Hydrologic Model Parameter Estimation in Ungauged Basins using Simulated SWOT Discharge Observations

Nicholas J. Elmer[1], James McCreight[2], Christopher Hain[3]

[1] NASA Postdoctoral Program, NASA Marshall Space Flight Center, Huntsville, Alabama

[2]National Center for Atmospheric Research, Boulder, Colorado

[3]Earth Science Office, NASA Marshall Space Flight Center, Huntsville, Alabama

Corresponding address: Nicholas J. Elmer, NASA Postdoctoral Program, NASA Marshall Space Flight Center, 320 Sparkman Drive, Huntsville, AL, 35805. Email: nicholas.j.elmer@nasa.gov

Key Points:

- SWOT observations are critical for calibrating hydrologic models in regions devoid of in situ observations
- For an expected SWOT discharge error of 35%, multi-point parameter estimation is successful for 90% of polar and 70% of tropical sub-basins.
- Multi-point parameter selection is preferred over single-point parameter selection, offering more robust results with less sensitivity.

Keywords: SWOT, WRF-Hydro, Alaska, Parameter Estimation, Calibration

Abstract

In situ gauge networks are often used in hydrological model calibration, but these networks are limited or nonexistent in many regions. The upcoming Surface Water Ocean Topography (SWOT) mission promises to fill this observation gap by providing discharge estimates for rivers with widths greater than 100 meters. Proxy SWOT discharge estimates derived from an observing system simulation experiment and Monte Carlo methods are used to assess SWOT observation utility for model parameter selection in regions devoid of in situ gauges the sensitivity of the parameter selection to measurement error and observation temporal frequency is also evaluated. Single-point and multi-point parameter selection is performed for ten sub-basins within the Susitna River and upper Tanana River basins in Alaska. SWOT is expected to observe Alaskan river points 4-7 times per 21-day repeat cycle with 120 km swath coverage. For an expected SWOT discharge error of 35%, parameter estimation is successful for 60% and 90% of sub-basins using single-point and multi-point selection, respectively. Decreasing observation frequency to simulate lower latitudes resulted in success for only 20% of midlatitude and 10% of tropical sub-basins for single-point selection, whereas multi-point selection was successful in 80% of midlatitudes and 70% of tropical sub-basins. Single-point parameter selection was much more sensitive to SWOT discharge error than multi-point parameter selection. The results strongly support the use of multi-point parameter selection over single-point parameter selection, yielding robust results nearly independent of observation error with approximately half the sensitivity to observation frequency.

1. Introduction

For decades, in situ gauge networks have been monitoring stream hydrology and are considered a robust observation with well-understood errors (Hirsch and Costa 2004, Boning 1992), measurements of floods and droughts notwithstanding. Stream gauges aid in the modeling and forecasting of major hydrologic events by enabling model calibration and validation. Unfortunately, in situ stream gauge networks are concentrated to only a few regions globally, and these networks are on the decline (Pavelsky et al. 2014), limiting the availability of observations of streamflow. Furthermore, very few observations are available from satellite platforms since all current satellite missions, including Jason-3 and the second Ice, Cloud and land Elevation Satellite (IceSAT-2), theoretically capable of measuring river stage using radar and laser nadir altimetry (Kouraev et al. 2004, Papa et al. 2010, O'Loughlin et al. 2016, Biancamaria et al. 2017), have insufficient spatial and temporal resolutions for adequate sampling (Alsdorf et al. 2007, Biancamaria et al. 2016).

To fill this observation gap, the Surface Water Ocean Topography (SWOT) mission (Biancamaria et al. 2016) was designed and is expected to be launched in late 2021 to provide the first global inventory of Earth's surface water, including rivers, lakes, and wetlands. A joint venture between the National Aeronautics and Space Administration (NASA), Centre National d'Etudes Spatiales (CNES), Canadian Space Agency, and the United Kingdom Space, SWOT supports a nadir altimeter and a bistatic Ka-band (35.75 GHz) Radar Interferometer (KaRIn) (Fjortoft et al. 2014). The nadir altimeter allows intercomparison with Jason measurements, will help to continue the data record of nadir altimeters, and fills the gap between the two 60 km KaRIn swaths, one on each side of nadir. KaRIn provides high-resolution water surface elevation (WSE, the height of the river surface above a reference geoid), width, and slope measurements

across a combined 120 km swath for rivers with widths greater than 100 m (Biancamaria et al. 2016, Pavelsky et al. 2014, Rodriguez 2016). Since KaRIn uses Ka-band, instead of C- and Ku-band used by Jason and IceSAT, there is less signal penetration into soil, snow, and vegetation (Fjortoft et al. 2014, Biancamaria et al. 2016) enabling SWOT to collect measurements at finer spatial resolutions. Therefore, KaRIn will be the first satellite instrument that can fully resolve terrestrial surface water bodies with high altimetric accuracy.

In the United States (U. S.), U. S. Geological Survey (USGS) stream gauges measure stage data at 3 mm accuracy, which translates to discharge accuracy of 5-10% (Hirsch and Costa 2004) under normal flow conditions. Generally, a 1% error in the effective stage input is equivalent to a 3% error in the computed discharge (Boning 1992). In comparison, SWOT WSE is expected to have a minimum error of 10 cm for most rivers (Biancamaria et al. 2016) with estimated discharge errors around 35% (Durand et al. 2016). However, even though expected SWOT errors are much larger than the error of in situ gauges, in the absence of in situ gauges SWOT measurements will provide the best estimates of stage and discharge available. This work also highlights that SWOT observations along many points in the stream network have better error characteristics than a single observation, as errors are not assumed to be correlated.

Hydrologic models, including the National Oceanic and Atmospheric Administration (NOAA) National Water Model (NWM; OWP 2020) which is an instantiation of the Weather Research and Forecasting Hydrological extension package (WRF-Hydro; Gochis et al. 2018), are typically calibrated using in situ gauges. WRF-Hydro is a modeling framework that couples column land surface, overland and subsurface terrain routing, and channel routing models in a multiscale hydrologic process representation. WRF-Hydro is fully-distributed with multi-physics options and multi-scale capabilities, enabling it to represent processes on a wide range of spatial

scales (Yucel et al. 2015, Senatore et al. 2015, Arnault et al., 2018, Gochis et al. 2018). Since many parameterizations are used to characterize sub-scale processes in numerical models, parameter values are often hard-coded or contained in parameter tables. For example, 139 hard-coded parameters and 71 standard parameters were identified within Noah-MP by Cuntz et al. (2016). Running Noah-MP coupled with WRF-Hydro, Cuntz et al. (2016) found that hydrologic output fluxes are sensitive to two-thirds of the standard parameters and surface runoff is sensitive to many parameters of snow processes, soil, and vegetation. Even after calibration, many parameter values can vary widely from basin to basin, even between neighboring watersheds. Calibration seeks to minimize an objective function as a measure of physical realism by optimizing the parameter values of the most sensitive model parameters. Traditionally, in situ observations are used to calibrate hydrologic models. Few, if any, alternatives are available if *in situ* observations are lacking. Following launch, SWOT will provide an additional source of discharge observations from a satellite platform, potentially providing more observations per basin than even some of the most robust in situ networks. This paper assesses the ability of SWOT discharge estimates to enable hydrological model parameter selection in regions devoid of in situ gauges. This paper also compares multi-point parameter selection (e.g., Cao et al. 2006, Niraula et al. 2012), which will be made possible with SWOT observations, to the traditional single-point calibration approach.

2. Methodology

2.1 Experimental Design

The design of our fraternal twin parameter selection experiment is shown in Figure 1. This Observing System Simulation Experiment (OSSE) is based on Elmer et al. (2020). The

fraternal twins, the “truth run” and “calibration run”, simulate model error by employing different hydrologic model representations in the model chain that generates streamflow. The experiment addresses whether the unknown, best parameters for the calibration run can be reliably selected (purple box in Figure 1) from observations of the true run streamflow imparted with the expected observation error characteristics of the SWOT sensor. Successfully identifying the best parameters from observations is the core of model calibration. In this experiment, because we know the true streamflow values, we can evaluate under what conditions parameter selection is successful.

Parameter, model, and observation errors are all ingredients of the experimental design. The parameter error is the quantity we seek to minimize in parameter selection and calibration via the objective function. The “truth run” was pre-calibrated to a single subdomain of the study and has a single realization (yellow box, Figure 1). In contrast, the calibration runs span the space of thirteen model parameters plus Manning’s roughness coefficient using 80 parameter sets. This is represented by the stack of calibration runs (red boxes) in Figure 1. Model error of the calibration runs relative to the truth run is generated by differences summarized by text in the respective boxes in Figure 1. These differences produce errors between the runs in terms of 1) the fixed boundary conditions or geometry for different land surface model (LSM) resolutions and channel routing networks, 2) the LSM and channel parameters, particularly infiltration parameters, which depend on spatial and temporal model resolutions, and 3) streamflow physics. We note that both the atmospheric forcing variables and LSM models are identical between the runs but that errors or differences in the model runs start with and accumulate over time within the soil moisture representation and its two-way coupling to the overland and subsurface runoff

133 models (Gochis and Chen 2003), which feeds back to LSM behavior and parameter differences.
134 The differences between the fraternal twins are described in further detail below.

135 The SWOT observation errors are considered random. Therefore 10,000 realizations of
136 observation errors are applied to the true states before use in parameter selection to avoid
137 drawing conclusions from a particular set of errors. This set of 10,000 possible observation
138 realizations is represented by the stack of observations (blue ovals) in Figure 1. Although results
139 using as few as 100 realizations would have led to similar conclusions, 10,000 realizations are
140 used for this study to ensure robust results. Over the 10,000 observation sets, the probability of
141 selection (identification as the best parameter set via a version of Nash-Sutcliffe Efficiency based
142 on the observations) is computed for each of the 80 parameter sets (purple diamond, Figure 1).
143 Finally, in the evaluation step in (green diamond, Figure 1), NSE is computed from the true
144 model states and the true rank of each parameter set is assigned, from best (low) to worst (high).
145 The cumulative probability of parameter selection (under observation error) is plotted against
146 rank. Do the true, best parameter sets have a reasonably high likelihood of selection in the
147 presence of SWOT observation characteristics (and model error)?

148 This paper focuses on sub-basins within the upper Tanana River and Susitna River basins
149 in Alaska, which will be regularly observed by SWOT (Biancamaria et al. 2016) but have few in
150 situ observations. These watersheds are delineated in Figure 2. The following subsections
151 provide additional details for each step of the experimental design shown in Figure 1.

152

153 2.2 Model configurations and parameters

154 For both truth and calibration model runs in Figure 1, this study uses WRF-Hydro version
155 5.0 (Gochis et al. 2018), which includes the NWM v2.0 configurations (OWP 2020). The Noah

land surface model with Multi-Parameterization options (Noah-MP; Niu et al. 2011) with a 1 km spatial resolution is used as the WRF-Hydro land surface model in both models, as well as Global Land Data Assimilation System (GLDAS) Version 2 (Rodell et al. 2004) meteorological forcing. Further details of the model configuration and physics parameterizations used are listed in Table 1, which also lists the differences between the truth run and the calibration runs.

The truth (calibration) run configuration has a model timestep of 1 (3) hours, performing overland and subsurface routing on a 100m (250m) grid, and uses the Muskingum-Cunge (diffusive wave) routing scheme for simulating streamflow within a channel model. GLDAS forcing data is available at three-hour increments and ingested into both configurations equally. The WRF-Hydro terrain routing grids (100m and 250m) and channel networks were derived from the WRF-Hydro GIS Pre-processing Toolkit v5.1 (Sampson and Gochis 2015) using the Weather Research and Forecasting (WRF) (Skamarock et al. 2008) Preprocessing System geogrid file and the National Elevation Dataset (NED) (U. S. Geological Survey 2017) Digital Elevation Model (DEM) as inputs. Both the Muskingum-Cunge and diffusive wave schemes represent channels with an infinite depth, preventing overbank flow. However, the diffusive wave scheme allows backwater effects, whereas the Muskingum-Cunge scheme does not. Importantly, the channel networks are derived using different DEM spatial resolutions, leading to different spatial representations of the channel routing.

Whereas we calibrate the truth run to in situ streamflow observations (described below), the calibration run configuration is uncalibrated: the point of our experiment being selection of parameters for the calibration run that most accurately simulate the truth run. Eighty calibration parameter sets were created by perturbing Manning's roughness coefficient (as a function of stream order) along with the thirteen most sensitive WRF-Hydro parameters (Cuntz et al. 2016;

Elmer 2019). As shown in Table 2, these parameters span the LSM, overland/subsurface routing, groundwater bucket, and channel routing components of the model (model variable names shown in parentheses): the Clapp-Hornberger B exponent (bexp), soil moisture maximum (smcmax), saturated soil conductivity (dksat), soil infiltration parameter (refkdt), soil drainage parameter (slope), retention depth (RETDEPRTFAC), saturated soil lateral conductivity (LKSATFAC), groundwater bucket model max depth (Zmax), groundwater bucket model exponent (Expon), canopy wind parameter (CWPVT), maximum carboxylation at 25°C (VCMX25) which is related to the vegetation height (HVT), the Ball-Berry conductance relationship slope (MP), and the snowmelt parameter (MFSNO). The ranges assigned to these parameters make up the calibration parameter space. From this parameter space, a sample of parameter sets were obtained by randomly assigning values within the valid parameter ranges listed in Table 2 using a uniform distribution. The result is a good representation of parameter space, as shown by depiction of the distribution of the sampled parameter sets using multidimensional scaling in Figure 3.

The truth model is calibrated using in situ USGS stream gauge observations at the basin outlet using the parameter space described above (Table 2). The mean of the Nash Sutcliffe Efficiency (NSE; McCuen et al. 2006) and the logarithmic NSE (NSE_{ln}) was used as the calibration metric, denoted as the mean NSE (NSE_{mean}). NSE_{mean} is akin to the metric used in the calibration of the NWM and is given by:

$$NSE_{mean} = (w)NSE + (1 - w)NSE_{ln} \qquad (1)$$

where w is the weight of 0.5. NSE_{mean} ranges from negative infinity to unity, where a value greater than zero indicates that the model provides a better estimate than the observation mean. Thus, the maximum NSE_{mean} is sought. For single point parameter selection, NSE_{mean} is the

metric for evaluation. For multipoint parameter selection, a basin average NSE_{mean} is calculated for evaluation, given by:

$$\overline{NSE_{mean}} = \frac{\sum_{i=1}^{P} NSE_{mean_i}}{P} \qquad (2)$$

where P is the number of observed points in the sub-basin or watershed.

As the calibration process is computationally expensive, calibration of the truth simulation was only performed for the Chena River watershed (within the upper Tanana River basin; watershed outlet denoted by point I in Figure 2) and halted after 75 model iterations. The parameters identified using the Chena River calibration were transferred to the full domain. Although the Chena River calibration may not transfer well to the entire upper Tanana River and Susitna River basins, the resulting model output is treated as truth for this experiment and therefore a perfect calibration is not necessary. Certain parameters (e.g., infiltration parameters) are strongly scale dependent, so the calibration of the truth model, in which the model resolution and streamflow physics differ from the 250-m model, is not directly transferrable to the 250-m model. The truth run provides the "true" geolocation and discharge (q) for the purposes of this experiment.

A spin-up period of eight years (March 2009 - March 2017) using default parameter values (Table 2) was performed, designed to allow for adequate accumulation of groundwater and snowpack and permitting each 250-m simulation to reach equilibrium. The March 2017 restart files from the spin-up were used to restart the 250-m simulations at March 2011 using their respective parameter set and integrated forward in time in an open loop configuration. The periods of March – September 2012 for the Susitna River basin and March – September 2014 for the upper Tanana River basin were used during analysis to determine the utility of SWOT observations in model calibration.

223

224 2.3. Generating Proxy SWOT Discharge

225 Since real SWOT data are not yet available, proxy SWOT observations were generated
226 for this analysis. Proxy SWOT data has been used by multiple studies to quantify assimilation
227 impacts on river modeling and reservoir management (Andreadis et al. 2007; Biancamaria et al.
228 2011, Munier et al. 2015) and develop procedures for estimating river bathymetry (Durand et al.
229 2008, 2010, 2014; Yoon et al. 2012). Furthermore, Pedinotti et al. (2014) used synthetic SWOT
230 data to optimize Manning roughness coefficients in the Interactions between Soil, Biosphere, and
231 Atmosphere-Total Runoff Integrating Pathways System (ISBA-TRIP) continental hydrologic
232 system using data assimilation, demonstrating that SWOT data can be used for calibration.

233 The truth model q is corrupted with random white noise following a Gaussian distribution
234 (N) with a zero mean and standard deviation σ to represent measurement error:

$$q' = q + N(0, \sigma) = q + N(0, \epsilon q), \qquad (3)$$

235 where q' is the corrupted discharge (i.e., the proxy SWOT discharge estimate) and ϵ is the
236 representative discharge error. For the analysis in Section 3.1, $\epsilon = 0.35$ is used, which is roughly
237 equivalent to the relative root mean squared error (RMSE) of instantaneous discharge estimated
238 by Durand et al. (2016). Since an equivalent increase in water surface height will yield a larger
239 increase in discharge for a river with a larger cross-section, uncertainty in *q'* naturally increases
240 as q increases.

241 However, a single dataset of *q'* does not provide an adequate sampling of measurement
242 error by which to assess calibration potential. Rather, it gives a snapshot of only one possible
243 scenario. Figure 3 illustrates this quite well. The blue dots indicate a single data set of *q'*,
244 containing some points in which the measurement errors are small and depart very little from the
245 truth value but also points that extend into the 2-σ error range. A single scenario may be biased if

246 only small random errors are present ($q' \approx q$), especially at key points along the time series,
247 which would enable superior parameter selection and suggest better results than could actually
248 be expected. Conversely, a scenario containing frequent large errors, parameter selection would
249 underperform. To sample a broad spectrum of the possible outcomes stemming from a SWOT
250 observation set laden with error, Monte Carlo methods were employed to create 10,000
251 randomly-perturbed sets of *q'* per sub-basin. Based on the outcome of each of the 10,000 sets,
252 the probability of successful sub-basin calibration was calculated.

253 To obtain *q'* with appropriate orbit characteristics, it was spatially sampled based on the
254 CNES proxy SWOT orbit (Aviso+ 2015). First, the cross-track distance of each WRF-Hydro
255 reach from the proxy SWOT orbit at each overpass was calculated. For each pass, only reaches
256 with cross-track distances of 10-60 km (i.e., within the SWOT measurement range) and with a
257 Strahler streamorder greater than or equal to five (used to approximate rivers with widths greater
258 than 50 m) were extracted, following the methodology of Elmer et al. (2020), which showed that
259 Alaskan rivers with a streamorder greater than or equal to five will generally be observable by
260 SWOT. Figure 3 compares the truth and *q'* for sub-basin E, where *q'* is used to calibrate the 250-
261 m model following the method described in Section 2.3.
262
263 2.3. Parameter selection from proxy SWOT discharge observations
264 We purposely use the term "parameter selection" to differentiate our overall approach
265 from calibration for the following reasons. The parameter sets run through the model and
266 ultimately judged by the objective function are generated a priori through combinations of
267 uniformly sampled distributions on each parameter. As such, the parameters sets being
268 discriminated via the objective function are generally not "close" in parameter space (Figure 3).

A true calibration approach would consider points with a similar spacing in parameter space, but would also include parameter sets much closer together in the quest to find minima of the objective function. This paper does not study the ability to accurately find local minima of the objective function using SWOT observations. Doing so would require understanding the relative sensitivities of the objective function to observation error and to distance in parameter space. Rather, we examine the potential for SWOT observations to give a more regional, less detailed picture of the objective function. Given the observation and error characteristics of SWOT, including its spatially distributed nature, and a finite collection of parameter sets, can we accurately select the best parameter set in this collection? Our results provide an encouraging basis for pursuing model calibration using SWOT observations.

To review and summarize the experiment design (Figure 1), the 250-m model described in Section 2.1 represents an uncalibrated hydrologic model of an ungauged basin (a single red box, Figure 1). For this basin, an infinite number of parameter sets can be selected from the parameter space for calibration and the correct solution (parameter set) is unknown. The goal is to find the best simulation of streamflow over a finite sample of parameter space. A sample size of 80 parameter sets (red stack of boxes, Figure 1) was chosen for this study to minimize computational requirements. While the parameter set can certainly be expanded to more fully represent the whole parameter space, the increased computational requirements may reduce the feasibility of this method for users without access to large computing systems.

This study ensures the 250-m model is blind to the calibration of the truth model so that the calibrated parameter set used by the truth model does not inform the selection of the parameter set sampling for the 250-m model. In the absence of in situ gauges, the only source of regular discharge observations for ungauged basins will be from the SWOT mission. The q'

values derived in Section 2.3 are representative of the SWOT discharge observations that will be available post-launch, and are used to calibrate the model by finding the best parameter set from the 80 sample parameter sets. A comparison between the 100-m truth model, the 250-m simulations, and the proxy SWOT discharge estimates is provided in Table 3, with example data shown in Figure 4. For this particular point, there are 38 SWOT overpasses from March 1 – September 15, yielding an observation approximately every 5 days, or 4 observations per repeat cycle.

Channel reaches were spatially matched between the truth and calibration model channel networks for evaluation, eliminating any matches separated by a 1 km or greater which are considered unrelated. Thus, a total of 10 basins and 991 channel reaches were evaluated. All basins were modeled simultaneously, but evaluated separately. Single-point parameter selection for the uncalibrated 250-m model is performed using NSE_{SWOT}, given as the NSE_{mean} between the discharge for each simulation and q' at each sub-basin outlet in Figure 2. Multi-point parameter selection is evaluated with $\overline{NSE_{SWOT}}$, the basin average NSE_{SWOT} for all observed points P (Equation 2). The 250-m simulation with the maximum NSE_{SWOT} or $\overline{NSE_{SWOT}}$ (the best match parameter set) is chosen, and the parameter set used by that simulation for the thirteen most sensitive WRF-Hydro parameters is selected as the "correct" parameter set.

2.4. Evaluation of parameter selection

The NSE_{mean} was also calculated between each 250-m simulation and the truth model q (NSE_{TRUTH}) at each basin outlet for comparison with NSE_{SWOT} for single point parameter selection evaluation. Multi-point parameter selection is similar to single-point selection except that $\overline{NSE_{SWOT}}$ and $\overline{NSE_{TRUTH}}$ (the basin-average values for all observable channel reaches within

each sub-basin) is used. The 250-m simulations are separately ranked according to their NSE_{SWOT} ($\overline{NSE_{SWOT}}$) and NSE_{TRUTH} ($\overline{NSE_{TRUTH}}$) values for single point (multi point) parameter selection with the member with the best (maximum) value being assigned a ranking of one, and the member with the worst (minimum) value assigned a ranking of 80. Expressing the cumulative rank of NSE_{SWOT} ($\overline{NSE_{SWOT}}$) as a function of NSE_{TRUTH} ($\overline{NSE_{TRUTH}}$) reveals whether single (multi) point parameter selection using SWOT observations is effective. For the purposes of drawing conclusions in this paper, a successful parameter selection is achieved for a watershed if the selected parameter set is contained within the best ten sets according to the NSE_{TRUTH} ($\overline{NSE_{TRUTH}}$) rank with a >80% probability. This criteria is subjective, thus plots showing the full range of probabilities are included. For example, in Figure 5a the y-axis shows the cumulative probability whereas the x-axis shows the parameter set rank. The cumulative probability is essentially the fraction of observation sets (10,000 sets in total). Thus, we see that for sub-basin G (pink line), approximately 0.95 (95%) of the 10,000 observation sets rank the true best parameter set in the top ten (indicated by vertical black dashed line). Since this value is above the 0.8 (80%) threshold, the parameter selection is successful.

2.5 Sensitivity to measurement error and temporal frequency

Additional analysis examines the sensitivity of the parameter selection results to measurement error (Section 3.2) and the temporal frequency of SWOT observations (Section 3.3). Although the measurement error of SWOT instantaneous discharge is estimated to have a relative RMSE of 35%, the incorporation of ancillary data in the discharge algorithms to reduce error is being considered (Durand et al. 2016). Thus, determining the sensitivity of these results to measurement error is useful in evaluating the range of possible impacts for SWOT, in

particular with respect to model calibration. Thus, q' is recalculated with $\epsilon = 0.20$ rather than 0.35. The temporal frequency of SWOT observations is inherently tied to latitude due to SWOT orbit characteristics (relatively narrow swath compared to satellite imagers and high inclination angle). Thus, polar regions are observed more frequently than the midlatitudes, and the midlatitudes are observed more frequently than the tropics. Biancamaria et al. (2016) show that SWOT will observe the tropics (0-30°) 1-2 times per repeat cycle, the midlatitudes (30-60°) 2-4 times per repeat cycle, and polar regions (60-90°) 3-7 times per repeat cycle.

To assess the sensitivity of parameter selection results to temporal frequency and determine whether this process is viable at lower latitudes, the same Alaskan sub-basins are considered but the observation frequency of q' is reduced to mimic SWOT observation of mid- and low-latitudes basins. For the midlatitudes, the observation frequency of q' was halved with respect to that of Alaska. For the tropics, the observation frequency was reduced by a factor of four. The sensitivity of model parameter selection to measurement error and the temporal frequency of observations is calculated by:

$$S = \frac{\partial Y}{\partial X}, \qquad (4)$$

where Y is the probability of selection and X is the measurement error ε or number of observations per repeat cycle. Error sensitivity is likely non-linear; thus, in evaluating error sensitivity these results are likely only valid for ε between 0.20 and 0.35.

3. Results and discussion

3.1. Parameter selection

Figure 5a presents the cumulative probability that the true best parameter set is selected via NSE_{SWOT} at or above (equal or higher ranking) each rank position for a measurement error

360 ϵ=0.35 using single-point parameter selection. For example, if the rank 10 likelihood for a given
361 basin is 80%, then the parameter set selected by NSE_{SWOT} is one of the ten highest-ranked sets
362 with respect to the truth for 8,000 of the 10,000 Monte Carlo simulations. For six of the ten sub-
363 basins, the selected parameter set is ranked in the top ten with a >80% probability, meeting our
364 criteria for success. These six basins also rank in the top five with a >60% probability. The
365 selected member for two additional sub-basins (E and I) is ranked in the top ten (five) with a
366 >50% (>40%) probability. Notably, two sub-basins (B and D) display much poorer results than
367 the other sub-basins. Sub-basin B selects a parameter set in the top 70 with less than a 30%
368 probability, indicating that the best parameter set as determined by NSE_{TRUTH} is regularly ranked
369 as one of the worst ten sets by this calibration approach. A major difference between B and D
370 and the other better performing sub-basins is the ability of the sampled parameter set to represent
371 the truth.

372 Figure 6 compares the NSE_{TRUTH} for each sub-basin shown in Figure 2. The NSE_{TRUTH}
373 curves for both B and D are flatter than for the other sub-basins, indicating that the sample
374 parameter set spread is narrower. Thus, there is less variation among the simulations, resulting in
375 lower ranked parameter sets based on NSE_{TRUTH} to be more frequently ranked highly based on
376 NSE_{SWOT}. Additionally sub-basin B has the lowest NSE_{TRUTH} of any sub-basin for its highest
377 ranked set with a value less than 0.75. Figure 7 maps the true rank of each parameter set selected
378 by NSE_{SWOT} and displays a histogram of these ranks for all SWOT observable channel reaches in
379 the domain. Results are generally very good for the entire upper Tanana basin and most of the
380 Susitna basin, with the selection of a highly-ranked (top 10) parameter set for most of channel
381 reaches. However, the worst performance occurs in the same outlying sub-basins B and D from
382 Figure 6, with sub-basin B clearly demonstrating the poorest results. Thus, it is apparent that the

383 simulation does not capture the physical processes of these sub-basins as well as for the other
384 sub-basins, indicating that the sample parameter set does not contain the true parameter set and
385 resulting in the inability to achieve a good parameter selection using the approach in this paper.
386 However, this shortcoming highlights several potential issues with parameter selection. Adequate
387 model spin-up and configuration are necessary to ensure significant physical processes are being
388 adequately captured by the model and that physical realism is adequate. Second, parameter sets
389 which appropriately cover the parameter space may also be key to differentiating model
390 simulations (Sharma et al. 2019, Hagedorn et al. 2012, Weigel et al. 2008, 2009). A larger
391 parameter set or a parameter sampling strategy that undertakes large searches across parameter
392 space may benefit the parameter selection at several sites in this study.

393 Figure 8a, interpreted in the same manner as Figure 5a, presents the cumulative
394 probability of successful parameter selection for a measurement error $\epsilon = 0.35$ using multi-point
395 parameter selection. Results improve compared to single-point parameter selection. For nine out
396 of ten of (70%) of the sub-basins, the selected parameter set has a true rank in the top ten with a
397 >80% probability, meeting our criteria for successful parameter selection. The only sub-basin
398 that again does not achieve successful calibration is sub-basin B. Seven basins also identify
399 parameter sets in the top five true parameter sets with >60% probability.

400

401 3.2. Sensitivity to measurement error

402 Reducing the discharge error ε from 0.35 to 0.20 results in the success probabilities
403 shown in Figures 5b and 8b for single-point and multi-point parameter selection, respectively. As
404 expected, the probability that the selected parameter set is highly ranked increases as error
405 decreases, with the exception of sub-basin B for reasons discussed in Section 3.1. For single-

406 point parameter selection, 70% of sub-basins are assigned a top-ten ranked parameter set with a
407 >95% cumulative probability, while 80% of sub-basins meet the definition of successful
408 parameter selection. 60% of sub-basins also display a >95% probability of selecting a parameter
409 set ranked in the top five. These statistics are improved further using multi-point parameter
410 selection with 90% of sub-basins achieving successful parameter selection all with a >95%
411 cumulative probability. Table 4 summarizes the effect of decreasing ϵ, showing that as error
412 decreases the likelihood of successful calibration increases, regardless of observation frequency.
413 The sensitivity of the probability of successful parameter selection is evaluated by
414 considering a 0.10 decrease in observation error ε, centered on ε=0.35. The mean sensitivity is
415 calculated over all sites. The mean sensitivity for each rank position is shown in Figure 9 in blue
416 with single-point and multi-point selection sensitivity as solid and dashed lines, respectively. For
417 single-point parameter selection, the probability is most sensitive between ranks 3-10, with a
418 mean sensitivity of 7% per 0.10 decrease in ϵ. Thus, a sizeable improvement in parameter
419 selection can be expected if the SWOT observation error can be reduced through the use of
420 ancillary datasets. Sensitivity gradually declines for ranks beyond 10, which is to be expected
421 since the cumulative probability, as shown in Figure 5, begins to level off for lower rank
422 positions. Sensitivity to reductions in ϵ decreases with for multi-point parameter selection.
423 Sensitivity peaks at rank 3 with a sensitivity of 5% per 0.10 decrease in ϵ, but quickly drops
424 below 2% by rank 10. Thus, the use of multi-point parameter estimation yields a more robust
425 result nearly independent of observation error given sufficient observations, as shown in Table 4,
426 assuming observation error is uncorrelated.
427
428 3.3 Sensitivity to temporal frequency of SWOT observations

Single-point parameter selection is very sensitive to the temporal frequency of SWOT observations. As shown in Table 4, successful parameter selection is reduced from 60% to 10% of sub-basins for $\epsilon = 0.35$ and from 80% to 40% of sub-basins for $\epsilon = 0.20$. For multi-point parameter selection, probability of success is reduced from 90% to 70% of sub-basins for $\epsilon = 0.35$ and from 90% to 80% of sub-basins for $\epsilon = 0.20$. Figure 10 compares the cumulative probability curves for the midlatitudes and the tropics using multi-point parameter selection, which alongside Figure 8a, shows that most sub-basins are unaffected by the decrease in observation frequency.

The sensitivity due to a single additional SWOT observation per repeat cycle for each rank position is shown in Figure 9 in orange with single-point and multi-point parameter selection identified by the orange solid and dashed lines, respectively. In calculating sensitivity to observation frequency, five observations per repeat cycle (21 days) is assigned to the polar region, three observations per repeat cycle is given to the midlatitudes, and 1.5 observations per repeat cycle is used for the tropics. For example, in calculating sensitivity between the midlatitudes and tropics, $\partial X = 3 - 1.5 = 1.5$. The magnitudes and patterns are similar to that of observation error sensitivity, in which maximum sensitivity is observed between ranks 3-10, peaking near 7% per additional observation for single-point parameter selection and 5% for multi-point parameter selection. While the number of SWOT observations per repeat cycle strongly affects the likelihood of successful parameter selection using single-point parameter selection, multi-point selection is much more robust, in that a large majority of sub-basins achieved successful parameter selection regardless of observation frequency. This is particularly true when observation error is reduced, limiting the effect of lower observation frequency.

4. Conclusions

Using Monte Carlo methods, we evaluate parameter selection for an uncalibrated 250-m WRF-Hydro model. We examine single- and multi-point objective function parameter selection using simulated SWOT observations in regions unserved or underserved by in situ gauges. The model parameter space is sampled to create an assortment of parameter sets for which the 250-m model is run. Proxy SWOT discharge estimates were derived from an OSSE following the methodology of Elmer et al. (2020). As the true values of streamflow are known, we can evaluate the selection of model parameters based on the comparison of model discharge simulations with proxy SWOT streamflow observations.

The results indicate that the use of multi-point parameter selection is advantageous over single-point parameter selection. In effect, the spatially distributed nature of the SWOT observations compensates for its observation errors. In fact, successful parameter selection is largely independent of observation error with approximately half its sensitivity attributable to observation frequency. The high spatial coverage of observations from SWOT also compensates for the lack of their temporal frequency in mid-latitude and tropical basins, perhaps due to spatiotemporal correlations in streamflow (Paiva et al. 2015, Yang et al. 2019, Fisher et al. 2020). Even with larger errors than in situ gauges, this study shows that SWOT discharge estimates provide adequate accuracy and temporal sampling to enable parameter selection for SWOT observable river basins globally. In regions devoid of in situ observations or with relatively scarce stream gage networks, this study demonstrates that SWOT will provide valuable observations for calibrating hydrologic models.

This study does not account for reprocessing of discharge estimates occasionally throughout the SWOT mission, which is a planned activity to improve accuracy. Actual SWOT

observations are expected by mid-2022, so these results are timely in preparing to apply SWOT data immediately following launch. While SWOT has many societal and research applications that rely on near-real-time SWOT measurements (e.g., data assimilation, inundation mapping), the use of SWOT observations for model parameter selection or calibration is not constrained by product latency or mission lifetime, but extend beyond the mission end.

Acknowledgements. This work was funded by NASA Headquarters under the NASA Postdoctoral Program at NASA Marshall Space Flight Center administered by Universities Space Research Association. All software and WRF-Hydro input data can be obtained freely online and are cited in the references. Data used in analysis and the creation of figures are available at doi: 10.5281/zenodo.4434566.

References

Alsdorf, D. E., E. Rodriguez, and D. P. Lettenmaier (2007), Measuring surface water from space, Rev. Geophys., 45, RG2002, doi:10.1029/2006RG000197.

Anderson, J. T., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano (2009), The Data Assimilation Research Testbed: A community facility, Bull. Amer. Meteor. Soc., vol. 90, pp. 1283-1296, doi:10.1175/2009BAMS2618.1.

Andreadis, K. M., E. A. Clark, D. P. Lettenmaier, and D. E. Alsdorf (2007), Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model, Geophys. Res. Lett., 34(L10403), 1-5, doi:10.1029/2007GL029721.

Arnault, J., S. Wagner, T. Rummler, B. Fersch, J. Bliefernicht, S. Andresen, and H. Kunstmann (2016), Role of Runoff–Infiltration Partitioning and Resolved Overland Flow on Land–

499 Atmosphere Feedbacks: A Case Study with the WRF-Hydro Coupled Modeling System
500 for West Africa. J. Hydrometeor., 17, 1489–1516, doi:10.1175/JHM-D-15-0089.1.
501 Aviso+ (2015), SWOT orbit, https://www.aviso.altimetry.fr/en/missions/future-
502 missions/swot/orbit.html, accessed 8 September 2015.
503 Biancamaria., S., M. Durand, K. M. Andreadis, P. D. Bates., A. Boone, N. M. Mognard, and E.
504 A. Clark (2011), Assimilation of virtual wide swath altimetry to improve Arctic river
505 modeling, Remote Sens. Environ., 115(2), 373-381, doi:10.1016/j.rse.2010.09.008.
506 Biancamaria, S., F. Frappart, A.-S. Leleu, V. Marieu, D. Blumstein, J. D., Desjonqueres, F. Boy,
507 A. Sottolichio, and A. Valle-Levinson (2016), Satellite radar altimetry water elevations
508 performance over a 200 m wide river: Evaluation over the Garonne River, Advances in
509 Space Research, 59(2017), 128-146, doi:10.1016/j.asr.2016.10.008.
510 Biancamaria, S., F. Frappart, A.-S. Leleu, V. Marieu, D. Blumstein, J. D., Desjonqueres, F. Boy,
511 A. Sottolichio, and A. Valle-Levinson (2017), Satellite radar altimetry water elevations
512 performance over a 200 m wide river: Evaluation over the Garonne River, Advances in
513 Space Research, 59(2017), 128-146, doi:10.1016/j.asr.2016.10.008.
514 Boning, C. W., (1992), Policy statement on stage accuracy. Tech. rep., Office of Surface Water
515 Technical Memorandum No. 93.07, available at
516 https://water.usgs.gov/admin/memo/SW/sw93.07.html.
517 Cao, W., W. B. Bowden, T. Davie, and A. Fenemor (2006), Multi-variable and multi-site
518 calibration and validation of SWAT in a large mountainous catchment with high spatial
519 variability. Hydrol. Process., 20, 1057-1073, doi:10.002/hyp.5933.
520 Cunge, J. A. (1969), On the subject of a flood propagation computation method (Muskingum
521 method), J. Hydraul. Res., 7, 205-230.

Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S. Thober (2016), The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, J. Geophys. Res. Atmos., 121, 10,676–10,700, doi:10.1002/2016JD025097.

Durand, M. K., K. M. Andreadis, D. E. Alsdorf, D. P. Lettenmaier, D. Moller, and M. Wilson (2008), Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. Geophys. Res. Lett., 35 (L20401), 1-5, doi:10.1029/2008GL034150.

Durand, M., E. Rodriguez, D. E. Alsdorf, and M. Trigg (2010), Estimating river depth from remote sensing swath interferometry measurements of river height, slope and width, IEEE. J. Sel. Top. Appl. Earth Observ. Remote Sens., 3(1), 20-31, doi:10.1109/JSTARS.2009.2033453.

Durand, M., J. Neal, E. Rodriguez, K. M. Andreadis, L. C. Smith, and Y. Yoon (2014), Estimating reach-averaged discharge for the River Severn from measurements of river water surface elevation and slope, J. Hydrol., 5111, 92-104, doi:10.1016/j.jhydrol.2013.12.050.

Durand, M., et al. (2016), An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope, Water Resour. Res.,52, 4527–4549, doi:10.1002/2015WR018434.

Elmer, N. J., C. R. Hain, F. Hossain, D. Desroches, C. Pottier (2020), Generating proxy SWOT water surface elevations using WRF-Hydro and the CNES SWOT Hydrology Simulator, Water Resour. Res., Early Online Release, doi:10.1002/essoar.10502399.1.

544 Fisher, C. K., M. Pan, and E. F. Wood (2020), Spatiotemporal assimilation-interpolation of
545 discharge records through inverse streamflow routing, Hydro. Earth. Syst. Sci., 24, 293-
546 305, doi:10.5194/hess-24-293-2020.
547 Fjørtoft R, Gaudin JM, Pourthie N, Lalaurie JC, Mallet A, Nouvel JF, Martinot-Lagarde J, Oriot
548 H, Borderies P, Ruiz C, Daniel S (2014), KaRIn on SWOT: characteristics of near-nadir
549 Ka-band interferometric SAR imagery. IEEE Trans Geosci Remote Sens 52(4):2172–
550 2185, doi:10.1109/TGRS.2013.2258402.
551 Gochis, D.J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J.
552 Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, W. Yu, (2018). The WRF-Hydro
553 modeling system technical description, (Version 5.0). NCAR Technical Note. 107 pages.
554 Available online at https://ral.ucar.edu/sites/default/files/public/WRF-
555 HydroV5TechnicalDescription.pdf. Source Code DOI:10.5065/D6J38RBJ.
556 Gochis, D.J. and F. Chen, 2003: Hydrological enhancements to the community Noah land
557 surface model. NCAR Technical Note, NCAR/TN-454+STR, 68 pp.
558 Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., and Palmer, T. (2012). Comparing
559 TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts.
560 Quarterly Journal of the Royal Meteorological Society, 138(668), 1814–1827.
561 Hirsch, R. and J. E. Costa (2004), U.S. streamflow measurement and data dissemination
562 improve. Eos Trans. AGU, 85 (20), 197-203, doi:10.1029/2004EO200002.
563 Kouraev, A. V., E. A. Zakharova, O. Samain, N. M. Mognard, and A. Cazenave (2004), Ob river
564 discharge from TOPEX/Poseidon satellite altimetry (1992–2002), Remote Sens.
565 Environ., 93, 238-245.

McCuen, R. H., Z. Knight, and A. G. Cutter (2006), Evaluation of the Nash–Sutcliffe efficiency index, J. Hydrol. Eng., 11(6), 597–602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597).

Munier, S., A. Polebistki, C. Brown, G. Belaud, and D. P. Lettenmaier (2015), SWOT data assimilation for operational reservoir management on the upper Niger River Basin, Water Resour. Res., 51, 554-575, doi:10.1002/2014WR016157.

Niraula, R., L. M. Norman, T. Meixner, and J. B. Callegary (2012), Multi-gauge calibration for modeling the semi-arid Santa Cruz watershed in Arizona-Mexico border area using SWAT. Air, Soil and Water Research, 5, 41-57, doi:10.4137/ASWR.S9410.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116(D12109), doi:10.1029/2010JD015139.

O'Loughlin, F. E., J. Neal, D. Yamazaki, and P. D. Bates (2016), ICESat-derived inland water surface spot heights, Water Resour. Res., 52, 3276-3284, doi:10.1002/2015WR018237.

OWP (cited 2020), The National Water Model, http://water.noaa.gov/about/nwm.

Paiva, R. C. D., M. T. Durand, and F. Hossain (2015), Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data, Water Resour. Res., 51, 430–449, doi:10.1002/2014WR015618.

Papa, F., F. Durand, W. B. Rossow, A. Rahman, and S. K. Bala (2010), Satellite altimeter-derived monthly discharge of the Ganga-Brahmaputra River and its seasonal to interannual variations from 1993 to 2008, J. Geophys. Res., 115(C12013), doi:10.1029/2009JC006075.

589 Pavelsky, T. M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. D. Paiva, G. H. Allen,
590 and Z. F. Miller (2014), Assessing the potential global extent of SWOT river discharge
591 observations, J. Hydrol., 519, 1516-1525, doi:10.1016/j.jhydrol.2014.08.044.
592 Pedinotti, V., A. Boone, S. Ricci, S. Biancamaria, and N. Mognard (2014), Assimilation of
593 satellite data to optimize large-scale hydrological model parameters: a case study for the
594 SWOT mission, Hydrol. Earth Syst. Sci., 18, 4485-4507, doi:10.5194/hess-18-4485-
595 2014.
596 Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B.
597 Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D.
598 Toll (2004), The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc.,
599 85(3), 381-394.
600 Rodriguez, E. (2016), Surface Water and Ocean Topography Mission (SWOT) Project, Science
601 Requirements Document, Rev A, 28 pp., https://swot.jpl.nasa.gov/files/swot/D-
602 61923_SRD_Rev%20A_20160318%20with%20signatures1.pdf.
603 Sampson, K. and D. Gochis (2015), WRF-Hydro GIS pre-processing tools: Version 2.2
604 documentation, Tech. rep., Boulder, Colorado, 39 pages.
605 Senatore, A., G. Mendicino, D. J. Gochis, W. Yu, D. N. Yates, and H. Kunstmann (2015), Fully
606 coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of
607 enhanced hydrological parameterization for short and long time scales, J. Adv. Model.
608 Earth Syst., 7, 1693-1715, doi:10.1002/2015MS000510.
609 Sharma, S., Siddique, R., Reed, S., Ahnert, P., and Mejia, A. (2019). Hydrological model
610 diversity enhances streamflow forecast skill at short- to medium-range timescales. Water
611 Resources Research, 55, 1510–1530, doi:10.1029/2018WR023197.

612 Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang,
613 W. Wang, and J. G. Powers (2008), A description of the Advanced Research WRF
614 version 3, NCAR Technical Note, 475, 113 pp.,
615 http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.
616 U. S. Geological Survey (2017), National Elevation Dataset (NED),
617 https://nationalmap.gov/elevation.html.
618 Weigel, A. P., Liniger, M., and Appenzeller, C. (2008). Can multi-model combination really
619 enhance the prediction skill of probabilistic ensemble forecasts? Quarterly Journal of the
620 Royal Meteorological Society, 134(630), 241–260.
621 Weigel, A. P., Liniger, M. A., and Appenzeller, C. (2009). Seasonal ensemble forecasts: Are
622 recalibrated single models better than multimodels? Monthly Weather Review, 137(4),
623 1460–1479.
624 Yang, Y., C. Fisher, and J. T. Reager (2019), Enhancing SWOT discharge assimilation through
625 spatiotemporal correlations, Remote Sensing of Environment, 234, 111450,
626 doi:10.1016/j.rse.2019.111450.
627 Yoon, Y., M. Durand, C. J. Merry, E. A. Clark, K. M. Andreadis, and D. E. Alsdorf (2012),
628 Estimating river bathymetry from data assimilation of synthetic SWOT measurements, J.
629 Hydrol., 464-465, 363-375, doi:10.1016/j.jhydro.2012.07.028.
630 Yucel, I., A. Onen, K. K. Yilmaz, and D. J. Gochis (2015), Calibration and evaluation of a flood
631 forecasting system: Utility of numerical weather prediction model, data assimilation, and
632 satellite-based rainfall, J. Hydrol., 523, 49-66, doi:10.1016/j.jhydrol.2015.01.042.
633

List of Tables

List of Figures

Figure 9. Mean sensitivity (% change per unit) of the probability that the selected parameter set is ranked at or above each rank position with respect to changes in proxy SWOT discharge error ε (blue) and SWOT observation frequency (orange) for single-point (solid) and multi-point (dashed) parameter selection. Since error sensitivity is likely non-linear, note that this evaluation is only valid for ε between 0.20 and 0.35. Units are shown in the legend in parentheses, and rank is determined by NSE_{TRUTH} and $\overline{NSE_{TRUTH}}$. The vertical dashed line marks the 10th-ranked member.

Figure 10. As in Figure 8a, but for proxy SWOT mimicking observation frequency for a) midlatitudes and b) tropics as opposed to Alaska. Results are for multipoint parameter selection.

Tables

Table 1. Noah-MP and WRF-Hydro parameterization options used. More information about these options are available in Niu et al. (2011) and Gochis et al. (2018).

Noah-MP Namelist Option	Namelist Value
Dynamic Vegetation Option	4 (table leaf area index, maximum GVF)
Canopy Stomatal Resistance Option	1 (Ball-Berry)
BTR Option	1 (Noah)
Runoff Option	3 (free drainage)
Surface Drag Option	1 (M-O)
Frozen Soil Option	1 (linear effects)
Supercooled Water Option	1 (no iteration)
Radiative Transfer Option	3 (two-stream applied to vegetated fraction)
Snow Albedo Option	2 (CLASS)
PCP Partition Option	1 (Jordan 1991)
TBOT Option	2 (original Noah)
Temp Time Scheme Option	3 (semi-implicit)
Glacier Option	2 (original Noah)
Surface Resistance Option	4 (Sakaguchi and Zeng for non-snow, snow surface resistance for snow)
WRF-Hydro	
Channel Routing Option	Truth run: 2=Muskingum-Cunge, Calibration runs: 3=Diffusive Wave (gridded)
Overland Flow Routing Option	1 (D8)
Groundwater/Baseflow Routing Option	1 (Exponential Bucket)
Resolutions	
LSM Timestep	Truth run: 1 hr Calibration runs:3 hr
LSM Spatial	1 km
Overland/Suburface Spatial	Truth: 100 m Calibration runs: 250 m

686 Table 2. Parameter table listing the model parameters comprising parameter space. The listed
687 value was applied either as a multiplicative factor (Mult) or as an absolute (substituted) value
688 (Abs).

Component	Parameter (Variable name)	Variable name	Scaling	Minimum Value	Maximum Value	Default value
LSM	Clapp-Hornberger B exponent	bexp	Mult	0.4	1.9	1.0
	Soil moisture maximum	smcmax	Mult	0.8	1.2	1.0
	Saturated soil conductivity	dksat	Mult	0.2	10.0	1.0
	Soil infiltration parameter	refkdt	Abs	0.1	4.0	0.6
	Soil drainage parameter	slope	Abs	0.0	1.0	0.1
	Canopy wind parameter	CWPVT	Mult	0.5	2.0	1.0
	Maximum carboxylation at 25°C	VCMX25	Mult	0.6	1.4	1.0
	Ball-Berry conductance relationship slope	MP	Mult	0.6	1.4	1.0
	Snowmelt parameter	MFSNO	Abs	0.5	3.5	2.0
Overland/ subsurface	Retention depth	RETDEPRTFAC	Abs	0.1	10.0	1.0
	Saturated soil lateral conductivity	LKSATFAC	Abs	10	10 000	1000
Bucket	Groundwater bucket maximum depth	Zmax	Abs	10	250	25
	Groundwater bucket exponent	Expon	Abs	1.0	8.0	1.75
Channel	Manning's roughness coefficient	MannN	Abs	**Minimum Value**	**Maximum Value**	**Default value**
			Order			
			1	0.45	0.65	0.55
			2	0.25	0.45	0.35
			3	0.125	0.25	0.15
			4	0.085	0.125	0.10
			5	0.060	0.085	0.07
			6	0.045	0.060	0.05
			7	0.035	0.045	0.04
			8	0.025	0.035	0.03
			9	0.015	0.025	0.02
			10	0.005	0.015	0.01

689

690 Table 3. Comparison between truth model, 250-m model, and proxy SWOT time series.

	Truth (100-m) model	**250-m model**	**Proxy SWOT**
Overland Routing Spatial Resolution	100 m	250 m	Not Applicable
Channel Routing Scheme	Muskingum-Cunge (vector)	Diffusive Wave (gridded)	Not Applicable
Temporal Resolution	1-hour	3-hour	Irregular
Calibration	Based on Chena River watershed calibration using USGS gauges	Uncalibrated	Not Applicable
Sets	1	80	10,000

691

692 Table 4. Percent of sub-basins successfully calibrated using SWOT observations for single-point
693 and multi-point calibration methods, with a successful parameter selection defined as a >80%
694 probability that the selected parameter set is ranked within the top ten sets (top 12.5%) as
695 determined by NSE_{TRUTH} and $\overline{NSE_{TRUTH}}$ rank.

Calibration Method	**ε**	**Alaska / Polar**	**Midlatitudes**	**Tropics**
Single-point	**0.20**	80	60	40
	0.35	60	20	10
Multi-point	**0.20**	90	80	80
	0.35	90	80	70

696

Figures

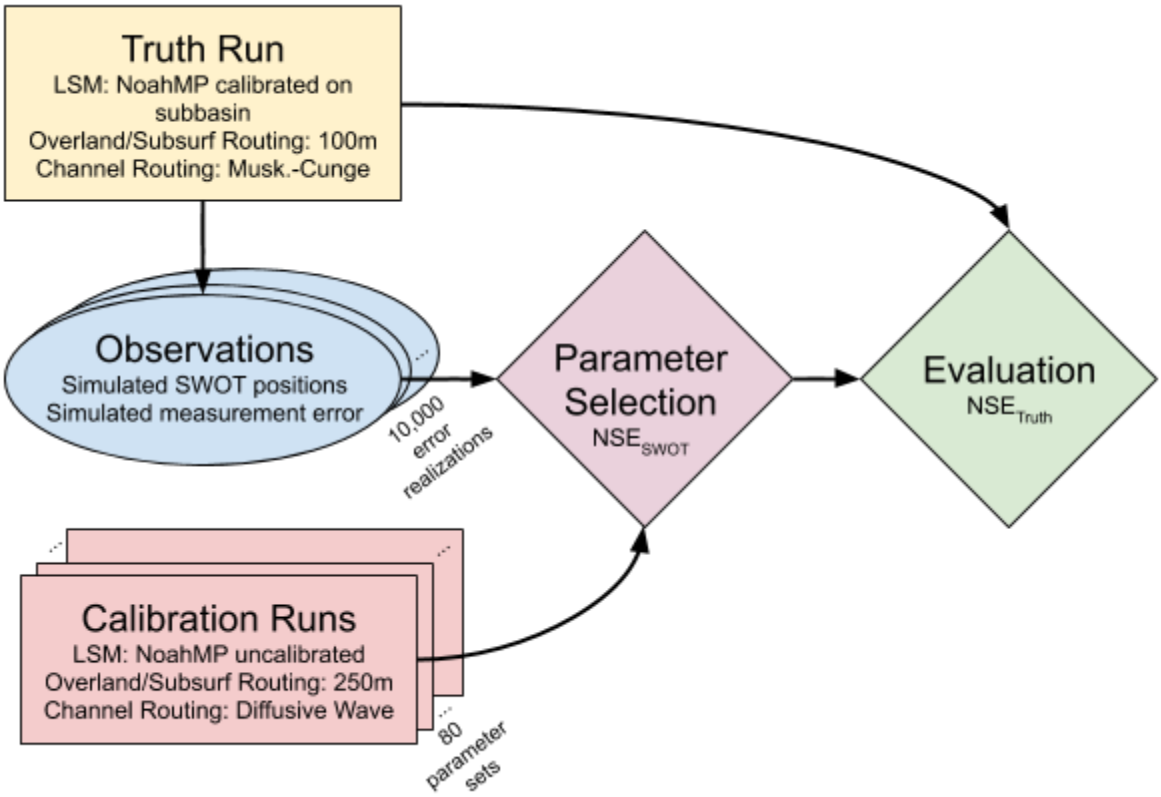


Figure 1. Design of "fraternal twin" experiment for evaluating the utility of (simulated) SWOT for hydrologic parameter selection with the WRF-Hydro model: Do the space-time sampling and observation error characteristics of SWOT permit accurate calibration?

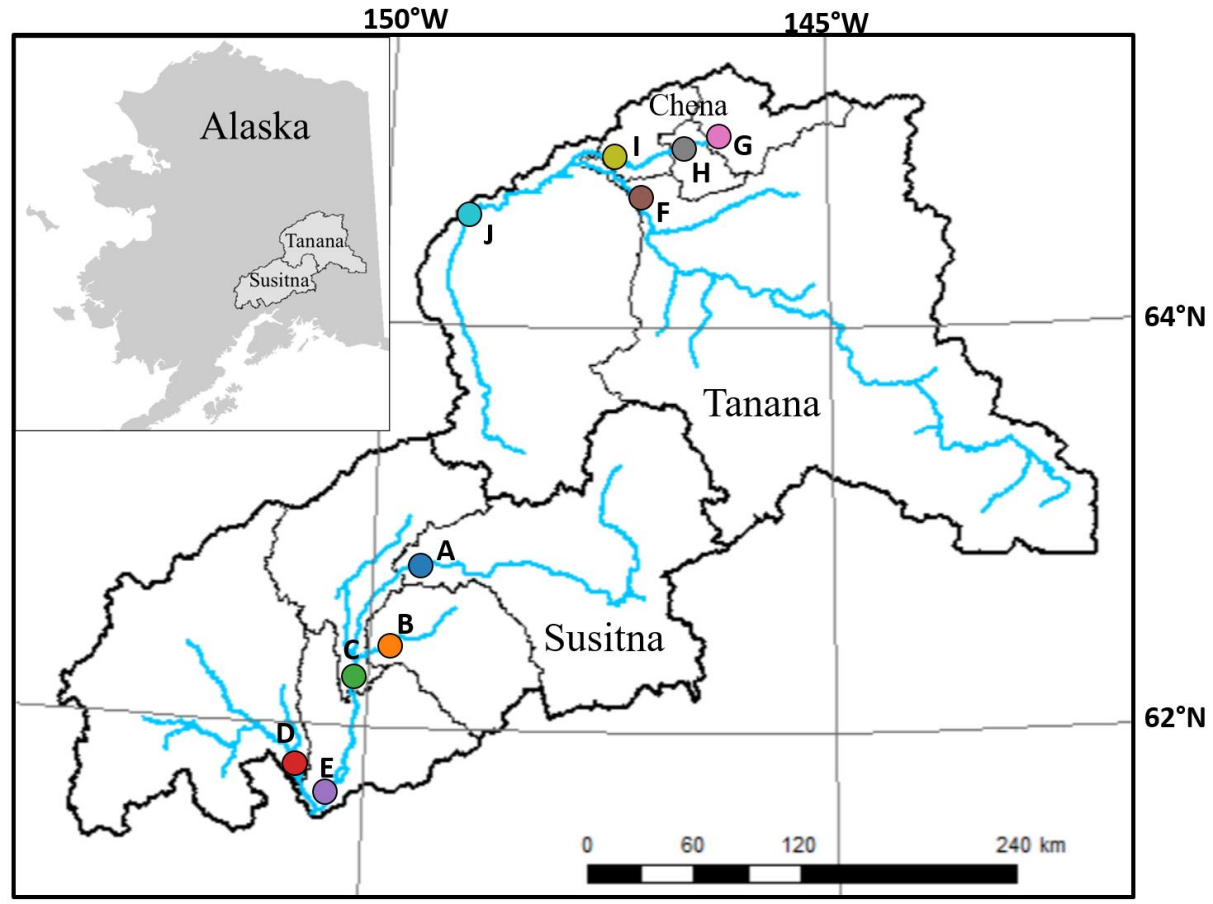


Figure 2. Study area within the Susitna River and upper Tanana River basins. Sub-basins A-J indicated by the letters and colors correspond with Figure 6. Streams of order five and greater are designated in blue.

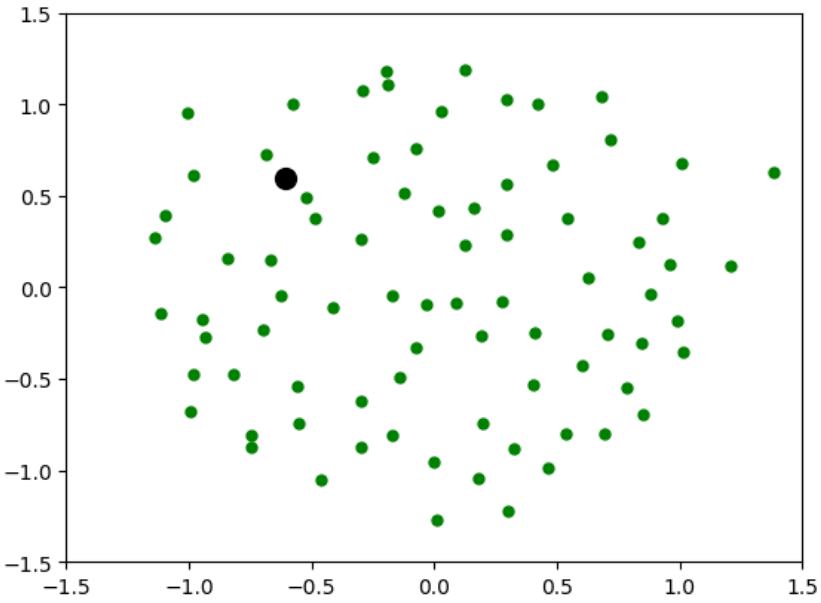
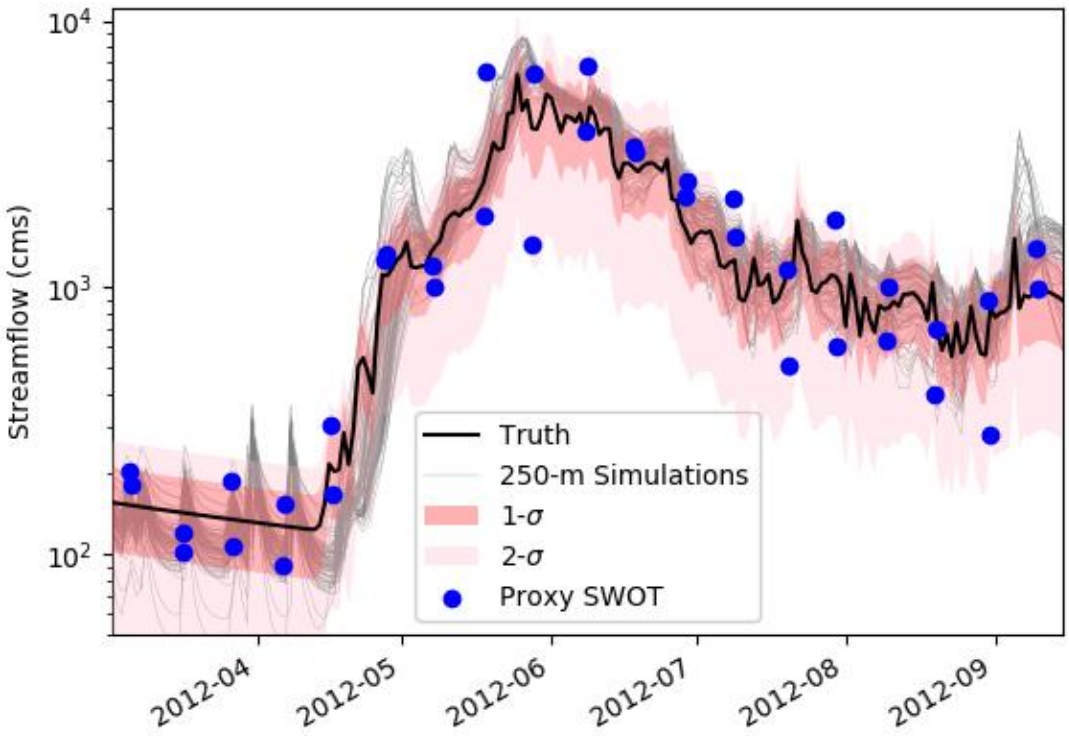


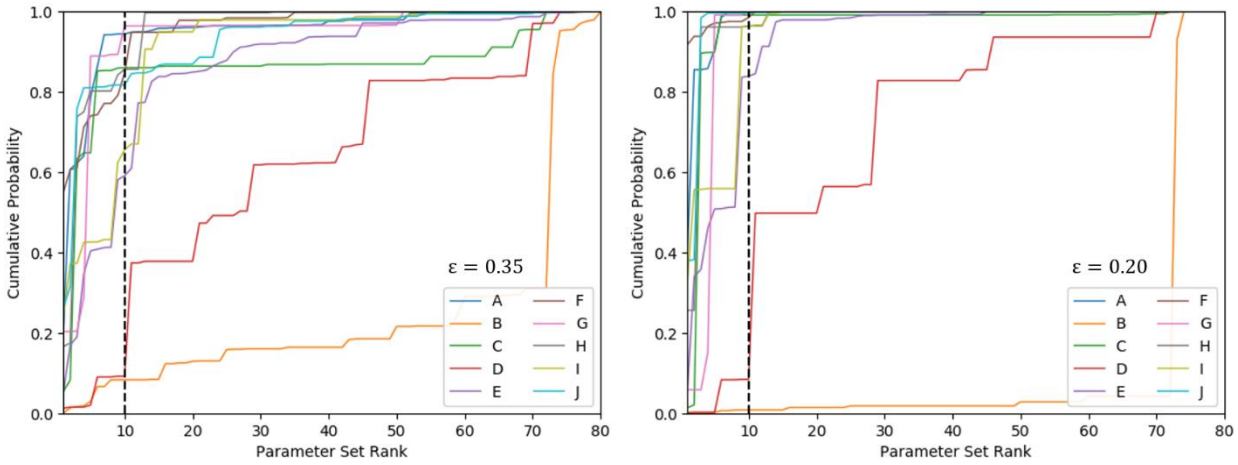
Figure 3. Distribution of the sampled parameter sets (green) used in this study visualized using multidimensional scaling (MDS). For reference, the calibrated parameter set for the truth model is shown in black.



713

714 Figure 4. 250-m simulations, truth simulation, 1-σ and 2-σ discharge error ranges, and a sample
715 proxy SWOT discharge observation set (i.e., hydrograph) for a single point corresponding to the
716 sub-basin E outlet.

717



(a) (b)

Figure 5. Percent of sub-basins with successful single-point parameter selection for a given cumulative probability and parameter set rank for a discharge error ε of a) 0.35 and b) 0.20. Rank is determined by NSE_{TRUTH}. The vertical dashed line marks the 10th-ranked member.

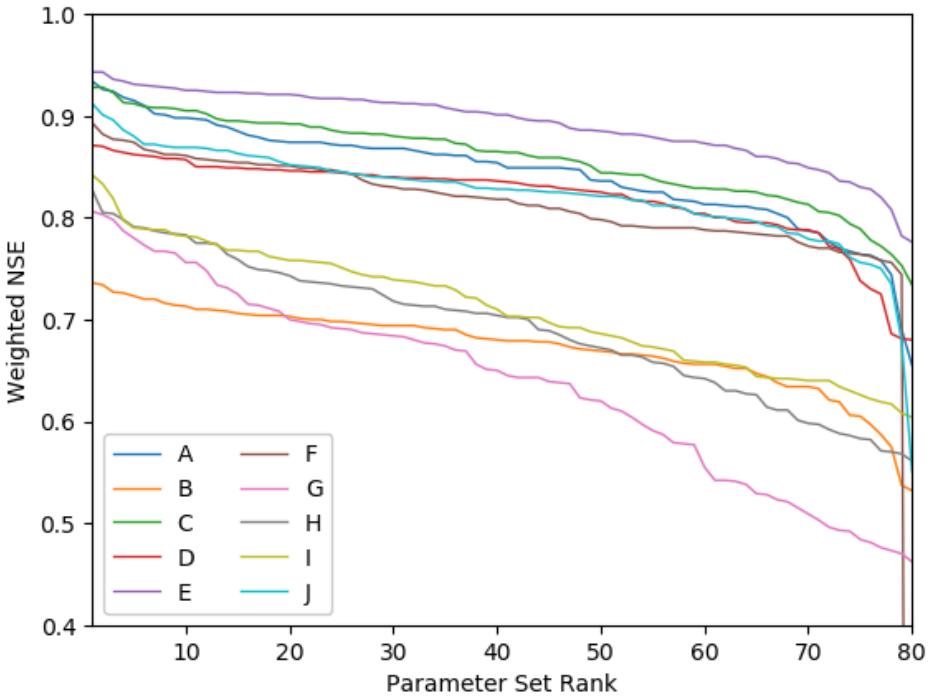
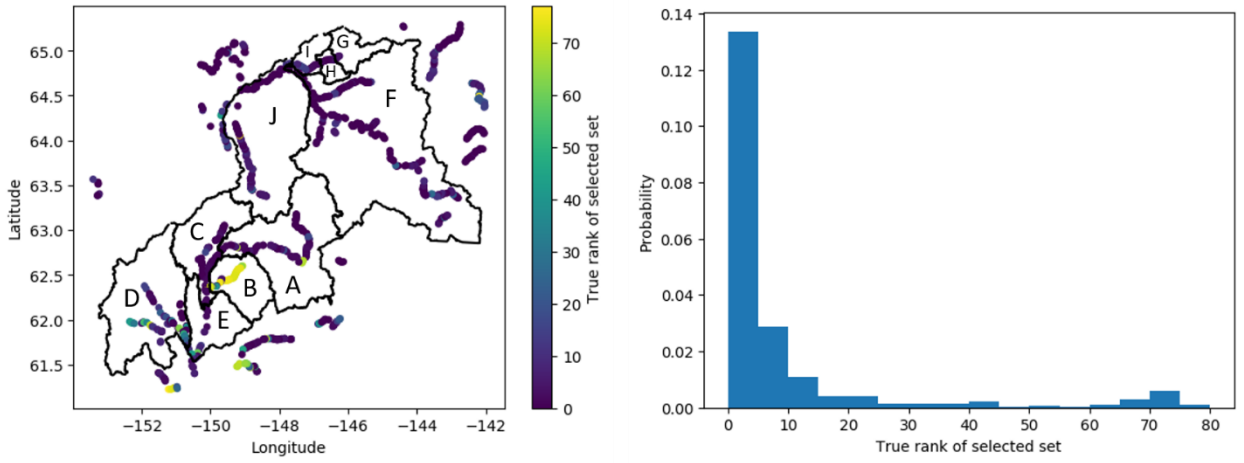
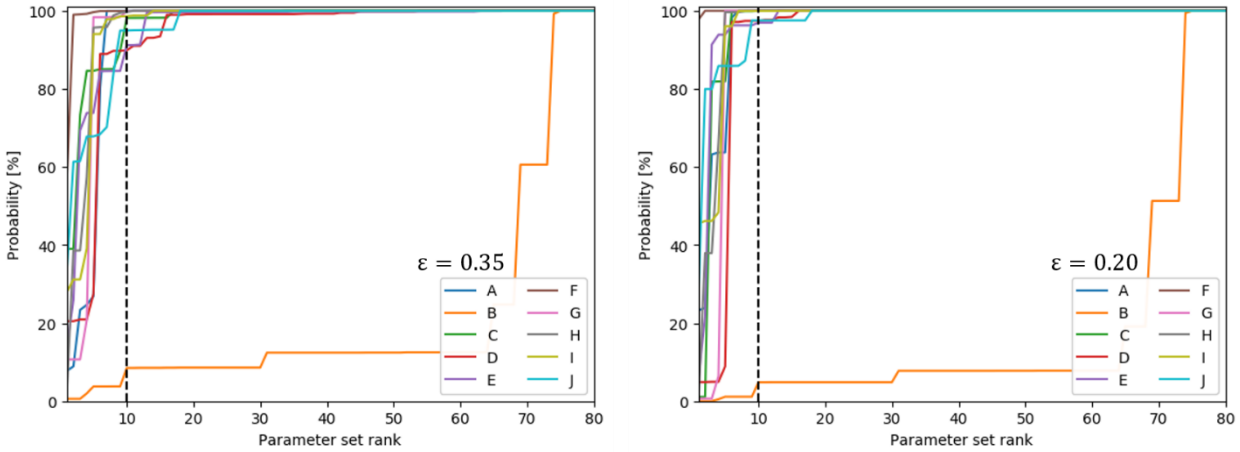


Figure 6. NSE_{TRUTH} for each point A-J in Figure 2. Parameter set rank is determined by NSE_{TRUTH} at the basin outlet (single-point).



(a) (b)

Figure 7. a) Map and b) histogram depicting the true rank of selected parameter set for $\epsilon = 0.35$ in Alaska. All SWOT observable channel reaches are shown. Basin borders match those in Figure 2. Zero indicates the best true rank.



(a) (b)

Figure 8. Same as Figure 5, but for multi-point parameter selection showing results for a) ε=0.35 and b) ε=0.20.

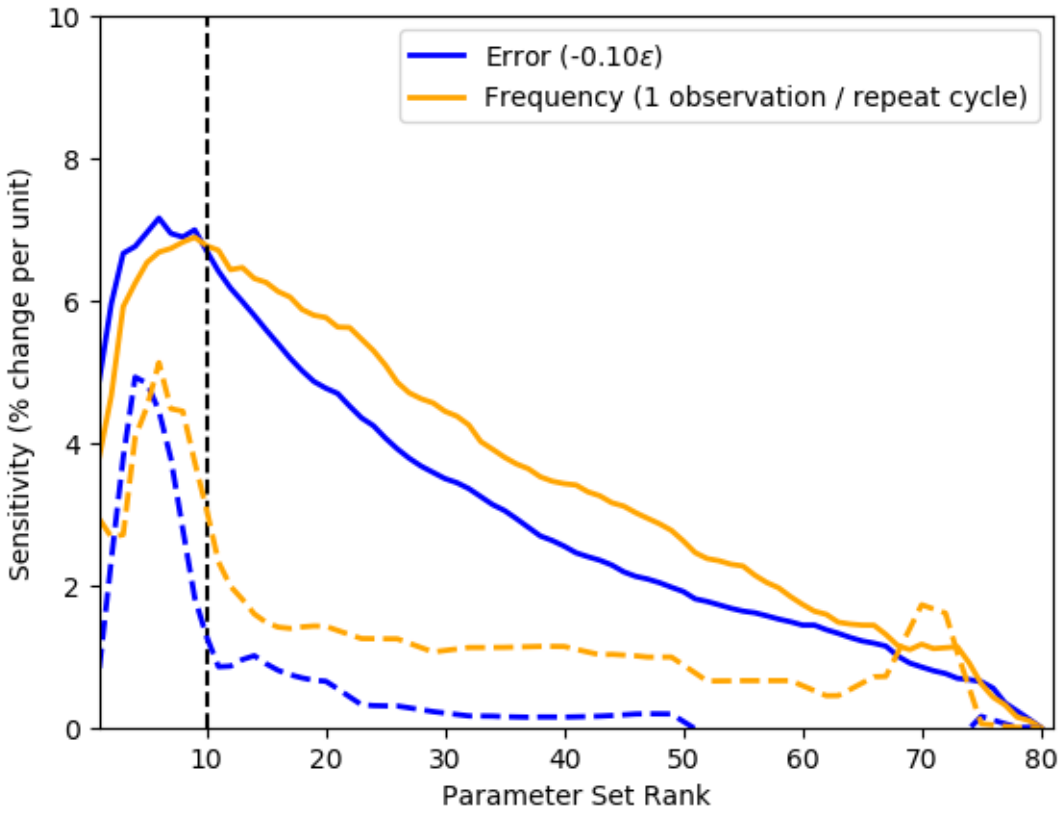
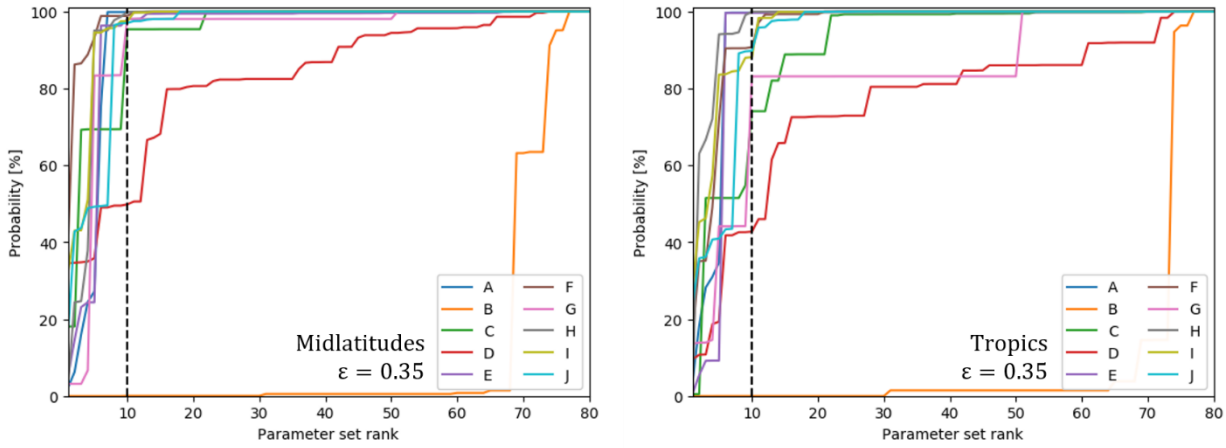


Figure 9. Mean sensitivity (% change per unit) of the probability that the selected parameter set is ranked at or above each rank position with respect to changes in proxy SWOT discharge error ε (blue) and SWOT observation frequency (orange) for single-point (solid) and multi-point (dashed) parameter selection. Since error sensitivity is likely non-linear, note that this evaluation is only valid for ε between 0.20 and 0.35. Units are shown in the legend in parentheses, and rank is determined by NSE_{TRUTH} and $\overline{NSE_{TRUTH}}$. The vertical dashed line marks the 10th-ranked member.



(a) (b)

Figure 10. As in Figure 8a, but for proxy SWOT mimicking observation frequency for a) midlatitudes and b) tropics as opposed to Alaska. Results are for multipoint parameter selection.