0.948 and zero mean. Therefore, A_0 values can be generated by simulating a normal random variable with mean equal to the right-hand side of Hagemann's regression equation, except for the error term, and a standard deviation of 0.948, and taking its exponent. As our method uses both width and height measurements, we expect it to be able to predict A_0 with lower error than a solely width-based method. We applied the NLSB method to the HydroSWOT data as a way to both validate the method on a new dataset and compare its accuracy to that of another A_0 prediction method.

HydroSWOT is a USGS dataset composed of acoustic Doppler current profiler measurements of channel and flow characteristics in USGS stream gage network. It includes measurements of flow rate, width, area, and velocity at 10,081 sites in the U.S., although not all variables are measured at all sites. There were 4,274 sites with at least 10 height and width measurements, which is required for the NLSB method. There were also 989 sites where all the width observations were zero, even though height varied. After excluding these sites and sites with fewer than 10 height and width measurements, we applied H17's regression (Equation 16) and NLSB to the remaining 4,253 sites. We calculated A_0 prediction error as $\hat{A}_0 - A_0$, and plotted error histograms for the mean A_0 predictions from each method in Figure 2.8. The NLSB method far outperformed the H17 method, predicting the mean A_0 with an average error of -105 m^2 compared to 2570 m^2 for H17. This is a significant difference in performance, considering that the mean true A_0 value was about 1300 m^2. The NLSB method failed to fit a model at 48 sites, where the observed height-width data were in very poor agreement with the NLSB model. Since the distributions have such different shapes, we compare modes, as well. Using the Freedman-Diaconis rule for the number of bins, which we used to generate Figure 2.8, we calculated the mode of the H17 prediction errors as 250 m^2 and the mode of the NLSB prediction errors -10 m^2.

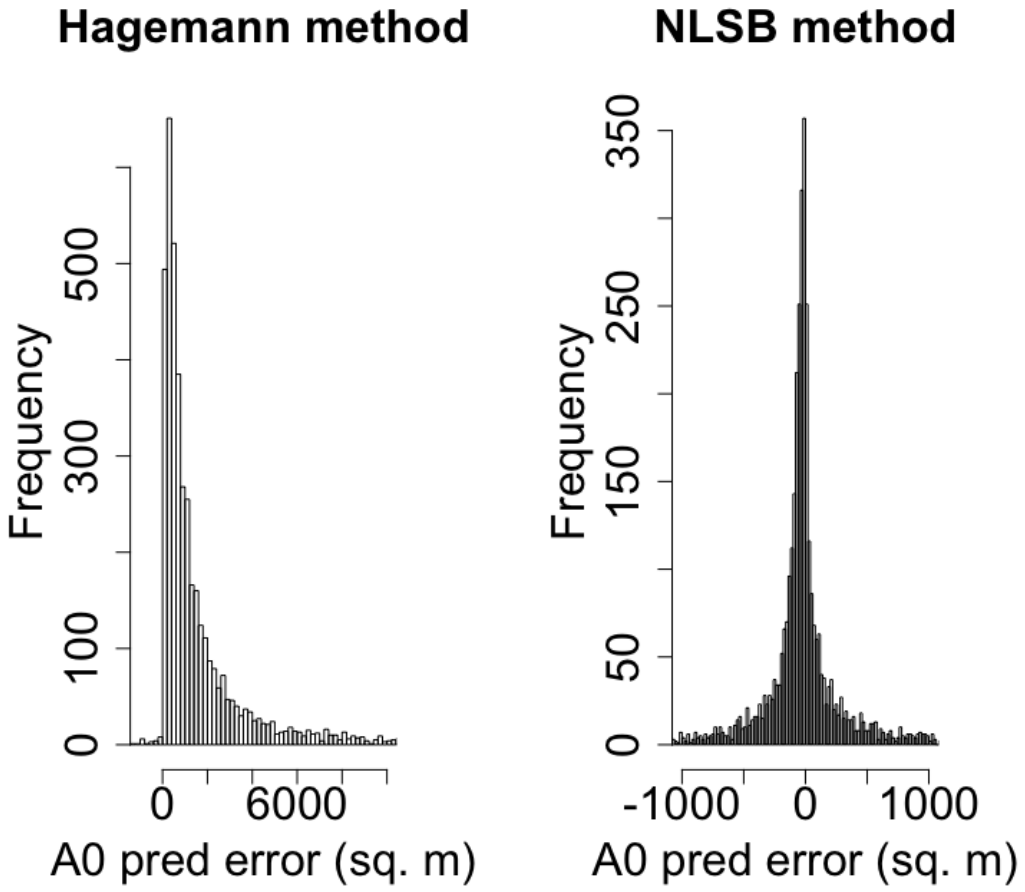

Figure 2.8: Histograms of A_0 prediction error for the Hagemann regression model vs. the NLSB method. There is one prediction for each of about 4000 HydroSWOT sites.

The results of applying the NLSB method to the HydroSWOT dataset are encouraging because of their accuracy but also because they demonstrate that the method has potential to work in rivers besides the Upper Mississippi River, and without any sort of bank modification.

2.4.4 Prior distributions

As discussed in Section 1, SWOT discharge algorithms can use prior information about A_0 to constrain their posterior estimates of discharge, as well as bathymetry and roughness parameters. The NLSB method predicts A_0 with lower error than the L, SB, or NL methods, so we developed a way to easily set up an A_0 prior based on the NLSB fit, but which could also be applied using the L, SB, or NL methods.

From looking at Figure 2.5, we might expect that the distribution of prediction errors should vary with exposure level. Thus, we separated the ε_{A_0} data by exposure level and fit normal and lognormal distributions for each exposure level using method of moments estimators. Visual inspection of the histograms showed the lognormal distribution to be the much better fit. When fitting lognormal distributions to ε_{A_0}, we shifted the data by $\gamma = -1$ because the lognormal distribution cannot produce negative values, and ε_{A_0} can range from -1 to ∞. ε_{A_0} must take values within this range because $\varepsilon_{A_0} = \frac{\hat{A}_0}{A_0} - 1$, and areas are required to be positive. Thus, we assume ε_{A_0} is distributed as:

$$(\varepsilon_{A_0} - \gamma) \sim LN(\mu', \sigma') \tag{2.17}$$

Or, equivalently:

$$ln(\varepsilon_{A_0} - \gamma) \sim N(\mu, \sigma) \tag{2.18}$$

where γ is the shift parameter for the shifted lognormal distribution (Singh, 1998).

We added one more layer of parameterization by plotting the lognormal parameters μ'

and σ' vs. exposure level and fitting the observed relationship with a linear regression. The resulting parameterization is given by:

$$\mu'_k = 1.212 - 2.438k \tag{2.19}$$

$$\sigma'_k = 0.5901 + 1.1827k \tag{2.20}$$

where μ'_k and σ'_k are lognormal location and scale parameters at exposure level k, where k ranges from 0 to 1. The variable μ'_k decreases with exposure level, while σ'_k appears to increase: this could be due to the tendency of ε_{A_0} to occasionally take on very high values at high exposure levels, when the denominator, A_0, is close to zero. The R^2 values for the regression equations (2.19) and (2.20) are 0.83 and 0.72, respectively. Further details of the fitting and subsequent parameterization can be found in the SI (Section S5).

As an example of how we can use the NLSB method with equations (2.19) and (2.20) to generate an A_0 prior, let us look at reach 1 of Pool 21 (with height and width measurements down to 40% channel exposure). We fit the NLSB model to the height and width data and used it to predict $\hat{A}_0 = 385m^2$. To incorporate error into our estimate of A_0, we rearrange the formula for relative A_0 error (Equation 2.14) as:

$$A_0 = \frac{\hat{A}_0}{1 + \varepsilon_{A_0}} \tag{2.21}$$

While ε_{A_0} is lognormally-distributed, A_0 is distributed as $\frac{1}{\varepsilon_{A_0}}$, with a reciprocal lognormal distribution. We simulated 10,000 realizations of $\varepsilon_{A_0} \sim LN(\mu'_k, \sigma'_k, \gamma = -1)$, calculated A_0 realizations using Equation (2.21), and used the simulated values to plot the histogram in Figure 2.9, which is the prior distribution for A_0. For this reach, the mean value of the prior was close to the true value of A_0. The median of the A_0 prior was 342 m^2.

We evaluated the performance of exposure-level dependent priors for every pool, reach, and exposure level by calculating bias and the 95% exceedance ratio (ER95; Moradkhani et

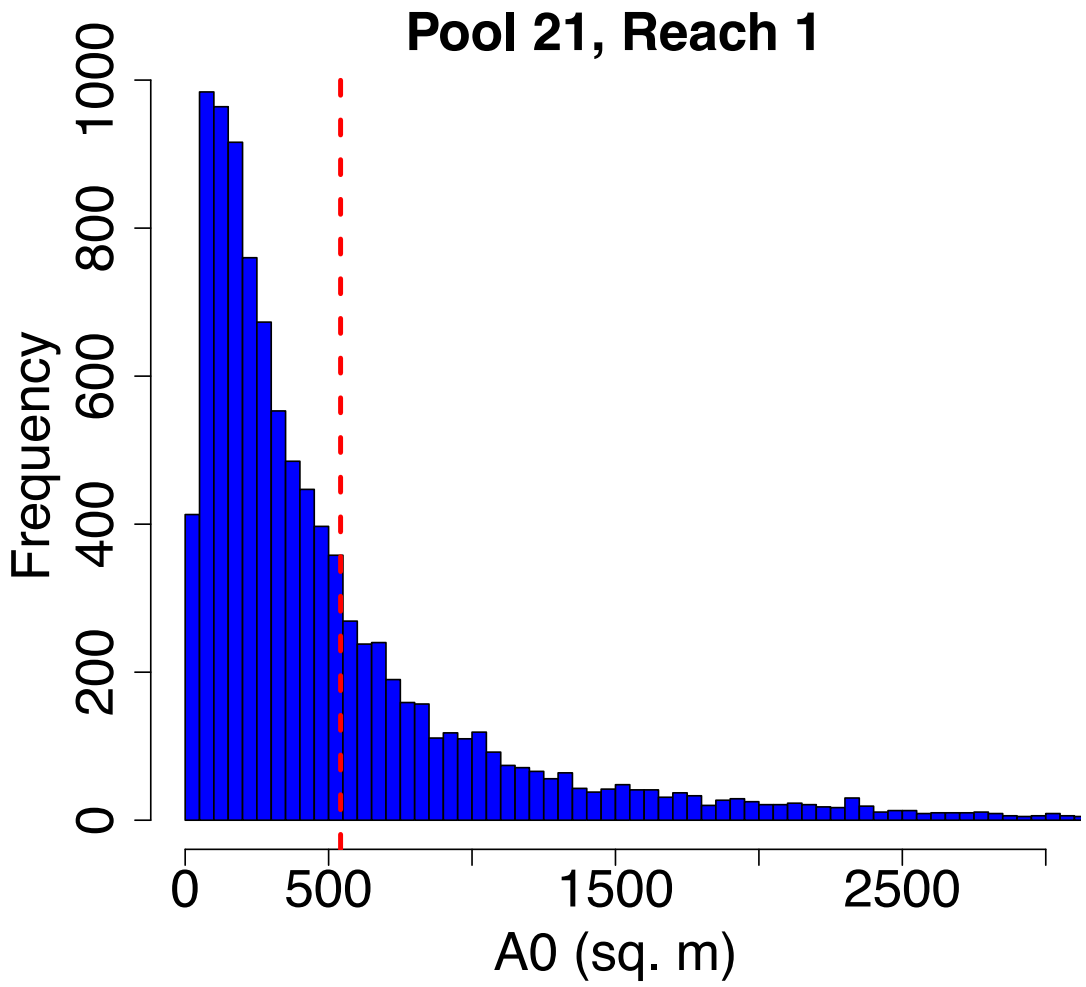

Figure 2.9: Example A_0 prior for Pool 21, reach 1, at 40% channel exposure, using the NLSB parameterization for the lognormal prior. The true value (red dashed line) is 542 m^2.

al., 2006) for each A_0 prior. We computed bias as the difference between the sample mean and the true value, and ER95 the number of times the truth fell outside the 95th percentile confidence range:

$$Bias_i = mean(A_0 - A_0^{true}) \tag{2.22}$$

$$ER_{95} = \frac{1}{N}\sum_{i=1}^{N}\left(A_0^{true} > A_0^{97.5} | A_0^{true} < A_0^{2.5}\right) \tag{2.23}$$

for the i^{th} pool, reach, and exposure level.

The bias between the A_0 priors and the true values for each pool, reach, and exposure level where an NLSB model could be fitted ranged from -2620 m^2 to 40.9×10^3 m^2, but with a much smaller mean value of 810 m^2. The ER95 was 5.3%; the majority of the priors encompassed the true value.

So far, we have assumed the maximum observed exposure level (e_1) is known. Once SWOT begins making observations, we will not necessary know the observed exposure level for a particular observation because we may not know the bankfull width. Therefore, we also evaluated how well A_0 could be estimated with a prior that is not dependent upon exposure level. We did so by fitting a single distribution to all the ε_{A_0} values as a whole. The lognormal distribution was a good fit for the data, with parameters $\mu' = -0.02$ and $\sigma' = 1.30$. Essentially, we are saying that there is some uncertainty described by a lognormal random variable ε_{A_0} that does not vary with exposure level. The exposure-level varying component of the A_0 prior comes from the NLSB prediction $\hat{A}_0$. We again calculated bias and ER_{95} and found that this assumption fares quite well. Bias ranged from -1970 m2 to $97.9\times10^3 m^2$, with a mean value of 3140 m^2. ER_{95} was 5.0%, slightly lower than when we used the additional knowledge about exposure level. Because the exposure level-dependent prior had lower bias than the more general prior, we recommend using it when w_{bf} (and thus exposure level) is known.

2.5 Conclusions

We evaluated four methods for using SWOT measurements of water surface height and flow width to predict low flow area A_0, as well as minimum bed elevation z_0 and bed slope s_0. The nonlinear (NL) and nonlinear slope break (NLSB) methods recognize that reach-average height-width relationships tend to have an exponential shape. They exploit this assumption to predict depth with a median error of 1.1% and A_0 with a median error of 13% at channel exposure as low as 40%, evaluated in thirty 10-km reaches of the Upper Mississippi River. Unlike a recently developed approach meant to be used only at "optimal" locations, the NL and NLSB methods developed here can be applied over any nonbraided river reach. As such, they could be used, for instance, to help produce global bathymetry maps based on SWOT data.

We expect our methods to be useful for SWOT discharge estimation algorithms that use prior A_0 distributions in a Bayesian parameter estimation framework, such as MetroMan ([Durand et al., 2014]; [Yoon et al., 2016]), BAM ([Hagemann et al., 2017]), or data assimilation methods ([Oubanas et al., 2018b]; [Oubanas et al., 2018c]). Equifinality makes it difficult for estimation algorithms to determine unknown bathymetry, roughness, and boundary inflows simultaneously. Prior knowledge of bathymetry can be used to constrain ill-posed hydraulic inverse problems where bathymetry, roughness, and boundary inflows are unknown. We showed an example of how the NLSB method can be used to generate a realistic A_0 prior from a limited number of height and width measurements. Such A_0 priors could better constrain the estimation problem and help overcome equifinality limitations.

Although we developed the curve-fitting methods specifically using data for the Upper Mississippi River, the HydroSWOT database allowed us to evaluate them at 4000 sites across the continental U.S., which has a broad range of climatologies. In general, we expect the NL and NLSB methods to apply anywhere that hydraulic geometry relationships hold true.

Globally-available bathymetry predictions for large rivers have applications beyond pa-

rameterizing SWOT discharge algorithms. They could also be used to improve the National Water Model (NWM; [Gochis et al., 2016]), which currently assumes a trapezoidal cross section of infinite depth with constant side slope of 1:20 for rivers in the continental United States — an assumption that hinders NWM's ability to forecast river stage and discharge ([Brackins et al., 2018]). Other potential applications include bathymetry mapping for navigation and understanding how human disturbance such as dam construction affects river bathymetry over time. Finally, this methodology was designed with SWOT in mind, but it could potentially be applied using other flow width and river height data, such as spaceborne lidar measurements from ICESat or ICESat-2.

2.6 Supporting Information

2.6.1 Tables of acronyms and variable names

Acronym	Explanation
AMHG	At-many-stations hydraulic geometry algorithm for discharge estimation
BAM	Bayesian-AMHG Manning algorithm for discharge estimation
L	Linear height-width model
McFLI	Mass-conserved flow law inversion
MetroMan	Metropolis-Manning algorithm for discharge estimation
NL	Nonlinear height-width model
NLSB	Nonlinear height-width model with one breakpoint
SB	Slope break, linear height-width model with one breakpoint
SWOT	Surface Water and Ocean Topography Mission
UMESC	Upper Midwest Environmental Science Center

Table 2.2: Acronyms.

Variable	Description
A	Wetted area
A_0	Low flow wetted area
u	Coefficient in the nonlinear height-width model
d	Flow depth
e	Exposure level
e_1	Exposure level at minimum observed flow width
ε_{A_0}	A_0 *prediction error*
ε_d	Depth prediction error
ε_{S_0}	Bed slope prediction error
h	Water surface height
h_1	Minimum observed water surface height
s	Shape parameter
s_0	Bed slope
w	Flow width
w_1	Minimum observed flow width
z_0	Minimum bed elevation

Table 2.3: Selected variable names.

2.6.2 Pre-processing UMESC bathymetry data

The UMESC bathymetry data consist of a dense grid of depth measurements. We wish to characterize the river in terms of "reaches," each with a uniform cross section along its length, and derive height-width relationships for each reach from which to obtain error-corrupted, SWOT-like, measurements. The preprocessing steps required to do so are described in this section and illustrated in Figure S1. We used the R package WSEw (`https://github.com/jschap1/WSEw`), to go from 2D bathymetry data to 1D reach-average height-width data.

1. Evenly-spaced transects are drawn perpendicular to the river centerline. We took river

centerlines from HydroSHEDS ([Lehner et al., 2008]), and drew perpendicular bisectors every 5 m (the highest resolution possible given the 5 m gridded bathymetry data) along the centerline. We smoothed the river centerline using Gaussian kernel smoothing before drawing the transects to avoid jagged portions of the original centerline that otherwise caused many transects to be misaligned relative to the river channel. Figure S1(a) shows transects drawn every 500 m along the centerline for Pool 21 of the Mississippi River.

2. Rivers often consists of multiple channels, i.e. they can be braided. Also, rivers can sometimes run parallel to lakes or wetlands. The UMESC data used in this study do not provide a way to distinguish the main river channel from other nearby water bodies. To avoid this issue, we restricted the analysis to nonbraided portions of the river. Figure S1(b) shows examples of cross sections classified as "braided" and "nonbraided." We extracted depths along each nonbraided transect. We calculated bed elevations at each cross section by subtracting depth from the reference water surface height at the time of the sounding, provided with the UMESC data.

3. Some of the cross sections constructed this way have uneven bank elevations, that is, the bank elevation differs between the left and right banks. Uneven bank elevations may occur naturally due to different soil materials on either bank, or they could be spurious, a result of the resolution of the bathymetry data near the banks not being fine enough to account for the steep changes in bed elevation occurring there. In order to determine bankfull width and depth, we imposed zero depth at the right and left banks to make them even. The example in Figure S1(c) shows the bank adjustment for a particular cross section; averaged over a reach, the bank adjustments are less severe. Sensitivity tests showed that this modification did not strongly affect the performance of the curve-fitting methods (see Section S3 for details).

4. After correcting uneven banks, we calculated maximum bankfull depth as the difference

in elevation between the tops of the banks and the thalweg, the point of lowest elevation in the channel. We generated height-width data by entering heights from the bed elevation to the bankfull depth, at 5 percent increments of exposure, and using the cross section geometry to calculate the corresponding width.

5. SWOT observations will be reported at a reach scale, so we computed average cross sections at the reach scale. We did this by taking the average cross section of all nonbraided cross sections within a 10 km reach, but only for reaches where a significant amount of the cross-sections within the reach are classified as nonbraided. Figure S1(d) shows nonbraided cross sections within each of Pool 21's three reaches. We performed our analysis on the 30 least braided reaches, out of 51 reaches in the UMESC domain. All these reaches were at least 30% nonbraided, with an average braided fraction of 42%. Figure S1(e) shows the average cross section shape and height-width relationship for reach 1 of Pool 21.

2.6.3 Sensitivity analysis to bank adjustment during UMESC preprocessing

As noted above, the UMESC cross sections could have different bank heights on the left and right sides. In order to define bankfull quantities, we assumed that the depth at both of the cross section banks is zero. This required adjusting the cross section as shown in Figure 2.10(c). We were concerned that adjusting the cross sections might introduce artificial breakpoints in the height-width relationship, and thus promote curve-fitting methods that account for breakpoints (SB, NLSB) over methods that do not account for breakpoints (L, NL). We present the following sensitivity analysis to support our claim that this adjustment does not seriously impact our results.

$$\Delta b = \frac{max(\delta b_l, \delta b_r)}{d_{bf}} \tag{2.24}$$

where δb_l, δb_r, and d_{bf} are defined in Figure 2.10(c). Δb measures how much the bank

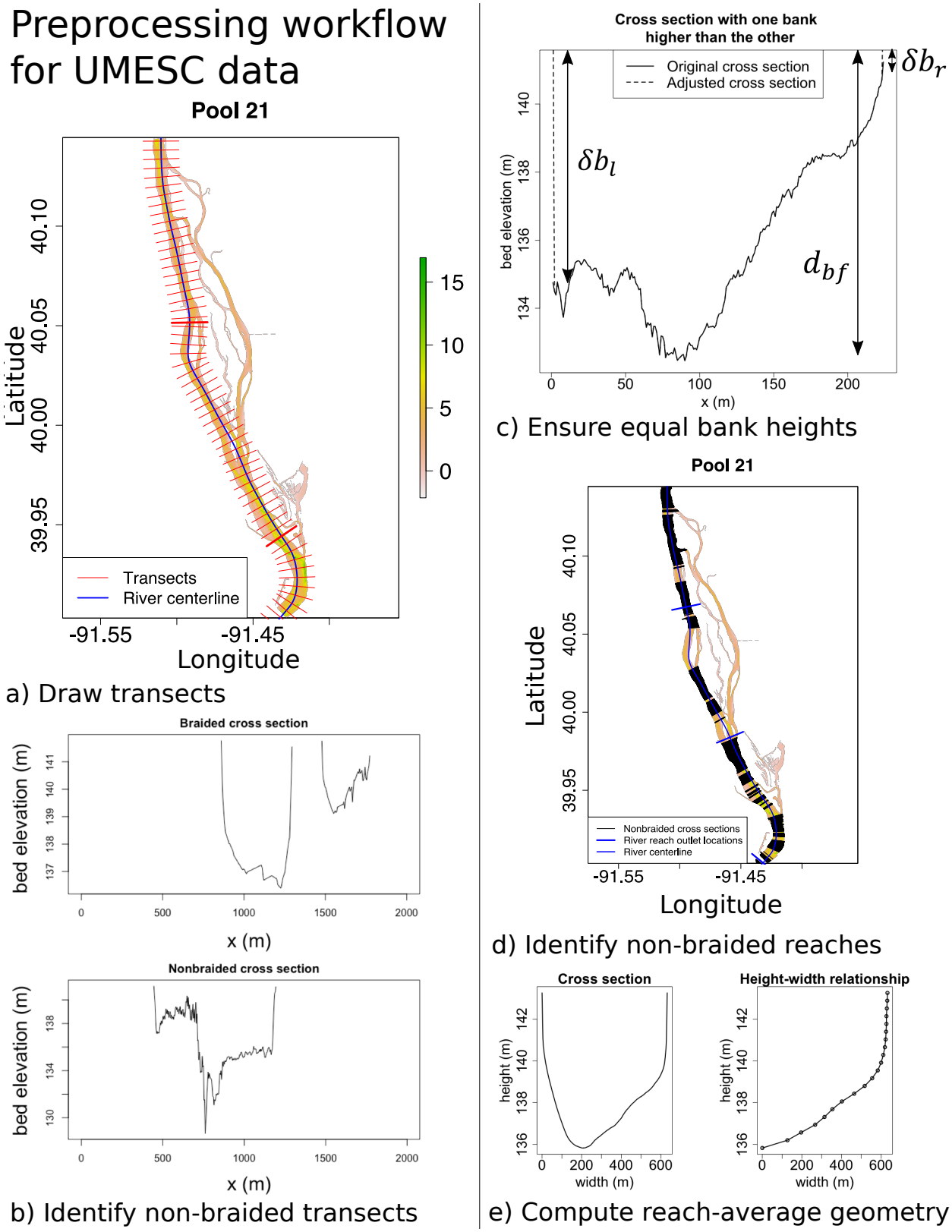

Figure 2.10: Illustration of preprocessing workflow used to compute reach-average cross section geometry from 2D UMESC bathymetry data for Pool 21 of the Upper Mississippi River. The two bolded transects in panel (a) are shown plotted in panel (b).

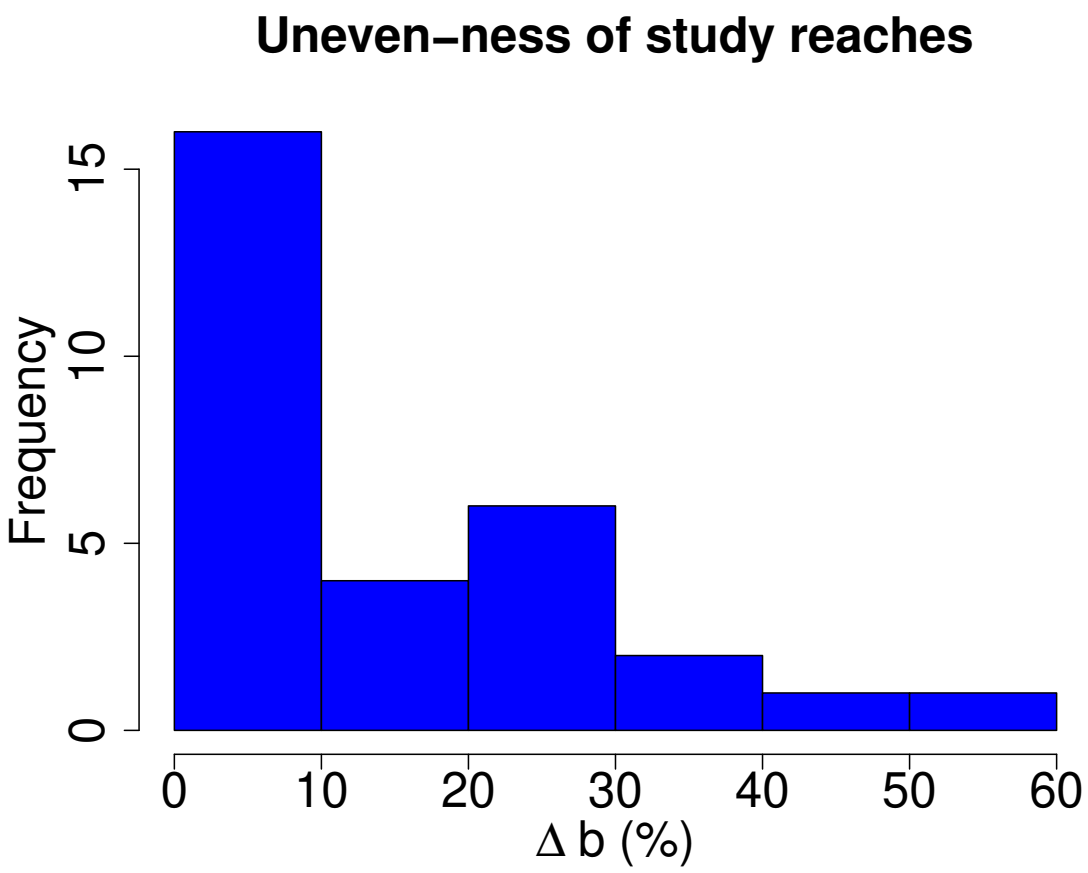

Figure 2.11: Histogram showing Δb values for the 30 study reaches, calculated as cross-section averages over each 10-km reach.

heights were adjusted, as a percent of bankfull depth; its average value for the 30 study reaches was 15%. Figure 2.11 shows the distribution of Δb values for the 30 study reaches. Ten of the reaches have $\Delta b \geq 20\%$. The highest Δb value was 51%.

Although most of the reaches have fairly small bank modifications, we were concerned about the potential impact of bank modifications on the performance of the curve-fitting methods, so we performed a sensitivity analysis, splitting the reaches into two groups: those with high Δb values ($\Delta b \geq 20\%$) and those with low Δb values ($\Delta b \leq 5\%$). We used the L, SB, NL, and NLSB methods to predict bathymetry for both groups of reaches and we

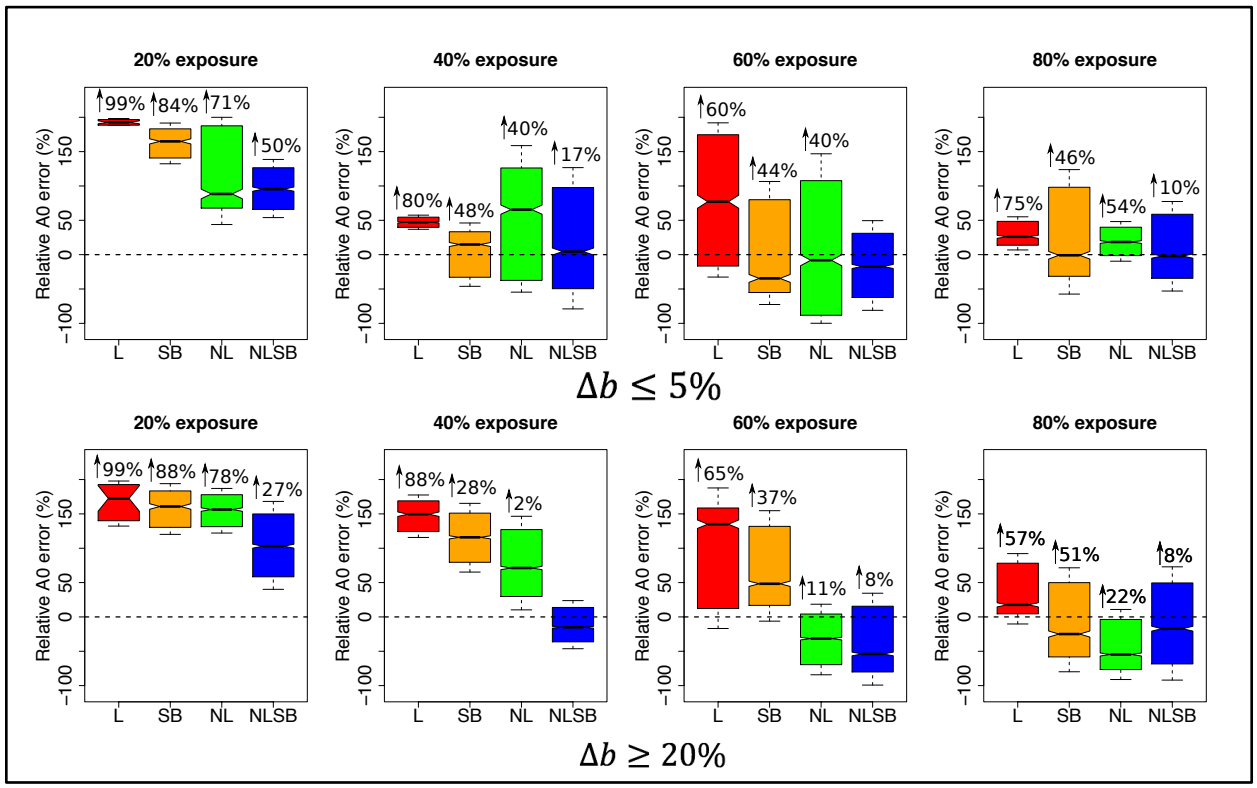

Figure 2.12: Box and whisker plots showing relative A_0 prediction error for each curve-fitting method. The top row shows ε_{A_0} calculated for the five reaches with the lowest Δb values, while the bottom row shows ε_{A_0} calculated for the eight reaches with the highest Δb values. Errors higher than 200% have been truncated and excluded from the plots, but they are indicated by annotations and arrows.

plotted the resulting ε_{A_0} values in Figure 2.12. For both groups of reaches, error is lower for the NLSB method than for the NL method and lower for the SB method than for the L method, so we conclude that the bank adjustments do not bias our experiment in favor of the SB or NLSB methods. Also, the magnitudes of the prediction errors do not differ greatly between reaches with small bank adjustments vs. reaches with large bank adjustments.

2.6.4 Relative errors for depth, slope, and low flow area prediction

In case the reader desires a closer look at the information shown in Figure 2.5, we have tabulated the 25th, 50th, and 75th percentiles of the prediction errors in Tables 2.4, 2.5, and 2.6, below.

Exposure		**L**			**SB**			**NL**			**NLSB**	
Percentile	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
20%	74.4%	151%	256%	65.8%	114%	173%	2.0%	38.8%	77.6%	-0.6%	24.3%	55.9%
40%	65.8%	90.6%	128%	28.0%	52.0%	73.3%	-6.5%	9.8%	28.7%	-15.5%	1.1%	20.2%
60%	20.7%	52.4%	82.6%	-3.7%	26.5%	42.1%	-23.1%	-7.8%	3.6%	-20.0%	-7.0%	2.2%
80%	-0.6%	25.9%	46.4%	-5.0%	6.5%	21.0%	-22.0%	-12.0%	3.5%	-15.6%	-3.9%	0.2%

Table 2.4: Depth prediction errors (ε_d) relative to bankfull maximum depth at 25th, 50th, and 75th percentiles for each method at 20, 40, 60, and 80% channel exposure

Exposure		**L**			**SB**			**NL**			**NLSB**	
Percentile	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
20%	-615%	-208%	15.8%	-651%	-276%	25.8%	-320%	-156%	54.5%	-283%	-143%	21.6%
40%	-334%	-90.3%	110%	-195%	-51.2%	103%	-159%	-36.3%	120%	-129%	-8.1%	79.2%
60%	-196%	-99.5%	199%	-148%	-34.5%	99.5%	-99.1%	10.9%	65.0%	-72.0%	-13.9%	82.9%
80%	-232%	-67.1%	172%	-59.3%	-0.7%	113%	-59.8%	-18.7%	34.9%	-20.1%	4.2%	39.1%

Table 2.5: Bed slope prediction errors (ε_{S_0}) at 25th, 50th, and 75th percentiles for each method at 20, 40, 60, and 80% channel exposure.

Exposure		**L**			**SB**			**NL**			**NLSB**	
Percentile	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
20%	82.0%	335%	571%	81.2%	239%	430%	13.2%	137%	280%	0.1%	83.6%	203%
40%	150%	298%	524%	60.9%	141%	265%	10.8%	79.8%	158%	-33.4%	12.8%	88.4%
60%	137%	342%	830%	12.9%	176%	327%	-60.7%	-2.6%	109%	-69.5%	-28.2%	48.2%
80%	9.0%	337%	1660%	-46.3%	84.5%	568%	-72.0%	-5.0%	207%	-72.6%	-23.4%	49.8%

Table 2.6: Low flow wetted area prediction errors (ε_{A_0}) at 25th, 50th, and 75th percentiles for each method at 20, 40, 60, and 80% channel exposure

2.6.5 Parameterizing the ε_{A_0} distributions as a function of channel exposure

This section describes our method for postulating A_0 priors based on the distribution of NLSB bathymetry prediction errors, going into more detail than Section 4.4 of the main text. The problem is as follows: Given h and w measurements down to a particular maximum observed exposure level, we wish to postulate a prior distribution on A_0. It is also possible that we will not know the exposure level, so we handle that case, too. Given that we have calculated ε_{A_0} replicates using Monte Carlo simulation based on 30 Upper Mississippi river reaches (the prediction error distributions are shown, for example, in Figure 2.5), our strategy is to use the distribution of these ε_{A_0} values calculated for the Upper Mississippi River study area to postulate prior distributions that we can use for predicting bathymetry in other rivers, where the true bathymetry is unknown.

Plotting histograms ε_{A_0}, we find that the prediction errors well-described by a lognormal distribution. Figure 2.13 shows histograms of ε_{A_0} values calculated using the NLSB method over all 30 study reaches at 20, 40, 60, and 80 percent channel exposure. Fitted normal and lognormal distributions are shown in the figure, as well. Clearly, the lognormal distributions are the better fit. The shifted, or three-parameter, lognormal distribution needed to be used, since ε_{A_0} can range from -1 to ∞.

We go one step further to develop ε_{A_0} priors that depend on how much of the channel is exposed. After obtaining lognormal distribution parameters for 5, 10, ..., 95% maximum observed channel exposure, we plot the parameters vs. channel exposure (Figure 2.14), and parameterize the parameters, so to speak. We tested linear, semilog, and loglog (where possible) models for $\hat{\mu}'$ and $\hat{\sigma}'$, the lognormal distribution parameters. We chose a linear model for its simplicity and because it provided a good fit for both parameters, in terms of R^2. The resulting regression equations are shown in the main text (Equations 2.19 and 2.20).

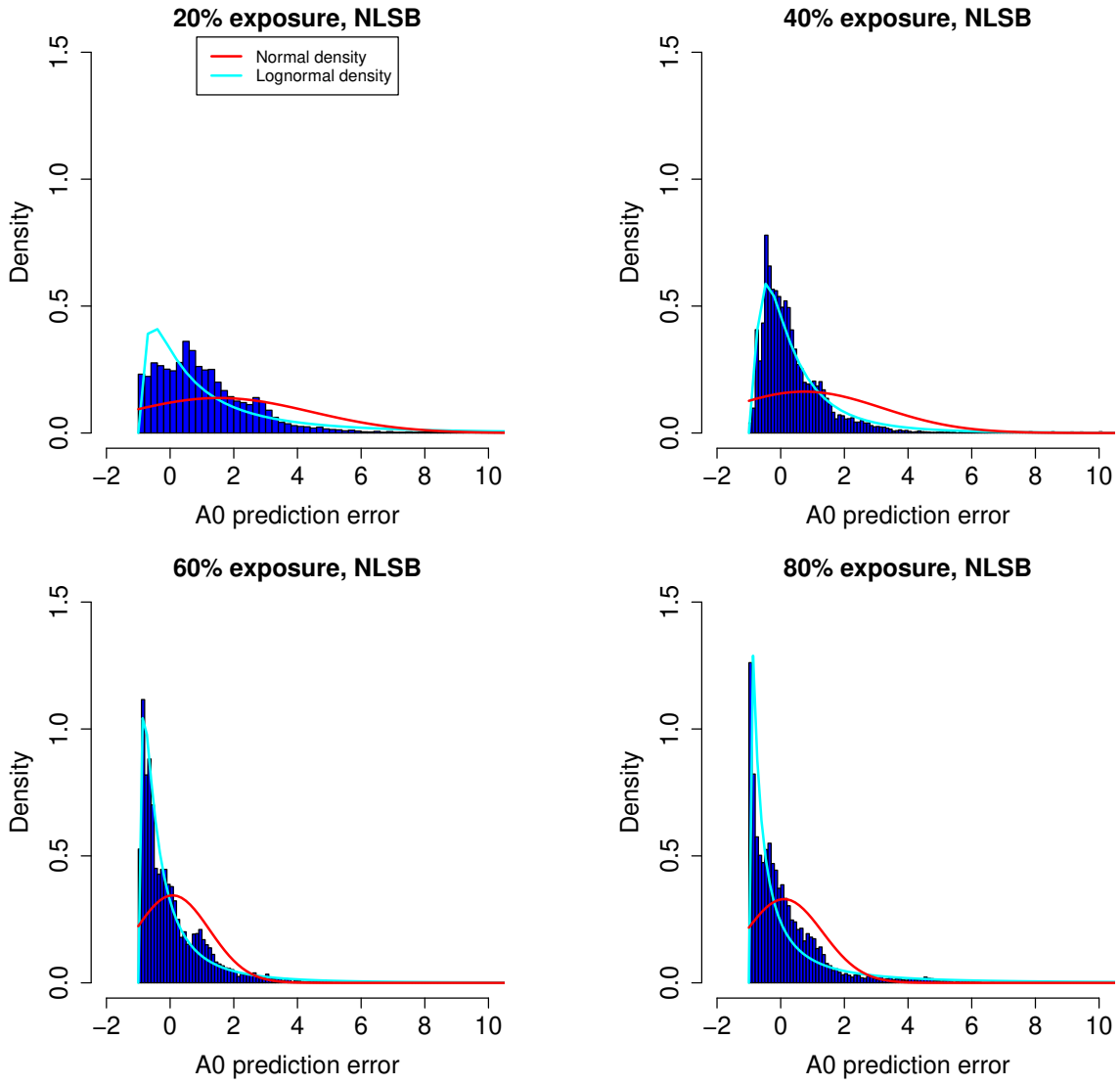

Figure 2.13: Histograms of ε_{A_0} values calculated with respect to the NLSB method, over all 30 study reaches.

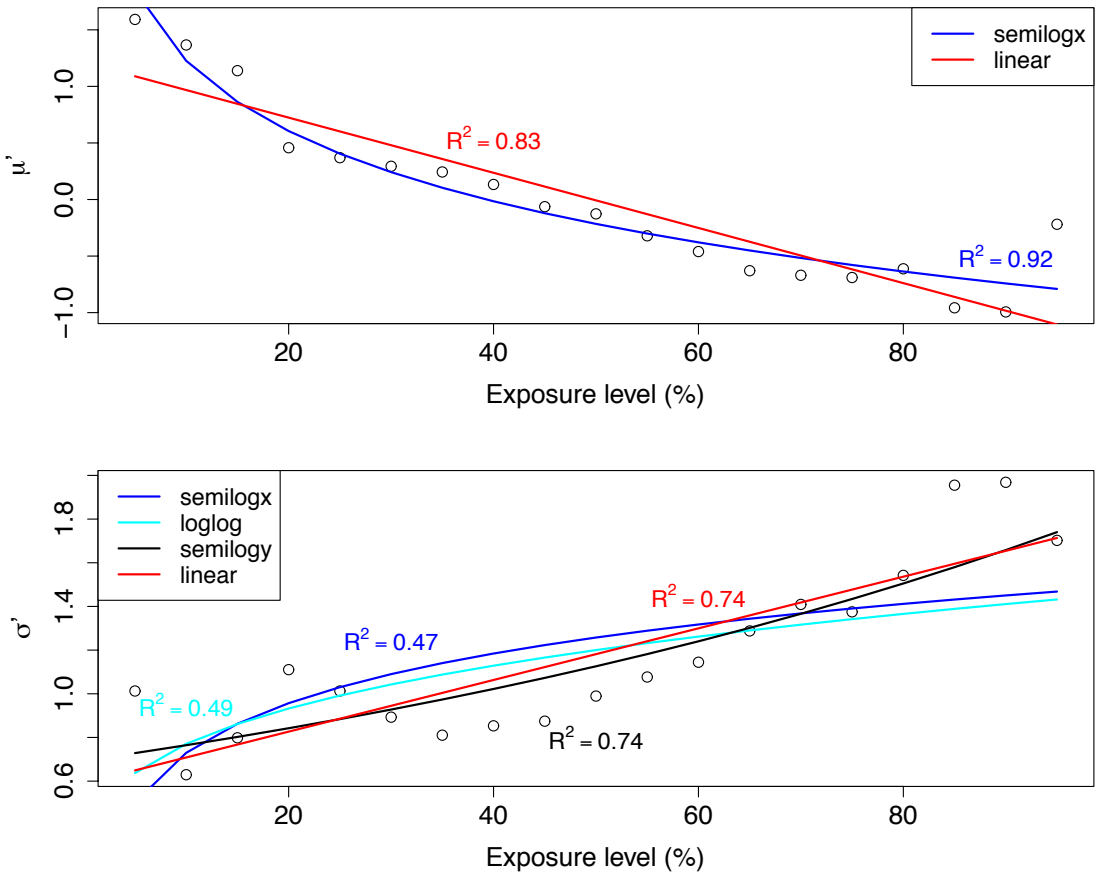

Figure 2.14: Comparison of different parameterizations for lognormal distribution parameters for describing the ε_{A_0} distribution. Each data point represents a parameter value for the distribution of prediction errors from all 30 study reaches at a particular exposure level.

2.6.6 Average cross sections for the 30 study reaches

Of the 51, 10-km long Upper Mississippi River reaches defined for this study, we tested our curve-fitting methods on a subset of 30 non-braided reaches. Figures S6-S15 show the average cross section geometry and height-width relationships for each reach. Reaches are identified by the Upper Mississippi River pool in which they are located. The dotted red lines show the breakpoint location defined for the nonlinear slope break (NLSB) method.

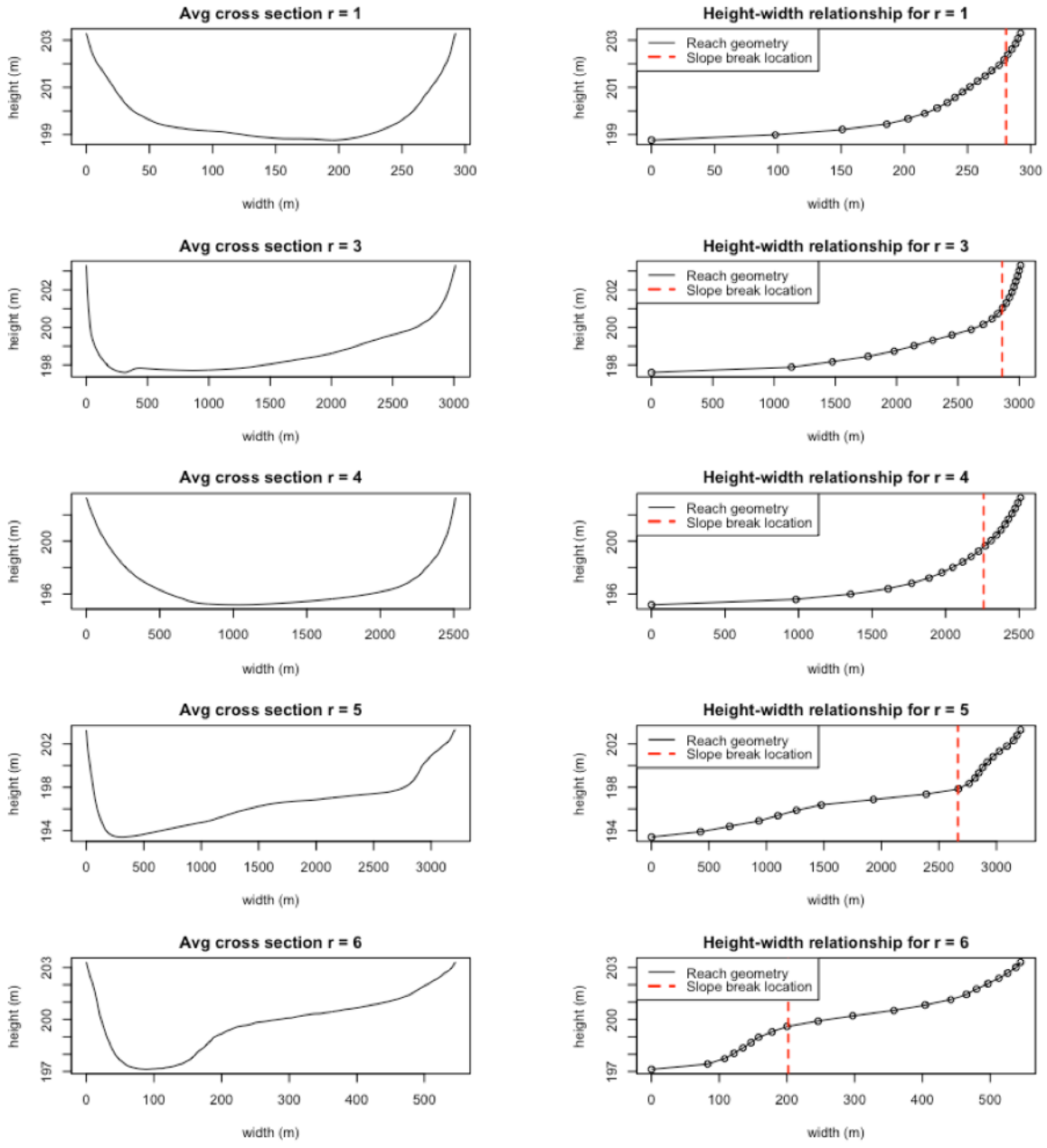

Figure 2.15: Average cross section geometry for non-braided reaches in Pool 4.

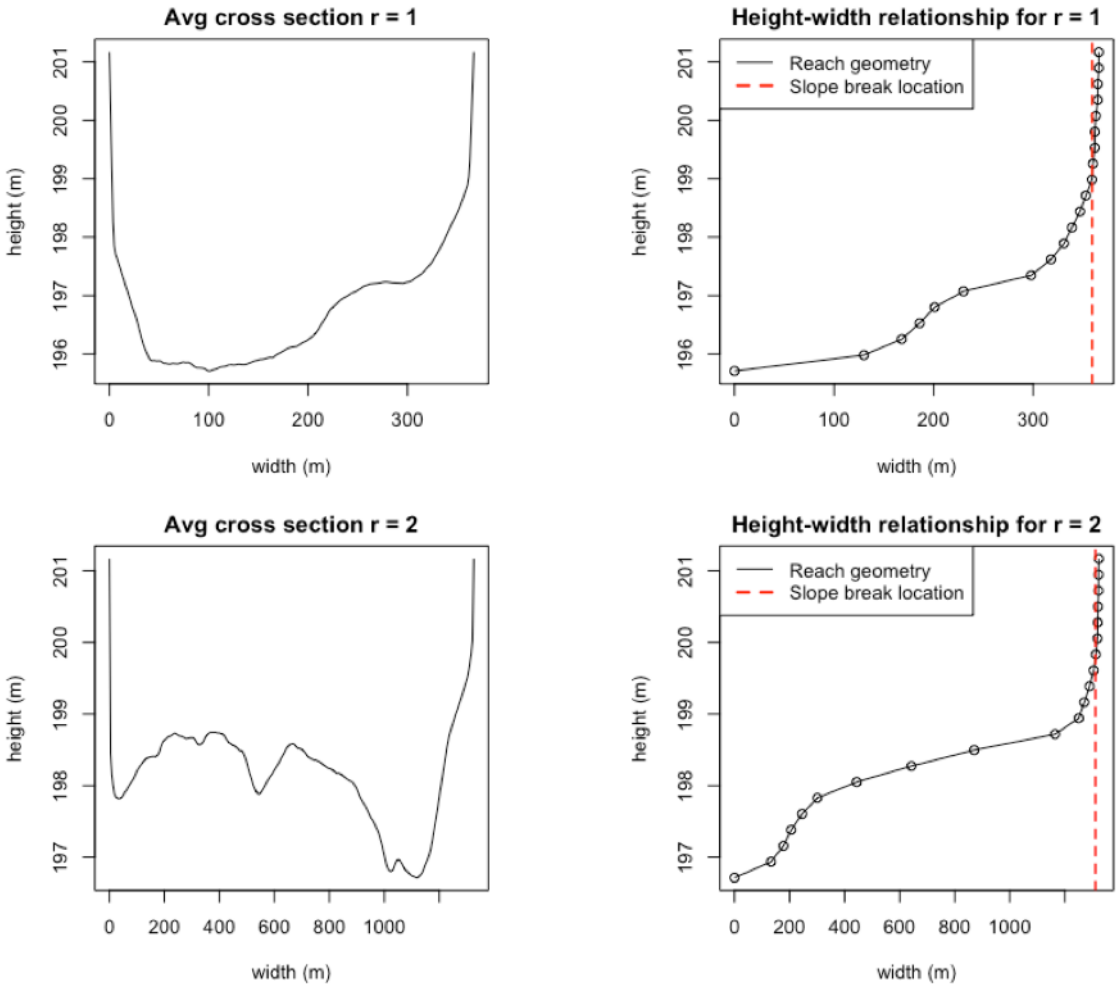

Figure 2.16: Average cross section geometry for non-braided reaches in Pool 5.

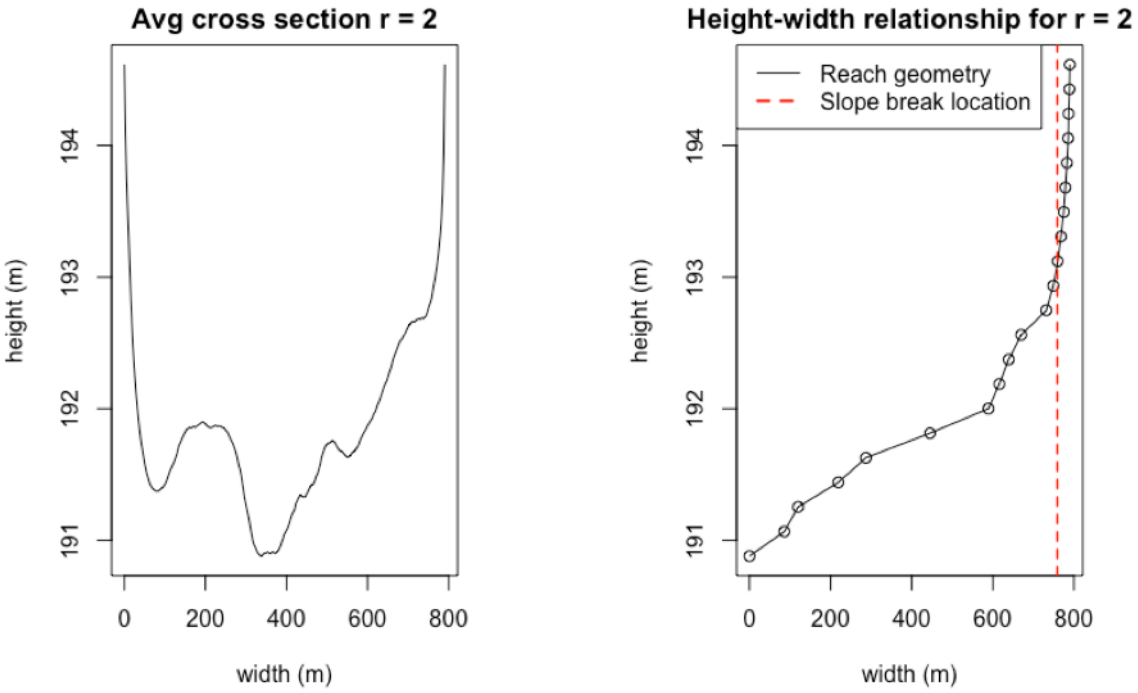

Figure 2.17: Average cross section geometry for non-braided reaches in Pool 7.

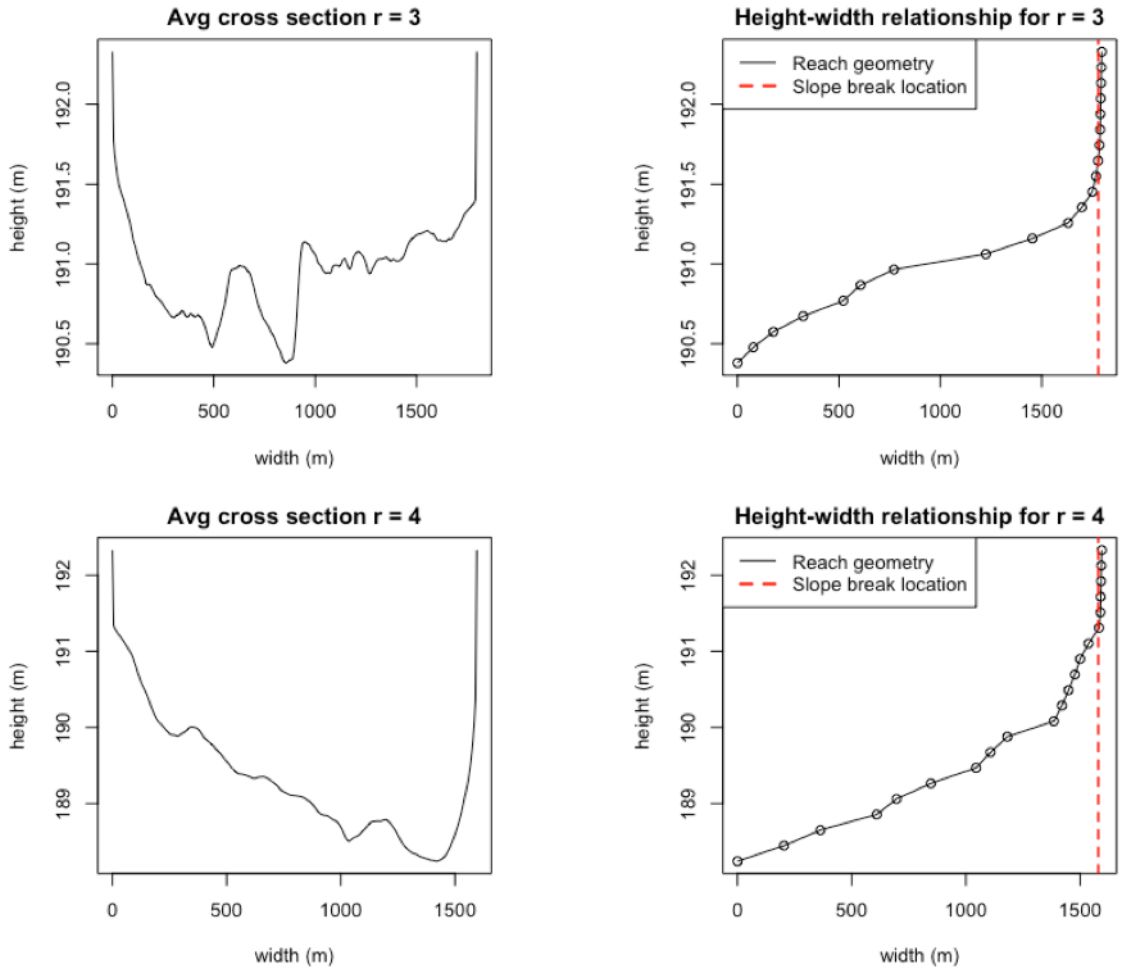

Figure 2.18: Average cross section geometry for non-braided reaches in Pool 8.

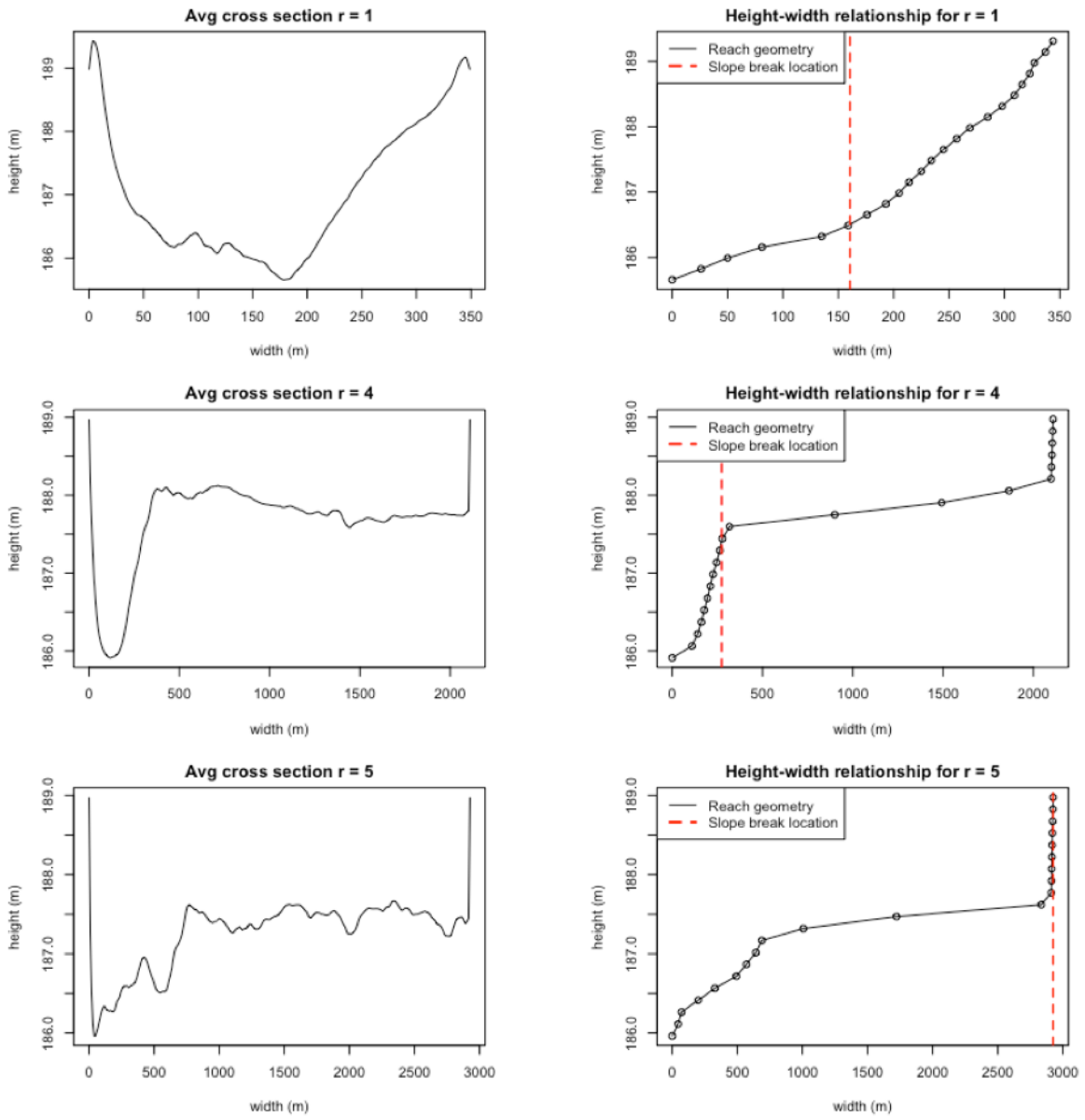

Figure 2.19: Average cross section geometry for non-braided reaches in Pool 9.

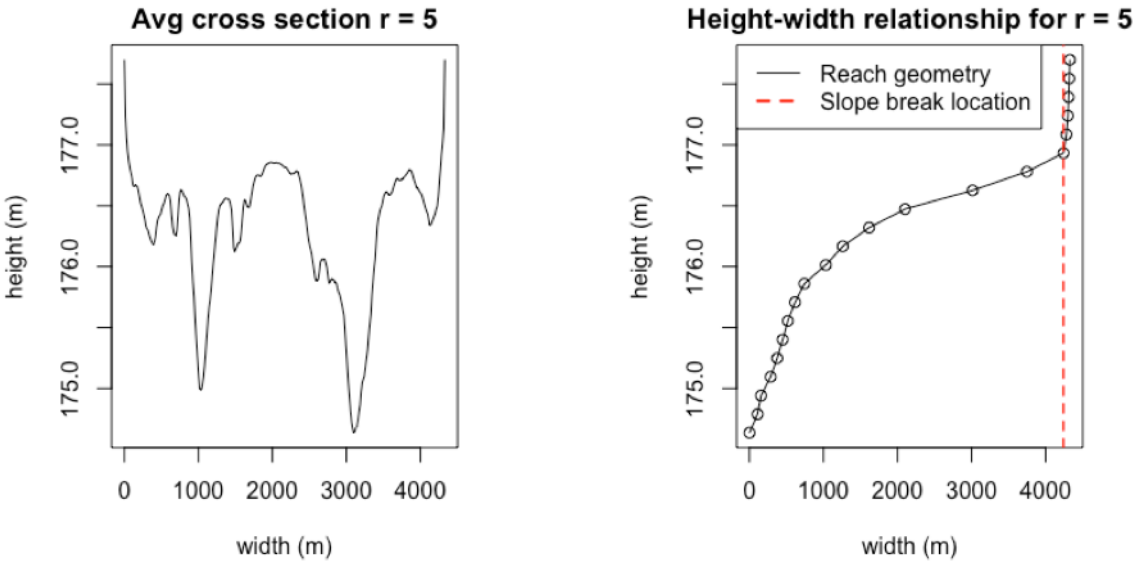

Figure 2.20: Average cross section geometry for non-braided reaches in Pool 13.

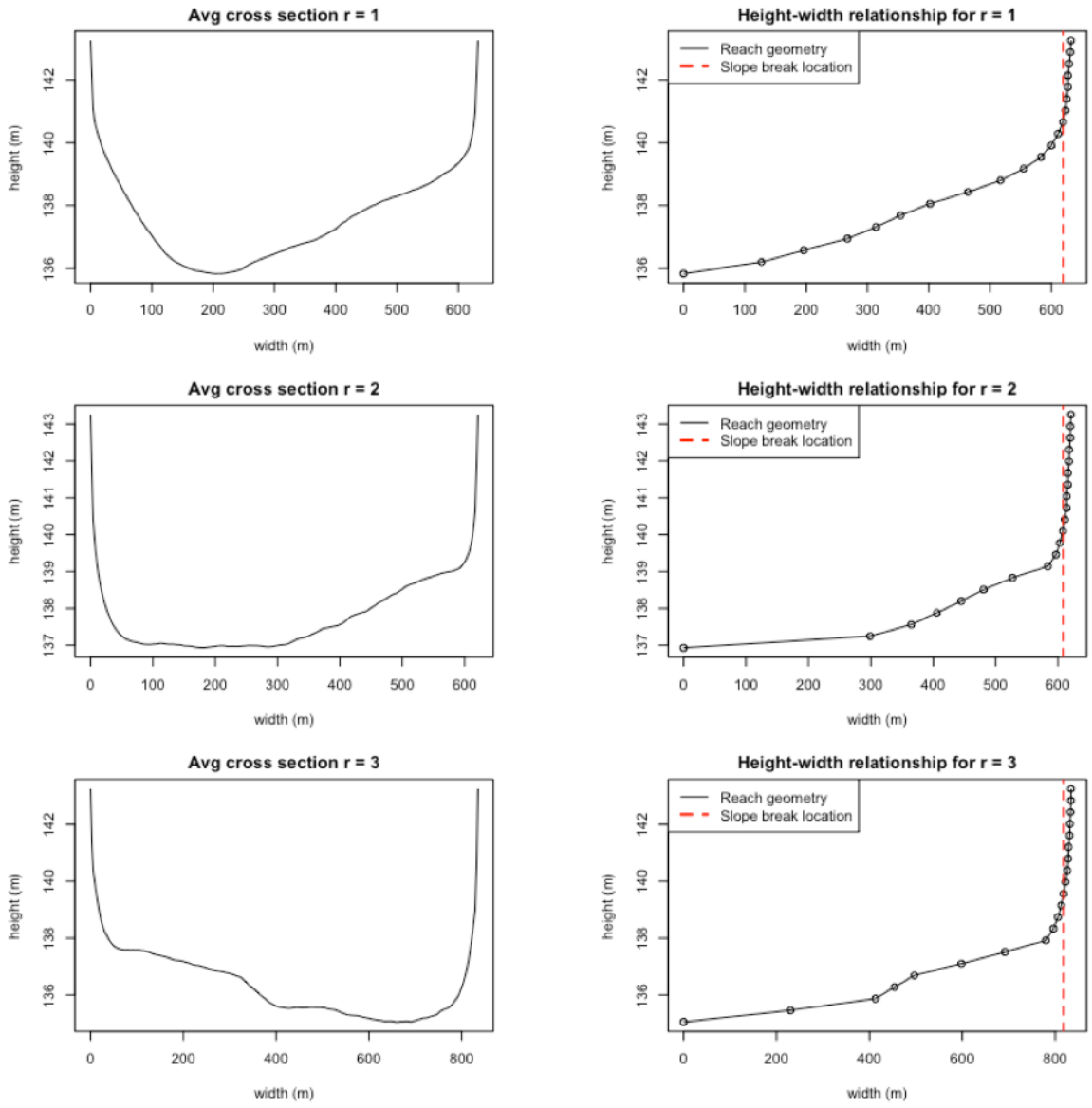

Figure 2.21: Average cross section geometry for non-braided reaches in Pool 21.

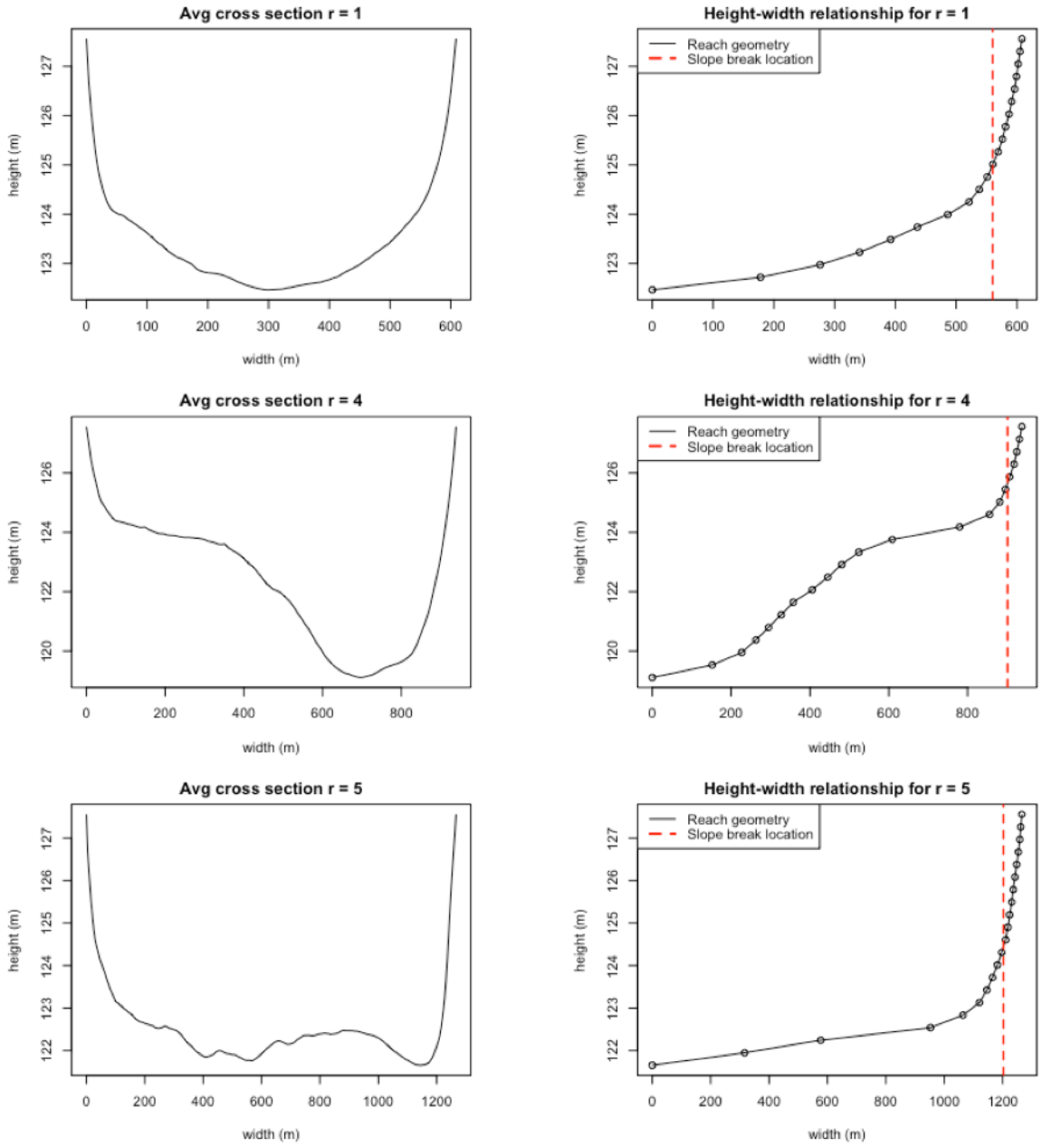

Figure 2.22: Average cross section geometry for non-braided reaches in Pool 26.

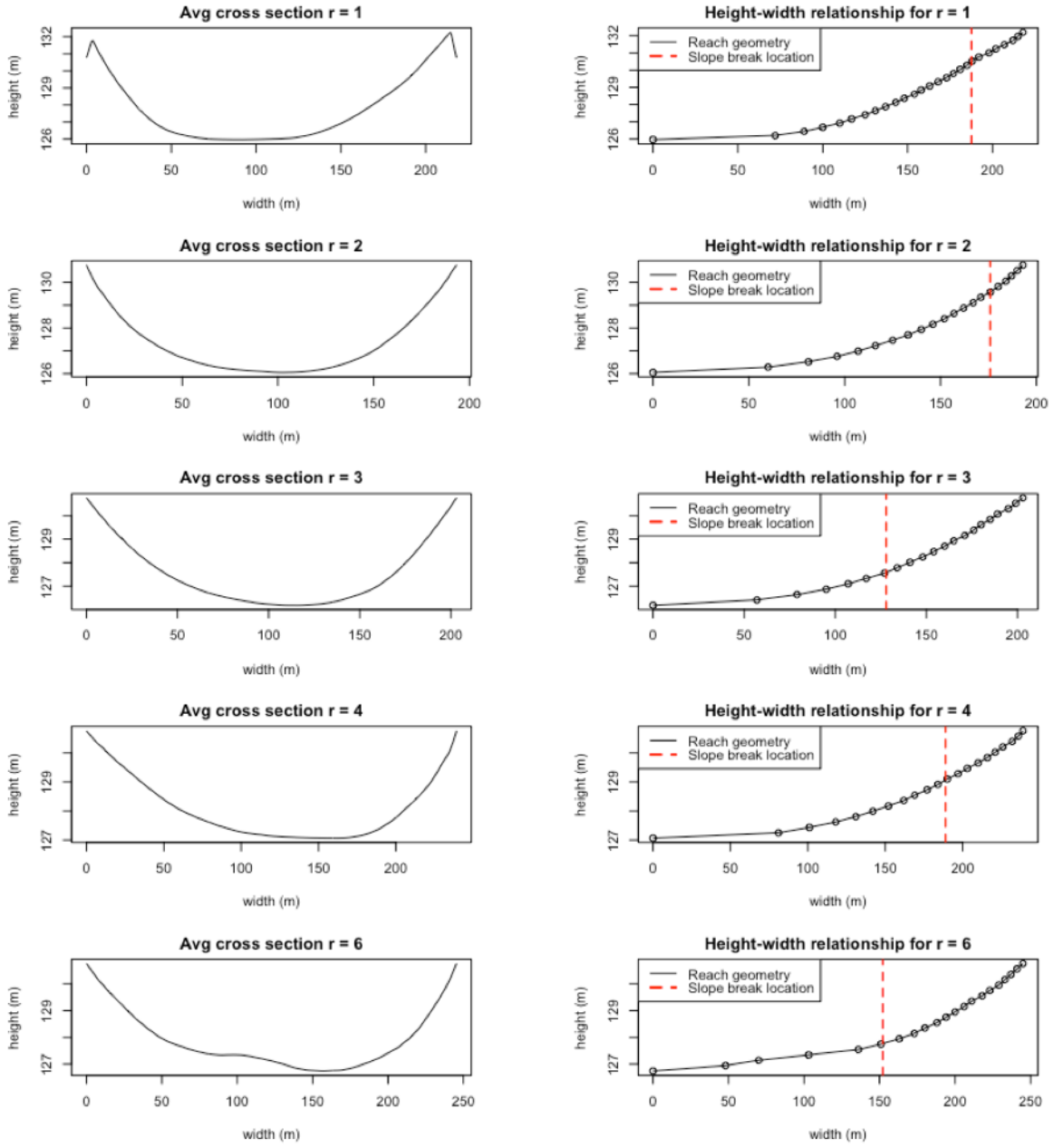

Figure 2.23: Average cross section geometry for non-braided reaches in the La Grange pool (split over two figures).

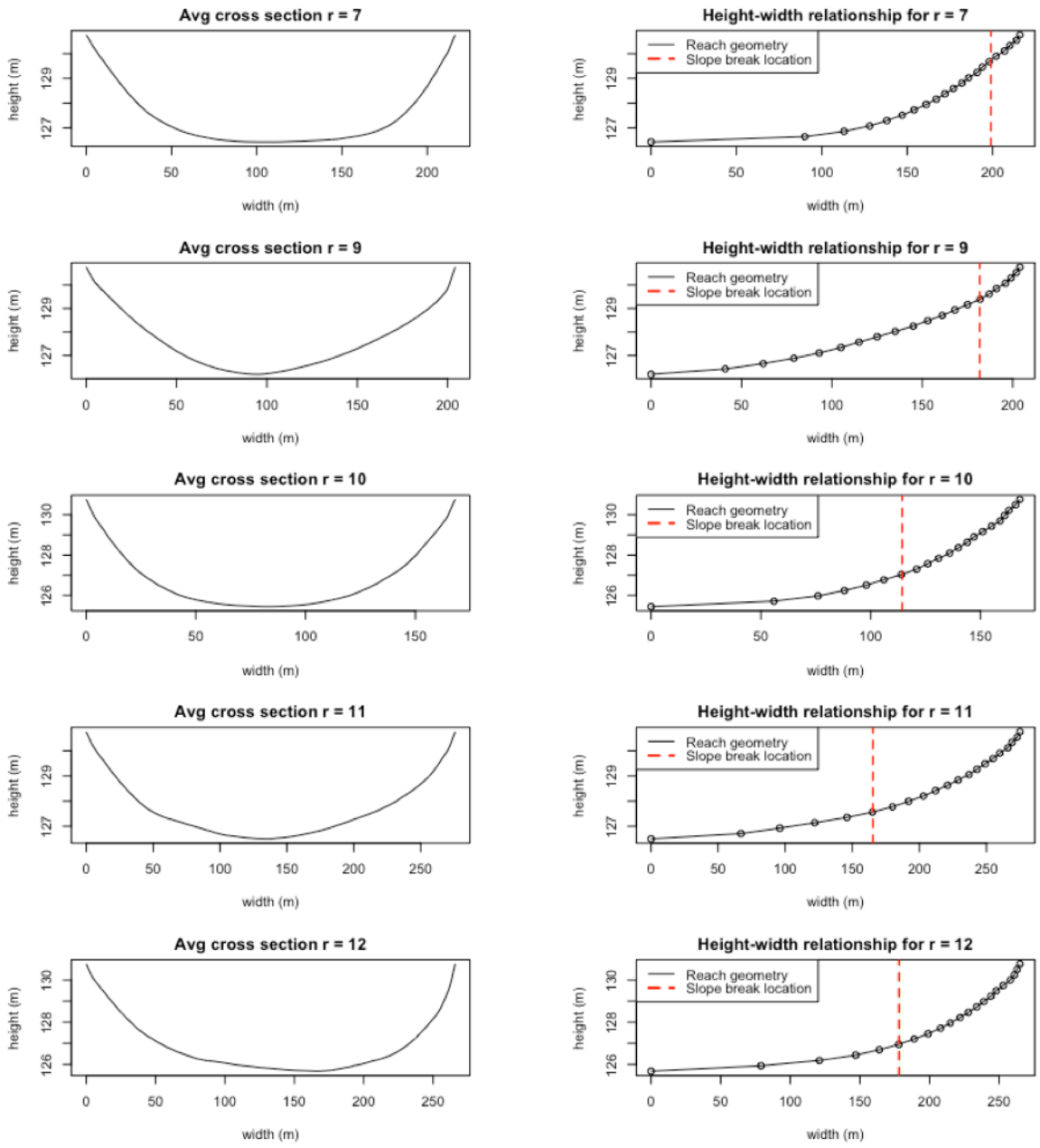

Figure 2.24: Average cross section geometry for non-braided reaches in the La Grange pool (split over two figures).

CHAPTER 3

A near-global, high resolution land surface parameter dataset for the variable infiltration capacity model

Hydrologic models predict the spatial and temporal distribution of water and energy at the land surface. Currently, parameter availability limits global-scale hydrologic modelling to very coarse resolution, hindering researchers from resolving fine-scale variability. With the aim of addressing this problem, we present a set of globally consistent soil and vegetation parameters for the Variable Infiltration Capacity (VIC) model at 1/16° resolution (approximately 6 km at the equator), with spatial coverage from 60°S to 85°N. Soil parameters derived from interpolated soil profiles and vegetation parameters estimated from space-based MODIS measurements have been compiled into input files for both the Classic and Image drivers of the VIC model, version 5. Geographical subsetting codes are provided, as well. Our dataset provides all necessary land surface parameters to run the VIC model at regional to global scale. We evaluate VICGlobal's ability to simulate the water balance in the Upper Colorado River basin and 12 smaller basins in the CONUS, and their ability to simulate the radiation budget at six SURFRAD stations in the CONUS.

3.1 Background and Summary

The Variable Infiltration Capacity (VIC, `https://github.com/UW-Hydro/VIC`) model is a macroscale, semi-distributed hydrologic model ([Hamman et al., 2018b];[Hamman et al., 2018a]; [Liang et al., 1994]) that calculates land surface states and fluxes by solving the surface water

and energy balances. The model has a wide user base — the citation index Web of Science shows the original VIC paper ([Liang et al., 1994]) has been cited nearly 2000 times, with contributing authors from at least 56 countries. Despite the model's popularity, there are only a few ready-made soil and vegetation parameter datasets that modelers can use to run VIC outside the continental United States. Previous global input datasets ([Nijssen et al., 2001]; [Su et al., 2005]; [Lin et al., 2019]; [Zhou et al., 2016]; [Adam et al., 2009]) have been compiled for VIC at resolutions ranging from 2° to 1/4°. Many studies, including [Su et al., 2005], [Zhou et al., 2016], and [Adam et al., 2009] use parameters based on the 2° soil and vegetation parameters developed by [Nijssen et al., 2001] (henceforth N2001). As useful as the N2001 dataset has been over the years, the VIC-modeling community would be well-served by a higher-resolution update. The N2001 dataset and its derivatives are limited by the dataset's coarse resolution, geographically sparse subset of leaf-area index observations, and assumptions of temporally-invariant albedo and 100 percent canopy coverage for all land cover classes, as noted by [Bohn and Vivoni, 2019], who developed a new VIC parameter dataset for North America that addresses these issues. Our dataset, VICGlobal, emulates their approach at a global-scale.

VICGlobal's predecessor, the N2001 soil and vegetation parameters, may be appropriate for continental-scale modelling, but its coarse resolution makes it less useful for parameterizing VIC at smaller scales. Coarse resolution land surface models miss topographic variability, distort river networks, and prevent proper representation of land-atmosphere interactions in coupled land-atmosphere models. At coarse resolution, topographic characteristics such as elevation and vegetation cover are averaged over a large grid cell, so the model will miss key details such as the effect of terrain on the radiation balance and the effect of vegetation on ET-soil moisture partitioning. This is particularly important in mountainous regions, where there are large changes in topography across relatively small areas. While VIC does not represent fluxes from one grid cell to another, it is frequently coupled to a routing model to simulate how runoff flows between grid cells. At very coarse resolutions, the modelled

river network loses its resemblance to the true river network, necessitating upscaling algorithms to obtain usable coarse-resolution river networks (e.g. [Wu et al., 2012]). Finally, high resolution land surface modelling could improve our ability to simulate land-atmosphere interactions that occur over relatively small spatial scales ([Wood et al., 2011]). With 1/16° grid cells and representation of up to 17 land cover classes within each grid cell, VICGlobal is a step toward addressing each of these resolution-related challenges.

Regional-scale VIC inputs at 1/16° resolution already exist but have limited coverage outside North America. [Livneh et al., 2013] (henceforth L2013) set up the VIC model over the conterminous United States (CONUS) using soil and vegetation data compiled from sources including the Food and Agriculture Organization (FAO)/UNESCO Soil Map of the World and the Advanced Very High Resolution Radiometer (AVHRR). The L2013 VIC parameterization is based on that of [Maurer et al., 2002], with calibration for a better match with streamflow data. [Bohn and Vivoni, 2019] (henceforth BV2019) released an updated 1/16° vegetation parameter dataset for the CONUS, Mexico, and part of Canada, improving on of the limitations of the L2013 dataset, such as its assumption of temporally-invariant albedo. They estimated time-varying albedo, leaf-area index (LAI), and fractional canopy cover using observations from the Moderate Resolution Imaging Spectroradiometer (MODIS).

Drawing on the L2013 and BV2019 VIC parameterizations, we developed VICGlobal, a near-global dataset of soil and vegetation parameters for the VIC model at 1/16° resolution, which VIC users can download and subset to their region of study. We estimated soil parameters based on the 30 arc-second FAO Harmonized World Soil Database (HWSD; [Nachtergaele et al., 2012]). Vegetation parameters are based on 500 m resolution MODIS observations. VICGlobal includes all the necessary parameters to run regional- to global-scale VIC simulations. We provide MATLAB® codes to subset the VICGlobal parameters to a particular domain. In addition to parameters, meteorological forcing data are required to run VIC. We do not include meteorological forcing data as part of VICGlobal. Instead, we direct readers to existing forcing datasets with near-global coverage, such as the re-

analysis datasets MERRA-2 ([Gelaro et al., 2017]) and GLDAS ([Rodell et al., 2004]), or real-time measurement-based datasets — see e.g. [Xiao et al., 2016]; [Livneh et al., 2013]; [Livneh et al., 2015]; and [Bohn et al., 2019].

Finally, a note on the file format: The upgrade from VIC version 4 (VIC-4) to VIC version 5 (VIC-5) introduced two "drivers" for running the model. The Image driver takes NetCDF files as inputs, while the Classic driver takes ASCII text files. The VICGlobal parameter files are available in two formats: one for VIC-5 Classic, and one for VIC-5 Image.

3.2 Methods

This section describes how we used freely-available data to compile Classic driver input files for the VIC model. First, we created parameter files for VIC-5 Classic, then we converted them to NetCDF format for VIC-5 Image. VIC-5 Classic requires three parameter files: a soil parameter file, a vegetation parameter file, and a vegetation library file. An optional elevation band file can be provided to resolve sub-grid variability in elevation, which is important in regions with complex topography. The parameters are arranged as a relational database: each grid cell has a unique identifier, called a grid cell number, in the soil parameter file, that VIC uses to find the corresponding rows of data in the vegetation parameter and elevation band files. The Image driver uses a different setup, with all parameters stored in a single NetCDF file.

3.2.1 Soil parameters

The soil parameter file for VIC-5 Classic is an ASCII text file that includes soil parameters such as hydraulic conductivity and porosity, but also other kinds of static parameters, such as average precipitation and time zone offset from GMT. Each row of the soil parameter file represents one grid cell, and each column represents a different variable. We compiled the soil parameter file using MERIT ([Yamazaki et al., 2017]) elevation data, soil texture

data from the FAO HWSD, pedotransfer tables relating soil texture to other soil properties, and interpolated weather station data (WorldClim; [Fick and Hijmans, 2017]). Any remaining parameters were set to suggested values from the VIC model's documentation ([Hamman et al., 2018a]). The following sections describe the estimation of each variable in the soil parameter file, summarized in Table 3.1.

Table 3.1: VIC model parameters for the soil parameter file. Parameters whose source is "VIC documentation" were set to suggested values from the VIC documentation ([Hamman et al., 2018a]). Bold parameters have distinct values in each of the three soil layers. Sources: WorldClim ([Fick and Hijmans, 2017], HWSD ([Nachtergaele et al., 2012]), MERIT ([Yamazaki et al., 2017]). *values are calculated based on HWSD soil texture data, using the pedotransfer table.

Soil parameter	Description	Source
lat	Latitude	MERIT
lon	Longitude	MERIT
b_{infilt}	Variable infiltration capacity parameter	VIC documentation
ds	Fraction of Dsmax where nonlinear baseflow occurs	VIC documentation
dsmax	Maximum velocity of baseflow (mm/day)	HWSD*
ws	Soil moist. at which nonlinear baseflow starts	VIC documentation2
c	Exponent used in baseflow curve	VIC documentation
expt	Campbell's eqn. exponent for hydraulic conductivity	HWSD*
Ksat	Saturated hydraulic conductivity (mm/day)	HWSD*
$\mathbf{phi_s}$	Soil moisture diffusion parameter (not used)	Not used

initm	Initial moisture content (mm)	HWSD*
elev	Elevation (m)	MERIT
depth	Thickness of each soil layer (m)	HWSD
avg_T	Average soil temperature (°C)	WorldClim
dp	Soil thermal damping depth (m)	VIC documentation
bubble	Bubbling pressure of soil (cm)	HWSD*
quartz	Quartz content	HWSD*
bulk_density	Bulk density of soil (kg/m3)	HWSD
soil_density	Soil particle density (kg/m3)	VIC documentation
offgmt	Time zone offset relative to GMT (hr)	Based on longitude
$\mathbf{wcr_{fract}}$	Fractional soil moisture content at critical point	HWSD*
$\mathbf{wpwp_{fract}}$	Fractional soil moisture content at wilting point	HWSD*
Rough	Surface roughness length of bare soil (m)	VIC documentation
snow_rough	Surface roughness length of snowpack (m)	VIC documentation
annual_prec	Average annual precipitation (mm)	WorldClim
resid_moist	Residual moisture fraction	Assumed zero
fs_active	Flag for whether to run frozen soils module	
July_Tavg	Average July temperature (°C)	WorldClim

3.2.2 Elevation and land mask

The VICGlobal soil parameter file uses the Multi-Error-Removed Improved-Terrain (MERIT; [Yamazaki et al., 2017]) digital elevation model (DEM) to define the elevations, latitudes, and longitudes of each land grid cell. The MERIT DEM is an error-corrected and extended version of the SRTM DEM, with 3 arc-second resolution and coverage from 60°S to 85°N and 180°W to 18 °E. Specifically, MERIT is a combination of the SRTM, AW3D, and

Viewfinder Panoramas' DEMs, corrected for striping, speckle, absolute bias, and tree height bias. We used bilinear interpolation to aggregate MERIT to 1/16° resolution and derive a 1/16° MERIT-based land mask and DEM (Figure S1).

3.2.3 Soil texture data

Soil texture (percent sand, silt, and clay) and bulk density were obtained from the FAO HWSD, a gridded soil parameter dataset derived from in-situ measurements of the soil column. We used a 0.05° resolution NetCDF dataset converted from the original HWSD Microsoft Access database by [Wieder et al., 2014]. We resampled the HWSD soil data from 0.05° to 1/16° resolution using bilinear interpolation with the MATLAB® function griddedInterpolant. While HWSD has near global coverage, there are missing data in some places around the world, notably Greenland and northern Africa. We filled in these missing data using inpainting, a gap-filling method from the field of image processing. We used the MATLAB® function inpaintnans ([D'Errico, 2012]), which uses a partial differential equation method to fill in missing data, to fill gaps in the HWSD data over the MERIT land mask. Figure 3.1 shows the HWSD bulk density data before and after inpainting.

The HWSD data are divided into "topsoil" and "subsoil" parameters. The first 30 cm of the soil column are considered topsoil and the lower 70 cm subsoil. VIC is typically run with three soil layers, so we created a three-layer soil parameter file by breaking up the 30 cm HWSD topsoil layer into two soil layers: one of 10 cm and one of 20 cm, so the final soil parameter file has three layers, with thicknesses of 10 cm, 20 cm and 70 cm, from top to bottom of the soil column. Ten centimeters has been a common choice for the uppermost layer soil depth in VIC modeling applications since its use by [Liang et al., 1996]. Soil layer depths are typically used as calibration parameters. VICGlobal values should be considered a starting estimate.

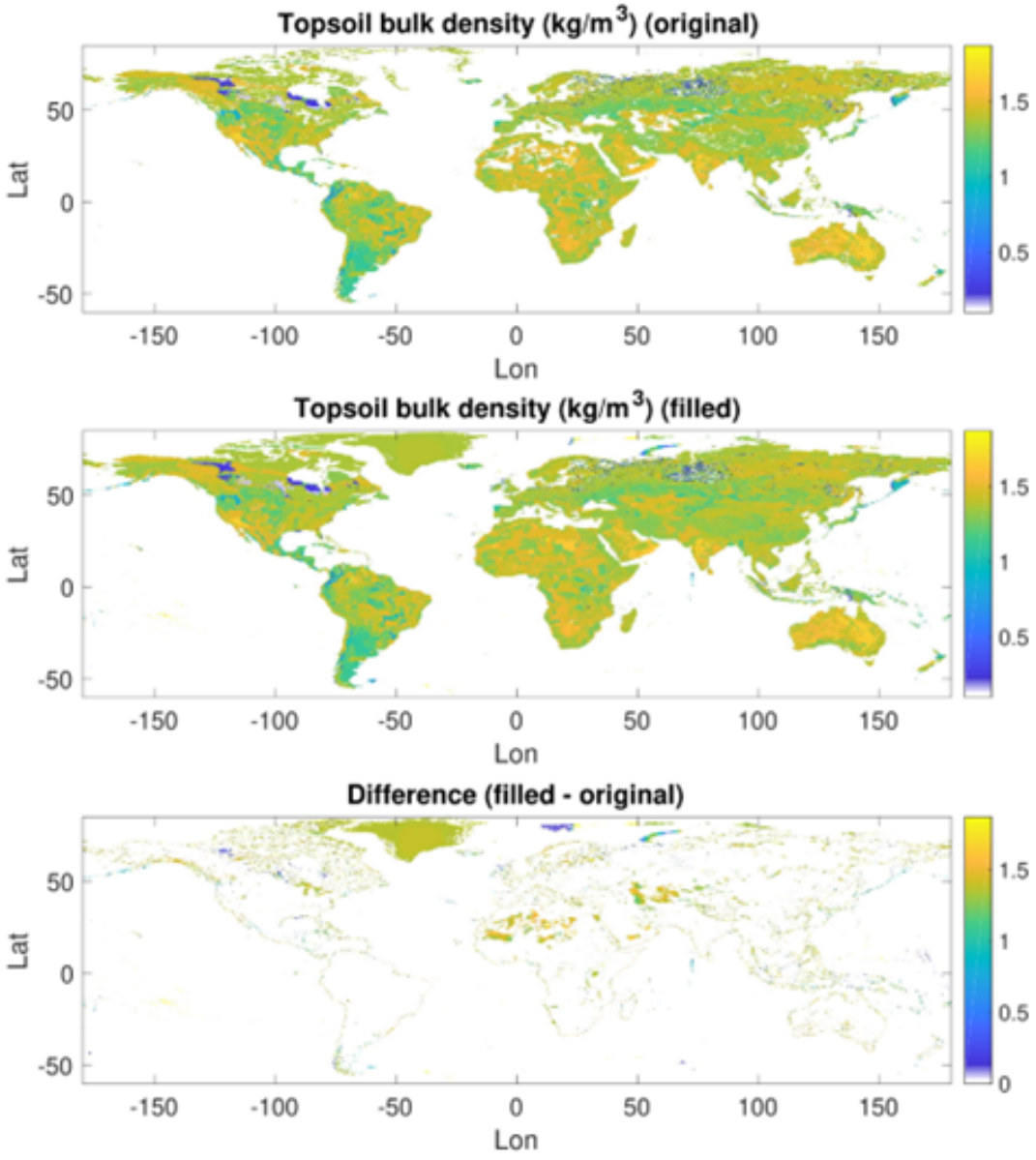

Figure 3.1: Bulk density data from the Harmonized World Soils Database (HWSD). The top panel shows HWSD bulk density data resampled to 1/16° resolution, the middle panel shows bulk density after infilling holes, and the bottom panel shows the difference.

3.2.4 Calculating soil parameter values based on soil textures

Pedotransfer functions (e.g. [Cosby et al., 1984]) relate readily available soil properties, such as soil texture, to less easily-observed properties, such as hydraulic conductivity. After re-sampling the HWSD data from 1/4° to 1/16° resolution, we estimated soil parameters by classifying each grid cell's USDA soil texture class and assigning physical soil properties based on a lookup table included with the VIC documentation ([Hamman et al., 2018a]; [Schaake, 2000]). The lookup table (Table 3.2.4) relates the 12 USDA soil texture classes to bulk density, field capacity, wilting point, porosity, saturated hydraulic conductivity, and slope of the soil water retention curve in Campbell's equation. We classified soil textures using the USDA soil texture triangle, as implemented by the MATLAB® function soil_classification ([Hoffman, 2016]). Figure 3.2 shows the derived USDA soil texture map. We used these along with the lookup table to estimate saturated hydraulic conductivity (K_{sat}), the exponent in Campbell's equation for hydraulic conductivity ($expt$), fractional soil moisture at the critical point (wcr_{fract}), where the critical point is about 70% of field capacity, fractional soil moisture at the wilting point ($wpwp_{fract}$), quartz content, and porosity for each soil layer. The lookup table ([Schaake, 2000]) did not include quartz content, so we supplemented it with the soil texture-quartz content lookup table from Peters-Lidard et al. ([Peters-Lidard et al., 1998]).

USDA class	Field capacity	Wilting point	Porosity	K_{sat} *(cm/hr)*	b	Quartz content
Sand	0.08	0.03	0.43	38.41	4.1	0.95
Loamy sand	0.15	0.06	0.42	10.87	3.99	0.85
Sandy loam	0.21	0.09	0.4	5.24	4.84	0.69
Loam	0.32	0.12	0.46	3.96	3.79	0.19
Silt loam	0.28	0.08	0.52	8.59	3.05	0.05
Silt	0.29	0.14	0.43	1.97	5.3	0.41
Sandy clay loam	0.27	0.17	0.39	2.4	8.66	0.61
Clay loam	0.36	0.21	0.48	4.57	7.48	0.09
Silty clay loam	0.34	0.21	0.46	1.77	8.02	0.25
Sandy clay	0.31	0.23	0.41	1.19	13	0.5
Silty clay	0.37	0.25	0.49	2.95	9.76	0.08
Clay	0.36	0.27	0.47	3.18	12.28	0.25

Table 3.2: USDA soil texture class lookup table. Field capacity, wilting point, porosity, saturated hydraulic conductivity (K_{sat}), and b are taken from a lookup table ([Schaake, 2000]), which is provided with the VIC version 4 software documentation. b is the slope of the retention curve, in log space.

We set the variable infiltration capacity parameter $b_{infilt} = 0.2$, the maximum baseflow fraction threshold $d_s = 0.001$, and maximum soil moisture threshold $w_s = 0.9$, their suggested values from the VIC documentation. These parameters, along with maximum baseflow velocity ($dsmax$) and soil depth, are typically calibrated. We set the baseflow curve exponent c=2, the soil thermal damping depth $dp = 4$ m, soil density = 2685 kg/m^3, surface roughness = 0.001 m, and snow roughness = 0.0005 m, also based on guidance from the VIC documentation. The soil moisture diffusion parameter phi_s is not used in the current version of VIC, so we set it to the no-data value (-999). The final few soil parameters — $dsmax$, initial soil moisture ($initm$), and bubbling pressure ($bubble$)— were calculated using the following equations, based on guidance from the VIC documentation.

$$dsmax = slope * \overline{K}_{sat} \tag{3.1}$$

$$initm = wcr_{fract} * porosity * t_l \tag{3.2}$$

$$bubble = 0.32 * expt + 4.3 \tag{3.3}$$

Equation (3.1) estimates $dsmax$ for each grid cell as the product of soil-column average K_{sat} and land surface slope, which was calculated from the elevation data using the MATLAB® function *gradientm* ([MathWorks, 2019]). Equation (3.2), where t_l is the thickness of soil layer l, assumes that initial soil moisture is equal to the fractional soil moisture content at the critical point. Equation (3.3) calculates bubbling pressure as a function of $expt$, based on linear regression of bubbling pressure vs. $expt$ ([Rawls, 1993]). Figures S2-S9 in the Supplementary Information show maps of each soil parameter. We assumed residual soil moisture, the amount of soil moisture that cannot be removed from the soil by drainage or evapotranspiration, was zero.

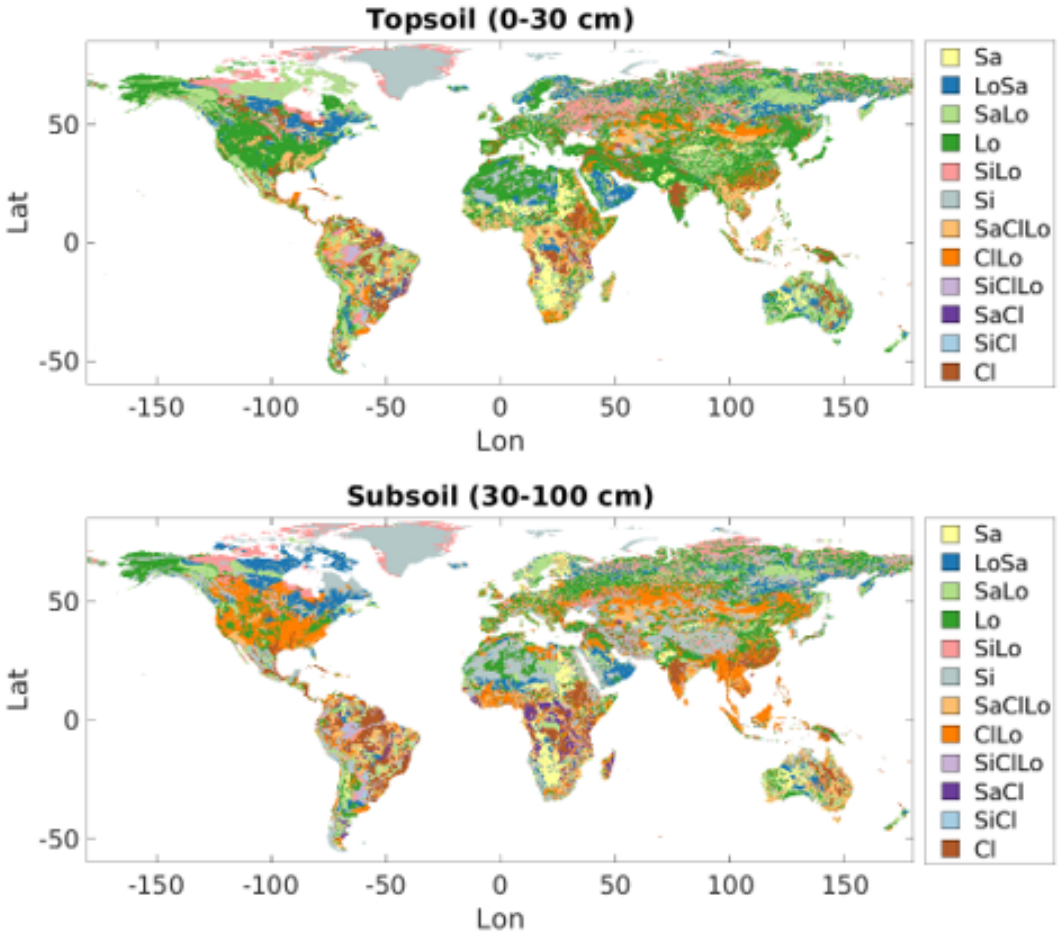

Figure 3.2: USDA soil texture classifications based on HWSD. Topsoil is soil from 0-30 cm below the surface, and subsoil is soil between 30-100 cm deep.

3.2.5 Elevation bands

VIC uses an elevation band file (also called a snow band file) to account for subgrid heterogeneity in grid cell elevations. The assumption of uniform elevation over an entire grid cell can lead to modeling errors in mountainous regions, where higher topography is associated with cooler temperatures and higher precipitation rates. The elevation band file accounts for subgrid variability in topography by dividing each grid cell into a number of elevation bands, each of which is simulated separately. VIC adjusts temperature, pressure, and precipitation depending on the elevation in each band. We prepared an elevation band file with five elevation bands by comparing the 1/16° DEM used for the soil parameter file with a 30 arc-second DEM. Both DEMs were derived by aggregating MERIT data. For simplicity, we assumed precipitation was evenly distributed among elevation bands within a grid cell. The elevation band file is provided with the caveat that using elevation bands requires more computing power; users may wish to turn elevation bands on or off (via the VIC global parameter file) depending on their needs.

3.2.6 Vegetation parameters

VIC-5 Classic uses a vegetation parameter file to define the fractional cover of different vegetation types within each grid cell and some of their physical properties. Other vegetation parameters are stored in the "vegetation library" file. (VIC-5 Image simply stores all parameters in a single "parameter" file.) The VIC-5 Classic vegetation parameter file consists of information about fractional cover of each land cover type in each grid cell, and their corresponding root zone depths and root fractions within each root zone. The vegetation parameter file can optionally include time-varying LAI, fractional canopy cover, and albedo data, but it is simpler to specify these in the vegetation library (at the cost of not representing some spatial heterogeneity).

We used MODIS land cover data from the 0.05° MODIS MCD12C1 Collection 6 data

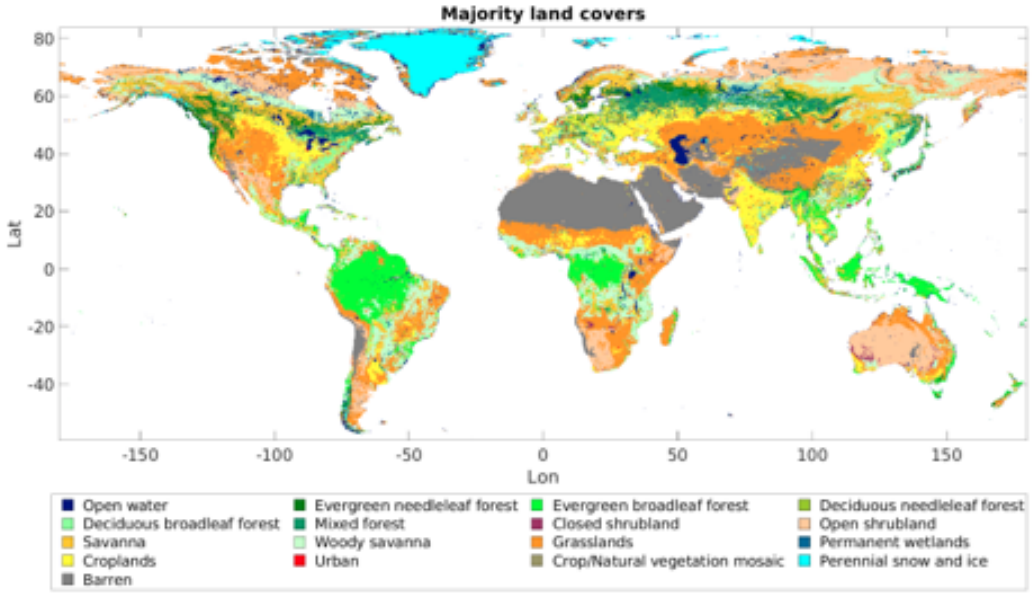

Figure 3.3: MODIS MCD12C1 majority land cover types (IGBP classifications).

product ([Friedl and Sulla-Menashe, 2015]) to assign fractional land cover values to each grid cell by calculating the average land cover for MCD12C1 observations over the 2017 calendar year. We chose 2017 because it was the most recent year with data in all the MODIS-based datasets used for this study, and there is very low interannual variability of land cover ([Sulla-Menashe et al., 2019]) in MCD12C1 Collection 6. Figure 3.3 shows majority land cover types from the 2017 MCD12C1 observations.

Like all global land cover data products, MCD12C1 makes classification errors. Sulla-Menashe et al. ([Sulla-Menashe et al., 2019]) reported 67% overall IGBP classification accuracy for 2001 land cover. Classification errors are more common in the "mixed" land covers, such as cropland/natural vegetation mosaic, shrublands, grasslands, and savannas. Fortunately for our purposes, the vegetation parameters for commonly-confused land covers tend to be fairly similar themselves, which reduces the impact of misclassification on land surface modelling results. For example, the LAI of open shrubland is not too different from the LAI of closed shrubland.

We calculated root fraction as a function of land cover class following the method of [Zeng, 2001], who defined the following formula (Equation 3.4) for use in parameterizing land surface models:

$$Y = 1 - \frac{1}{2}\left(e^{-ad} + e^{-bd}\right) \tag{3.4}$$

where Y = cumulative root fraction, d = depth, and a and b are empirical parameters defined by ([Zeng, 2001]) for each International Geosphere–Biosphere Programme (IGBP) land cover type, based on a rooting depth database compiled from more than 200 field surveys. We used this formula with depths of 0.1 m, 0.7 m, and dr, corresponding to three root zones. The value of dr, the maximum rooting depth for each IGBP land cover type, was taken from [Zeng, 2001]. This method assumes that the depth and distribution of roots depends only on the land cover type; we assume that land cover type is the primary control on root characteristics. Table 3.3 shows root fractions and root zone depths for each IGBP land cover type.

	Depth(1)	Depth(2)	Depth(3)	Fract(1)	Fract(2)	Fract(3)
Open water	0.1	0.6	0.8	0.44	0.45	0.11
Evergreen needleleaf forest	0.1	0.6	1.1	0.34	0.51	0.14
Evergreen broadleaf forest	0.1	0.6	2.3	0.32	0.44	0.23
Deciduous needleleaf forest	0.1	0.6	1.3	0.34	0.5	0.16
Deciduous broadleaf forest	0.1	0.6	1.3	0.31	0.52	0.17
Mixed forest	0.1	0.6	1.7	0.25	0.52	0.22
Closed shrublands	0.1	0.6	1.8	0.31	0.49	0.21
Open shrublands	0.1	0.6	2.4	0.33	0.43	0.24
Savanna	0.1	0.6	1.7	0.36	0.45	0.19
Woody savanna	0.1	0.6	1	0.37	0.5	0.13
Grasslands	0.1	0.6	0.8	0.44	0.45	0.11
Permanent wetlands	0.1	0.6	0.8	0.44	0.45	0.11
Cropland	0.1	0.6	0.8	0.33	0.55	0.12
Urban	0.1	0.6	0.8	0.44	0.45	0.11
Cropland/natural vegetation mosaic	0.1	0.6	0.8	0.33	0.55	0.12
Permanent snow and ice	0.1	0.6	0.8	0.44	0.45	0.11
Barren	0.1	0.6	3.3	0.22	0.46	0.31

Table 3.3: Root zone depths (m) and fraction of roots in each zone for IGBP land cover classes.

Like previous large-scale VIC vegetation cover datasets, our vegetation parameter file neglects land cover change over time. However, it does have a few other advantages over past vegetation parameter datasets. The land cover classification used in the N2001 and L2013 VIC parameter sets is referred to as "UMD-NLDAS" because it is a modified version of the AVHRR-based University of Maryland (UMD) land cover product ([Hansen et al., 2000]). The UMD-NLDAS classification was modified for the North American Land Data Assimilation project (NLDAS; [Mitchell et al., 2004]) to exclude open water, urban, and snow and ice land cover classes (see BV2019). VICGlobal uses 17 IGBP land cover classes, including urban, barren, perennial snow and ice, and inland water bodies, permitting better description of land cover variability than the 11 UMD-NLDAS classification.

3.2.7 Vegetation library file

The vegetation library maps each land cover type to a set of vegetation parameters (Table 3.4). We adapted the LDAS vegetation library ([Cherkauer, 1999]) for use with the 17 IGBP land cover classes, taking monthly average LAI, fractional canopy cover ($fcanopy$), and albedo values obtained from recent MODIS data products. We set architectural resistance (r_0) and minimum stomatal resistance (r_{min}) to values from literature (described below). The rest of the parameters, which are described in the N2001 paper, were left to their original LDAS vegetation library values. This section describes how we estimated LAI, $fcanopy$, albedo, r_0, and r_{min}, and how we transferred the remaining parameters from the 11 UMD-NLDAS land cover classes to the 17 IGBP land cover classes.

Parameter	Description	Source
Overstory	Flag for whether the land cover type has an overstory	LDAS
R_0 *(s/m)*	Architectural resistance	SECHIBA
R_{min} *(s/m)*	Minimum stomatal resistance	SiB
LAI	Monthly average leaf-area index	2017 GLASS LAI
Canopy cover fraction	Monthly average partial vegetation cover fraction	2017 MODIS NDVI
Albedo	Monthly average albedo	2017 GLASS albedo
Roughness length (m)	Average roughness length	LDAS)
Displacement height (m)	Average displacement height	LDAS
WindH	Wind measurement height	LDAS
RGL	Minimum incoming shortwave radiation for transpiration to occur	LDAS
Solar attenuation	Radiation attenuation factor	LDAS, set to 0.5
Wind attenuation	Wind attenuation factor through the overstory	LDAS, set to 0.5
Trunk fraction	Ratio of total tree height that is trunk	LDAS, set to 0.2

Table 3.4: VIC model parameters for the vegetation library file. Parameters remapped from UMD-NLDAS to IGBP classification, following BV2019, are assigned the source "LDAS" ([Cherkauer, 1999]). GLASS citations: [Liang et al., 2013a]; [Xiao et al., 2016]; [Liang et al., 2013b]. SECHIBA ([Ducoudré et al., 1993]). SiB ([Sellers et al., 1986]). MODIS ([Didan et al., 2015])

We used MODIS observations from the year 2017 to calculate monthly average LAI, $fcanopy$, and albedo for each IGBP land cover type. We calculated LAI and albedo from the MODIS-based Global LAnd Surface Satellite dataset (GLASS; [Liang et al., 2013a]; [Xiao et al., 2016]; [Liang et al., 2013b]) and fcanopy from NDVI observations (MCD13C1; [Didan et al., 2015]) The expression used for fcanopy follows BV2019:

$$fcanopy = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2 \tag{3.5}$$

where $NDVI_{min}$ and $NDVI_{max}$ are the minimum and maximum values of NDVI observed for that month. Monthly LAI, $fcanopy$, and albedo values were calculated by averaging over all grid cells of the same land cover type, counting only cells that were at least 90% homogenous, to avoid noise from grid cells with multiple land covers. Excepting perennial snow and ice land cover, the vegetation parameters in the VIC vegetation library should describe snow-free vegetation. Therefore, before calculating LAI, fcanopy, and albedo for each land cover class, we used fractional snow cover data from MOD10CM ([Hall and Riggs, 2015]), a global 0.05 degree monthly snow cover dataset, to exclude grid cells with more than 90% snow cover. Additionally, we set albedo to 0.05 for open water, and we set LAI and $fcanopy$ to 0 for open water and perennial snow and ice.

The resistances r_{min} and r_0 play a role in determining how much plant transpiration occurs. Higher resistance means less transpiration. Stomatal resistance is resistance to the release of water through the plant stomata, and architectural resistance is the aerodynamic resistance between the leaves and the canopy top ([Ducoudré et al., 1993]). Two sets of resistance parameters have been used in past large-scale VIC implementations. N2001 ran VIC over the entire globe using r_{min} values adapted from Dorman and Sellers' global database of r_{min} values ([Dorman and Sellers, 1989]) computed using the Simple Biosphere Model (SiB; [Sellers et al., 1986]). The [Nijssen et al., 2001] r_0 values were taken from Ducoudre et al.'s SECHIBA land surface parameterization ([Ducoudré et al., 1993]). The other set of r_{min} and r_0 parameters are those used in the LDAS vegetation library and in studies such

as [Livneh et al., 2013]. This set of r_{min} values comes from [Mao and Cherkauer, 2009] and [Mao et al., 2007]. We used the r_{min} values from SiB ([Sellers et al., 1986]) and the r_0 values from SECHIBA ([Ducoudré et al., 1993]) for VICGlobal as they appeared to be the better documented values.

For the other parameters in the vegetation library file (displacement height, roughness length, etc.), we assigned values using the existing LDAS vegetation library. Since there are 17 IGBP land cover classes, and only 11 UMD-NLDAS land cover classes in the LDAS vegetation library, we re-assigned some IGBP land cover classes to take the parameters of UMD-NLDAS land cover classes. We remapped barren land, permanent wetlands, snow and ice, urban land, and water bodies to take the parameters of "grasslands" from the LDAS vegetation parameter file. While the characteristics of the barren, snow and ice, urban, and water land cover types clearly differ from those of grasslands, their low LAI and $fcanopy$ values, corresponding to sparse vegetation, essentially "turns off" the other vegetation parameters in the VIC model, as pointed out by BV2019. The other remappings were more straightforward. Croplands and croplands/natural vegetation mosaics inherited values from "croplands," savannas became "wooded grasslands," and woody savannas became "woodlands." We were thus able to assign vegetation parameter values to the each of the 17 IGBP land cover classes.

To calculate global average time series of seasonally-varying vegetation parameters would be of limited interest as the seasonal cycle would average out across the equator. Therefore, we calculated average monthly $fcanopy$, LAI, and albedo for each vegetation type in each hemisphere, and we developed two separate vegetation library files: one for the northern hemisphere and one for the southern hemisphere. Maps of January and July LAI, $fcanopy$, and albedo are shown in Figure 3.4. For illustrative purposes, the parameter values in this figure have been averaged over the 17 IGBP land cover classes using area-based weighting. Figures S14-S19 show maps of the remaining vegetation parameters. Figures S20-S22 show the cycle of LAI, fractional canopy cover, and albedo for each vegetation type, averaged

separately over each hemisphere.

3.3 Data Records

Soil and vegetation parameters for the VIC model are available for download at Zenodo ([Schaperow and Li, 2020]) in NetCDF format for version 5 of the VIC model. The files are stored as zip archives. Classic.zip contains ASCII text files with soil parameters, vegetation parameters, elevation bands, and two vegetation library files — one of the northern hemisphere and one for the southern hemisphere — for VIC-5 Classic. Image.zip contains a NetCDF "parameter" file with all the soil and vegetation parameters described above and a NetCDF "domain" file describing the VICGlobal domain (all land mass between 60°S and 85°N) for VIC-5 Image. MATLAB® codes for subsetting either set of parameters from the entire VICGlobal extent to a subregion of interest.

3.4 Technical Validation

3.4.1 Streamflow and snow-water equivalent in the Upper Colorado Basin

Having created input files for the VIC model, we tested the parameters in a large, well-studied river basin. We used the VICGlobal parameters to run VIC in water balance mode over the Upper Colorado River Basin (UCRB), a 293,600 km^2 basin in the western United States. We ran VIC once using the VICGlobal parameters and once using the L2013 parameters. Both simulations used the meteorological forcing data from L2013, at a six-hourly timestep, in water balance mode, for the 6-year period from Oct. 1, 2005 to Sept. 30, 2011.

We compared estimated streamflow from the VICGlobal and L2013 simulations with naturalized streamflow estimates from the U.S. Bureau of Reclamation (USBR; [Prairie and Callejo, 2005]) at Lees Ferry, Arizona (Figure 3.5). Naturalized flow is measured streamflow adjusted for the effects of reservoir storage and management and consumptive uses such as irrigation. We

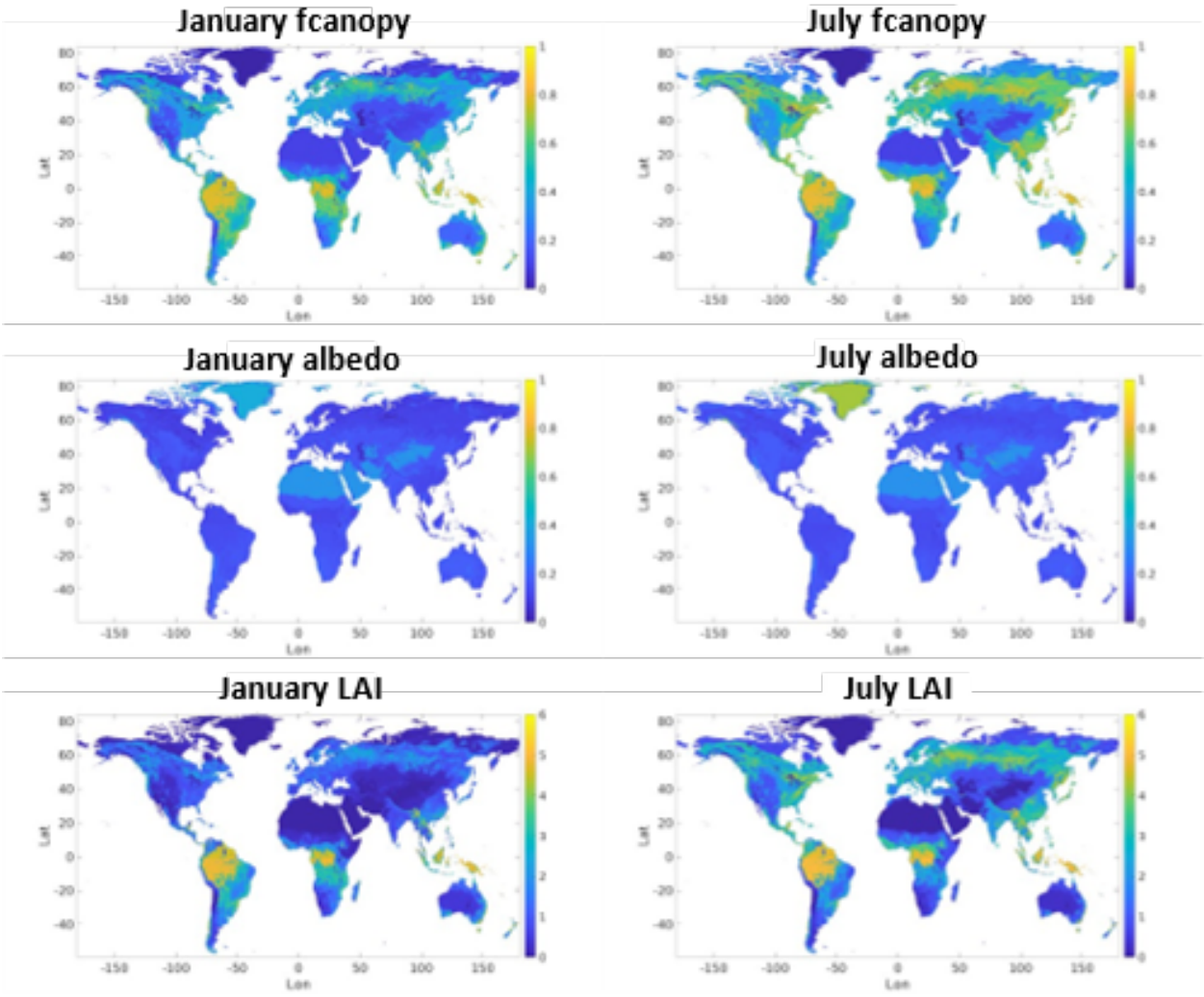

Figure 3.4: Maps of leaf-area index, albedo, and fractional canopy cover values. Parameter values have been averaged over the 17 IGBP land cover classes using area-based weighting.

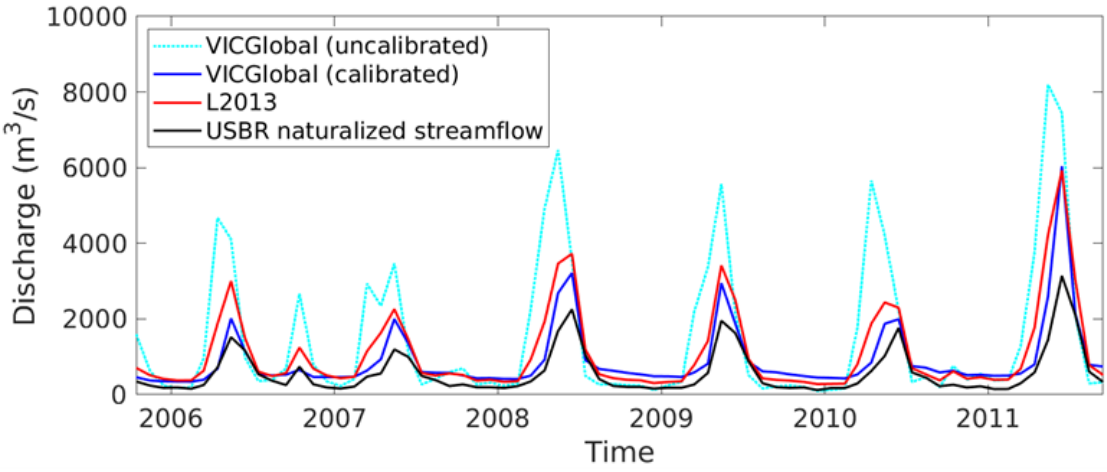

Figure 3.5: Monthly discharge estimates at Lee's Ferry (the outlet of the Upper Colorado River Basin). The dotted cyan line shows the uncalibrated VICGlobal estimate, the dark blue line shows the calibrated VICGlobal estimate, the red line shows the L2013 estimate, and the black line shows the U.S. Bureau of Reclamation naturalized streamflow data.

compared our VIC model outputs with naturalized streamflow because our VIC implementation does not simulate consumptive water use or reservoir storage. Due to differences in soil and vegetation parameters between the two sets of input files, there are notable differences in the hydrographs from each simulation. Relative to the L2013 results, the uncalibrated VICGlobal peak flows' timing is too early and their magnitude is too high. This is expected given that the L2013 parameters have been calibrated to get a good match to gauge data.

To understand the cause of this mismatch, we examine seven commonly calibrated soil parameters, which are difficult or impossible to estimate from measurements: ds, ws, $dsmax$, b_{infilt}, and the thicknesses of each soil layer (t_1, t_2, t_3). Taking a closer look at these soil parameters in the UCRB (Figure 3.6), we see that the infiltration capacity parameter b_{infilt} is considerably higher in the VICGlobal parameters than it is in the L2013 parameters, which would tend to cause higher runoff rates. ds is lower for VICGlobal than for L2013, so nonlinear baseflow occurs at a lower fraction of $dsmax$, tending to make baseflow peaks occur earlier. $dsmax$ is considerably higher for VICGlobal than for L2013 in much of the UCRB,

so the maximum baseflow rate is higher for VICGlobal. Finally, the thicker soil layers in L2013 mean that more water can infiltrate into the soil before baseflow occurs. Table 3.5 describes each of the seven calibration parameters and their influence on VIC model outputs. VIC users seeking more guidance on calibrating soil parameters should consult the VIC model documentation and relevant literature ([Demaria et al., 2007]; [Shi et al., 2008]; [Gou et al., 2020]). We calibrated the VICGlobal parameters b_{infilt}, $dsmax$, and t_3, the same parameters calibrated by L2013, to get a good match between predicted and observed (USBR naturalized) streamflow. However, decreasing b_{infilt} on its own was not enough to reduce the high runoff estimates produced by the VICGlobal parameters (VIC's sensitivity to b_{infilt} depends on the water-holding capacity of the upper two soil layers). By introducing the thickness of the second soil layer t_2 as a fourth calibration parameter, we were able to reduce runoff and increase transpiration. The final set of calibrated parameters was $b_{infilt} = 0.038$, $dsmax = 0.60$ mm/day, $t_2 = 1.5$ m, and $t_3 = 1.6$ m.

Variable	Description	Effect	Min	Max
ds	Fraction of $dsmax$ where nonlinear (NL) baseflow begins.	For soil moisture (SM), linear baseflow term increases with ds, while NL term decreases	0	1
$dsmax$	Max baseflow vel. in lowest soil layer	Baseflow increases α $dsmax$.	0	About 30 mm/day
ws	Fract of max SM at which NL baseflow occurs.	Higher value inc. the SM required for rapid increase in baseflow (NL baseflow term), causes delayed baseflow peaks.	0	1
b_{infilt}	Infilt. capacity param. Defines shape of VIC curve.	Higher b, lower infilt, higher surf. runoff.	0	0.4
t1	Thickness of upper soil layer	Thicker soil, more water can be stored before baseflow. More evap. occurs given more water stored in soil.	0.01 m	0.5 m
t2	Thickness of middle soil layer	Controls water avail. for transpiration. Thicker soils store more water.	0.05 m	1 m
t3	Thickness of bottom soil layer	Thicker soil, more water can be stored before baseflow. Controls water avail. for baseflow.	0.5 m	2.5 m

Table 3.5: Commonly calibrated soil parameters in the VIC model and their effects on model outputs. The baseflow equations are found in [Liang et al., 1994].

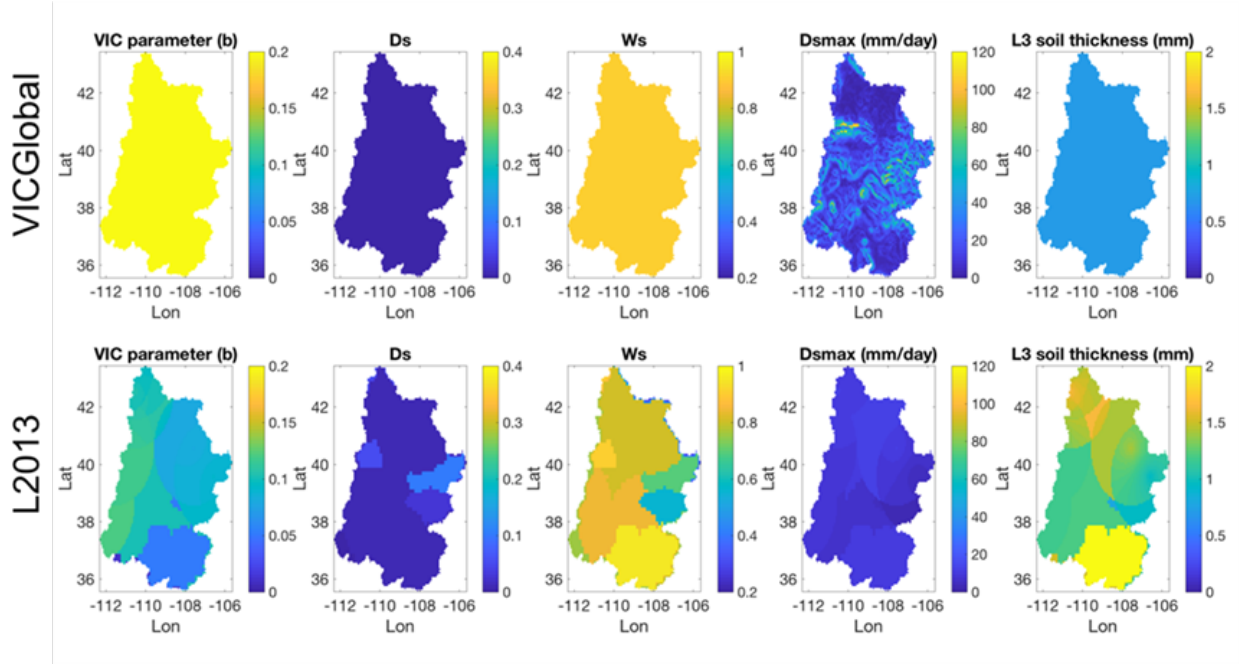

Figure 3.6: Maps of VICGlobal and L2013 soil parameters in the Upper Colorado River Basin.

For this analysis, we used manual calibration because the model run time made automated methods, which require hundreds to thousands of model runs, impractical. We used a custom MATLAB® application — a graphical user interface for running the VIC model, tuning its parameters, and displaying its outputs — to assist with manual calibration. We assumed the calibrated parameters were uniform over the basin. In addition to calibration by trial and error, VICGlobal users may also wish to explore automated calibration methods such as the Shuffled Complex Evolutionary algorithm (SCE-UA; [Duan et al., 1992]) or Dynamically-Dimensioned Search (DDS; [Tolson and Shoemaker, 2007]) when practical.

The calibrated VICGlobal simulation outperformed the L2013 simulation, with a Kling-Gupta efficiency (KGE; [Gupta et al., 2009]) of 0.24, compared to -0.26 for L2013 and -1.7 for the uncalibrated VICGlobal simulation. We also compared simulated snow-water equivalent (SWE) between the VICGlobal and L2013 simulations. Spatial patterns of simulated SWE were consistent between the two simulations (Figure 3.7a-c), as were patterns of snow

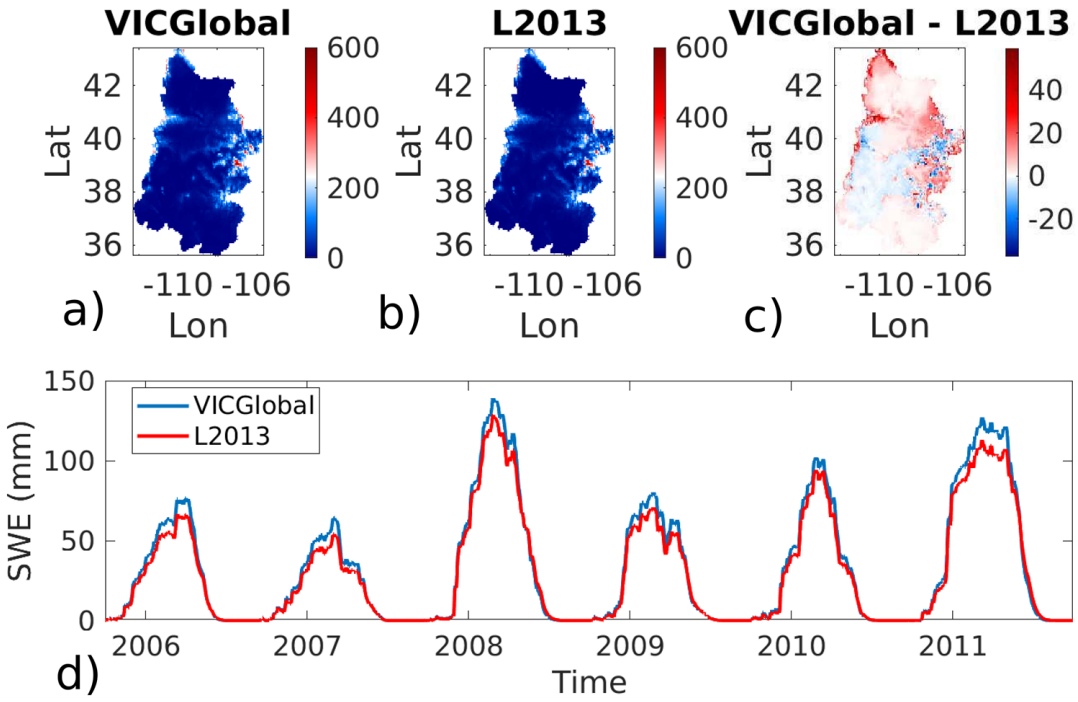

Figure 3.7: Time-average snow-water equivalent (SWE) maps (a-c) and basin-average SWE time series (d) for the Upper Colorado River Basin comparing simulations using the L2013 and VICGlobal parameters.

accumulation and melt (Figure 3.7d). VICGlobal SWE was 3 mm higher than L2013 SWE, on average. Snow sublimation, including canopy sublimation, was higher for L2013 than for VICGlobal, which helps explain the slight overestimate of SWE by VICGlobal relative to L2013. Overall, VICGlobal is able reproduce the timing, magnitude, and spatial pattern of L2013-simulated SWE in the UCRB, with no need for parameter calibration.

3.4.2 Water balance in 12 unregulated CONUS basins

Beyond the Upper Colorado Basin, we evaluated the VICGlobal parameters' potential for modelling the water and energy balance in 12 basins, ranging from 1500-25000 km^2, chosen for good spatial coverage of the CONUS. Modelled discharge was compared with monthly observations at USGS reference stream gages ([Falcone, 2011]) at each basin outlet; we used

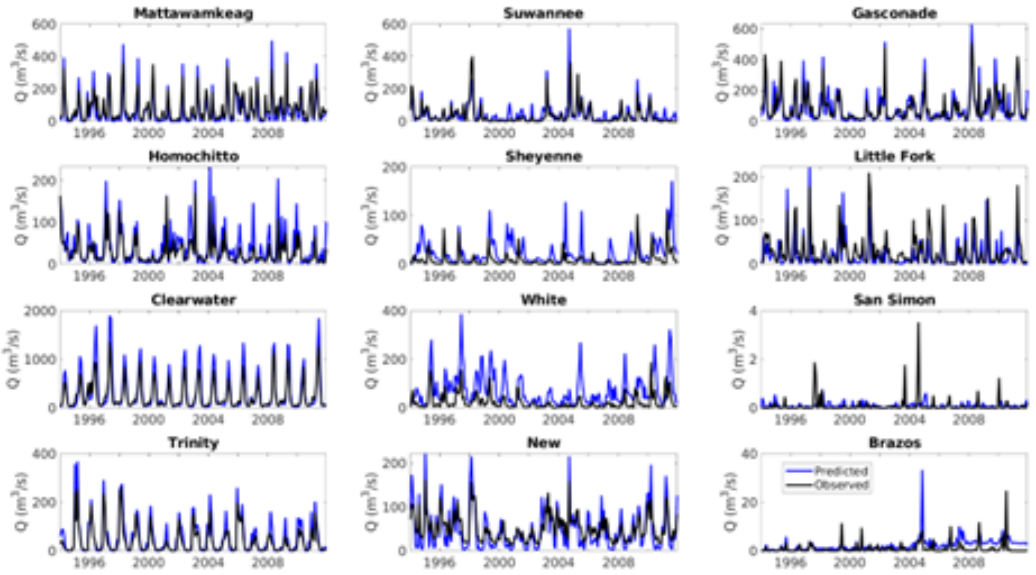

Figure 3.8: Monthly discharge predictions for 12 CONUS basins compared to USGS reference gauge measurements.

the DDS method to calibrate b_{infilt}, $dsmax$, t_2, and t_3. (These basins are small enough for automatic calibration to be practical.) The calibration was performed with L2013 meteorological forcing data, for calendar year 1993, with the VIC model run from 1990-1992 as spin-up. After 500 model evaluations (50 for the Clearwater River), the average discharge calibration KGE was 0.47, with a maximum of 0.82 for the Clearwater River in Idaho and a minimum of 0.00 for the White River in Arkansas. Using the calibrated parameters, we performed a validation run from 1994-2011. Table 3.6 shows goodness of fit between modelled and measured discharge for the 18-year validation run. Figure 3.8 shows discharge plots for each basin over the validation period. Good matches can be seen for the Clearwater, New, Homochitto, Mattawamkeag, Gasconade, Trinity, and Little Fork rivers, while the White, Sheyenne, Brazos, and San Simon rivers did not respond well to calibration, suggesting that parameters other than the four calibrated here are to blame. See e.g. [Demaria et al., 2007] for more insight on VIC calibration.

	Calibration (1993)		Validation (1994-2011)		
	KGE	RMSE	KGE	RMSE	DA (km^2)
Clearwater River	0.83	167	0.50	214	14268
Homochitto River	0.72	60	0.49	58	2073
Suwannee River	0.71	42	0.37	81	6136
New River	0.69	38	0.57	47	2952
Mattawamkeag River	0.64	86	0.70	66	3676
Gasconade River	0.62	235	0.72	115	8265
Sheyenne River	0.60	9	-0.87	28	7582
Trinity River	0.46	76	0.51	56	1980
Little Fork River	0.24	43	0.52	44	4383
San Simon Wash	0.12	1	-0.04	1	1482
Brazos River	0.05	1	-0.65	10	3348
White River	0.01	60	-1.47	117	25791

Table 3.6: Goodness of fit metrics for the 1994-2011 validation run over 12 CONUS basins.

3.4.3 Surface radiation budget validation with SURFRAD

To evaluate how well the (uncalibrated) VICGlobal parameters simulate the surface radiation balance, we ran VIC over six SURFRAD ([NOAA,]) sites, using soil parameters from the 1/16° grid cell containing the SURFRAD sites. We ran hourly simulations in energy balance mode from 1995-2011, with 1994 as a spin-up year. Meteorological inputs were taken from meteorological stations at the sites provided input data for the model, except for precipitation, which we took from L2013.

The first row of Figure 9 shows modelled and observed net radiation, upwelling longwave, downwelling longwave, upwelling shortwave, and downwelling shortwave radiation averaged over each day from 1995-2011 for six SURFRAD sites in the CONUS. Downwelling shortwave and longwave radiation predictions similar to the observations because the SURFRAD data were used as inputs for the VIC model (but not identical because precipitation inputs were taken from L2013 due to lack of ground measurements at the sites). There is a positive bias for net radiation resulting from a slight low bias for upwelling shortwave radiation. Overall, the bias is small.

The second and third rows of Figure 9 show scatterplots of predicted vs. observed upwelling shortwave and longwave radiation, with one-to-one lines shown in black. The correlation between predicted and observed upwelling shortwave and longwave radiation is close to 1 (ranges from 0.96-0.99 for all sites). Running VIC with VICGlobal parameters allows simulation of upwelling longwave radiation with an RMSE of 25 W/m^2 and RMSE of 15 W/m^2 for upwelling shortwave radiation. In Figure 3.9, we have excluded data at times when snow covers the ground to address the scale-issue — the spatial scale of the VIC simulations (a 1/16° grid cell) is much larger than that of the SURFRAD measurement — because of snow's large role in determining upwelling solar radiation, we excluded times when either the VIC model or SURFRAD measurements had snow on the ground using an albedo threshold of 0.4; none of the VICGlobal albedos for non-snowy land surfaces are this

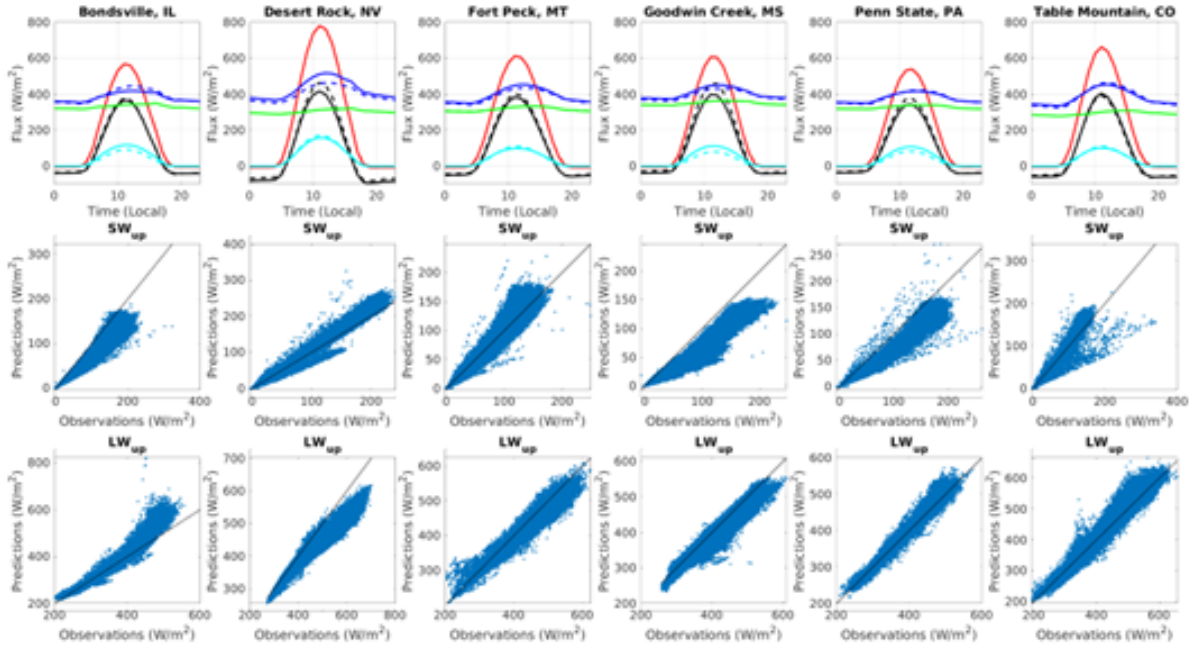

Figure 3.9: First row: Hourly modelled and observed net radiation, upwelling longwave, downwelling longwave, upwelling shortwave, and downwelling shortwave radiation from 1995-2011 for six SURFRAD sites in the CONUS. Dashed lines show predictions, while solid lines show observations. Second and third rows: scatterplots of predicted vs. observed upwelling shortwave and longwave radiation, with one-to-one lines shown in black. Key: Black = net radiation, red = downwelling shortwave, cyan = upwelling shortwave, green = downwelling longwave, blue = upwelling longwave.

large.

3.5 Usage Notes

We have described VICGlobal, a globally-consistent 1/16° VIC parameter dataset with soil and vegetation parameters derived from the latest satellite-based remote sensing datasets (MODIS and MERIT, which is based on SRTM data) and in-situ soil data from the FAO HWSD. In addition to its higher resolution, VICGlobal has an advantage over previous

global VIC setups due to its inclusion of seasonally-varying fractional canopy cover, LAI, and albedo, and because it explicitly accounts for barren, wetland, open water, and perennial snow and ice land covers. VICGlobal is provided in geographic coordinates, referenced to the WGS84 ellipsoid and datum.

VICGlobal has a few limitations. Its parameters are uncalibrated, so users must calibrate sensitive yet hard-to-measure parameters such as soil depth and the variable infiltration capacity parameter to get a good match between simulated and observed discharge. Several of the vegetation parameters, such as roughness length and displacement height, are assumed constant in time, even though realistically these parameters change as vegetation blooms and senesces throughout the year. And while we believe our monthly, hemisphere-average fractional canopy cover, LAI, and albedo are a major improvement over past global datasets, the most realistic parameter set would have them vary from grid cell to grid cell, even for the same vegetation type. Despite its limitations, we hope that VICGlobal, with its relatively high spatial resolution, wide coverage, and easy availability will be a valuable resource for VIC users.

3.6 Supporting Information

This Supporting Information includes:

- A plot of the resampled 1/16°MERIT DEM (Figure 3.10) and plots of each spatially-varying soil parameter (Figures 3.11 to 3.22).
- Plots (Figures 3.23 to 3.28) of vegetation parameters not shown in the main text. The vegetation parameters have been averaged over the 17 IGBP land cover classes.
- Plots (Figures 3.29 to 3.31) of the monthly LAI, albedo, and fractional canopy cover for each of the 17 IGBP land cover types calculated based on several MODIS data products, described in the main text.

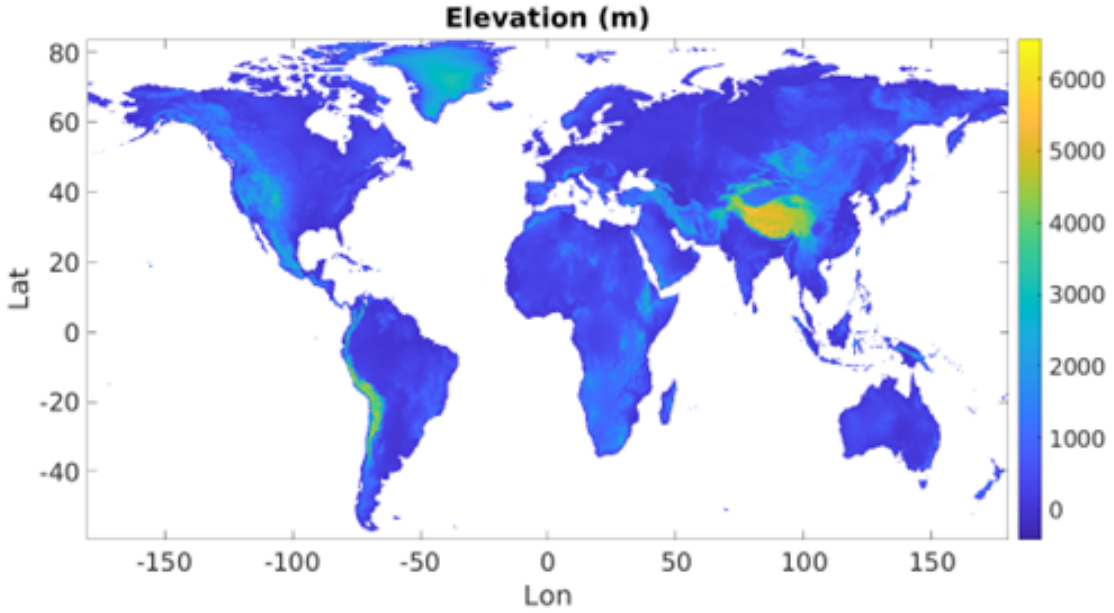

Figure 3.10: Elevation map showing the MERIT DEM resampled to 1/16°resolution. The spatial domain covers all the world's land, except Antarctica.

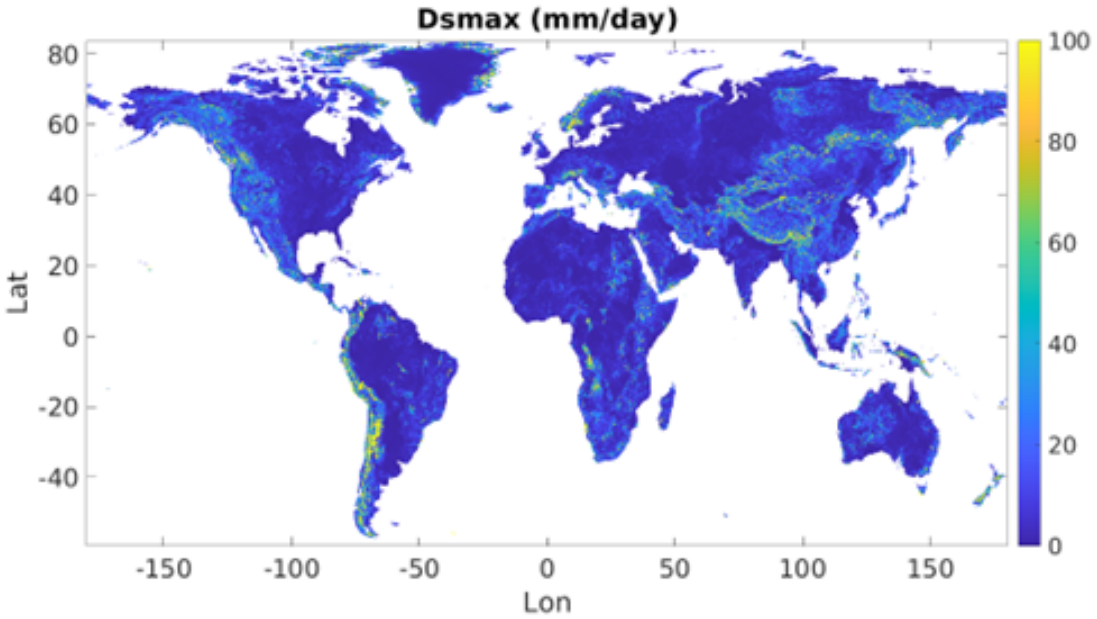

Figure 3.11: Maximum baseflow velocity *dsmax*. Values greater than 100 mm/day have been censored to show spatial variability.

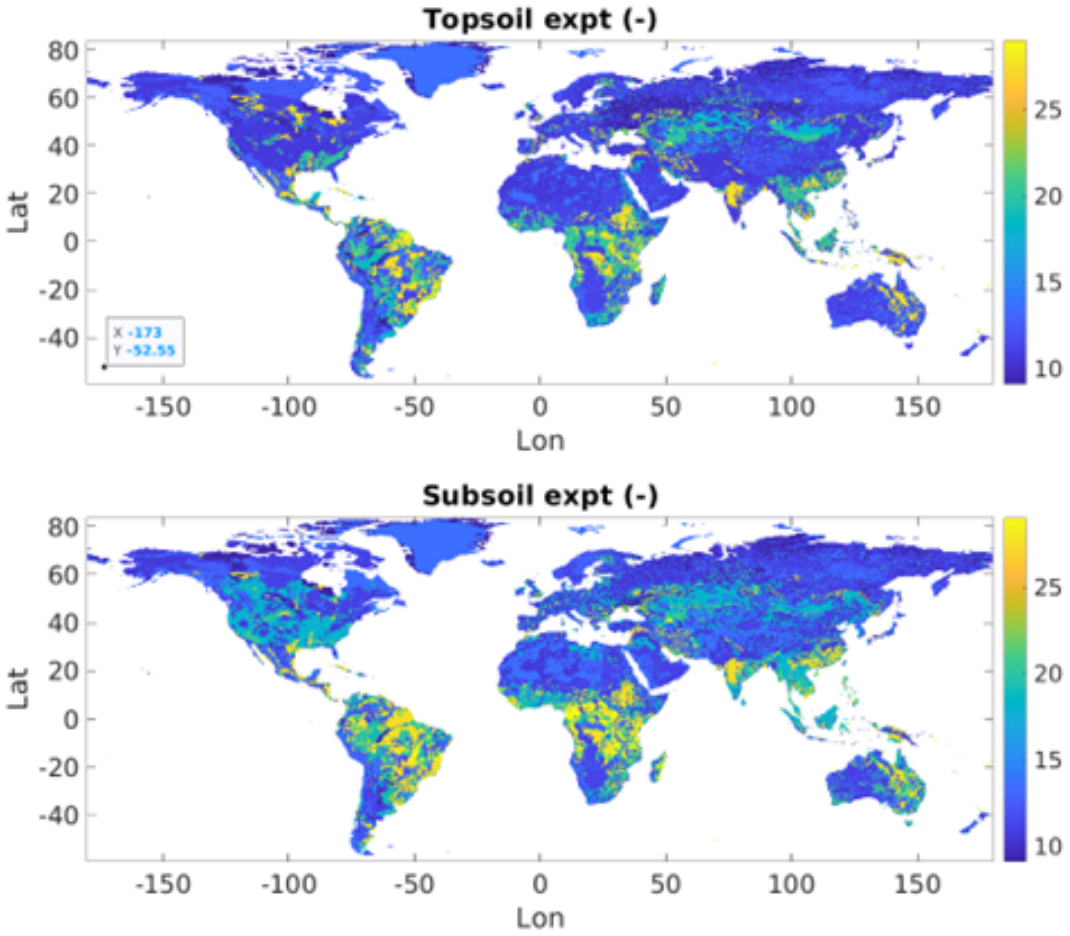

Figure 3.12: The exponent (*expt*) in Campbells's equation for hydraulic conductivity, calculated based on the FAO's Harmonized World Soils Database (HWSD). Topsoil values represent the first 30 cm of the soil column, and subsoil values represent soil from 30-100 cm deep.

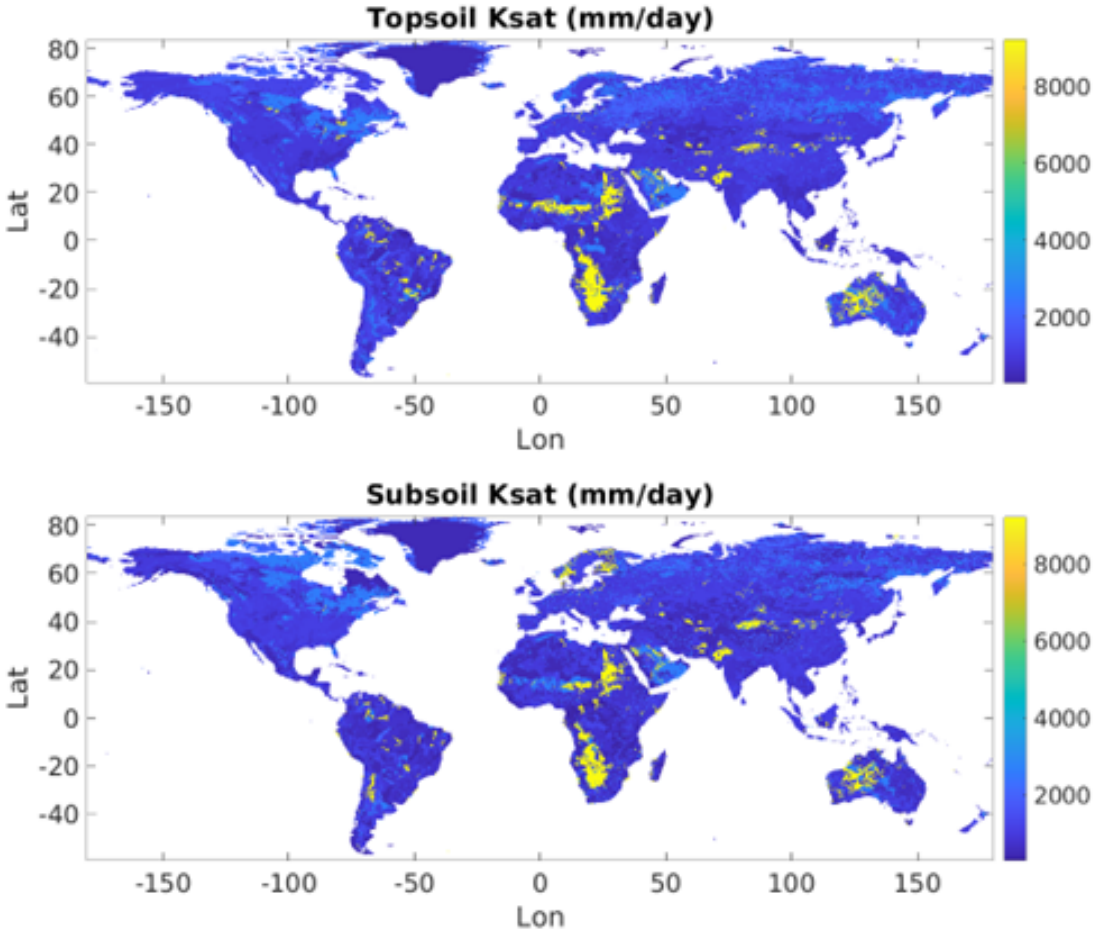

Figure 3.13: Saturated hydraulic conductivity (K_{sat}) for topsoil and subsoil, calculated based on HWSD data.

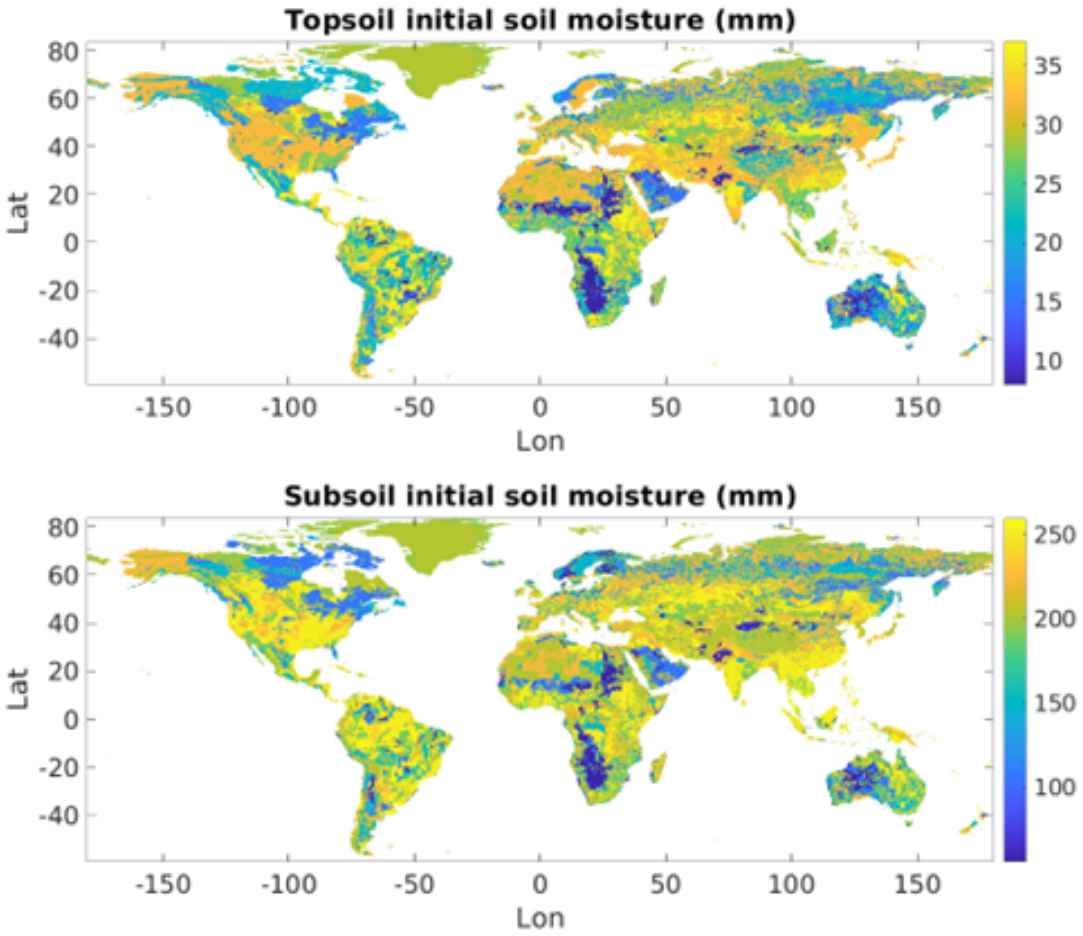

Figure 3.14: Initial soil moisture for topsoil and subsoil, calculated based on HWSD data, assuming the soil is initially at the critical point.

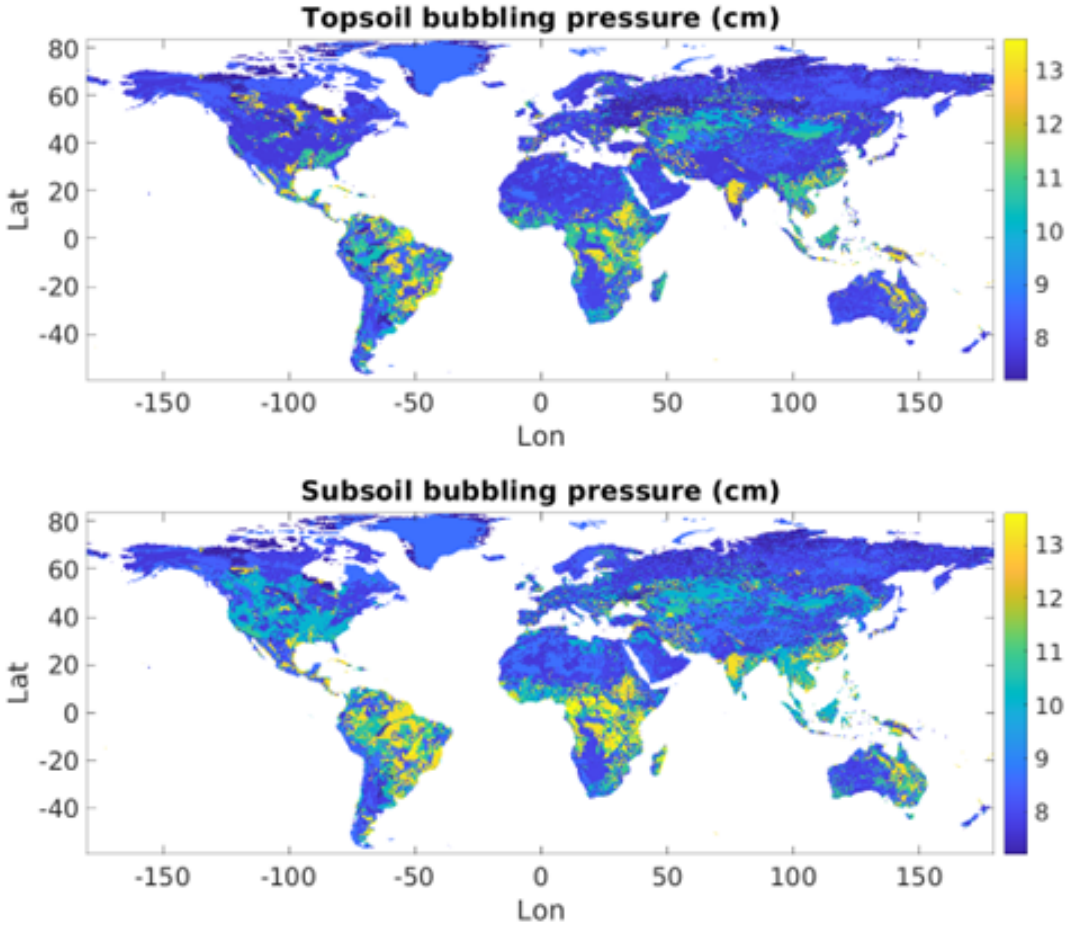

Figure 3.15: Bubbling pressure, calculated as a function of the Campbell's equation exponent, *expt*.

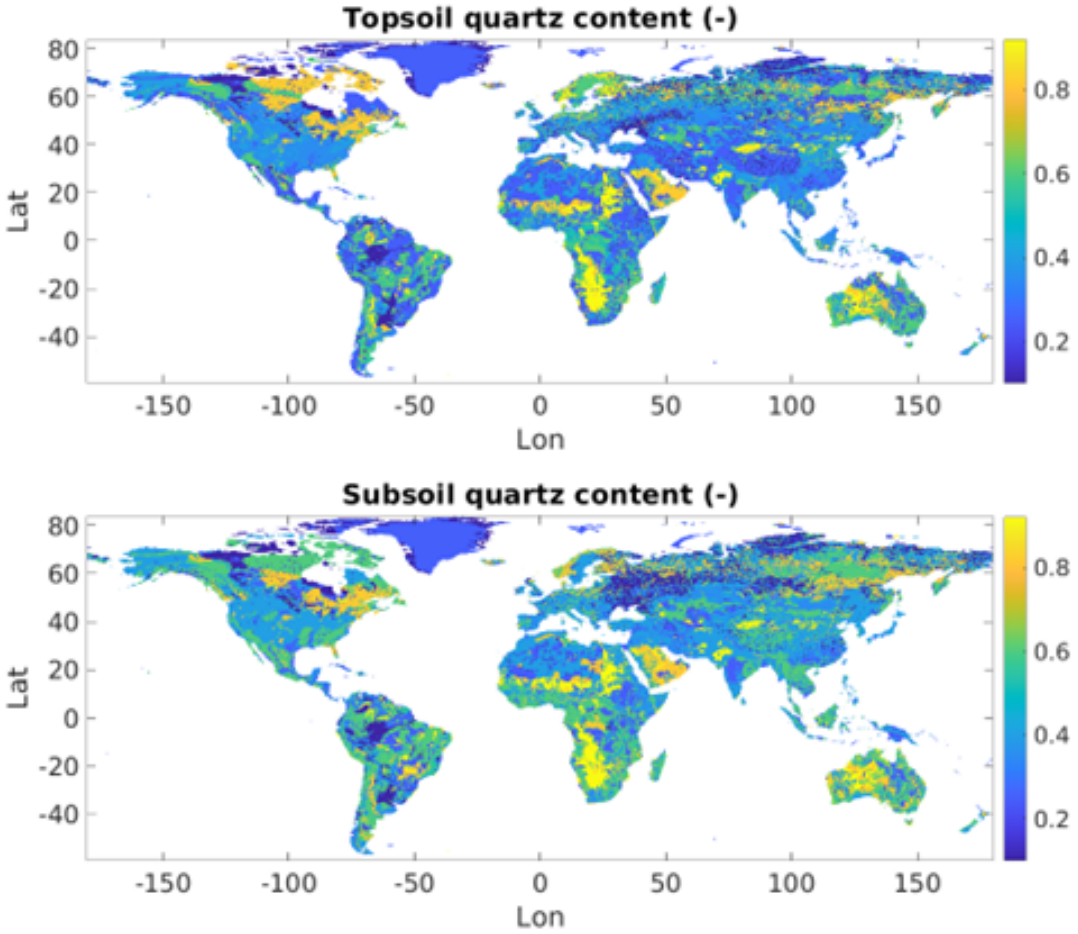

Figure 3.16: Quartz content, estimated using a pedotransfer table and HWSD soil texture data.

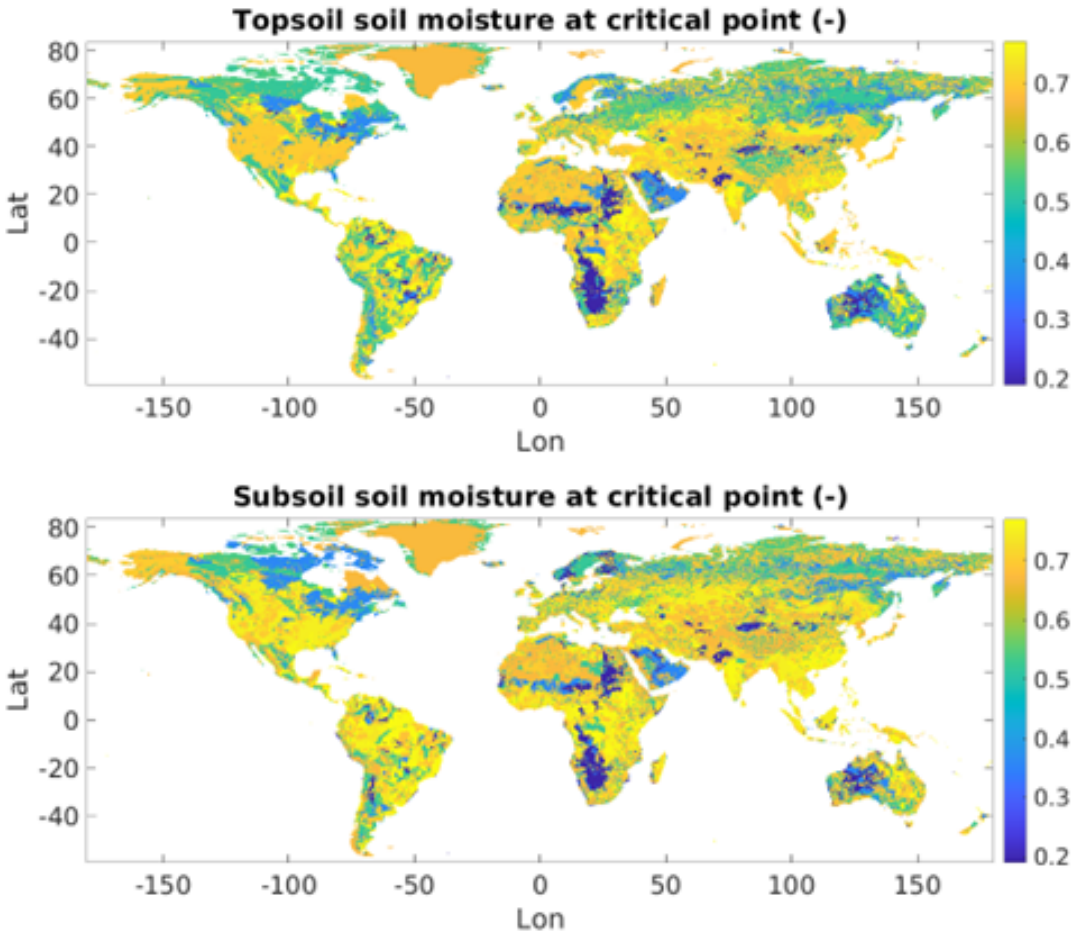

Figure 3.17: Fractional soil moisture content at the critical point (about 70% of field capacity) wcr_{fract}.

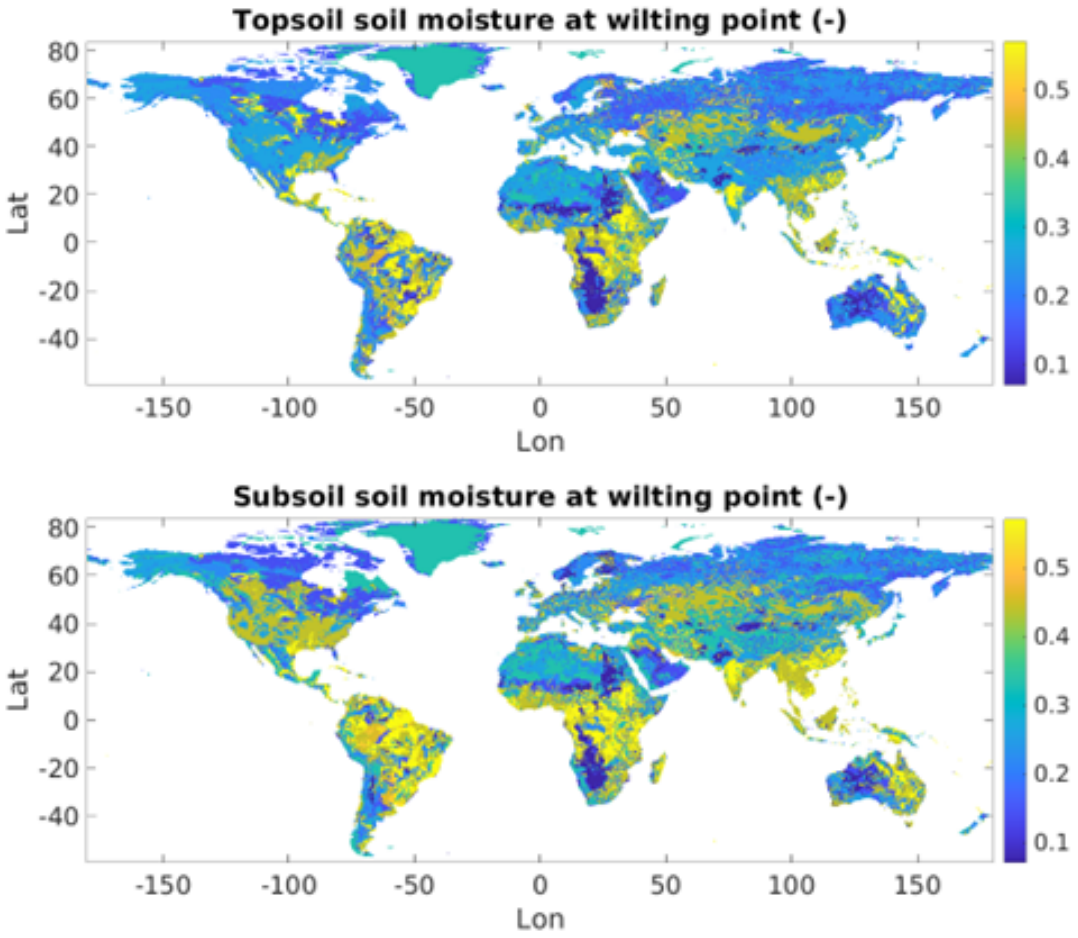

Figure 3.18: Fractional soil moisture at the wilting point $wpwp_{fract}$.

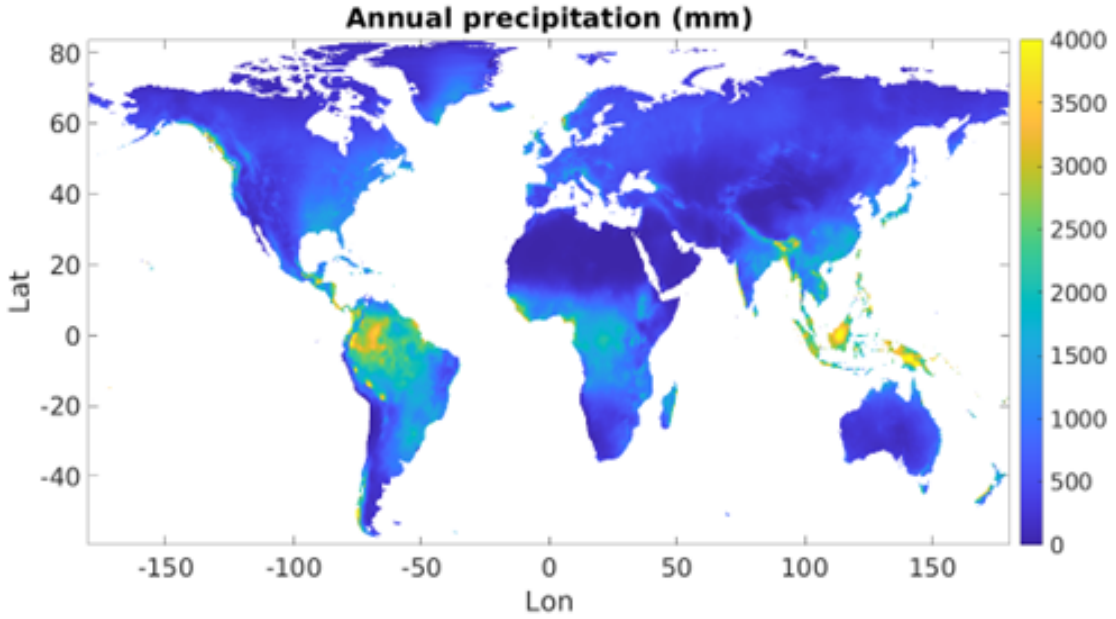

Figure 3.19: Average annual precipitation from WorldClim. Values greater than 4000 mm have been censored to show spatial variability.

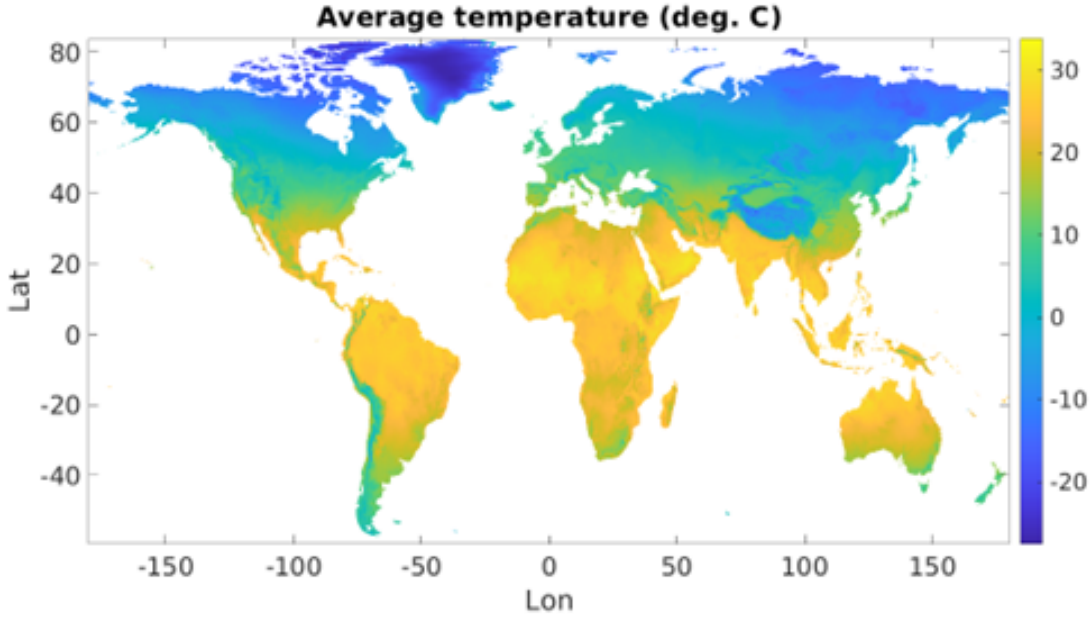

Figure 3.20: Average temperature from WorldClim, used as the bottom boundary condition for the VIC model's soil heat flux algorithms.

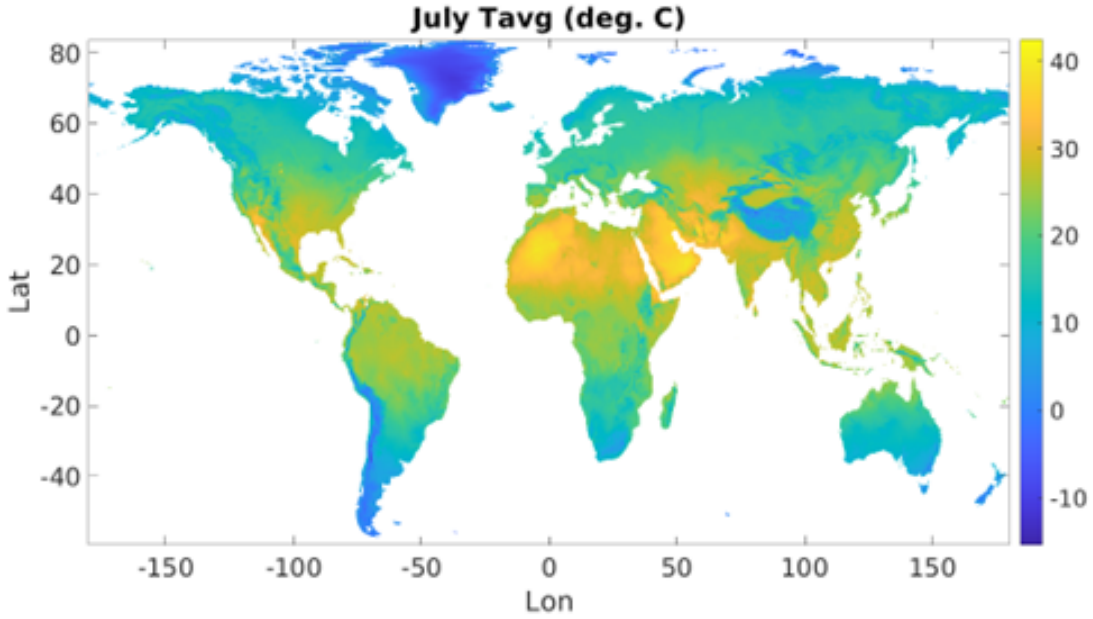

Figure 3.21: Average July air temperature, used for treeline computations if this option is chosen in the VIC global parameter file..

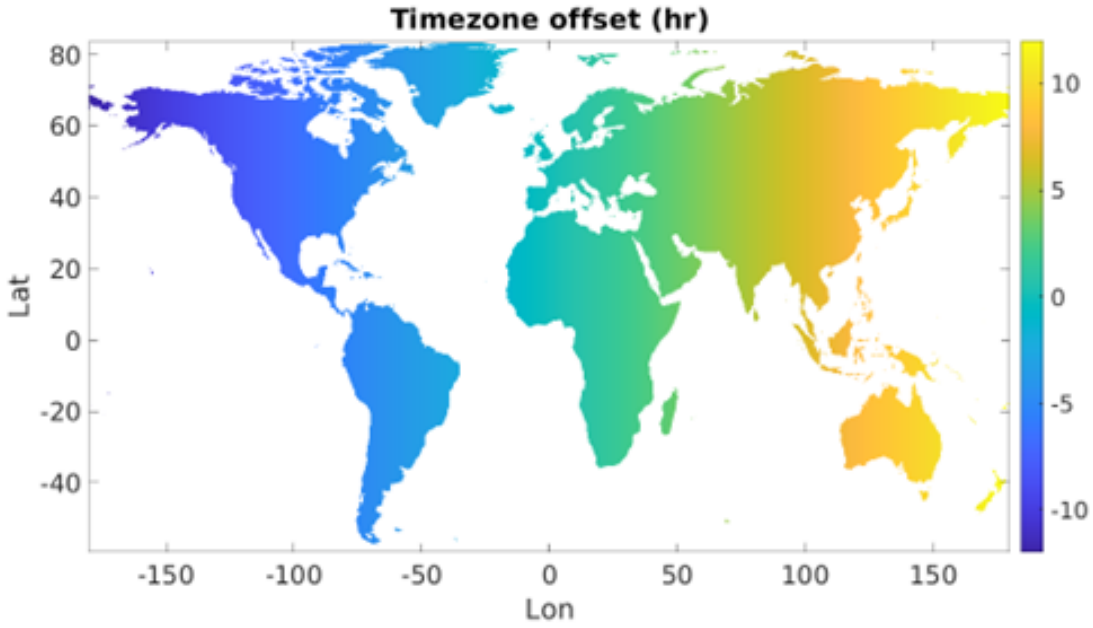

Figure 3.22: $offgmt$, the time zone offset from Greenwich Mean Time/UTC.

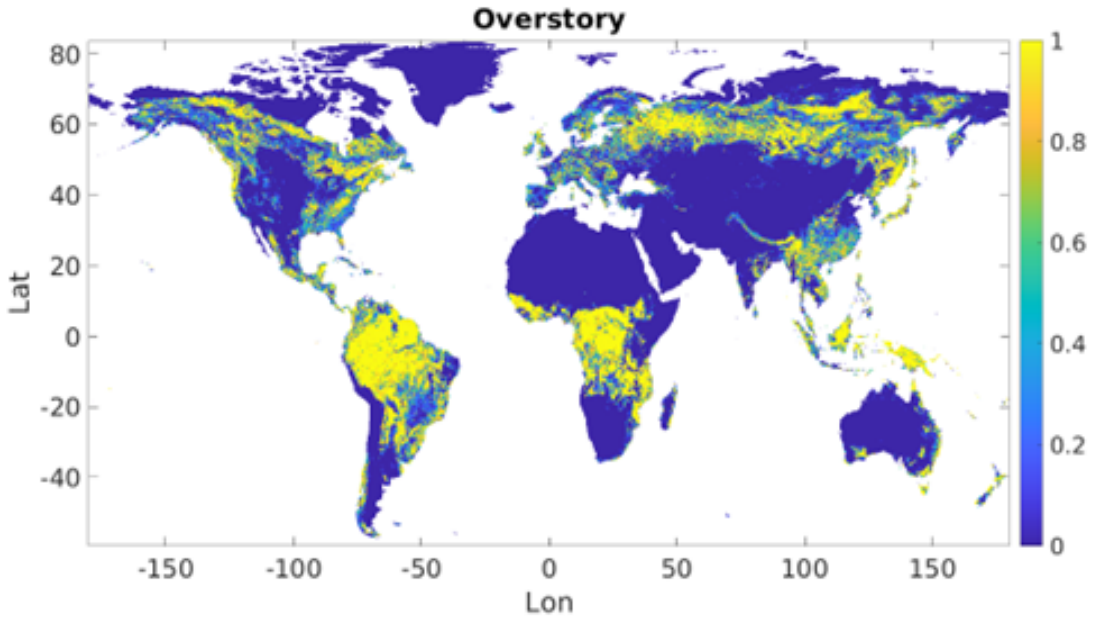

Figure 3.23: This map shows the area-weighted overstory fraction averaged over all the land cover classes within a single 1/16°grid cell. Overstory equals one if a vegetation cover class has an overstory and zero otherwise.

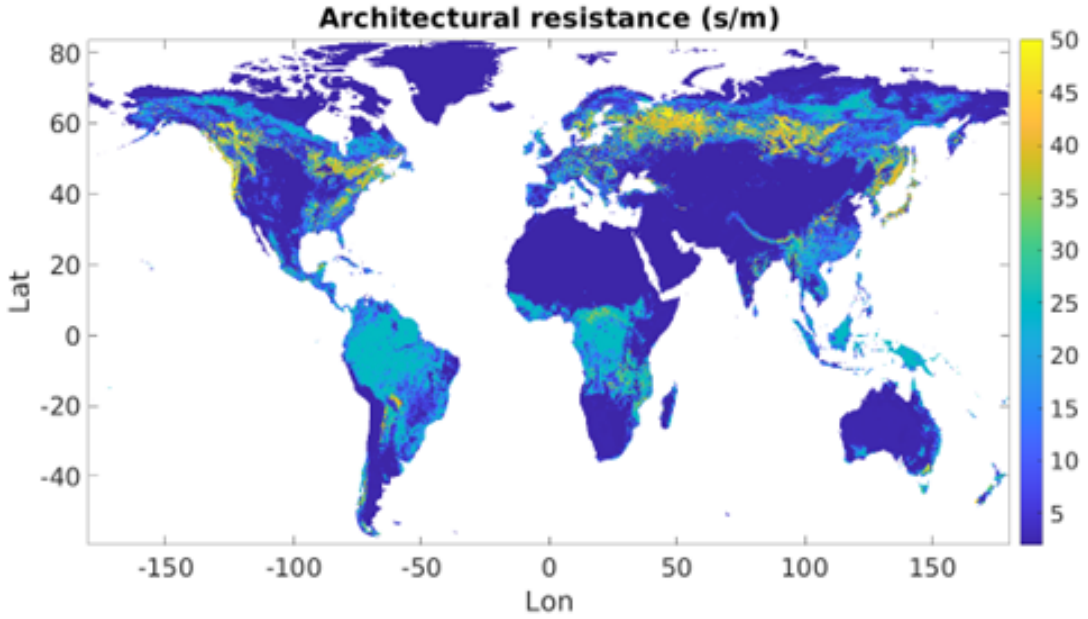

Figure 3.24: Architectural resistance $r0$, shown averaged over the different land cover classes for a given grid cell.

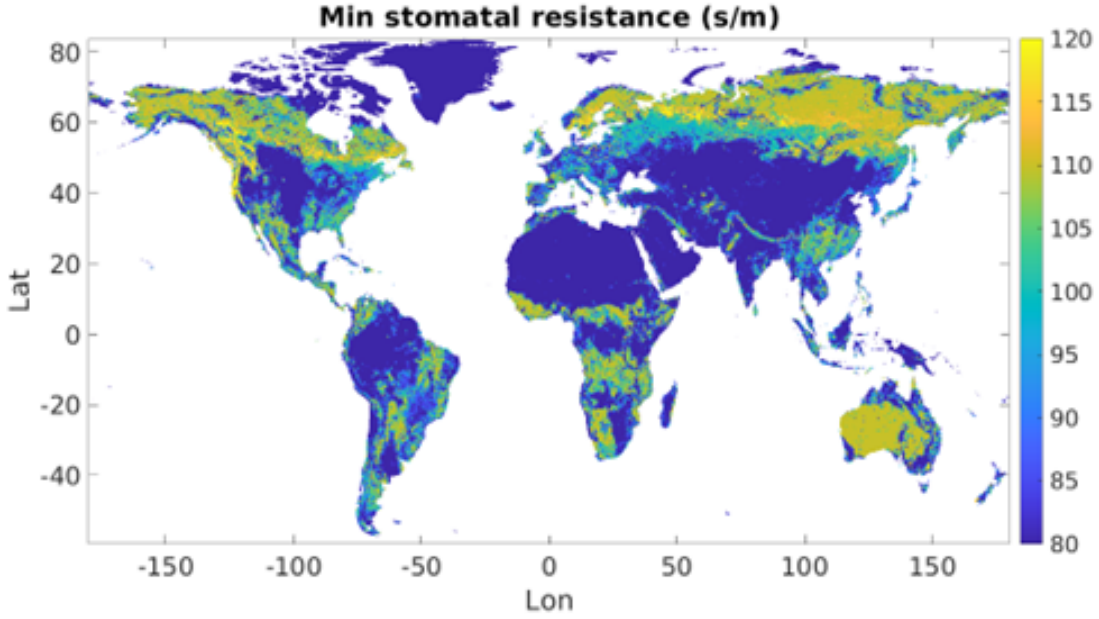

Figure 3.25: Minimum stomatal resistance $rmin$ for a land cover class, shown averaged over the different land cover classes for a given grid cell.

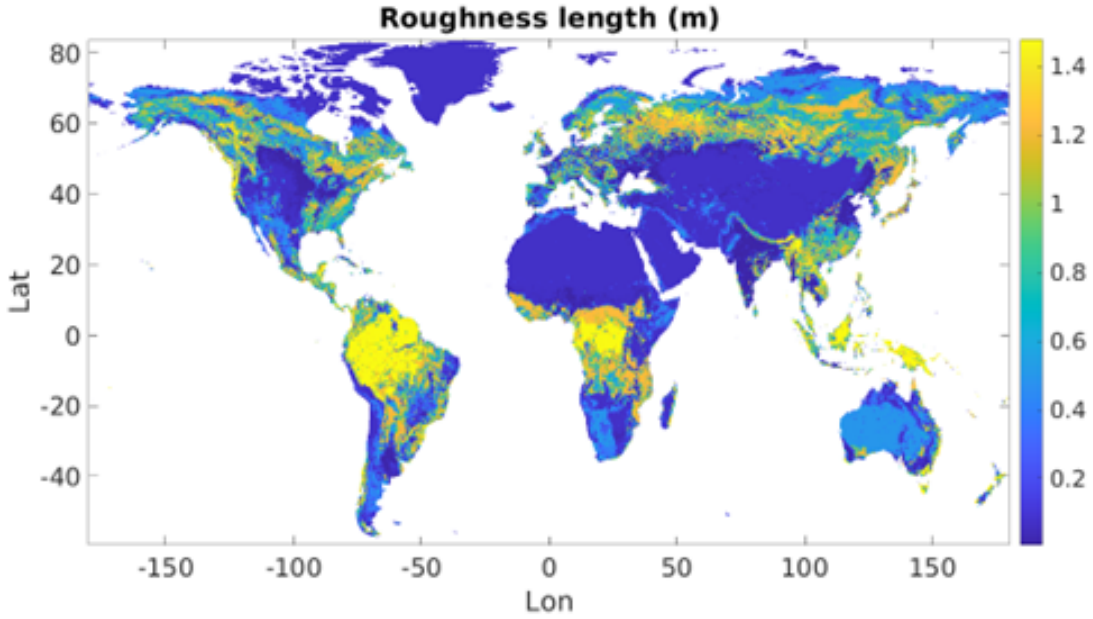

Figure 3.26: Vegetation roughness length, shown averaged over the different land cover classes for a given grid cell.

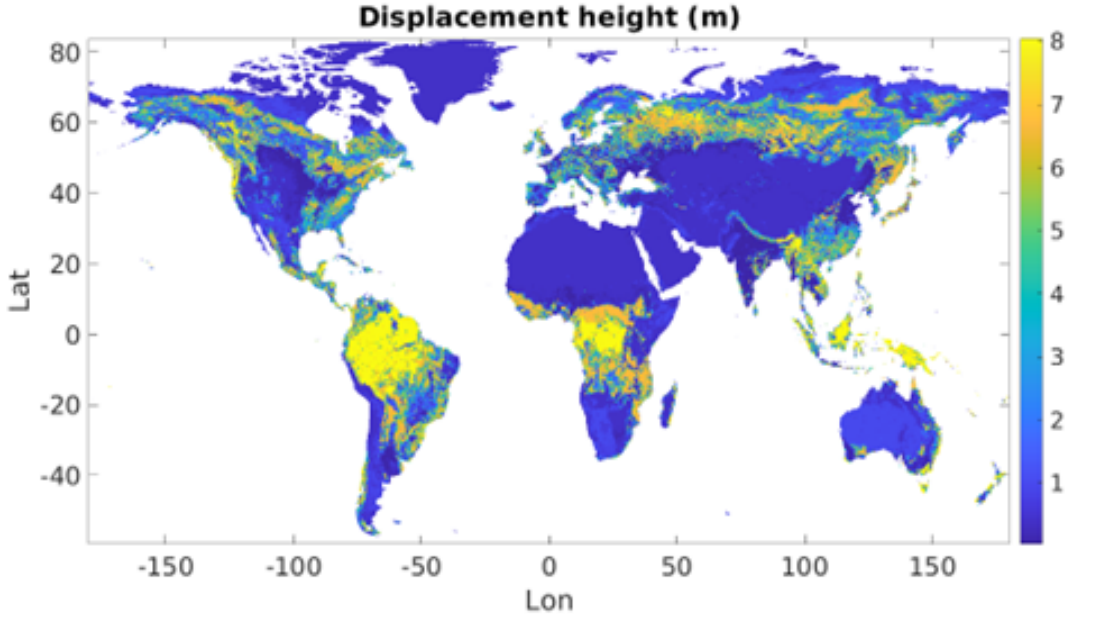

Figure 3.27: Vegetation displacement height, shown averaged over the different land cover classes for a given grid cell.

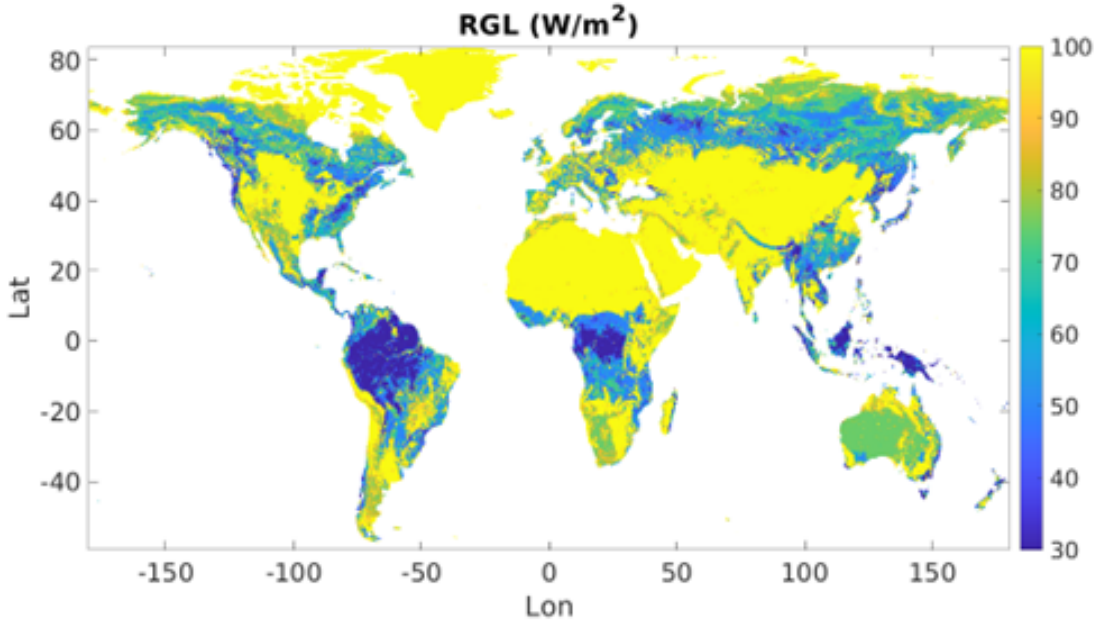

Figure 3.28: Minimum incoming shortwave radiation in order for transpiration to occur (RGL), shown averaged over the different land cover classes for a given grid cell. Typically, this is about 30 W/m^2 for trees and 100 W/m^2 for crops, according to the VIC documentation.

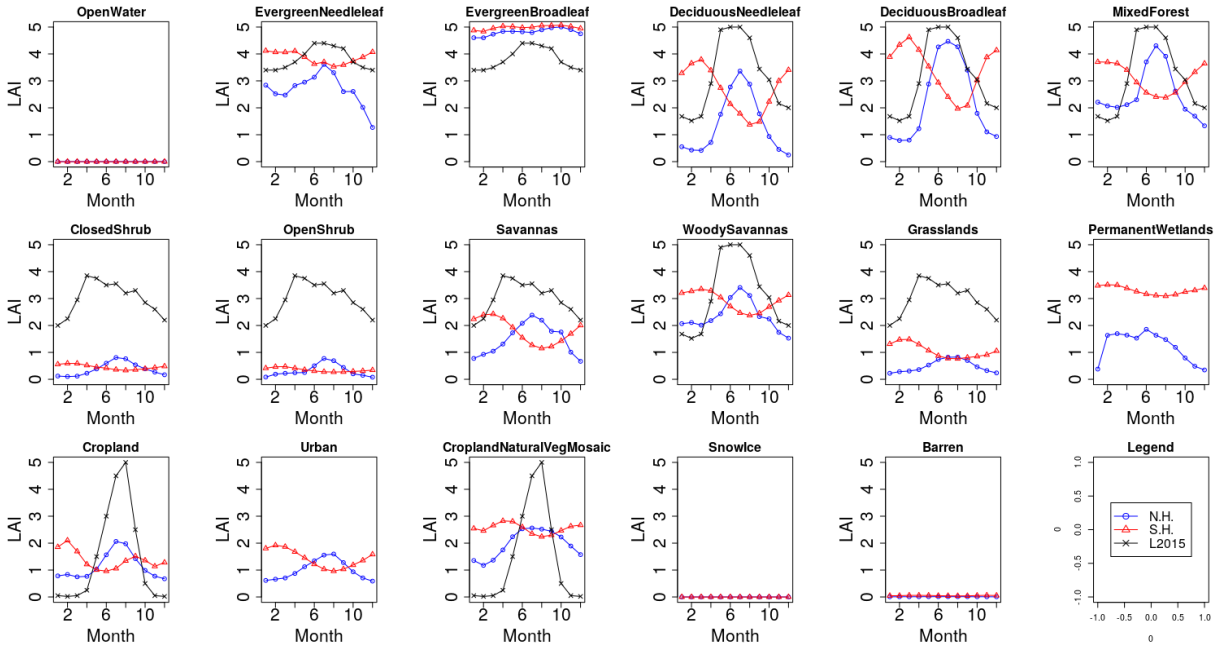

Figure 3.29: Hemisphere-average monthly leaf-area index (LAI) values for each IGBP land cover type. Gray lines show comparison to L2013 parameter values, when available. LAI values calculated using snow-free MODIS observations from 2017.

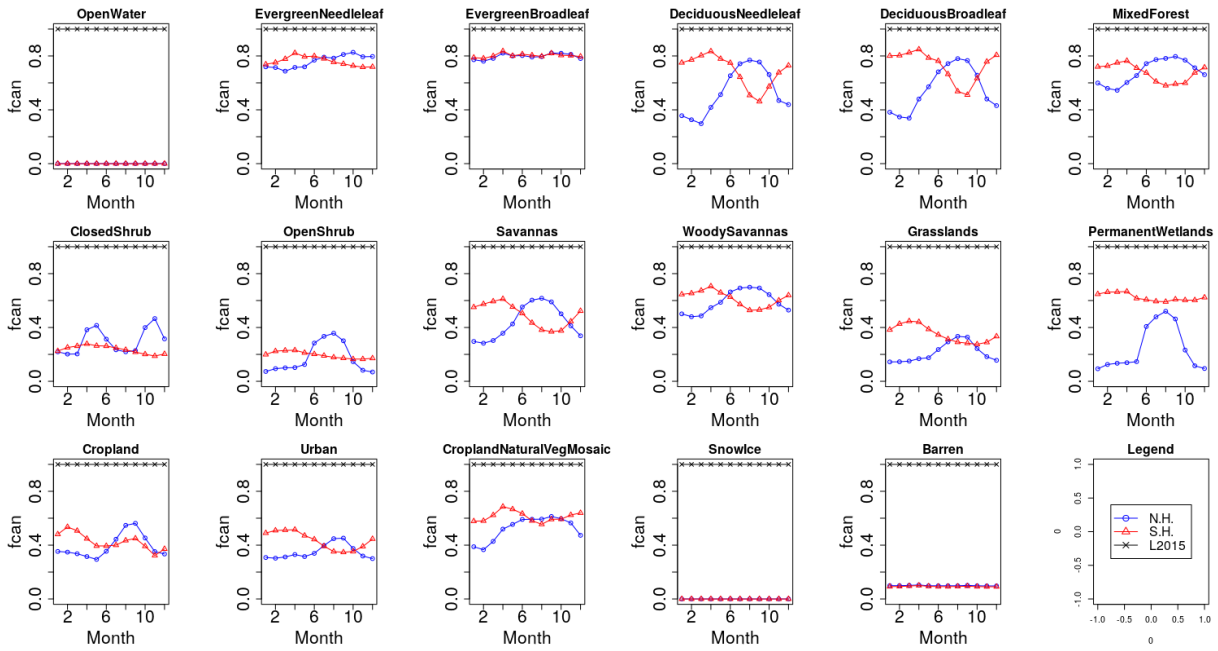

Figure 3.30: Hemisphere-average monthly canopy cover fraction ($fcan$) for each IGBP land cover type. Gray lines show comparison to L2013 parameter values, when available. $fcan$ values calculated using snow-free MODIS observations from 2017.

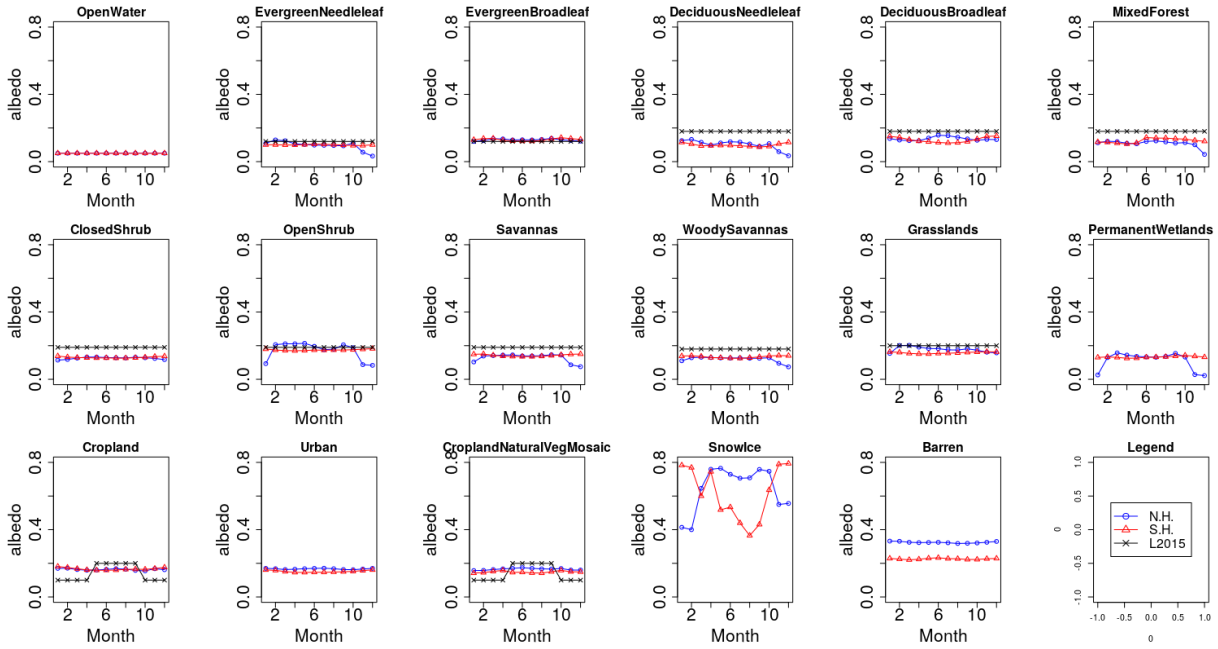

Figure 3.31: Hemisphere-average monthly albedo values for each IGBP land cover type. Black lines show comparison to L2013 parameter values, when available. Albedo values calculated using snow-free MODIS observations from 2017.

CHAPTER 4

An ensemble-based formulation for inverse streamflow routing and its application to the Allegheny Basin

Inverse streamflow routing (ISR) algorithms use point measurements of streamflow to estimate spatially-distributed upstream runoff. Since direct measurement of runoff is impractical at large scale, ISR provides a framework for estimating this key hydrologic variable from measurements of more easily observed point streamflow measurements. We build on the existing ISR algorithm to make the method more easily applicable to realistic problems, with large, high resolution basins and prior runoff estimates generated through ensemble hydrologic modeling.

In this paper, we present an ensemble formulation for ISR that (1) provides a consistent probabilistic framework for estimating runoff and discharge, as well as the error of the estimate, (2) builds in flexibility to the method by relaxing the assumption of linear forward and measurement models, and (3) eliminates the need to store extremely large runoff covariance matrices (smaller RAM requirement). We compare and contrast the new method, which we call "Ensemble ISR (EnISR)" with the existing ISR framework in the literature. Also, we implement a sub-domain localization method to further reduce computational needs, thus making the method more easily applicable to large, high resolution domains.

We find that EnISR runoff and discharge estimates are comparable with ISR runoff and discharge estimates from previously developed methods, but with the added advantages of straightforward uncertainty propagation and lower RAM storage requirements. We highlight two scenarios, one where daily discharge measurements are available at the basin outlet, and

one where 10-day measurements are available at multiple grid cells along the river channel, representing the expected SWOT-based discharge measurement scenario. In both cases, the posterior runoff and discharge estimates are a better match to the observations than the prior estimates, with narrowed uncertainty bounds.

4.1 Background and Motivation

4.1.1 Introduction

Streamflow data play an important role in shaping infrastructure and public policy decisions: reservoir management, flood mitigation, flood risk analysis, and water allocation decisions all rely on accurate streamflow estimates. Streamflow data, which can be thought of as the sum of runoff from each point in the river basin, are also widely used for calibrating land surface models (LSMs), which help us understand the impacts of climate change on the water cycle. Traditionally, streamflow data, also known as discharge, has been derived from in-situ gage stations. Unfortunately, these gage networks are sparse globally ([Hannah et al., 2011]), and the available information is often not openly shared between countries ([Gleason and Hamdan, 2017]). The recent launch of the Surface Water and Ocean Topography (SWOT) mission in December 2022 offers a potential solution.

SWOT orbits the Earth with a 21-day repeat cycle, observing rivers wider than 100 m every 2-4 times per cycle, depending on latitude ([Biancamaria et al., 2016]). Various retrieval algorithms have been developed to estimate discharge using SWOT's water surface elevation measurements. One leading method involves using the Manning equation to estimate discharge at each SWOT overpass location and time ([Durand et al., 2023]). This method is simple, but it only provides discharge estimates at overpass locations and times, and it relies on good estimates of Manning's roughness coefficient and average cross sectional area of flow. Discharge estimates are expected to have errors on the order of 15% ([Durand et al., 2023]). While these discharge estimates may have higher uncertainty and less frequent temporal

sampling compared to gage measurements, they provide spatial coverage far beyond that of existing gage networks.

Various algorithms, including interpolation ([Paiva et al., 2015]) and data assimilation ([Li et al., 2020], [Larnier et al., 2019], and [Oubanas et al., 2018a]) have been proposed to fill the gaps in between spatiotemporally sparse streamflow measurements in the context of SWOT. One method, inverse streamflow routing (ISR, [Pan and Wood, 2013], [Fisher et al., 2020], [Yang et al., 2020]) is particularly interesting because of its capability to simultaneously estimate both spatially-distributed runoff and river discharge. ISR estimates upstream runoff from point discharge measurements ([Fisher et al., 2020] calls this "the first sweep"), which may be sparse in space and/or time), then the runoff estimates are routed forward through the algorithm ("the second sweep") to produce spatiotemporally continuous streamflow estimates. Generating discharge estimates in this way is appealing because it ensures that discharge estimates will be consistent with the flow network (i.e. increasing downstream) and we can use the routing model to produce discharge estimates at any location within the basin. Furthermore, spatially-distributed runoff estimates are interesting in their own right, since they could provide us with a more effective method for calibrating spatially-distributed LSMs compared to the traditional, discharge-based calibration.

To clarify terms, since there may be different definitions of runoff in the literature, we are specifically referring to lateral flow into streams that is generated through surface and subsurface flow processes ([Beven, 2012]). Figure 4.1 shows a sample river basin schematic illustrating this concept. "Runoff" at each grid cell flows along the flow direction network until it reaches the river channel, where it becomes "streamflow." In ISR, we seek to estimate gridded runoff given streamflow measurements at one or more gage locations. These runoff estimates can then be used to estimate streamflow in between measurement times and at ungaged locations.

Large-scale, spatially distributed runoff estimation has mainly been done using hydrological models. Models such as the Variable Infiltration Capacity (VIC) model ([Liang et al., 1994]),

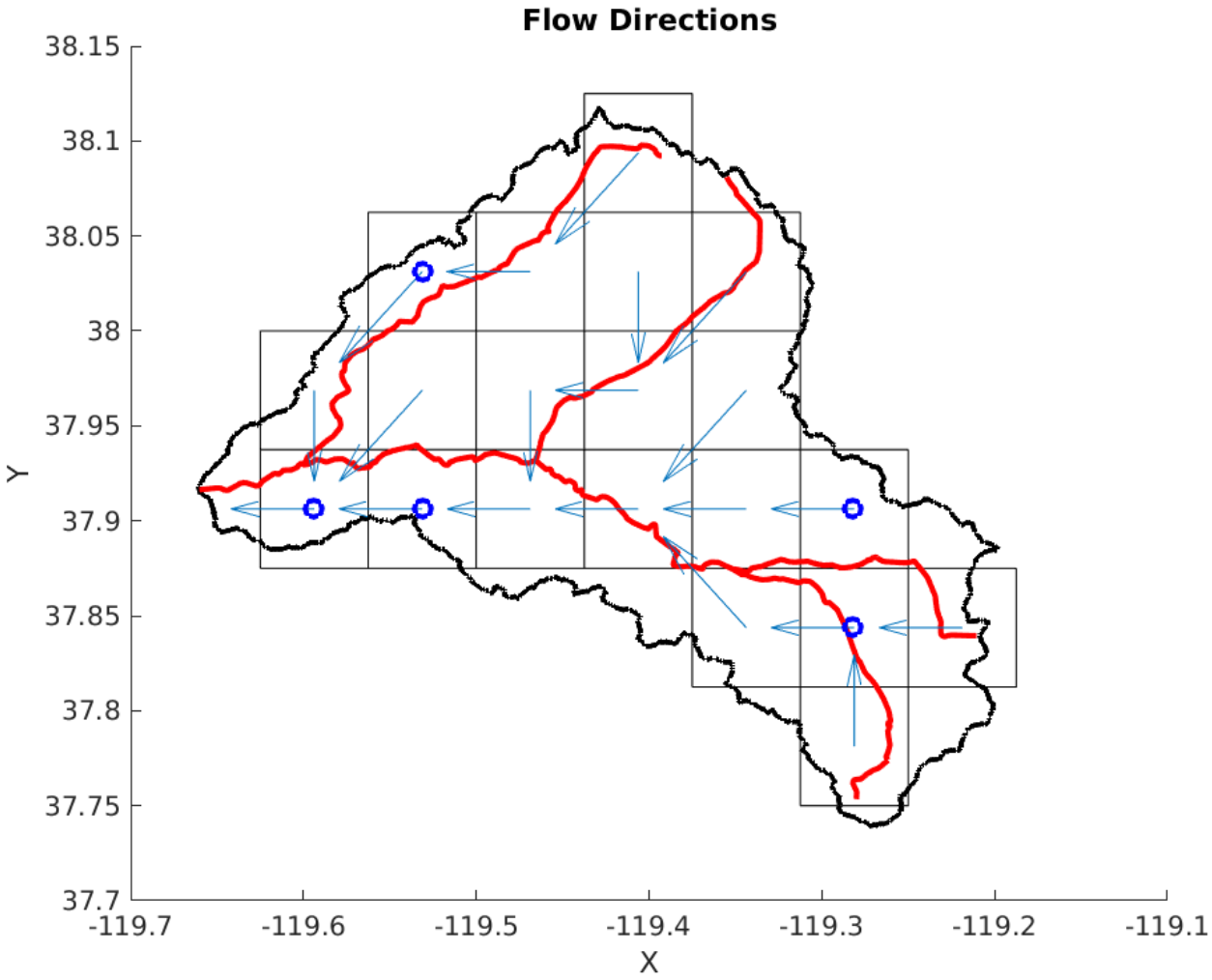

Figure 4.1: A schematic of a small river basin. The basin is modeled as a collection of grid cells, with flow directions (blue arrows), river centerlines (red lines), and gages (blue circles) superimposed.

Noah-MP ([Niu et al., 2011]), and the Community Land Model ([Lawrence et al., 2019]) convert precipitation, air temperature, and other atmospheric forcings into predictions of runoff, evaporation, and other water balance states and fluxes. These models can estimate runoff at high spatial and temporal resolution — however, they are limited by the accuracy and precision of the inputs (i.e. soil hydraulic properties, vegetation cover, and many others). Precipitation data, whether from gage, ground-based radar, or satellite measurements, con-

tain significant uncertainty ([Kidd and Huffman, 2011]) and have a strong effect on the accuracy of model-based runoff estimates. For example, [Bárdossy et al., 2022] found that up to 50% of error in streamflow prediction could be attributed to precipitation input uncertainty. In addition, topography, land cover, and soil properties play a role in determining how precipitation and snowmelt are partitioned into the various components of the water balance. For example, tree-covered areas slow down runoff via canopy interception, and soil hydraulic conductivity determines how quickly water travels through the soil, which influences how quickly hydrograph peaks occur. These parameters, particularly soil parameters, are spatially heterogenous and difficult to measure ([Heuvelink and Webster, 2001]), leading to model uncertainty. Routing these model-based runoff estimates to estimate discharge can lead to significant errors between predicted and true discharge.

Since ISR uses discharge measurements to revise an initial guess of runoff and produce an updated estimate, ISR can be viewed as a type of data assimilation algorithm called a fixed lag Kalman Smoother ([Bruhwiler et al., 2005]). This approach balances uncertainty in hydrological model predictions with uncertainty in streamflow measurements to make a combined estimate that leverages information from each source.

4.1.2 Inverse streamflow routing literature review

This section summarizes the contributions of past ISR research. A complete mathematical description of ISR methods follows in Section 2. The ISR literature consists of three studies: [Pan and Wood, 2013] (henceforth PW13) originally developed the ISR method for runoff estimation, and tested it in the Ohio River basin in the context of an observing system simulation experiment (OSSE). [Fisher et al., 2020] (henceforth F20) conducted experiments to test ISR's ability to estimate discharge at ungaged locations and also conducted sensitivity tests regarding gage configuration within the Ohio basin. [Yang et al., 2020] (henceforth Y20), the most recent contribution to the ISR literature, introduces a novel correlation model in an effort to take advantage of spatial and temporal correlations in actual runoff

errors, and evaluates ISR's performance in the context of the SWOT mission.

PW13's experimental design set the stage for both (F20 and Y20) subsequent ISR studies. PW13 ran the VIC model once with daily precipitation data from the National Land Data Assimilation System project phase 2 (NLDAS-2) ([Xia et al., 2012a], [Xia et al., 2012b]) to produce synthetic true runoff ("NLDAS runoff") and once with Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) version 3B42RT ([Huffman et al., 2007]) precipitation to produce an initial guess for runoff ("TMPA runoff"). In this controlled (synthetic) experiment the goal is to assess whether runoff and streamflow can be estimated when the initial guess contains expected (but known) levels of error. F20 and Y20 use the same data and experimental setup for the Ohio River basin.

PW13 updated the runoff initial guess (which can also be thought of as the "prior" runoff estimate in a Bayesian context) using (a) synthetic truth discharge produced by routing the NLDAS runoff through a river routing model (no measurement error) and (b) real USGS gage discharge measurements. PW13 assumed that prior runoff errors, a key input for the method, were uncorrelated with variance proportional to the square of the runoff initial guess. Given error-free discharge measurements from 75 gages for the 490,600 km^2 Ohio basin, PW13's attempt at recovering synthetic truth runoff succeeded in decreasing bias from 0.553 mm/day to 0.145 mm/day and reducing RMSE from 1.962 mm/day to 1.283 mm/day. While the posterior runoff estimate showed skill in estimating true runoff for the no measurement error case, its performance was significantly degraded in the real USGS gage discharge case, where PW13 assumed error-free discharge measurements, but used USGS discharge measurements to estimate posterior runoff. Neglecting to account for errors between USGS discharge and the synthetic truth discharge may have contributed to degraded runoff estimates in the real USGS discharge case.

F20 explored the potential of using ISR as an interpolator for estimating discharge at ungaged locations along the river channel using the "two-sweep" approach discussed above. We use the term "interpolator" because in both PW13 and F20, discharge errors were set

to zero, even if the actual discharge measurements (e.g. USGS gages) being used for the inversion do, in fact, have error. F20 notes that ISR produces improved discharge estimates even when using a "null" initial guess, which is set equal to the long-term mean runoff within the basin. This is meant to serve as a noninformative prior, meaning the updated (posterior) runoff estimate depends entirely on the routing model and the discharge measurements. We note that this null prior is only truly noninformative when discharge measurement error is assumed zero and when prior uncertainty is very large compared to the true underlying errors between the prior and the truth. F20 did not show runoff estimation results, since their focus was on streamflow interpolation only.

Y20 used SWOT-like discharge "measurements" to implement ISR in the Ohio River basin and in 16 large river basins around the world, ranging in size from 218,000 km^2 (the Senegal River basin) to 4,680,000 km^2 (the Amazon River basin). Y20 introduced a runoff error correlation model that takes advantage of prior runoff error correlations to estimate runoff more effectively. Given a good choice of prior runoff error correlation parameters, they obtained improved discharge estimates in each basin, with Kling-Gupta Efficiency (KGE; [Gupta et al., 2009]) increasing, for example, from 0.291 to 0.527 for the Yenesei River basin, after calibrating the prior runoff error model to achieve the best discharge estimates at calibration gages.

Y20's calibration of runoff error correlation parameters addresses an inconsistency in the PW13 ISR method. PW13 use a runoff initial guess equal to the output from forcing the VIC model with TMPA precipitation and assume, at the same time, that the runoff errors are uncorrelated, which is unlikely to be true. By estimating runoff error spatial and temporal correlation parameters, Y20 attempt to better represent the prior runoff errors.

Y20 used the same proportional runoff error variance model as PW13, and generated SWOT-like measurements assuming 30% relative discharge error. Unfortunately, they do not state what proportionality constant they used for their prior runoff error model, which could be problematic, i.e. if prior runoff error variance is too large, more trust will be placed

on the measurements than is appropriate.

4.1.3 Contributions of this paper

In this paper, we examine the state of the art for ISR with more theoretical rigor than has been done previously, and we develop Ensemble ISR (EnISR), which is based on the ensemble batch smoother (EnBS), to address limitations of the Y20 ISR method. We fully characterize Y20's ISR setup in a data assimilation context, investigate the effects of Y20's runoff error model assumptions, and determine when and why ISR produces negative runoff values. Ultimately, we set up a framework for a full end-to-end ensemble data assimilation approach, where the prior runoff ensemble could be generated by running an LSM with uncertain meteorological and land surface parameters, more effectively representing runoff uncertainty as a function of input uncertainty, rather than arbitrarily pre-specifying runoff uncertainty, as has been done in previous ISR papers. For comparison and consistency with previous ISR papers, in this paper we maintain the PW13/Y20 ISR paradigm of specifying runoff errors a priori. Our intention here is to propose a framework for performing ISR with realistic, ensemble representations of runoff error, which could then be used in the context of a full end-to-end data assimilation approach in future work.

While ISR is a promising method for (potentially SWOT-based) runoff and discharge estimation, several aspects of ISR deserve further investigation, or could be improved:

1. Although ISR is a data assimilation method and should be capable of providing confidence bounds around its estimates, no past ISR paper has explored this aspect of its potential.

2. ISR can produce physically unrealistic negative runoff estimates - more could be done to understand how and why these negative estimates arise.

3. Past ISR implementations use an arbitrary runoff error covariance model that assumes

that prior runoff error is proportional to the prior mean runoff - this may not hold true in realistic scenarios.

4. Storing the error covariance can be extremely computationally intensive, particularly for larger basins or at higher spatial resolutions.

Previous ISR papers (PW13, F20, Y20) have described ISR as a fixed-interval Kalman smoother, however, the assimilation problem has not been fully characterized in that context. We recognize that the "runoff initial guess" in PW13 should be the mean of the prior runoff distribution in the Kalman smoother, which, in order to be fully characterized should also have a clearly specified covariance matrix. This allows generation of probabilistic posterior runoff estimates that account for errors from both information sources: the model (or "initial guess") and the measurements. Analytical propagation of runoff error to discharge through the linear routing model is difficult due to the need to account for cross correlations between runoff at different locations and timesteps. [David et al., 2019] devised a scheme for runoff-to-discharge error propagation, but it holds for $\geq$monthly timesteps only. With EnISR, error propagation is as simple as routing each ensemble member through the routing model.

Viewed as a Kalman smoother, Y20 ISR assumes additive Gaussian errors and linear forward and measurement models (while there is technically no forward model, it can also be thought of as a trivial linear model). Non-Gaussian errors violate these Kalman filter assumptions and can theoretically lead to a suboptimal update. In realistic applications of ISR, the runoff prior may come from an ensemble of runoff estimates, either from different reanalysis datasets (see, for example, [Emery et al., 2020]), or by corrupting the precipitation inputs of a hydrological model with postulated errors. Neither of these cases is likely to produce Gaussian runoff errors. Instead of using an assumed runoff error covariance matrix, EnISR uses an ensemble batch smoothing approach to estimate the runoff error covariance from the ensemble. This nonparametric approach could make it a better choice for cases where runoff errors are known to be non-Gaussian.

EnISR also provides important computational benefits over Y20 ISR. Y20 ISR requires storing the runoff error covariance matrix, which can grow extremely large and may not fit in computer memory, limiting Y20 ISR's ability to estimate runoff in large basins or at high spatial/temporal resolution. EnISR does not need to store the full runoff error covariance matrix in order to calculate the Kalman gain. The largest matrix EnISR needs to store is orders of magnitude smaller than the largest matrix Y20 ISR stores. We also introduce a novel localization method that can further reduce the computational burden for both Y20 ISR and EnISR by breaking down the full basin into sub-basins and applying ISR to sub-basins.

4.2 Methods

4.2.1 Inverse streamflow routing method overview

River routing models estimate how long it takes for water to flow (over the hillslope) to the grid cell outlet. Water is presumed to travel from one grid cell to another according to the 1D diffusive wave equation and pre-specified wave celerity (flow velocity), following the flow direction network (see, for example, Figure 4.1). ISR uses a linear routing model (Equation 4.1) to route runoff (x) through a basin with n grid cells to discharge (y) at m gage locations (e.g. the basin model in Figure 4.1). This formulation is used in both the UW routing model ([Lohmann et al., 1998]) that is commonly coupled to VIC, and the RAPID river routing model ([David et al., 2011]). In state-space notation, the routing (runoff-to-discharge) model can be written

$$y_t = H_0 x_t + H_1 x_{t-1} + \ldots + H_k x_{t-k} \tag{4.1}$$

where:

- H_τ is the $m \times n$ matrix that adds the contribution of the runoff at time $t - \tau$ to

discharge at time t.

- x_t is an $n \times 1$ runoff vector at time t
- y_t is an $m \times 1$ discharge vector at time t
- $(k+1)$ is set to the maximum number of time steps for water to flow from the farthest point in the basin to the outlet, to account for contributions from all cells in the basin (essentially the time of concentration).

H_τ is determined based on the flow network and an assumed value of wave celerity (how quickly water moves through the network). We assume that the time step is long enough that a grid cell completely drains within one time step, such that $h_{i,j}$, an element of H_τ, has value 1 if cell i contributes to gage j at time lag τ and 0 otherwise.

Inverse Streamflow Routing (ISR) uses the relationship in Equation (4.1) to update an initial guess of runoff in a manner similar to the Kalman Filter. However, due to the contribution of runoff at previous time steps to discharge at the current time step, the update must be performed using overlapping time windows.

In order to capture information contained in future discharge measurements about runoff at the current time step, PW13 augmented the state vector (x_t) over a time window of length $s+1$, where $s+1 > k+1$ (where k is the time of concentration, rounded to the next timestep) is necessary to advance the algorithm forward in time. The current estimate of runoff (the state variable) is updated in batches in each time window given all discharge measurements within that window. This is done by augmenting the state vector to include runoff at times t to $t-s$ and rewriting Equation (4.1) for the augmented state vector. Because their values depend partially on discharge measurements in the next time window, runoff estimates at the last $k+1$ time steps within the current window are considered "partially updated." These runoff estimates are overwritten in the next time window. The augmented state vector is given by

$$\vec{x_t} = \Big[x_t, x_{t-1}, \ldots, x_{t-s}\Big]^T \tag{4.2}$$

The augmented state vector $\vec{x_t}$ (Equation 4.2) has length $n \times (s+1)$ and each component vector x_t has length n, representing runoff at each grid cell for the time step denoted in its subscript. The augmented routing model is derived by writing out measurement model for a single time step at all $s+1$ time steps in the interval $[t-s, t]$:

$$\begin{aligned} y_t &= H_0 x_t + H_1 x_{t-1} + ... + H_k x_{t-k} \\ y_{t-1} &= H_0 x_{t-1} + H_1 x_{t-2} + ... + H_k x_{t-k-1} \\ &... \\ y_{t-s} &= H_0 x_{t-s} + H_1 x_{t-s-1} + ... + H_k x_{t-k-s} \end{aligned} \tag{4.3}$$

By defining time-augmented streamflow and runoff vectors $\vec{y_t}$ and $\vec{x_t}$ we can write the time-augmented routing model compactly as

$$\vec{y_t} = \vec{H}\vec{x_t} + \vec{L}\vec{x_{t-k}} \tag{4.4}$$

where $\vec{H}$ and $\vec{L}$ are augmented linear model operators defined to make Equation (4.4) equivalent to Equation (4.3). They are $m(s+1) \times n(s+1)$ block matrices composed of the H_τ matrices for different lag times τ. H is the upper triangular matrix given in Equation 4.5.

$$\vec{H} = \begin{bmatrix} H_0 & H_1 & \cdots & H_k & 0 & 0 \\ 0 & H_0 & \cdots & H_{k-1} & H_k & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & H_0 & H_1 \\ 0 & 0 & 0 & 0 & \cdots & H_0 \end{bmatrix} \tag{4.5}$$

and L is the lower triangular matrix given in Equation 4.6.

$$\vec{L} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & H_1 & \cdots & H_{k-1} & H_k & 0 \\ \cdots & 0 & H_1 & \cdots & H_{k-1} & H_k \end{bmatrix} \tag{4.6}$$

At each time window $[t-s, \ldots, t]$, the runoff initial guess is updated using an update similar to a Kalman filter (reviewed in more detail in the next section), as follows:

$$\vec{x_t^+} = \vec{x_t^-} + \vec{K}\left(\vec{y_t} - \vec{H}\vec{x_t^-} - \vec{L}\vec{x_{t-k}^p}\right) \tag{4.7}$$

where $\vec{K}$, the Kalman gain, is calculated based on the runoff and discharge error covariance matrices and $\vec{x_t^+}$ and $\vec{x_t^-}$ represent the updated runoff and the initial guess, respectively. $\vec{x_{t-k}^p}$ is the partially-updated runoff vector using the update from the previous window.

PW13 and follow-on papers ([Fisher et al., 2020]; Y20) showed that it is possible to estimate spatially-distributed runoff based on discharge measurements, but key questions remain about the estimation method and its associated uncertainties. Putting ISR in the context of the Kalman smoother can help us better understand the method and how we can improve it. In the next section, we provide background on the Kalman filter and explain how ISR can be seen in this framework. In the following sections, we drop the bold notation for time-augmented variables (e.g. $\vec{x_t}$) for simplicity of notation.

4.2.2 The Kalman filter/smoother

ISR seeks to estimate unknown runoff based on imperfect discharge measurements. The Kalman filter ([Kalman, 1960]) provides the best estimate of an uncertain state given a set of uncertain measurements, given that the state variables and measurements can be treated as Gaussian (normally-distributed) random vectors. The Kalman filter starts with

an estimate of the state at an initial time t_0 and uses a forward model to propagate that estimate forward in time. If the state vector is a normally-distributed random variable and the forward model is linear, then the state variable will continue to be normally-distributed as it is propagated forward in time, and can therefore be fully represented by its mean vector and covariance matrix.

The forward model f is given by

$$x_{t+1} = f(x_t, \alpha, u_t) \tag{4.8}$$

where x_t is the state vector, α are time-invariant parameters, and u_t are forcing data. Equation (4.8) is used to propagate the state forward in time until a measurement becomes available. The relationship between the (uncertain) measurements and the state variable is given by

$$y_t = M(x_t) + v_t \tag{4.9}$$

where the measurement model M relates the states (x_t) to the measurements (y_t), and v_t represents the measurement error. $M(x_t)$ is equivalent to the routing model (Equation 4.1). If the errors in the states and measurements are Gaussian and the forward model and measurement model are linear, the mean update is given by

$$\overline{x}_t^+ = \overline{x}_t^- + K_t(y_t - \hat{y}_t) \tag{4.10}$$

where $\overline{x}_t^+$ is the posterior mean, $\overline{x}_t^-$ is the prior mean, K_t is the Kalman gain, and $\hat{y}_t = M(x_t)$ are the predicted measurements.

Similarly, the covariance update is given by

$$P_t^+ = (I - KH)\, P_t^- \tag{4.11}$$

where $P_t^+ = cov(x_t^+)$ is the posterior error covariance, P_t^- is the prior error covariance, and H is the linear measurement model operator ($y_t = M(x_t) = Hx_t$). (It can be shown that the covariance of the errors are equal to the covariance of the state vector).

The Kalman gain is equal to

$$K_t = P_t^- H^T \left(H P_t^- H^T + R_t \right)^{-1} \tag{4.12}$$

where $P_t^- H^T$ is the cross covariance between the state and the predicted measurements, $H P_t^- H^T$ is the covariance of the predicted measurements, and R_t is the measurement error covariance.

4.2.3 The PW13 approach

PW13's inverse streamflow routing method can be viewed as a Kalman smoother with overlapping windows and no forward model. When viewed through the lens of the Kalman filter, the runoff estimate x_t^+ is a normally-distributed random variable with mean equal to the "runoff initial guess" (e.g. from a land surface model) and with covariance P_t^-. The PW13 Kalman gain (Equation 4.12) is computed assuming there are no measurement errors ($R_t = 0$). PW13 perform experiments under two sets of assumptions:

1. Uniform initial guess equal to long-term mean runoff and uncorrelated covariance matrix equal to long-term variance

2. Initial guess obtained by forcing the VIC model with TMPA precipitation and uncorrelated covariance matrix with entries proportional to the square of the prior mean runoff.

Here, we consider the consequences of these assumptions. In the first case, the runoff error covariance matrix $P_t^- = \sigma^2 I$ for all t, where σ^2 is the long-term runoff variance, which is calculated from the (known) synthetic truth. The Kalman gain simplifies as follows:

$$K_t = P_t^- H^T \left(H P_t^- H^T \right)^{-1} \tag{4.13}$$

$$K = \sigma^2 I H^T \left(H \sigma^2 I H^T \right)^{-1} \tag{4.14}$$

$$K = H^T (H H^T)^{-1} \tag{4.15}$$

Equation (4.15) shows that under this set of assumptions, the Kalman gain is time invariant. Neither the value of the long-term runoff mean nor variance has any effect on the update, and the posterior runoff estimate is simply the downstream discharge redistributed evenly over the contributing grid cells (which is guaranteed to be non-negative, as long as the discharge measurements are non-negative). The discharge is effectively spread over the upstream cells, with equal volume originating from each cell, spaced out over bands of travel time. Visually, the change in runoff resembles a bullseye, with concentric circles radiating out the from gage location.

In the second case, a runoff estimate produced by forcing the VIC model with TMPA precipitation is taken as the runoff initial guess (implicitly the prior mean). Runoff errors are assumed uncorrelated with variance proportional to the square of the prior mean. In other words, $\sigma_{x_t}^- = \alpha \times x_t^-$, where the proportionality constant α is the coefficient of variation $\alpha = \sigma_{x_t} / \mu_{x_t}$. Therefore, the Kalman gain K_t has the form

$$K_t = I \odot x_t^{-2} 1^T H^T \left(H I \odot x_t^{-2} 1^T H^T \right)^{-1} \tag{4.16}$$

where $\odot$ is the Hadamard product, 1 is the ones vector, and $I \odot x_t^{-2} 1^T$ is a diagonal matrix with the squared prior runoff values $(x_t^-)^2$ on the diagonal. Notably, the proportionality constant α cancels out (because there is no measurement error). In this case, the updated runoff estimates depend on the initial guess. The assumption is that cells with larger prior runoff values have more uncertainty and require larger corrections. It can be shown that each entry in the Kalman gain K_t is the ratio of the squared prior runoff for the current cell

divided by the squared prior runoff of all cells that contribute to discharge at that gage and time (minus discharge at any upstream gages contributing to the current gage). If desired, one could exploit this to calculate the Kalman gain without needing to perform matrix inversions, which could be useful for very large or very high-resolution domains:

$$K_{i,j} = \begin{cases} \frac{x_i^2}{\sum_{k \in j} x_k^2} & \text{if cell } i \text{ is upstream of gage } j \\ -\frac{x_i^2}{\sum_{k \in j} x_k^2} & \text{if cell } i \text{ is downstream of gage } j \\ 0 & \text{if cell } i \text{ is disconnected from gage } j \end{cases}$$

where the sums are over the contributing grid cells to gage j.

4.2.4 The Y20 approach

Y20 modified the PW13 ISR method by relaxing the assumption of uncorrelated runoff errors. They used the following runoff error covariance model:

$$P_{i,j} = \rho_{L,T}(\delta_{i,j}, \tau_{i,j}) * \sigma_i * \sigma_j \tag{4.17}$$

where $P_{i,j}$ is the covariance between a runoff at a particular location and time i and runoff at another location and time j. $\rho_{L,T}(\delta, \tau)$ is a correlation function in terms of space and time lag δ and τ between x_i and x_j, and σ_i and σ_j are the standard deviations of the x_i and x_j. The form of $\rho(\delta, \tau)$ is given by

$$\rho_{L,T} = e^{-\delta_{i,j}/L} e^{-\tau_{i,j}/T} \tag{4.18}$$

where L and T are space and time decorrelation lengths that determine how quickly correlation decays with space and time. This is a stationary model, as L and T do not change in space or time. The standard deviations σ_i are assumed to be proportional to the mean prior runoff via $\sigma_i = \alpha \times x_i^-$, the same assumption used by PW13. This leads to the

following form for the Kalman gain:

$$K_t = \left(\alpha^2 \Phi_t \odot e^{-\Delta/L} \odot e^{-\mathcal{T}/T}\right) H^T \left[\alpha^2 H(\Phi_t \odot e^{-\Delta/L} \odot e^{-\mathcal{T}/T}) H^T + R\right]^{-1} \tag{4.19}$$

where $\Phi_t = x_t^- x_t^{-T}$ is the $n(s+1) \times n(s+1)$ matrix of prior runoff values, Δ is the distance matrix, and $\mathcal{T}$ is the time lag matrix. A potential drawback of this approach is that if the mean prior runoff is inappropriately small, then the method will place too much emphasis on the prior and underestimate the uncertainty or vice versa.

Although they consider measurement errors by adding random errors to the discharge time series, it is not clear whether Y20 use R corresponding to 30% relative discharge error or whether they continue to assume $R_t = 0$ in their calculation of the Kalman gain. They also do not report what proportionality constant α they use to relate runoff error standard deviation to the mean runoff value. It is possible that their favorable runoff estimates are a consequence of assuming $R = 0$ or using a very large runoff error proportionality constant (equivalent because as $\alpha \to \infty$, R becomes insignificant in Equation 4.19).

4.2.5 EnISR: A ensemble framework for ISR

Now that we have provided background on ISR and how it relates to the Kalman filter, we can introduce our "Ensemble ISR" method, a new runoff estimation algorithm that addresses three key problems with the ISR method. Although ISR has been shown to provide reasonable runoff estimates under idealized conditions, it still has limitations for general application, namely:

1. Assumes additive Gaussian runoff and discharge errors, which is unlikely to be the case in realistic scenarios

2. Although ISR is a data assimilation method, there has been no discussion of error propagation in past ISR papers

3. Extremely large covariance matrices must be stored, which makes it impractical to run ISR for large basins at high resolution.

Our proposed method, EnISR, uses an ensemble Kalman smoother update. This has three benefits:

1. Like Y20 ISR, EnISR uses a Kalman smoother formulation to produce a posterior estimate of runoff. However, unlike Y20 ISR, EnISR estimates the Kalman gain, which determines the update, based on the full distribution of the prior runoff, rather than just based on their mean and covariance. This makes EnISR applicable even when runoff and discharge errors are non-Gaussian and therefore not fully characterized by their mean and covariance.

2. EnISR makes error propagation trivial, as both runoff and downstream discharge errors can be simply estimated from the ensemble.

3. EnISR dramatically reduces storage requirements compared to past ISR implementations.

EnISR is based on the ensemble Kalman filter method ([Evensen, 2003]), a more flexible method compared to the Kalman filter (or Kalman smoother) that allows specification of non-Gaussian random variables and nonlinear state transition and measurement models, and has been shown to work well in many contexts, even in the presence of non-Gaussian state errors and/or measurement errors (see, e.g., [Moradkhani et al., 2005], [Andreadis et al., 2007], [De Lannoy et al., 2012]). An added benefit of an ensemble update is that we do not have to store the (very) large augmented state error-covariance matrix C_{x_t}. Our ISR approach can be thought of as a moving-window ensemble Kalman batch smoother (EnBS; [Margulis et al., 2015]).

To set up the method, we start with a prior ensemble $x_t^{(i)}$ for $i = [1, 2, \ldots, M]$. The larger the ensemble size (M) compared to the length of the state vector ($n(s+1)$), the more accurate

the estimates of the sample covariances used to perform the ensemble Kalman update. For this reason, we recommend choosing the smallest possible assimilation window $s = k+1$. As the number of ensemble members increases, the EnISR estimate will approach the Y20 ISR estimate, given the Y20 assumptions of additive Gaussian runoff and measurement errors are satisfied. Equation (4.7) shows the ensemble update equation:

$$x_t^{(i)+} = x_t^{(i)-} + \hat{K}_t \left(y_t + v_t^{(i)} - H x_t^{(i)-} + L x_{t-k}^{(i)} \right) \tag{4.20}$$

where $v_t^{(i)}$ are simulated measurement errors, which are required in order to prevent unrealistic collapsing of the ensemble ([Burgers et al., 1998]). (The formulation in Equation (4.20) assumes discharge measurements errors are additive.)

Measurement errors are described by the additive model given in Equation (4.21). Let our measurement of discharge y_t be a random vector that is related to the true discharge by

$$y_t = v_t + y_t^{true} \tag{4.21}$$

where $v_t \sim N(\mu_{v_t}, \Sigma_{v_t})$ is a Gaussian error term with mean m_{v_t} and covariance C_{v_t}. This is the same additive Gaussian model used for discharge measurement error in Y20. Discharge errors are assumed unbiased ($\mu_{v_t} = 0$), uncorrelated, and with standard deivation proportional to the discharge measurement ($\sigma_{v_t} = c_v y_t$), where σ_{v_t} is the standard deviation of discharge error for observation y_t and c_v is the percent discharge error (e.g. for 5% discharge error, $c_v = 0.05$). Since EnISR does not require additive Gaussian discharge errors, other forms could be used, as well.

We compute the updated runoff value in Equation (4.20) using Equation (4.22) to estimate the Kalman gain using the sample covariances in Equation (4.22), which are estimated from the prior runoff ensemble $x_t^{(i)-}$ and the ensemble of predicted measurements $\hat{y_t}^{(i)} = H x_t^{(i)-} + L X_{t-k}^{(i)}$.

$$\hat{K}_t = \hat{C}_{xy} \left(\hat{C}_{yy} + C_v \right)^{-1} \tag{4.22}$$

This can be compared with Equation (4.12) from our description of the Kalman filter above. $C_{xy} = P_t^- H^T$ is the cross covariance between the state and the predicted measurements and $C_{yy} = H P_t^- H^T$ is the covariance of the predicted measurements. Note that these equalities are only true if runoff and discharge errors are additive and Gaussian.

We also wish to comment on the length of the assimilation window, $s + 1$, a topic that has not seen sufficient attention in the ISR literature. PW13, F20, and Y20 have used a window length of $s = 2(k + 1)$. At the lower limit, the window length can be $s = k + 1$ without affecting ISR estimates. This works for PW13 ISR. For Y20 ISR, however, choosing an assimilation window that is too short can cut off the influence of the a priori runoff error correlation time T, which influences the Kalman gain and could artificially reduce the update provided by measurements at a past or future time steps. We recommend choosing an assimilation window that is as short as possible, in order to reduce the length of the state vector (therefore minimizing the computational burden associated with ISR) while ensuring that the window is wide enough such that a priori temporal correlations will be taken into consideration for the update. A shorter assimilation window also has benefits for EnISR, as the number of degrees of freedom involved in estimating the runoff error cross correlation matrix $C_{x,y}^-$ will be smaller, enabling better Kalman gain estimates with smaller ensemble sizes.

4.2.6 Localization

Like other ensemble-based assimilation methods ([Houtekamer and Mitchell, 2001]), Ensemble ISR can suffer from spurious correlations in the sample covariance matrices used to compute the Kalman gain (Equation 4.22). Localization methods ([Houtekamer and Mitchell, 2001]; [Hamill et al., 2001]) can reduce the effect of spurious correlations by superimposing a pro-

posed covariance structure (where variables that are farther apart have less correlation than variables that are closer together). This is known as covariance localization. However, covariance localization is not a good approach for ISR because there are often non-spurious long-range correlations stemming from the flow network configuration — covariance localization would remove the physically-based correlations along with the spurious correlations. Instead, we look to another class of localization methods, known as "domain localization" (see e.g. [Ott et al., 2004], [Hunt, 2007]). These methods update the state vector only over a limited spatial domain, assuming that observations outside the domain have little influence over the state vector. Here, we demonstrate a domain localization method applied to EnISR. By reducing the size of the state vector, the localization avoids spurious correlations, reduces storage requirements, and speeds up processing time.

The approach is as follows:

1. We start by sorting the gages from upstream to downstream by arranging them from smallest to largest drainage area.

2. Next, we define the ISR problem for the upstream-most gage. The size of this problem will be determined by n_{sub}, the number of cells in the sub-basin, $m_{sub} = 1$, the number of gages in the sub-basin, and $s_{sub} + 1$, the length of the assimilation window for the sub-basin.

3. Step through each assimilation window for the upstream-most sub-basin, solving for updated runoff at each grid cell and time step.

4. Move to the sub-basin with the next smallest drainage area. If there are any cells in this sub-basin whose runoff values have already been updated due to an upstream gage, these values are considered fixed and do not come into the state vector for the current sub-basin. They are used to compute the predicted measurements, but they are not included in the update equation.

5. Repeat Steps 2-4 for the remaining gages, moving from upstream to downstream.

4.3 Study Area and Experimental Design

4.3.1 The Allegheny River Basin

We perform numerical experiments in the Allegheny River Basin to explain results from past ISR implementations and implement our new EnISR framework. We chose the Allegheny Basin as our study area because it is located within the greater Ohio River basin, which has been the study area for past ISR papers and of several SWOT-based discharge retrieval studies (e.g. [Andreadis et al., 2007]; [Durand et al., 2010]), yet it is small enough to quickly and repeatedly run the various ISR algorithms (233 grid cells at 1/8°resolution, compared to 3681 grid cells for the Ohio Basin overall).

Figure 4.2 shows a map of the 33750 km^2 Allegheny Basin. We consider any grid cell with an upstream drainage area greater than 5000 km^2 to be a "river pixel" (shown in red in the figure). Two gage locations are singled out for closer analysis: the outlet gage (black) and an upstream gage (green) in the headwaters of the basin.

We obtain a priori discharge predictions by routing the TMPA runoff through our routing model (Equation 4.1). Figure 4.3a-b shows scatterplots of TMPA vs. NLDAS runoff and discharge. Overall, the TMPA runoff prior is slightly low biased compared to the NLDAS synthetic truth, with an overall mean of 1.0 mm/day, compared to 1.3 mm/day for the NLDAS runoff. Likewise, on average, the prior discharge is low-biased compared to the true discharge, with mean discharge at the outlet equal to 242 m^3/s for TMPA vs. 301 m^3/s for NLDAS. Figure 4.3c-d shows discharge time series at the outlet and at the upstream gage highlighted in Figure 4.2. The TMPA prior is smaller than the true runoff, with a notable exception on June 18, 2009 when the prior discharge spikes much higher than the true discharge.

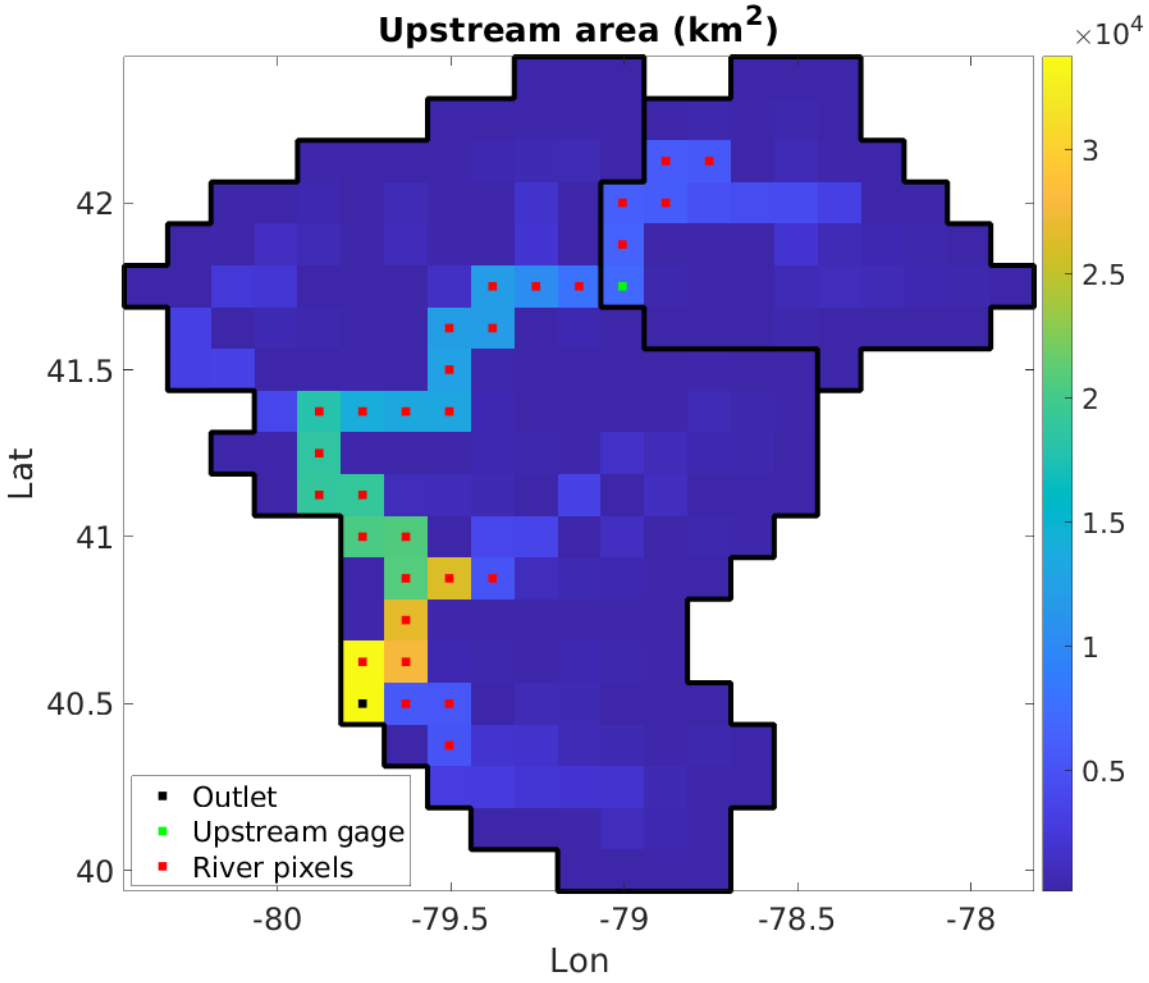

Figure 4.2: Map of the Allegheny River Basin as modeled by VIC at 1/8° resolution. The colormap indicates flow accumulation, the total upstream area of upstream cells contributing to current cell, with dark blue representing no upstream area and bright yellow the full basin upstream area, and corresponds to the theoretical river network.

We used GRWL ([Allen et al., 2018]) river widths to identify the width of each "river pixel" in the Allegheny Basin. The 14 downstream-most pixels are wide enough to be observed by SWOT (these are shown in a later figure). SWOT-based discharge estimates will be available each time the satellite passes over the river channel (about once every ten days for the Allegheny Basin). We assume here that discharge errors from SWOT retrieval algorithms

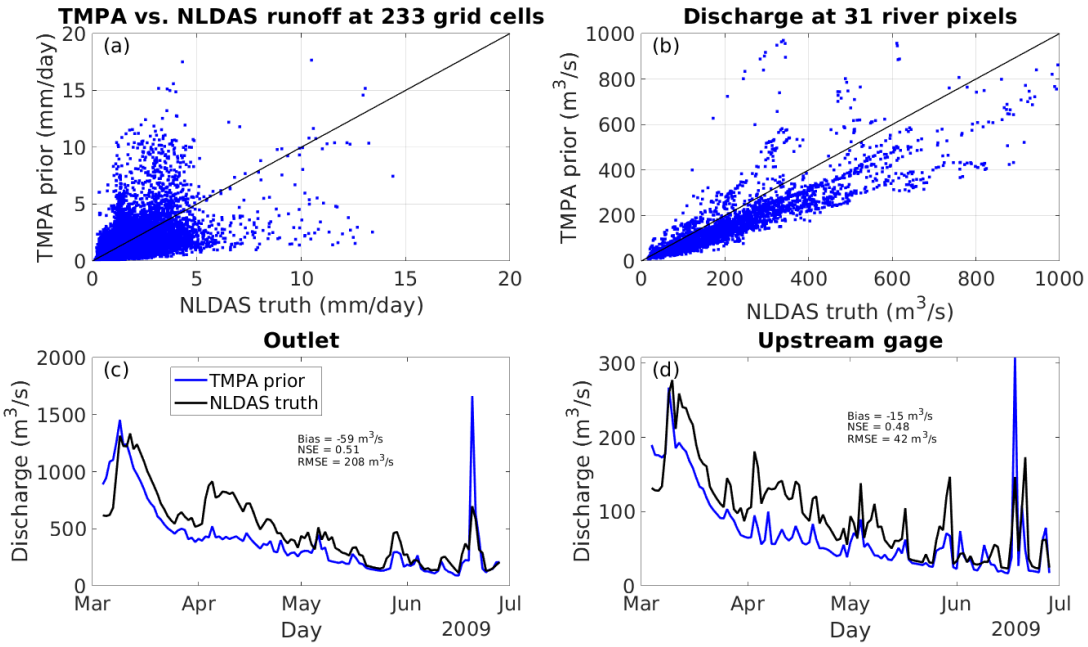

Figure 4.3: Panel (a) shows a scatterplot of predicted (TMPA) and true (NLDAS) runoff for each timestep and grid cell. Panel (b) shows a scatterplot of predicted (TMPA) vs. true (NLDAS) discharge at all river pixels. Panels (c) and (d) show discharge time series at the outlet and for the upstream gage.

are Gaussian with mean zero and standard deviation equal to 15% of the true discharge value. This is in line with expectations from recent SWOT publications ([Durand et al., 2023]). Discharge measurements may be SWOT-based discharge estimates or gage measurements. Throughout this paper, we assume that SWOT discharge measurements are available every 10 days with 15% error and that gage measurements are available daily with 5% error.

4.3.2 Experimental Design

We developed EnISR in order to address limitations of past ISR methodologies (PW13, Y20) and set the stage for future innovations: for example, relaxing the linear model requirement makes it possible to directly use a hydrological model as the forward model. We use a series

of observing system simulation experiments to compare the performance of EnISR with Y20 ISR, using the Allegheny River basin as a study area. We devised the first four experiments (E1-E4 in Table 4.1) to assess and compare the performance of EnISR vs. Y20 ISR under similar assumptions, where they are expected to perform similarly. The latter three experiments (E5-E7) provide insight on our localization method for reducing computational costs. The experiments are run either for the "gage" case, with daily streamflow at the outlet or for the "SWOT" case, with SWOT-like streamflow "measurements" at virtual gages for river pixels wider than 100 m, every 10 days, which is similar to the frequency with which SWOT will observe the Allegheny basin. For the gage case, we use a gage at the outlet as the "calibration" gage, whose measurements we use to estimate runoff via ISR, and we set aside an upstream gage (see Figure 4.2) as a "validation" gage, to check how well the ISR runoff estimates reproduce discharge at the upstream gage. For the SWOT case, we use 14 virtual gage locations, where SWOT discharge estimates are expected to be available. These SWOT virtual gages, which are shown in a later figure, consist of the 14 downstream-most "river pixels" shown in Figure 4.2 and do not include the aforementioned upstream validation gage.

Table 4.1: Summary of numerical experiments performed in this paper. All experiments are run from Mar-Jun, 2009. Prior mean runoff is set equal to the TMPA runoff, while runoff error covariance parameters (Equation 4.17) are described in the table.

Case	ISR method	Measurements, freq., error	Prior runoff error correlation
E1	Y20	gage at outlet, daily, 5%	L = 40 km, T = 5 days
E2	EnISR	gage at outlet, daily, 5%	L = 40 km, T = 5 days
E3	Y20	SWOT, 10-day, 15%	L = 40 km, T = 5 days
E4	EnISR	SWOT, 10-day, 15%	L = 40 km, T = 5 days
E5	EnISR	Outlet, upstream gage, or both, 5%	L = 0 km, T = 0 days
E6	EnISR	Outlet, upstream gage, or both, 5%	L = 40 km, T = 5 days
E7	EnISR with localization	SWOT, 10-day, 15%	L = 40 km, T = 5 days

For the Y20 ISR cases, mean prior runoff is set equal to the TMPA runoff, and the runoff errors are assumed to come from a covariance model (Equation 4.17) where errors are implicitly assumed to be correlated with spatial and temporal correlation lengths equal to 40 km and 5 days, respectively, and the standard deviation of prior runoff errors is assumed proportional to the prior mean runoff value (we used proportionality constant $\alpha = 1$). Y20 found these spatial and temporal correlation parameters produced good discharge estimates in the Ohio Basin, using the same TMPA-NLDAS experimental setup used here. For EnISR, we used the same prior runoff error covariance model to generate random errors with which we corrupted the TMPA runoff to produce a prior runoff ensemble with additive Gaussian errors. We limited our analysis to the 120-day period from March 1-June 28, 2009 for ease of computation.

In Sections 4.1 and 4.2, we seek to understand how our assumption of proportional runoff error variance affects ISR runoff estimates and how and when ISR runoff estimates can become negative. In Section 4.3, we compare EnISR to Y20 ISR under similar assumptions, in order to show that the new method is equivalent to the existing method when runoff and discharge errors are additive Gaussian. In Section 4.4, we demonstrate our localization method and seek to understand how our domain localization method will affect the ISR estimates.

4.4 Results and Discussion

4.4.1 How runoff error model assumptions affect ISR estimates

Viewed as a data assimilation algorithm, ISR updates an a priori "guess" of runoff based on discharge observations, by weighting the relative uncertainty between the predicted and actual discharge measurements. Since there are many more unknowns (runoff at each grid cell and time) than measurements (discharge at each gage), there are many possible combinations of runoff that can produce the same discharge. ISR uses the routing model (Equation 4.1)

to assign runoff to upstream grid cells in order to get a match to discharge (or as close to a match as is appropriate given errors in the discharge measurements). If we assume that runoff errors are correlated, as did Y20, then runoff in cells nearby a gage (and times nearby a measurement time) will be influenced by measurements at that gage, even if they are not hydrologically connected to the gage by the flow network.

We focus here on the effect of the assumption of proportional variance; that is, PW13 and Y20 both take the standard deviation of prior runoff error ($\sigma_{x_t}^-$) to be proportional to the mean prior runoff value. This prior uncertainty, along with the flow network and a priori correlations, dictates where ISR adjusts runoff from the prior to the posterior. Grid cells and times with larger a priori runoff values are considered more uncertain, and may receive a larger update — potentially leading to degraded estimates if the initial estimate was close to the truth to begin with. Grid cells and times with small a priori runoff are considered well-known, and receive little update.

Figure 4.4a shows the mean difference between true and prior runoff. Cells with prior runoff larger than true runoff are red, and cells with prior runoff smaller than true runoff are blue. Figure 4.4b shows the change in Nash Sutcliffe Efficiency (NSE, [Nash and Sutcliffe, 1970]) ($\Delta NSE = NSE_{post} - NSE_{prior}$). NSE measures the goodness of fit between two data series and ranges from $-\infty$ to 1, with NSE=1 indicating a perfect match and NSE=0 implying that the model is no better than using the mean of the data as the estimate. Therefore, when $\Delta NSE > 0$, we say that runoff estimates have improved, and when $\Delta NSE < 0$, we say that runoff estimates have degraded. Cells where prior runoff is larger than true runoff (red) in Figure 4.4a, for example, Cell 1, tend to experience decreases in NSE (blue) in Figure 4.5b. Cells where prior runoff is much smaller than true runoff (dark blue) in Figure 4.4a, for example, Cell 3, tend to see little change in NSE (white) in Figure 4.5b.

The assumption of proportional runoff variance gives rise to three possible situations, which are illustrated by Figure 4.5, which plots runoff time series for each of the three cells indicated in Figure 4.4. The posterior runoff appearing in Figure 4.5 is not produced in a

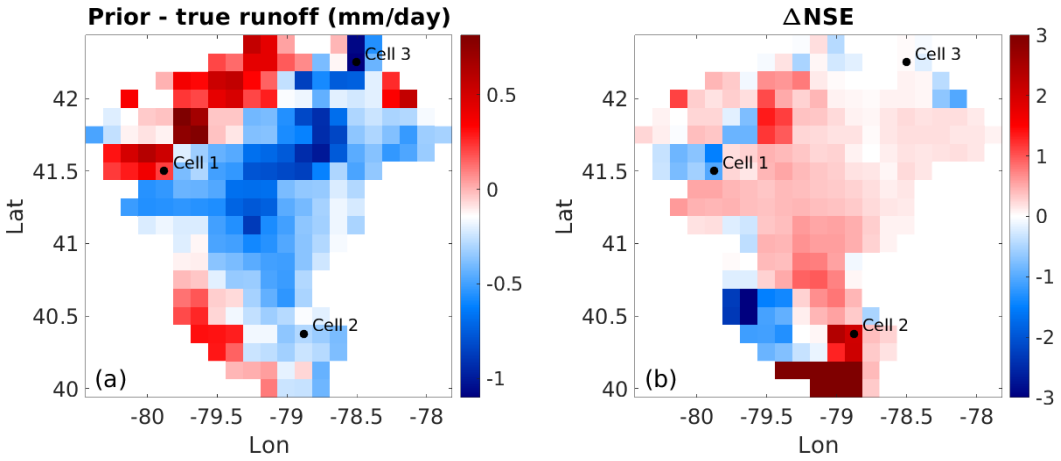

Figure 4.4: Runoff estimation results from experiment E1 (Y20, gage). Panel (a) shows temporal average prior runoff bias, calculated as $E[x_t^- - x_t]$. Panel (b) shows change in NSE for each grid cell, calculated as $NSE_{post} - NSE_{prior}$. When $\Delta NSE > 0$, ISR's runoff estimate is better than the prior. When $\Delta NSE < 0$, ISR's runoff estimate is worse than the prior. Colorbars have been fixed at the limits shown in order to resolve spatial variability more clearly.

vacuum — rather, all cells contributing to discharge at the gage are updated together, as a group.

Since ISR is a one-to-many problem, where one gage records runoff generated by many upstream cells, there is no guarantee that the correct grid cells will be updated in the correct direction. The proportion of the update that goes to a particular grid cell depends on its prior uncertainty, compared to the uncertainty associated with prior runoff estimates in other grid cells. The direction of the update is such that the posterior discharge will more closely match the measured discharge. So if prior runoff at a cell is large, it can be updated upward or downward depending on if measured discharge is larger or smaller (respectively) than the a priori discharge estimate.

For example, if there are two grid cells contributing runoff to the outlet, and both are located the same distance (for the sake of argument, assume they are equally distance along

the river network and also in terms of absolute distance) away from the gage, the grid cell with a higher prior runoff (and therefore more uncertain, according to the proportional variance assumption) will receive a larger update, in the direction required to make posterior discharge close to observed discharge (within measurement error) than the cell with the smaller (and therefore considered more certain) prior runoff.

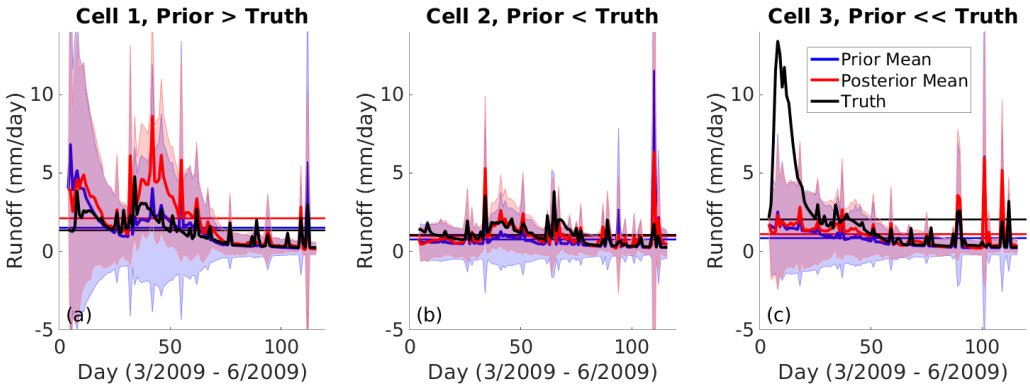

Figure 4.5: Time series of prior, true, and posterior (estimated using Y20 ISR) runoff, along with shaded 95% uncertainty bounds, for the three cells indicated in Figure 4.4. Vertical axis limits have been truncated to show details. The horizontal lines in the runoff time series plots show mean prior, posterior, and true runoff for each cell in blue, red, and black, respectively. Note that the runoff uncertainty bounds for the Y20 estimates come from the Kalman update (Equation 4.11) — whereas analytical propagation of runoff uncertainty to discharge uncertainty is not straightforward.

In Cell 1, the prior runoff is higher, on average, than the true runoff: the assumption of proportional variance results in an inappropriately large update to runoff in this cell, which makes the posterior runoff estimate worse than the prior, with lower NSE. The most common case is that of Cell 2. In this grid cell, the mean prior runoff is smaller than the true runoff, yet not so small that we underestimate its uncertainty (the true runoff falls within 2 standard deviations of the prior, such that $x_t^- - 2\sigma_{x_t}^- \leq x_t \leq x_t^- + 2\sigma_{x_t}^-$). This allows ISR to provide a useful update, with an improved match to true runoff. Finally, for Cell 3, prior runoff is

much lower, on average, than the true runoff: the proportional variance assumption places a lot of confidence in the prior estimate, and there is very little change in NSE. Table 4.2 gives an overview of the three cases illustrated by Cells 1-3 in Figure 4.5, along with ΔNSE values for each case.

Table 4.2: Average change in pixel-wise NSE (posterior - prior) for each prior runoff scenario, as illustrated by Cells 1-3 in Figure 4.5. Expected ΔNSE indicates our expectation, based on the discussion in the text, of whether there will be an improvement in NSE from the prior to the posterior estimate. In general, we expect a different runoff estimation outcome depending on whether the prior mean runoff is larger than, smaller than, or much smaller than true runoff. *Note that when prior mean > truth, runoff estimates can be, but are not always, degraded.

Prior vs. truth	Example grid cell	Expected ΔNSE	ΔNSE
Prior mean > Truth	Cell 1	<0	0.17*
Prior mean < Truth	Cell 3	>0	0.58
Prior mean << Truth	Cell 4	0	0.15

This is captured in more detail in the left-hand panel of Figure 4.6, where we plot absolute runoff for the posterior runoff estimates vs. absolute prior runoff error for each grid cell and time. The colors in the scatterplot denote instances where prior runoff is much smaller than true runoff (blue), instances where prior runoff is larger than true runoff (red), and other instances (cyan). When prior runoff is small (blue), posterior runoff is not much improved over the prior, while when prior runoff is large (red), it is more likely that there will be significant change: either improvement or degradation. We define "small" prior runoff (blue) such that the true runoff does not lie within 2 standard deviations of the prior mean, and we define large runoff (red) as instances where prior runoff is larger than true runoff, noting that for proportional error, it is impossible for prior runoff to be "too high" in the sense that the true runoff will always fall within 2 standard deviations of the prior mean. Runoff at

other cells and times (cyan) is smaller than the true runoff, but is still a "reasonable" guess in the sense that the true runoff falls within 2 standard deviations of the prior mean.

The histograms on the right-hand side of Figure 4.6 further illustrate this point, with posterior absolute error ($\hat{x}_t - x_t$) increasing by 0.22 mm/day for instances with large prior runoff. Grid cells and times with "small" prior runoff see a decrease in error of 0.34 mm/day. That this occurs despite the small prior runoff uncertainty for these grid cells likely is related to the values of uncertainty at other nearby grid cells. Gage discharge error is assumed to be fairly low (5%), so if all contributing grid cells have small prior uncertainty, then they may still receive an update. Other instances (when prior runoff is neither larger nor much smaller than true runoff) see a decrease in error of 0.12 mm/day.

In summary, under the proportional variance assumption, we expect the greatest reduction in runoff errors when prior runoff is smaller than true runoff, yet not so small that prior runoff uncertainty does not encompass true runoff (however, even in this scenario, a small runoff prior can lead to reduced runoff errors in the posterior estimate, if other contributing cells also have small prior runoff). When prior runoff is large, these grid cells and times are more volatile — when there is an update due to a mismatch between prior and observed discharge, these instances will receive the bulk of the update, even if the prior estimate was already close to the truth.

4.4.2 Occurrence of negative runoff values in ISR estimates

ISR as implemented by PW13, F20, and Y20 can produce physically unrealistic negative runoff estimates. Although these do not occur frequently (no more than 1.5% of runoff values were negative in our experiments), we wish to understand why negative estimates occur and what can be done to prevent them. Before considering the Allegheny case, let us consider an idealized basin with two grid cells (with runoff values $x_{t,1}$ and $x_{t,2}$) and one gage at the outlet (with discharge y_t). This lets us hone in on the circumstances under which negative runoff values can arise. The Y20 ISR update equation is

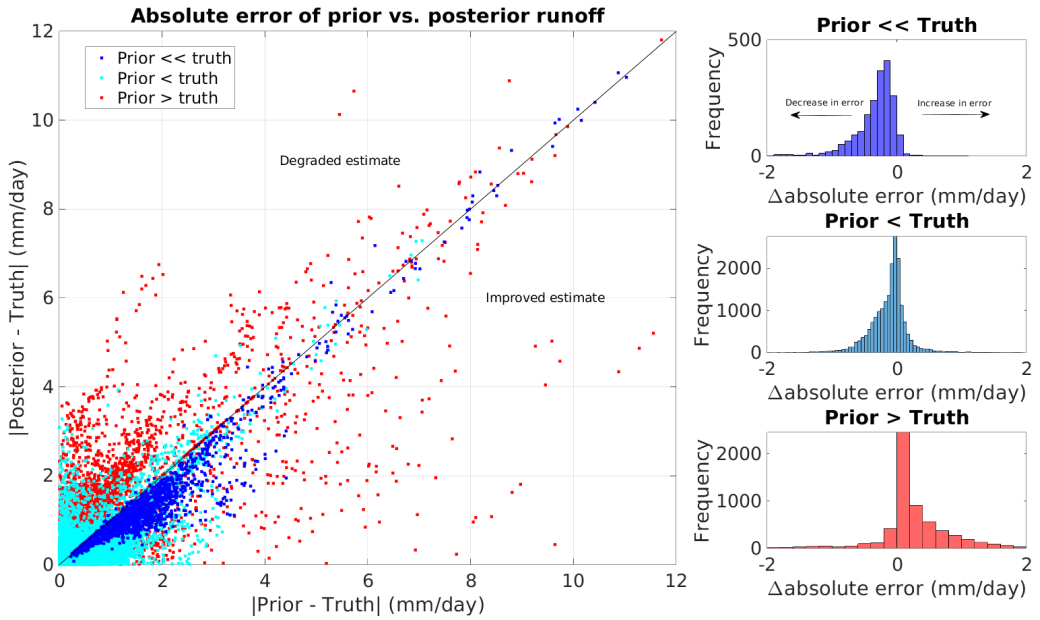

Figure 4.6: Left-hand plot shows absolute error of prior and posterior runoff, calculated as shown on the axis labels, for each grid cell and time in the E1 experiment (120 time steps, 233 grid cells) in the Allegheny basin. Instances where prior runoff is greater than the true runoff are shown in red, instances with prior runoff smaller than true runoff, but still capturing true runoff within their 2 standard deviation uncertainty bounds are shown in cyan, and instances with prior runoff much smaller than the true are shown in blue. Points to the right of the 1:1 line have an improved runoff estimate, while points to the left of the 1:1 line have a degraded runoff estimate. The histograms on the right-hand side change in absolute error (Δ absolute error) corresponding to each set of points on the scatterplot. Δ absolute error< 0 indicates that posterior errors have been reduced from the prior.

$$x_t^+ = x_t^- + K_t(y_t - \hat{y}_t) \tag{4.23}$$

where x_t^+ where is the posterior mean runoff vector, x_t^- is the prior mean runoff vector, K_t is the Kalman gain, y_t is the discharge measurement, and $\hat{y}_t$ is the predicted measurement. Written for one cell only, and assuming uncorrelated prior runoff errors for simplicity

(equivalent to PW13's ISR implementation), the update equation is

$$x_{t,1}^{+} = x_{t,1}^{-} + \frac{\sigma_{t,1}^2}{\sigma_{t,1}^2 + \sigma_{t,2}^2}(y_t - \hat{y}_t) \tag{4.24}$$

where $\sigma_{t,1}^2$ is the prior error variance for cell 1 and $\sigma_{t,2}^2$ is the prior error variance for cell 2. Note that we have assumed perfect measurements to simplify the expression for the Kalman gain. By no means does this update equation guarantee non-negativity. If we replace the expressions for prior error variance with the variance model proposed by PW13 (which states that runoff error standard deviation is proportional to the prior mean runoff), we obtain

$$x_{t,1}^{+} = x_{t,1}^{-} + \frac{x_{t,1}^2}{x_{t,1}^2 + x_{t,2}^2}(y_t - \hat{y}_t) \tag{4.25}$$

where $x_{t,1}^{2-}$ indicated the squared prior mean runoff. This equation shows that two conditions are required to produce negative runoff estimates. First, the predicted measurement $\hat{y}_t$ must be greater than the observation y_t in order to make the second term in Equation (4.25) negative — when the prior runoff (and consequently discharge) has a high bias. Second, the Kalman gain must not be so small such that $K_t(y_t - \hat{y}_t)$ is small compared to $x_{t,1}^{-}$. These conditions are likely to be satisfied when observed discharge y_t is small and when $x_{t,1}^{-}$ is either close to 0 or large compared to $x_{t,2}^{-}$.

In Figure 4.7, we show posterior mean runoff estimates for cell 1 in the idealized two-cell basin as a function of the mean prior runoff estimate in each grid cell. In the left-hand panel, negative runoff estimates occur when $x_{t,1}^{-}$ is larger than $x_{t,2}^{-}$ because the observed discharge is smaller than the predicted measurement. In the right-hand panel, there are no negative runoff estimates because y_t is larger, and the $(y_t - \hat{y}_t)$ term is not large enough to produce negative runoff estimates within the range of prior mean runoff values shown.

In general, negative runoff is likely to occur at a grid cell if that grid cell has a high prior runoff compared to other contributing cells and the measured discharge at the gage is relatively low. This fact should be considered when choosing a runoff prior. For example, the

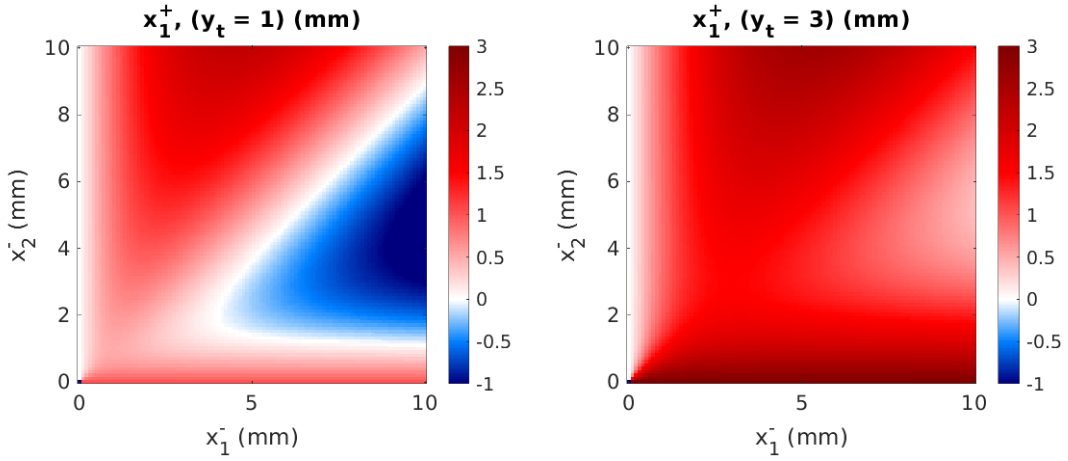

Figure 4.7: Shows the posterior mean runoff estimate for a single grid cell ("grid cell 1" in an idealized basin with two grid cells) as a function of prior runoff at each cell. The left-hand panel shows posterior runoff when the observed discharge is low ($y_t = 1$ mm/day), and the right-hand panel shows posterior runoff when the observed discharge is high ($y_t = 3$ mm/day). The colorbar shows the posterior mean runoff estimate (in mm/day) for grid cell 1, with blue indicating negative posterior runoff estimates.

TMPA-based prior uses TMPA precipitation data to drive a hydrological model and produce runoff estimates. The prior runoff will roughly correspond to storm tracks (for rainfall-runoff events) or snowy areas (for snowmelt runoff events). Due to Y20's assumption of proportional runoff error variance, having areas of unusually high runoff in one part of the basin on a given day can cause these grid cells to get a disproportionately large update compared to other cells, potentially going into the negative.

Negative runoff estimates occurred in the Allegheny basin on June 18, 2009, when the TMPA runoff prior was disproportionately high in the south portion of the basin. Runoff snapshots on this date (Figure 4.8) show that the 95th percentile a priori runoff estimate on this date is 10.7 mm, while the median runoff is only 0.36 mm. These high prior runoff values result in a prior discharge estimate that is much higher than the measured discharge. There is therefore a need for a significant negative update to the state vector. The grid cells

with a much higher prior runoff (and therefore uncertainty) than others will receive the vast majority of the runoff update, resulting in negative posterior runoff estimates in these cells to make up for the more "certain" runoff values in cells with lower prior mean runoff.

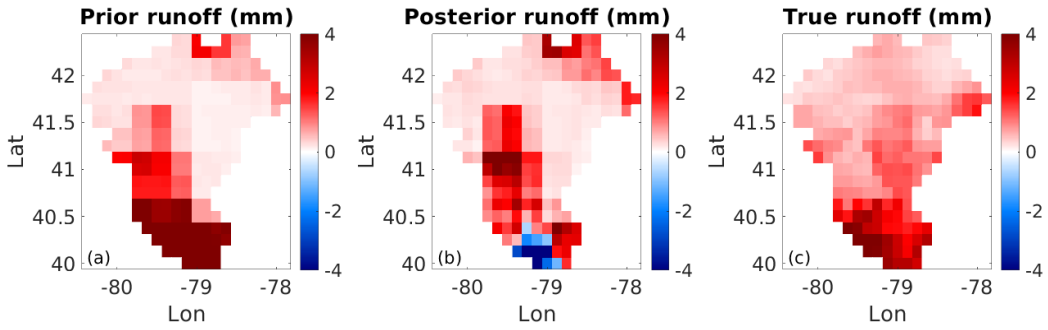

Figure 4.8: Snapshot on June 18, 2009 of TMPA runoff prior, posterior mean runoff, and true runoff from experiment E1 (Y20 ISR with daily gage measurements at the outlet). Very large prior runoff values in the southern portion of the basin (compared to other cells) lead to negative posterior runoff estimates there. Colorbar limits have been truncated to show spatial variability.

While negative runoff estimates rarely occurred in the Allegheny basin experiments (TMPA is overall low biased in the Allegheny basin, so there are few times when a a large downward adjustment of runoff is needed), we would like to propose options for handling negative runoff values should they occur. One approach is to run ISR with a so-called noninformative prior (PW13, [Fisher et al., 2020]), where prior runoff is set equal to some presumed long-term mean value. Since negative runoff estimates occur due to large spatial variations between prior runoff in neighboring cells, this guarantees that there will be no negative runoff estimates, at the cost of not using all available a priori runoff information and only relying on discharge measurements. Other possible approaches include using a constrained Kalman filter ([Simon, 2010]) or Gaussian anamorphosis techniques (e.g. performing the update in log space, as in ([Girotto et al., 2014]), and the Particle Smoother ([Margulis et al., 2015]). We performed experiments with updates in log space in order to

avoid negative runoff estimates (not shown), but we found that there are trade-offs (associated with violations of the Kalman Filter's Gaussianity assumption) in terms of optimality of the update.

4.4.3 Comparison of EnISR vs. Y20 ISR

We next demonstrate that given the same set of inputs (TMPA prior with Gaussian errors), EnsISR is equivalent to Y20 ISR. Both methods, Y20 and EnsISR, use the TMPA runoff prior used in past ISR papers (PW13, F20, Y20). We perform two sets of experiments: one with daily "gage" discharge and one with 10-day "SWOT" discharge measurements. Runoff and discharge estimates are compared to synthetic (NLDAS) truth. As shown in Table 4.1, experiments E1 and E3 use the Y20 ISR method, while experiments E2 and E4 use EnISR.

Figures 4.9 and 4.10 show prior and posterior discharge estimates at the outlet (the calibration gage) and at an upstream gage (a validation gage, which is not used for the ISR estimate). Gage locations are shown in the study area map (Figure 4.2). As expected, Figures 4.9 and 4.10 show that the EnISR and Y20 prior and posterior mean estimates are nearly identical — we attribute any differences to error in the EnISR ensemble representation of the prior and posterior runoff distributions. Importantly, EnISR, but not Y20 ISR, provides uncertainty bounds on the discharge estimates, a key piece of information for assessing real-world ISR implementations where true discharge data are not available.

Posterior discharge estimates at the outlet have very tight uncertainty bounds in Figure 4.9, corresponding to the daily high-accuracy discharge measurements at every time step at that gage. Figure 4.10 shows higher uncertainty for discharge estimates at the upstream validation gage, since there are many possible combinations of runoff that can produce the same discharge at the outlet, and not all of them are a good match for discharge at the upstream validation gage. In both cases, the true discharge data fall within the range of the posterior discharge uncertainty bounds.

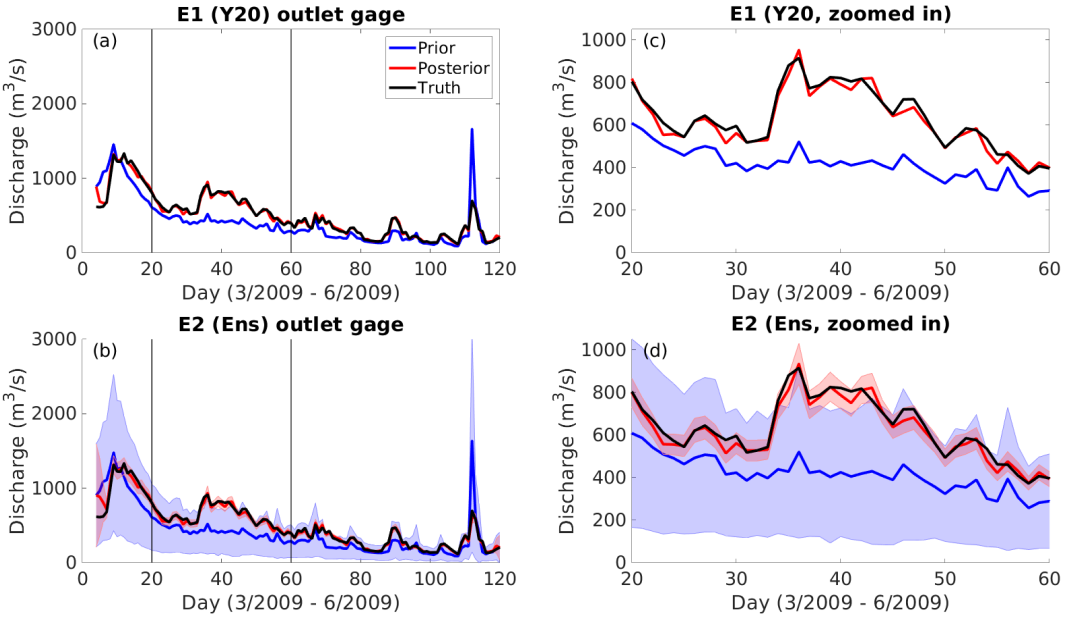

Figure 4.9: Discharge estimates at the Allegheny basin outlet (calibration gage) with daily, 5% error, discharge measurements corresponding to experiments E1 (using Y20 ISR) and E2 (using EnISR). Bold lines indicate the mean and shaded areas indicate 95% confidence intervals for EnISR (Y20 ISR does not produce discharge uncertainty estimates). Panels (a) and (b) show discharge estimates over the whole study period. Panels (c) and (d) showed a zoomed in view from days 20-60 - March 20 to April 29, 2009.

Figures 4.11 and 4.12 show Y20 and EnISR runoff estimates for Experiments E3 and E4, where we assume that discharge measurements are available every 10 days at all river pixels with channel width $>$ 100m (shown in Figure 4.13), with 15% relative error (the "SWOT" case). As in the "gage" case, we show posterior discharge estimates for the outlet and for an upstream validation gauge. Again, we see a reduction in error compared to the prior discharge estimates, both at the outlet and at the upstream validation gage, with uncertainty reduced most at observation times (denoted by gray vertical lines in Figures 4.11 and 4.12). The NSE of the mean discharge estimate at the outlet increases from 0.51 to 0.77 for both Y20 ISR and EnISR, with the time-average standard deviation of the discharge estimate

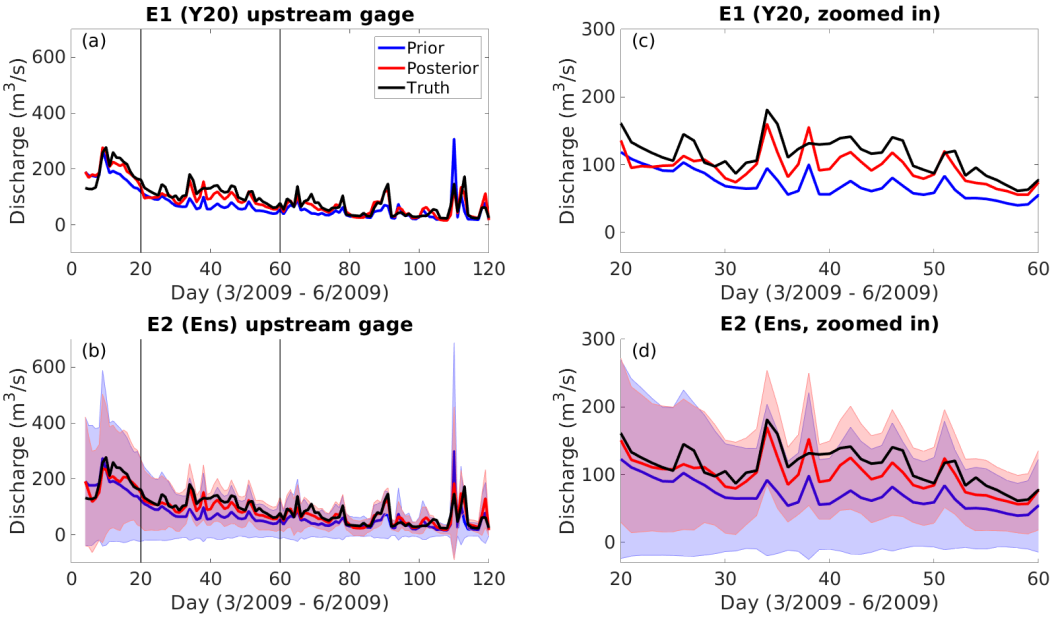

Figure 4.10: Discharge estimates at an upstream validation gage (not used as an input for ISR) with daily, 5% error, discharge measurements corresponding to experiments E1 (using Y20 ISR) and E2 (using EnISR). Bold lines indicate the mean and shaded areas indicate 95% confidence intervals for EnISR (Y20 ISR does not produce discharge uncertainty estimates). Panels (a) and (b) show discharge estimates over the whole study period. Panels (c) and (d) showed a zoomed in view from days 20-60 - March 20 to April 29, 2009.

decreasing 30% from 156 m^3/s to 109 m^3/s, while the NSE of the validation gage increases from 0.48 to 0.75, with discharge standard deviation decreasing from 44 m^3/s to 34 m^3/s.

Discharge uncertainty estimates are shown here by 95% confidence intervals about the mean discharge estimate. Uncertainty is reduced at measurement times, but increases as time passes between measurements, eventually reverting to the prior. Our ability to estimate discharge uncertainty is unique to EnISR — it is straightforward to estimate the variance of the discharge estimate by applying the method of moments to the prior or posterior discharge ensemble. For Y20 ISR, the propagation of error from runoff to discharge is a challenge due to the lag terms in the routing model.

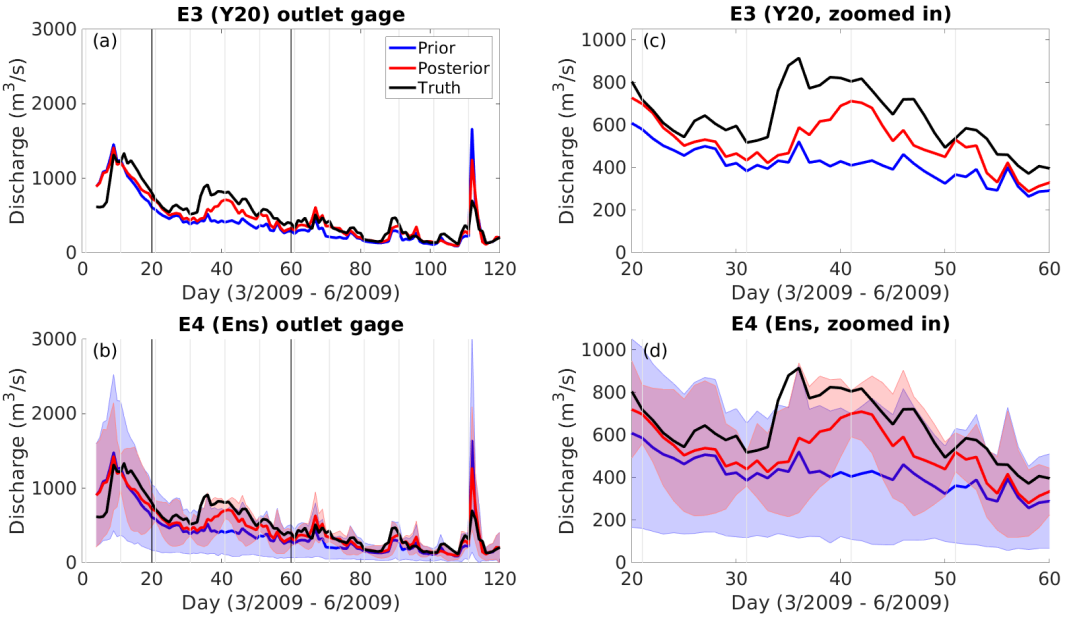

Figure 4.11: Discharge estimates at the Allegheny basin outlet, with 10-day, 15% error, discharge measurements ("SWOT case") corresponding to experiments E3 and E4. Bold lines indicate the mean and shaded areas indicate 95% confidence intervals. Vertical lines (gray) show measurement times. Panels (a) and (b) show discharge estimates over the whole study period. Panels (c) and (d) showed a zoomed in view from days 20-60 - March 20 to April 29, 2009.

The outlet and upstream gage shown in Figures 4.9 to 4.12 are illustrative of overall improvement in discharge estimates from the prior to the posterior. Table 4.3 shows the improvement in discharge NSE at the 31 "river pixels" (shown in the study area map Figure 4.2) for both the gage case and the SWOT case for experiments E1-E4. In each case, posterior NSE increased compared to the prior, from 0.36 to 0.88 for the gage case (E1-2) and from 0.36 to 0.75 for the SWOT case (E3-4).

We also compare runoff estimates between Y20 ISR and EnISR. Median runoff NSE increases significantly from 0.06 to 0.34 for Y20 ISR (E1 and E2). For the SWOT case, on the other hand, median runoff NSE only increases from 0.06 to 0.12 for Y20 ISR (E3)

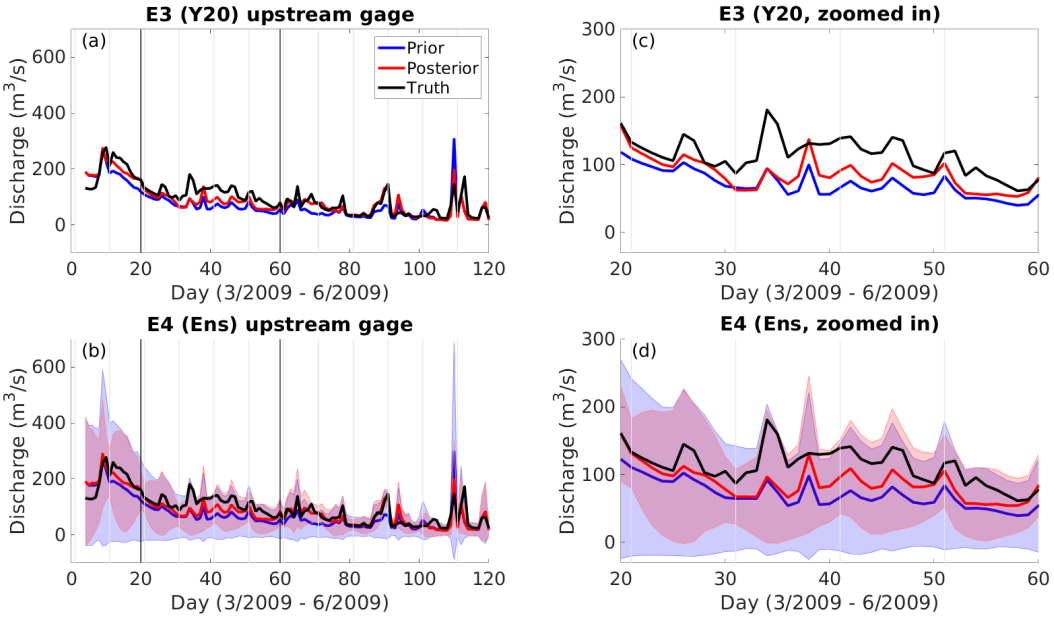

Figure 4.12: Discharge estimates at an upstream validation gage, which was not used as input for ISR, with 10-day, 15% error, discharge measurements ("SWOT case") corresponding to experiments E3 and E4. Bold lines indicate the mean and shaded areas indicate 95% confidence intervals. Vertical lines (gray) show measurement times. Panels (a) and (b) show discharge estimates over the whole study period. Panels (c) and (d) showed a zoomed in view from days 20-60 (March 20 to April 29, 2009).

Table 4.3: Prior and posterior discharge NSE, calculated for both the gage and SWOT measurement cases for experiments E1-4.

Experiment	Prior NSE	Posterior NSE
E1 (Y20, gage)	0.36	0.88
E2 (EnISR, gage)	0.36	0.87
E3 (Y20, SWOT)	0.36	0.75
E4 (EnISR, SWOT)	0.36	0.75

and 0.11 for EnISR (E4). We computed NSE by comparing the true runoff and the mean estimated runoff at each grid cell. Figure 4.13a-c shows prior and posterior runoff NSE maps. We show only the results for EnISR in the figure; results for Y20 ISR are essentially identical.

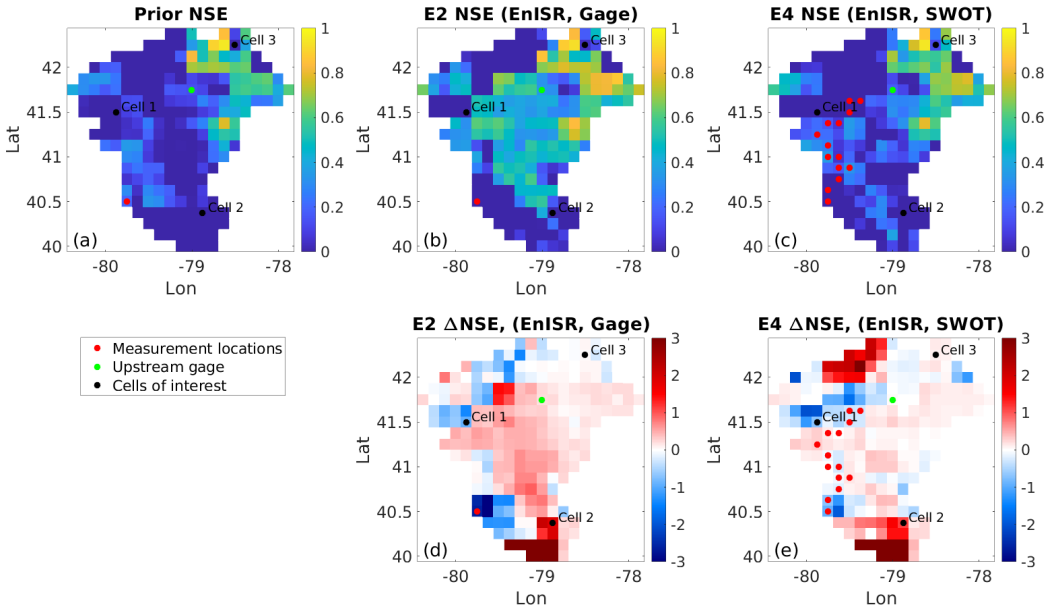

Figure 4.13: Maps of mean runoff NSE for experiments E2 (EnISR with gage measurements) and E4 (EnISR with SWOT measurements). The top row shows NSE, the bottom row shows change in NSE ($\Delta NSE = NSE_{post} - NSE_{prior}$). Red dots show either gage locations (panels (b) and (d)) or SWOT measurement locations (panels (c) and (e)). The green dot shows an upstream validation gage, which is not used as an input for ISR. Black dots show grid cells of interest.

Finally, we look at estimates of runoff in space and time. In Figure 4.5, we saw how grid cells with prior runoff larger than true runoff tended to have degraded posterior runoff estimates, cells with prior runoff somewhat less than true runoff tended to have improved posterior runoff estimates, and cells with prior runoff much smaller than true runoff tended to have no update. This result is borne out for EnISR as well as Y20 ISR, with identical

runoff time series for Experiments E1 and E2. Figure 4.14 shows EnISR runoff estimation results for the three grid cells with varying runoff priors for the SWOT measurement case (E4); the posterior runoff time series for each cell identical to the Y20 estimates in Experiment E3 (not shown). Runoff uncertainty is still reduced relative to the prior, particularly around measurement times, but there is less reduction in runoff uncertainty than in the daily measurement case.

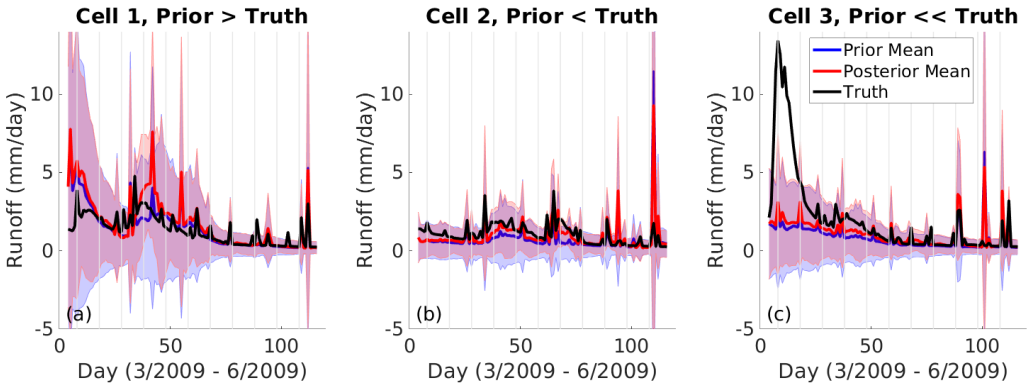

Figure 4.14: Prior and posterior runoff time series for two cells indicated in Figure 4.13. Solid lines indicate means and shaded areas indicate 95% confidence bounds. Gray vertical lines show measurement times.

Even though measurements are available only once every 10 days, the effect of the update lasts beyond the immediate observation time due to the assumption of a priori temporal error correlation of $T = 5$ days. It might seem unexpected that the SWOT scenario performs worse than the gage scenario, given that both experiments use a similar number of measurements. The SWOT scenario uses $14 \times 12 = 168$ measurements compared to 120 measurements at the outlet in the gage scenario. In both cases, this is a highly underconstrained problem, since there are 120×233 unknowns (runoff at each time and grid cell). Since the SWOT virtual gages are so close together (SWOT only observes rivers wider than 100 m), the discharge measurements at each gage are highly correlated with one another, providing similar information, and in this case, the 14 gages likely act more like one gage, providing only $1 \times 12 = 12$ independent discharge measurements. The higher measurement error (15%

vs. 5%) for the SWOT scenario plays a role, as well. In larger river basins where rivers wider than 100 m occur in a greater portion of the basin, SWOT measurements should be expected to provide more improvement in ISR runoff estimates.

In this section, we have shown that EnISR produces equivalent runoff and discharge estimates to Y20 ISR under similar assumptions of additive Gaussian errors, with the advantage of improved uncertainty characterization in EnISR. The ISR methods improved runoff and particular discharge estimates at ungaged locations, with discharge NSE increasing from 0.36 to 0.88 and 0.75 for the gage (outlet) and SWOT measurement cases, respectively. In future work, we could envision using EnISR to update an ensemble of a priori runoff estimates generated by forcing a hydrologic model, such as VIC, with error-corrupted precipitation and other input data. Although we use additive Gaussian runoff errors here in order to compare with Y20, as an ensemble data assimilation method EnISR can accommodate the non-Gaussian runoff errors expected to arise from this method of analysis.

4.4.4 EnISR: Addressing computational challenges

4.4.4.1 Constraints on memory and computation time

With larger basins and higher spatial resolutions, both Y20 and EnISR require increasing amounts of computational resources. Computational challenges take two forms: storage limitations and computation time. The augmented runoff error covariance matrix C_{x_t} can require a very large amount of RAM to store: its size increases with the square of the augmented runoff vector. For example the Ohio River basin, has a travel time of 16 days (PW13) and therefore a minimum assimilation window time of 17 days. At 1/8°resolution, there are 3681 grid cells. The length of the augmented state vector is $17 \times 3681 = 62577$, so C_{x_t} will be 62577×62577. Assuming each number is stored with double precision, this will require about 31 GB of RAM. The main computation time limitation is the time it takes to invert $HC_{x_t}^{-}H^T$ to calculate the Kalman gain. This $m(s+1) \times m(s+1)$ matrix tends to

be much smaller than C_{x_t} because the number of gages m is much smaller than the number of cells n. However, with hundreds of potential virtual gages from SWOT, the computation time associated with this matrix inversion could quickly grow.

Memory constraints are less of an issue for EnISR. Under EnISR, there is no need to store the runoff error covariance matrix. Instead, we calculate the approximate Kalman gain as the product of

$$\hat{K}_t = \hat{C}_{xy}(\hat{C}_{yy} + C_v)^{-1} \tag{4.26}$$

The largest matrix that needs to be stored is the runoff-discharge cross covariance matrix $\hat{C}_{xy}$, which is estimated from the ensemble of runoff and discharge vectors. Its size ($n(s+1) \times m(s+1)$) would be about 640 MB for the Ohio basin example above, assuming $m = 75$ gages (as in the Y20 study). However, we should note that generating an ensemble of correlated errors to produce the prior runoff ensemble can be computationally intensive. Depending on the error model and how the errors are generated, this can potentially require storage of the full runoff error covariance matrix for the entire length of the study period. We recommend using a random number generator, such as the copula-based random number generator CoSMoS ([Papalexiou and Serinaldi, 2020]), that does not require storage of C_x^+, to generate the correlated errors.

4.4.4.2 A localization method for applying ISR to high dimensional basins

ISR operates on an augmented state vector whose length ($n(s+1) \times n(s+1)$) can be very long. This leads to the computational challenges noted in the previous section and also means that a larger number of ensemble members are needed to accurately estimate the Kalman gain for EnISR. We can address both of these challenges using localization methods.

As described in Section 2.6, our localization method splits the basin into sub-basins, with one sub-basin per gage. (It is also possible to apply this localization to sub-basins with more

than one gage, at the cost of needing to store larger error covariance matrices.) Each sub-basin n_{sub} grid cells and travel time k_{sub}. Assimilation is done for each sub-basin, starting with the upstream-most sub-basin and continuing downstream to the outlet. It is important to understand whether there will be any drawbacks to subdividing the basin in this way. For simplicity, let us examine the idealized case where the runoff error covariance matrix is diagonal so that there is no a priori correlation between runoff error at one grid cell and time and another grid cell and time. Then the Kalman gain $K_t = PH^T(HPH^T + R)^{-1}$ only updates runoff at cells that are upstream of the gage. This information is embedded in the routing model operator H. However, if we assume that there is spatial or temporal correlation in the runoff, then cells that are upstream of the gage will receive an update, and also cells that are nearby to those cells in space or time. In this case, there is a trade-off between reducing computational demands and using all available information to estimate runoff. The portion of the update that cells near the sub-basin boundary would have received due to their correlation with cells in the neighboring sub-basin will be lost, with more information lost as the basin is broken down into progressively smaller sub-basins.

As a proof of concept for the localization method, we selected two virtual gages in the Allegheny basin, one at the outlet, and one in the headwaters (Figure 4.2), dividing the basin into two sub-domains. (Experiments E5-E6 in Table 4.1.) We ran EnISR three times: once with only the upstream gage, once with only the downstream gage, and once with both gages. Experiments E5 and E6 use the TMPA runoff prior and daily discharge measurements. In E5, we assume there is no a priori runoff correlation; this is considered an ideal scenario for localization, since there will be no loss of information across sub-basin boundaries. In E6, we use a spatial and temporal correlation parameters $L = 40$ km and $T = 5$ days.

The time series plots in Figure 4.15 illustrate how runoff in the upstream and downstream sub-basins can, to an extent, be estimated in a separate steps without loss of information. In the top row of Figure 4.15, we show sub-basin average runoff time series for E5, with no prior runoff error correlation. The top row of Figure 4.15 shows that runoff in the downstream sub-

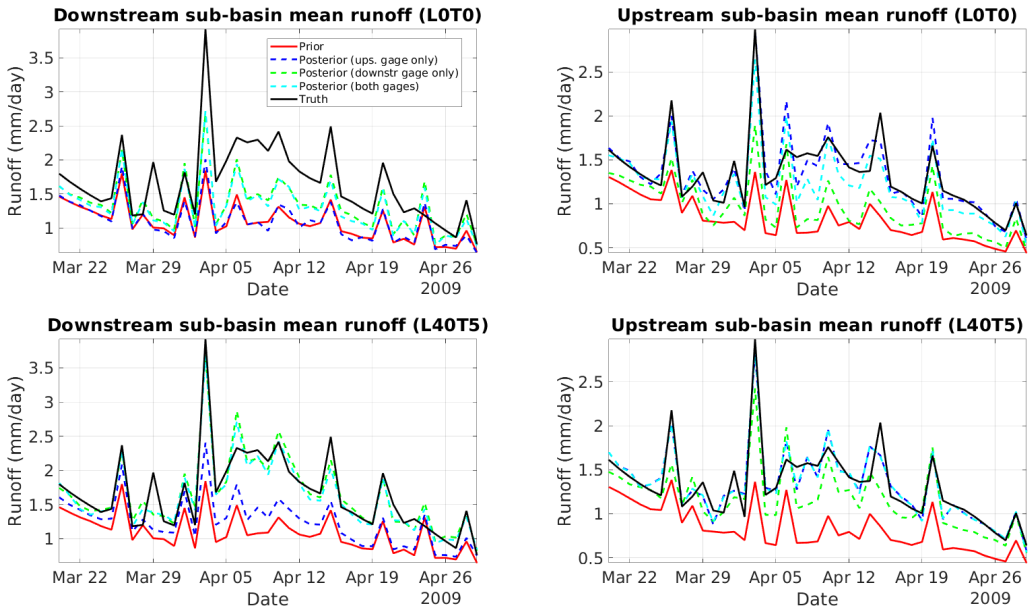

Figure 4.15: Sub-basin mean runoff estimates in the upstream and downstream sub-basins of the divided Allegheny basin. Top row shows results from experiment E5, with no prior runoff error correlations. Bottom row shows results from experiment E6, with correlation parameters $L = 40$ km, $T = 5$ days. Posterior mean estimates from EnISR with one gage at the outlet, one upstream gage, or both are shown, along with true and prior mean runoff.

basin is entirely unaffected by discharge measurements at the upstream gage, while posterior runoff is the upstream basin is the same regardless of whether or not the downstream gage is present. The bottom row of Figure 4.15 shows that when prior runoff correlation is present, runoff in the downstream sub-basin is affect by measurements at the upstream gage. Table 4.4 shows runoff NSE changes for each sub-basin average runoff time series in Figure 4.15. Note that there is no change in runoff NSE ($\Delta NSE = -0.02$) in the downstream basin given measurements at the upstream gage in E5, but there is an increase in NSE for E6, where there is a priori spatiotemporal correlation.

In experiment E7, we apply EnISR with localization to the Allegheny basin for the SWOT observation scenario, using the same setup as Experiment E4, but using domain localization,

Table 4.4: NSE improvement $\Delta NSE = NSE_{post} - NSE_{prior}$ for the upstream and downstream Allegheny sub-basins. Experiments E5 and E6 use the TMPA runoff prior and daily discharge measurements, with correlation parameters $L = 0$, $T = 0$ for E5 and $L = 40$ km, $T = 5$ days for E6.

Sub-basin ΔNSE	Upstream sub-basin	Downstream sub-basin
E5, upstream	0.38	-0.02
E5, outlet	0.19	0.24
E5, both	0.36	0.23
E6, upstream	0.35	0.15
E6, outlet	0.3	0.32
E6, both	0.35	0.32

splitting the basin into 14 sub-basins, one for each gage. While localization decreased the length of the state vector by 80%, the median NSE barely increased (from 0.06 to 0.07). Fortunately, we are still able to reduce uncertainty in prior discharge estimates and improve our estimates of discharge in between SWOT measurements, as can be seen in Figures 4.16a and 4.16b, where discharge NSE increases from 0.57 to 0.69 at the outlet, with the 95% confidence interval reduced 16% from 156 to 131 m^3/s. Overall, the mean discharge NSE at the 31 river pixels in Figure 4.2 increases from 0.35 to 0.57.

Due to the localization, which reduces the amount of influence the a priori runoff error correlations have on the update, the updated discharge moves back toward the prior more quickly here than in the non-localized case (shown in Figures 4.2c and 4.2d). Users should keep in mind this significant tradeoff in accuracy compared to the full, non-localized EnISR method. One possible solution would be to split the basin into a smaller number of sub-basins (with sub-basin size specified to satisfy RAM requirements), rather than using one gage for each sub-basin. We do not test this approach here, but it is likely that such a hybrid method could be used to navigate RAM constraints while minimizing loss due to a priori

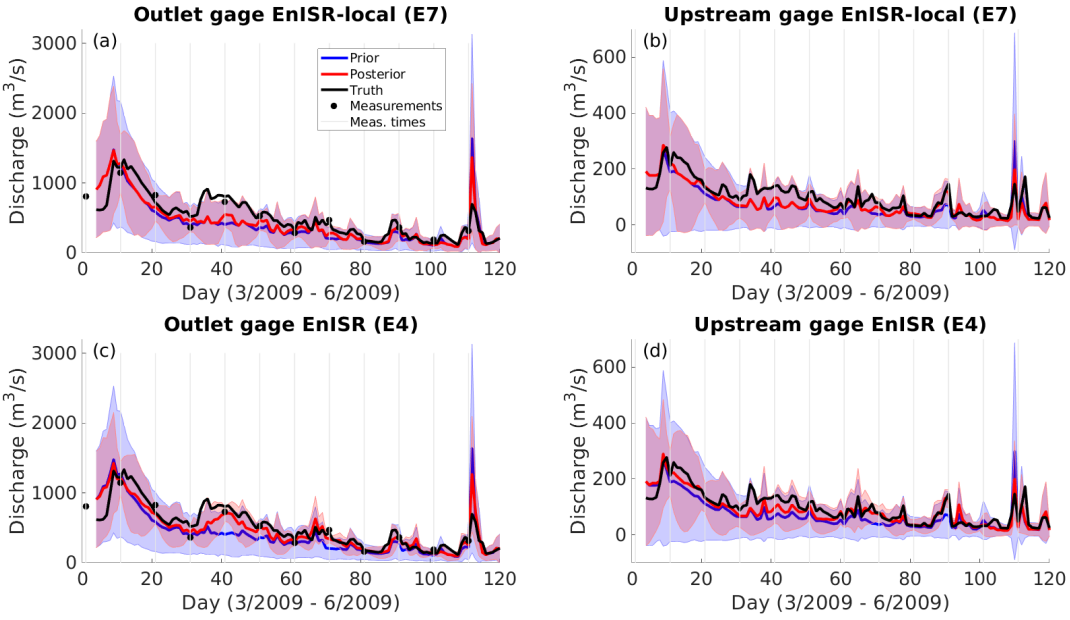

Figure 4.16: Discharge estimates at the outlet and upstream gage with 10-day, 15% error, discharge measurements ("SWOT case") corresponding to experiments E7 (using domain localization) and E4 (without localization).

runoff correlations in one sub-basin not affecting runoff estimates in nearby sub-basins.

4.5 Conclusions

We developed an alternative formulation for inverse streamflow routing based on the Ensemble Batch Smoother. This data assimilation-based approach has computational advantages over existing ISR algorithms, provides uncertainty quantification, and builds flexibility into a methodological framework by providing a path forward for a conventional data assimilation approach using a nonlinear forward (hydrological) model to produce the prior runoff estimate (although we leave implementation of such a model to future work). We also developed a domain localization method that can dramatically reduce the computational expense of ISR, making it practical to implement for large basins with many gage locations.

We showed that EnISR can match the performance of Y20 ISR under similar assumptions, with the added benefits of producing runoff discharge uncertainty estimates and using less computer memory due to its ensemble formulation. For the TMPA runoff prior and daily discharge measurements at the basin outlet, EnISR increased the median NSE of runoff at a grid cell from 0.06 to 0.34. For the SWOT-like measurement case, with higher errors and less frequent measurements, runoff NSE did not increase as much (from 0.06 to 0.11) but the method still has potential for larger basins, where SWOT measurements will be spread out and provide more spatial coverage, possibly in conjunction with gage measurements. Both methods produce greatly improved discharge estimated at ungaged basins, with NSE increasing from 0.36 to 0.88 for the gage case and 0.75 for the SWOT case.

ISR can be performed at any resolution where flow direction maps are available. Thus, runoff can be estimated at spatial scales limited only by the resolution of the underlying elevation data used to derive the flow network maps.

ISR-derived runoff estimates have utility in three realms:

1. Runoff estimates for understanding the water balance at varying scales
2. Interpolating discharge where gages are unavailable
3. Calibrating hydrological models to runoff vs. traditional discharge-based calibration

With the advent of SWOT-based discharge estimates, EnISR would be a prime candidate for generating spatiotemporally continuous discharge estimates from sparse SWOT virtual gages. EnISR represents a pathway for using SWOT to produce continental- to global-scale discharge estimates, along with uncertainty quantification. Future work could also use existing stream gage datasets, such as GRDC ([GRDC, 2023]) or the USGS's GAGES-II ([Falcone, 2011]) to produce discharge data at ungaged locations. The high temporal frequency and relative accuracy of gage discharge estimates could also be used to make runoff estimates in basins around the world, which could in turn be used as an alternative

method for hydrological model calibration. Calibrating hydrological models to runoff at individual grid cells offers an advantage over traditional discharge-based calibration, which can introduce spatial discontinuities in calibrated parameters. Also, calibrating one grid cell at a time makes it possible to use parallel computing to speed up processing with the potential to calibrate each grid cell on a different processor.

CHAPTER 5

Conclusions and Future Work

5.1 Summary and Original Contributions

We estimated bathymetry using satellite-based measurements of river width and water surface elevation. The method works by fitting simple mathematical relationships to river width and water surface elevation observations and extrapolating to the hypothetical center of the channel. We showed that the nonlinear model based on hydraulic geometry relationships outperformed a linear model for estimating the unobserved channel depth. Our method provides a way to estimate the cross sectional area of flow below the minimum observed water surface elevation, a critical parameter for the Manning's equation-based SWOT discharge retrieval algorithm.

The Variable Infiltration Capacity (VIC) model is a widely used hydrological model, but setting up input files for a new region (particularly outside the continental United States) is an arduous process. We produced a 1/16°resolution dataset of soil, vegetation, and other inputs to the VIC model and made them publicly available for other researchers to use freely. This dataset is considerably higher resolution than previously published datasets (the only other global VIC input dataset is at 2°resolution), allowing researchers to better resolve spatial variability in water and energy states and fluxes at the land surface.

Lastly, we developed a runoff estimation method based on the Inverse Streamflow Routing (ISR) method of Pan and Wood (2013) that accounts for errors in the runoff initial guess and measurements and propagates those errors to the final estimate. Our method, "Ensemble ISR

(EnISR)" which is based on the Ensemble Kalman Batch Smoother, is more computationally efficient than past ISR methods, provides runoff and discharge uncertainty estimates, and can estimate runoff without requiring Gaussian runoff errors, which could be a key consideration when applying ISR with a runoff prior generated from, e.g., an ensemble of hydrologic models. We show when and where we expect posterior estimates to be improved (or degraded) vs. the prior, and we demonstrate that EnISR is equivalent to past ISR methods under similar assumptions, in synthetic experiments with both in-situ (gage) and remote sensing-based (SWOT) discharge measurements.

5.2 Potential for Future Work

The SWOT satellite launched in 2022 will produce a 3-year record of water surface elevation measurements, which can be used to estimate discharge via various discharge retrieval algorithms ([Durand et al., 2016]). These "virtual gauge" measurements, along with streamflow measurements from in-situ gauges, can (1) be used to estimate streamflow at unobserved locations via the "two-sweep ISR" method, and (2) produce spatiotemporally continuous runoff maps, along with their associated uncertainty estimates, using our Ensemble ISR algorithm. The runoff prior for the Ensemble ISR algorithm could be produced using the VIC model with our VICGlobal parameters.

One exciting potential use of these runoff maps is for calibration of distributed hydrological models. Hydrologic model calibration is challenging for three reasons:

1. Discharge is a spatially-aggregated variable. Calibrating to a spatially-aggregated variable leaves out valuable information about spatially-distributed model parameters, and creates the potential to get "the right answer for the wrong reason" due to parameter equifinality.

2. Parameters calibrated to discharge tend to have discontinuities from one basin to an-

other.

3. Automated calibration methods require running the hydrologic model hundreds or thousands of times, which is computationally prohibitive for high resolution or large-scale models.

Calibrating to spatially-distributed runoff, rather than spatially-aggregated discharge, could solve each of these challenges. In particular, for a semi-distributed hydrologic model such as VIC, that can be run for individual grid cells, the calibration problem is easily parallelized by running (and calibrating) each grid cell on a separate processor.

REFERENCES

[Adam et al., 2009] Adam, J. C., Hamlet, A. F., and Lettenmaier, D. P. (2009). Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes*, 23(7):962–972.

[Allen et al., 2018] Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., and Famiglietti, J. S. (2018). Global Estimates of River Flow Wave Travel Times and Implications for Low-Latency Satellite Data. *Geophysical Research Letters*, 45(15):7551–7560.

[Andreadis et al., 2007] Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., and Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, 34(10):1–5.

[Aronica et al., 1998] Aronica, G., Hankin, B., and Beven, K. (1998). Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources*, 22(4):349–365.

[Bárdossy et al., 2022] Bárdossy, A., Kilsby, C., Birkinshaw, S., Wang, N., and Anwar, F. (2022). Is Precipitation Responsible for the Most Hydrological Model Uncertainty? *Frontiers in Water*, 4(March):1–17.

[Bates and De Roo, 2000] Bates, P. and De Roo, A. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, 236(1):54–77.

[Beven, 2006] Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320:18–36.

[Beven, 2012] Beven, K. (2012). Rainfall-Runoff Modelling. In *Rainfall-Runoff Modelling*, chapter 1.4 Percep. John Wiley & Sons, Ltd, Oxford, UK, 2nd edition.

[Biancamaria et al., 2010] Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodriguez, E., Mognard, N. M., Alsdorf, D. E., Lettenmaier, D. P., and Oudin, Y. (2010). Preliminary characterization of SWOT hydrology error budget and global capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1):6–19.

[Biancamaria et al., 2016] Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M. (2016). The SWOT Mission and Its Capabilities for Land Hydrology. *Surveys in Geophysics*, 37(2):307–337.

[Bjerklie et al., 2018] Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., and Garambois, P.-A. (2018). Satellite remote sensing estimation of

river discharge: Application to the Yukon River Alaska. *Journal of Hydrology*, 561(March 2017):1000–1018.

[Bohn and Vivoni, 2019] Bohn, T. J. and Vivoni, E. R. (2019). MOD-LSP, MODIS-based parameters for hydrologic modeling of North American land cover change. *Scientific Data*, 6(1):144.

[Bohn et al., 2019] Bohn, T. J., Whitney, K. M., Mascaro, G., and Vivoni, E. R. (2019). A deterministic approach for approximating the diurnal cycle of precipitation for use in large-scale hydrological modeling. *Journal of Hydrometeorology*, 20(2):297–317.

[Bonnema et al., 2016] Bonnema, M., Sikder, S., Hossain, F., Durand, M. T., Gleason, C. J., and Bjerklie, D. M. (2016). Benchmarking wide swath altimetry-based river discharge estimation algorithms for the Ganges river system. *Water Resources Research*, 52:1–20.

[Brackins et al., 2018] Brackins, J., Moragoda, N., and Rahman, A. (2018). An Approach for Incorporating Realistic Channel Geometry into Continental-Scale Hydrological Modeling. In Aristizabal, F., Grimley, L. E., Bales, J., Tijerina, D., Flowers, T., and Clark, E. P., editors, *National Water Center Innovators Program Summer Institute Report 2018*.

[Brisset et al., 2018] Brisset, P., Monnier, J., Garambois, P. A., and Roux, H. (2018). On the Assimilation of Altimetric Data in 1D Saint-Venant River Flow Models. *Advances in Water Resources*.

[Bruhwiler et al., 2005] Bruhwiler, L. M., Michalak, A. M., Peters, W., Baker, D. F., and Tans, P. (2005). An improved Kalman smoother for atmospheric inversions. *Atmospheric Chemistry and Physics*, 5(10):2691–2702.

[Buffington, 2012] Buffington, J. M. (2012). Changes in Channel Morphology Over Human Time Scales. In Church, M., Biron, P. M., and Roy, A. G., editors, *Gravel-Bed Rivers*, pages 433–463. John Wiley & Sons, Ltd, Chichester, UK.

[Burgers et al., 1998] Burgers, G., Van Leeuwen, P. J., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724.

[Canova et al., 2016] Canova, M., Fulton, J., and Bjerklie, D. M. (2016). USGS HYDRoacoustic dataset in support of the Surface Water Oceanographic Topography satellite mission (HYDRoSWOT).

[Cherkauer, 1999] Cherkauer, K. A. (1999). GLDAS vegetation parameters.

[Chow, 1959] Chow, V. (1959). *Open channel hydraulics*. McGraw-Hill Book Co., New York.

[Chow et al., 1988] Chow, V. T., Maidment, D., and Mays, L. (1988). *Applied Hydrology*. McGraw-Hill Book Co.

[Cosby et al., 1984] Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R. (1984). A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resources Research*, 20(6):682–690.

[David et al., 2019] David, C. H., Hobbs, J. M., Turmon, M. J., Emery, C. M., Reager, J. T., and Famiglietti, J. S. (2019). Analytical Propagation of Runoff Uncertainty Into Discharge Uncertainty Through a Large River Network. *Geophysical Research Letters*, 46(14):8102–8113.

[David et al., 2011] David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., and Eijkhout, V. (2011). River network routing on the NHDPlus dataset. *Journal of Hydrometeorology*, 12(5):913–934.

[De Lannoy et al., 2012] De Lannoy, G. J., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E., and Pauwels, V. R. (2012). Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*, 48(1):1–17.

[Demaria et al., 2007] Demaria, E. M., Nijssen, B., and Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research Atmospheres*, 112(11):1–15.

[D'Errico, 2012] D'Errico, J. (2012). inpaint_nans.

[Didan et al., 2015] Didan, K., Munoz, A. B., Solano, R., and Huete, A. (2015). MODIS Vegetation Index User's Guide (MOD13 Series) Version 3.0 Ccollection 6). 2015(May):38.

[Dorman and Sellers, 1989] Dorman, J. and Sellers, P. J. (1989). A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere Model (SiB). *J. Appl. Meteor.*

[Duan et al., 1992] Duan, Q., Sorooshian, S., and Gupta, V. (1992). Effective and Efficient Global Optimization. *Water Resources Research*, 28(4):1015–1031.

[Ducoudré et al., 1993] Ducoudré, N. I., Laval, K., and Perrier, A. (1993). SECHIBA, a New Set of Parameterizations of the Hydrologic Exchanges at the Land-Atmosphere Interface within the LMD Atmospheric General Circulation Model. *Journal of Climate*, 6(2):248–273.

[Durand et al., 2016] Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., Rodriguez, E., Bates, P. D., Pavelsky, T. M., Monnier, J., Chen, X., Di Baldassarre, G., Fiset, J. M., Flipo, N., Frasson, R. P., Fulton, J., Goutal, N., Hossain, F., Humphries, E., Minear, J. T., Mukolwe, M. M., Neal, J. C., Ricci, S., Sanders, B. F., Schumann, G., Schubert, J. E., and Vilmin, L. (2016). An intercomparison of remote

sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, 52(6):4527–4549.

[Durand et al., 2023] Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., Altenau, E. H., Tebaldi, N., Larnier, K., Monnier, J., Malaterre, P. O., Oubanas, H., Allen, G. H., Astifan, B., Brinkerhoff, C., Bates, P. D., Bjerklie, D., Coss, S., Dudley, R., Fenoglio, L., Garambois, P., Getirana, A., Lin, P., Margulis, S. A., Matte, P., Minear, J. T., Muhebwa, A., Pan, M., Peters, D., Riggs, R., Sikder, M. S., Simmons, T., Stuurman, C., Taneja, J., Tarpanelli, A., Schulze, K., Tourian, M. J., and Wang, J. (2023). A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission. *Water Resources Research*, 59(4):1–19.

[Durand et al., 2014] Durand, M., Neal, J., Rodríguez, E., Andreadis, K. M., Smith, L. C., and 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 et al., 2010] Durand, M., Rodriguez, E., Alsdorf, D., and Trigg, M. (2010). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1):20–31.

[Emery et al., 2020] Emery, C. M., David, C. H., Andreadis, K. M., Turmon, M. J., Reager, J. T., Hobbs, J. M., Pan, M., Famiglietti, J. S., Beighley, E., and Rodell, M. (2020). Underlying fundamentals of kalman filtering for river network modeling. *Journal of Hydrometeorology*, 21(3):453–474.

[Evensen, 2003] Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367.

[Fahnestock, 1963] Fahnestock, R. K. (1963). Morphology and hydrology of a glacial stream - White River, Mount Rainier Washington. *Geological Survey Professional Paper*, 422-A:1–70.

[Falcone, 2011] Falcone, J. A. (2011). GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow. Technical report, USGS, Reston, VA.

[Fick and Hijmans, 2017] Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

[Fisher et al., 2020] Fisher, C. K., Pan, M., and Wood, E. F. (2020). Spatiotemporal assimilation-interpolation of discharge records through inverse streamflow routing. *Hydrology and Earth System Sciences*, 24(1):293–305.

[Frasson et al., 2017] Frasson, R. P. d. M., Wei, R., Durand, M., Minear, J. T., Domeneghetti, A., Schumann, G., Williams, B. A., Rodriguez, E., Picamilh, C., Lion, C., Pavelsky, T., and Garambois, P. A. (2017). Automated River Reach Definition Strategies: Applications for the Surface Water and Ocean Topography Mission. *Water Resources Research*, 53(10):8164–8186.

[Friedl and Sulla-Menashe, 2015] Friedl, M. and Sulla-Menashe, D. (2015). MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG.

[Garambois and Monnier, 2015] Garambois, P. A. and Monnier, J. (2015). Inference of effective river properties from remotely sensed observations of water surface. *Advances in Water Resources*, 79:103–120.

[Gelaro et al., 2017] Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454.

[Gessese et al., 2013] Gessese, A., Wa, K. M., and Sellier, M. (2013). Bathymetry reconstruction based on the zero-inertia shallow water approximation. *Theoretical and Computational Fluid Dynamics*, 27(5):721–732.

[Gessese, 2011] Gessese, A. F. (2011). Reconstruction of river bed topography from free surface data using a direct numerical approach in one-dimensional shallow water flow Reconstruction of river bed topography from free surface data using a direct numerical approach in one-dimensional shallo.

[Girotto et al., 2014] Girotto, M., Cortés, G., Margulis, S. A., and Durand, M. (2014). Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern river watershed, Sierra Nevada. *Water Resources Research*, 50(8):6713–6734.

[Gleason et al., 2018] Gleason, C. J., Garambois, P. A., and Durand, M. T. (2018). Tracking river flows from space. *Eos*, 99(1):32–36.

[Gleason and Hamdan, 2017] Gleason, C. J. and Hamdan, A. N. (2017). Crossing the (watershed) divide: satellite data and the changing politics of international river basins. *Geographical Journal*, 183(1):2–15.

[Gleason and Smith, 2014] Gleason, C. J. and Smith, L. C. (2014). Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, 111(13):4788–4791.

[Gochis et al., 2016] Gochis, D. J., Dugger, A., and McCreight, J. (2016). Technical Description of the National Water Model Implementation WRF-Hydro.

[Gou et al., 2020] Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J., and Zhou, R. (2020). Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China. *Water Resources Research*, 56(1):1–19.

[GRDC, 2023] GRDC (2023). The global runoff data centre.

[Grimaldi et al., 2018] Grimaldi, S., Li, Y., Walker, J. P., and Pauwels, V. R. (2018). Effective Representation of River Geometry in Hydraulic Flood Forecast Models. *Water Resources Research*, 54(2):1031–1057.

[Gupta et al., 2009] Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91.

[Hagemann et al., 2017] Hagemann, M. W., Gleason, C. J., and Durand, M. T. (2017). BAM: Bayesian AMHG-Manning Inference of Discharge Using Remotely Sensed Stream Width, Slope, and Height. *Water Resources Research*, 53(11):9692–9707.

[Hall and Riggs, 2015] Hall, D. K. and Riggs, G. (2015). MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG, Version 6.

[Hamill et al., 2001] Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11):2776–2790.

[Hamman et al., 2018a] Hamman, J., Nijssen, B., Bohn, T., Franssen, W., Gergel, D. R., Mao, Y., Helgason, H., Peng, B., and Craig, T. (2018a). Variable Infiltration Capacity (VIC) Macroscale Hydrologic Model.

[Hamman et al., 2018b] Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., and Mao, Y. (2018b). The variable infiltration capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geoscientific Model Development*, 11(8):3481–3496.

[Hannah et al., 2011] Hannah, D. M., Demuth, S., Lanen, H. A. J. V., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M. (2011). Large-scale river flow archives : importance , current status and future needs. *Hydrological Processes*, 1200(May 2010):1191–1200.

[Hansen et al., 2000] Hansen, M. C., Defries, R. S., Townshend, J. R. G., and Sohlberg, R. (2000). *Global land cover classification at 1 km spatial resolution using a classification tree approach*, volume 21.

[HEC-RAS, 2008] HEC-RAS (2008). HEC-RAS: Hydraulic Reference Manual Version 4.0.

[Heuvelink and Webster, 2001] Heuvelink, G. B. and Webster, R. (2001). Modelling soil variation: Past, present, and future. *Geoderma*, 100(3-4):269–301.

[Hoffman, 2016] Hoffman, H. (2016). soil_classification.

[Houtekamer and Mitchell, 2001] Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137.

[Huffman et al., 2007] Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55.

[Hunt, 2007] Hunt, B. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126.

[Kalman, 1960] Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35.

[Kidd and Huffman, 2011] Kidd, C. and Huffman, G. (2011). Global precipitation measurement. *Meteorological Applications*, 18(3):334–353.

[Larnier et al., 2019] Larnier, K., Monnier, J., Garambois, P., Verley, J., Larnier, K., Monnier, J., Garambois, P., and River, J. V. (2019). River discharge and bathymetry estimations from SWOT altimetry measurements To cite this version : HAL Id : hal-01811683.

[Lawrence et al., 2019] Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12):4245–4287.

[Lawrence, Dingman, 2007] Lawrence, Dingman, S. (2007). Analytical derivation of at-a-station hydraulic-geometry relations. *Journal of Hydrology*, 334(1-2):17–27.

[Legleiter et al., 2009] Legleiter, C., Roberts, D., and Lawrence, R. (2009). Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms*.

[Lehner et al., 2008] Lehner, B., Verdin, K., and Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, 89(10):93–94.

[Leopold and Maddock, 1953] Leopold, L. B. and Maddock, T. J. (1953). The Hydraulic Geometry of Stream Channels and Some Physiographic Implications (USGS Numbered Series No. 252). *Professional Paper. U. S. Government Printing Office, Washington, D.C*, page 57.

[Lettenmaier et al., 2015] Lettenmaier, D. P., Alsdorf, D. E., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, pages 7309–7342.

[Lewis, 1966] Lewis, L. A. (1966). The adjustment of some hydraulic variables at discharges less than one CFS. *The Professional Geographer*, 18(4):230–234.

[Li et al., 2020] Li, D., Andreadis, K. M., Margulis, S. A., and Lettenmaier, D. P. (2020). A Data Assimilation Framework for Generating Space-Time Continuous Daily SWOT River Discharge Data Products. *Water Resources Research*, 56(6):1–22.

[Liang et al., 2013a] Liang, S., Zhang, X., Xiao, Z., Cheng, J., Liu, Q., and Zhao, X. (2013a). *Global LAnd Surface Satellite (GLASS) products: Algorithms, validation and analysis*. Springer.

[Liang et al., 2013b] Liang, S., Zhao, X., Liu, S., Yuan, W., Cheng, X., Xiao, Z., Zhang, X., Liu, Q., Cheng, J., Tang, H., Qu, Y., Bo, Y., Qu, Y., Ren, H., Yu, K., and Townshend, J. (2013b). A long-term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. *International Journal of Digital Earth*, 6(SUPPL1):5–33.

[Liang et al., 1994] Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models. *J. Geophys. Res.*, 99(D7):14415–14428.

[Liang et al., 1996] Liang, X., Wood, E. F., and Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change*, 13(1-4):195–206.

[Lin et al., 2019] Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., and Wood, E. F. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, pages 1–18.

[Livneh et al., 2015] Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., and Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950-2013. *Scientific Data*, 2:1–12.

[Livneh et al., 2013] Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., and Lettenmaier, D. P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, 26(23):9384–9392.

[Lohmann et al., 1998] Lohmann, D., Raschke, E., Nijssen, B., and Lettenmaier, D. P. (1998). Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrological Sciences Journal*, 43(1):131–141.

[Mao and Cherkauer, 2009] Mao, D. and Cherkauer, K. A. (2009). Impacts of land-use change on hydrologic responses in the Great Lakes region. *Journal of Hydrology*, 374(1-2):71–82.

[Mao et al., 2007] Mao, D., Cherkauer, K. A., and Bowling, L. C. (2007). Improved vegetation properties for the estimation of evapotranspiration in the Midwestern United States. In *ASABE Annual International Meeting*.

[Marcus and Fonstad, 2008] Marcus, W. A. and Fonstad, M. A. (2008). Optical remote mapping of rivers at sub-meter resolutions and watershed extents. *Earth Surface Processes and Landforms*.

[Margulis et al., 2015] Margulis, S. a., Girotto, M., Cortés, G., and Durand, M. (2015). A Particle Batch Smoother Approach to Snow Water Equivalent Estimation. *Journal of Hydrometeorology*, 16:1752–1772.

[MathWorks, 2019] MathWorks (2019). gradientm. In *Mapping Toolbox ™ Reference R2019a*, pages 680–682.

[Maurer et al., 2002] Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B. (2002). A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States. *J. Climate*, 15(22):3237–3251.

[Mersel et al., 2013] Mersel, M. K., Smith, L. C., Andreadis, K. M., and Durand, M. T. (2013). Estimation of river depth from remotely sensed hydraulic relationships. *Water Resources Research*, 49(6):3165–3179.

[Mitchell et al., 2004] Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple

GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research D: Atmospheres*, 109(7).

[Moradkhani et al., 2005] Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2):135–147.

[Moramarco and Singh, 2010] Moramarco, T. and Singh, V. P. (2010). Formulation of the Entropy Parameter Based on Hydraulic and Geometric Characteristics of River Cross Sections. *Journal of Hydrologic Engineering*, 15(10):852–858.

[Nachtergaele et al., 2012] Nachtergaele, F., van Velthuizen, H., Verelst, L., Wiberg, D., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., and Shi, X. (2012). Harmonized World Soil Database (version 1.2). *FAO, Rome, Italy and IIASA, Laxenburg, Austria*, pages 1–50.

[Nash and Sutcliffe, 1970] Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

[Neal et al., 2015] Neal, J. C., Odoni, N. A., Trigg, M. A., Freer, J. E., Garcia-Pintado, J., Mason, D. C., Wood, M., and Bates, P. D. (2015). Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *Journal of Hydrology*, 529:169–183.

[Nijssen et al., 2001] Nijssen, B., Schnur, R., and Lettenmaier, D. P. (2001). Global retrospective estimation of soil moisture using the variable infiltration capacity land surface modl, 1980-93. *Journal of Climate*, 14(8):1790–1808.

[Niu et al., 2011] Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y. (2011). The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, 116(D12).

[NOAA,] NOAA. SURFRAD (Surface Radiation) Network.

[Ott et al., 2004] Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56(5):415.

[Oubanas et al., 2018a] Oubanas, H., Gejadze, I., Malaterre, P., and Durand, M. (2018a). Discharge Estimation in Ungauged Basins Through Variational Data Assimilation : The Potential of the SWOT Mission. *Water Resources Research*, pages 2405–2423.

[Oubanas et al., 2018b] Oubanas, H., Gejadze, I., Malaterre, P. O., Durand, M., Wei, R., Frasson, R. P., and Domeneghetti, A. (2018b). Discharge Estimation in Ungauged Basins Through Variational Data Assimilation: The Potential of the SWOT Mission. *Water Resources Research*, 54(3):2405–2423.

[Oubanas et al., 2018c] Oubanas, H., Gejadze, I., Malaterre, P. O., and Mercier, F. (2018c). River discharge estimation from synthetic SWOT-type observations using variational data assimilation and the full Saint-Venant hydraulic model. *Journal of Hydrology*, 559:638–647.

[Paiva et al., 2015] Paiva, R. C. D. R., Durand, M. T., and Hossain, F. (2015). Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. *Water Resources Research*, 51(1):430–449.

[Pan and Wood, 2013] Pan, M. and Wood, E. F. (2013). Inverse streamflow routing. *Hydrology and Earth System Sciences*, 17(11):4577–4588.

[Pan et al., 2015] Pan, Z., Glennie, C., Legleiter, C., and Overstreet, B. (2015). Estimation of Water Depths and Turbidity from Hyperspectral Imagery Using Support Vector Regression. *IEEE Geoscience and Remote Sensing Letters*, 12(10):2165–2169.

[Papalexiou and Serinaldi, 2020] Papalexiou, S. M. and Serinaldi, F. (2020). Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity. *Water Resources Research*, 56(2).

[Peters-Lidard et al., 1998] Peters-Lidard, C. D., Blackburn, E., Liang, X., and Wood, E. F. (1998). The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *Journal of the Atmospheric Sciences*, 55(7):1209–1224.

[Prairie and Callejo, 2005] Prairie, J. and Callejo, R. (2005). Natural Flow and Salt Computation Methods, Calendar Years 1971-1995. *Bureau of Reclamation*, (December):1–112.

[Rawls, 1993] Rawls, W. (1993). Infiltration and Soil Water Movement. In *Handbook of Hydrology*. McGraw-Hill, Inc.

[Rodell et al., 2004] Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-j., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3):381–394.

[Rodriguez et al., 2016] Rodriguez, E., Fu, L.-L., Cherchali, S., Vaze, P., Lindstrom, E., Pavelski, T., Morrow, R., Cretaux, J.-F., and Lafon, T. (2016). SWOT Science Requirements Document. Technical report, Jet Propulsion Laboratory, Pasadena.

[Rogala, 1999] Rogala, J. T. (1999). Bathymetry - Chapter 2 extracted from James T. Rogala, 1999. Technical report, U.S. Geological Survey.

[Schaake, 2000] Schaake, J. C. (2000). Average hydraulic properties of ARS soil texture classes.

[Schaperow and Li, 2020] Schaperow, J. R. and Li, D. (2020). VICGlobal: soil and vegetation parameters for the Variable Infiltration Capacity hydrological model.

[Sellers et al., 1986] Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A. (1986). A Simple Biosphere Model (SIB) for Use within General Circulation Models. *Journal of the Atmospheric Sciences*, 43(6):505–531.

[Shi et al., 2008] Shi, X., Wood, A. W., and Lettenmaier, D. P. (2008). How essential is hydrologic model calibration to seasonal stream flow forecasting? *Journal of Hydrometeorology*, 9(6):1350–1363.

[Simon, 2010] Simon, D. (2010). Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *IET Control Theory and Applications*, 4(8):1303–1318.

[Su et al., 2005] Su, F., Adam, J. C., Bowling, L. C., and Lettenmaier, D. P. (2005). Streamflow simulations of the terrestrial Arctic domain. *Journal of Geophysical Research D: Atmospheres*, 110(8):1–25.

[Sulla-Menashe et al., 2019] Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., and Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, 222(November 2018):183–194.

[Tolson and Shoemaker, 2007] Tolson, B. A. and Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(1):1–16.

[Wieder et al., 2014] Wieder, W., Boehnert, J., Bonan, G., and Langseth, M. (2014). Regridded Harmonized World Soil Database v1 .2. Technical report, NASA Earthdata.

[Wood et al., 2011] Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47(5):W05301.

[Wu et al., 2012] Wu, H., Kimball, J. S., Li, H., Huang, M., Leung, L. R., and Adler, R. F. (2012). A new global river network database for macroscale hydrologic modeling. *Water Resources Research*, 48(9).

[Xia et al., 2012a] Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D. (2012a). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research Atmospheres*, 117(3):1–23.

[Xia et al., 2012b] Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D. (2012b). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research Atmospheres*, 117(3):1–27.

[Xiao et al., 2016] Xiao, M., Nijssen, B., and Lettenmaier, D. P. (2016). Drought in the Pacific Northwest, 1920–2013. *Journal of Hydrometeorology*, 17(9):2391–2404.

[Yamazaki et al., 2017] Yamazaki, D., Ikeshim, D., Tawatari, R., and Al, E. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, pages 181–189.

[Yang et al., 2020] Yang, Y., Lin, P., Fisher, C. K., Turmon, M., Hobbs, J., Emery, C. M., Reager, J. T., David, C. H., Lu, H., Yang, K., Hong, Y., Wood, E. F., and Pan, M. (2020). Enhancing SWOT discharge assimilation through spatiotemporal correlations. *Remote Sensing of Environment*, 234(September):111450.

[Yoon et al., 2012] Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., and Alsdorf, D. E. (2012). Estimating river bathymetry from data assimilation of synthetic SWOT measurements. *Journal of Hydrology*, 464-465(December 2014):363–375.

[Yoon et al., 2016] Yoon, Y., Garambois, P. A., Paiva, R. C., Durand, M., Roux, H., and Beighley, E. (2016). Improved error estimates of a discharge algorithm for remotely sensed river measurements: Test cases on Sacramento and Garonne Rivers. *Water Resources Research*, 52(1):278–294.

[Zeileis et al., 2003] Zeileis, A., Kleiber, C., Kramer, W., and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics*, 44(1):1–13.

[Zeng, 2001] Zeng, X. (2001). Global Vegetation Root Distribution for Land Modeling. *Journal of Hydrometeorology*, 2(5):525–530.

[Zhou et al., 2016] Zhou, T., Nijssen, B., Gao, H., and Lettenmaier, D. P. (2016). The contribution of reservoirs to global land surface water storage variations. *Journal of Hydrometeorology*, 17(1):309–325.