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

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

[1]NASA Postdoctoral Program
[2]National Center for Atmospheric Research (UCAR)
[3]Marshall Space Flight Center, NASA

November 30, 2022

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.

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

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

[1] Science and Technology Institute, Universities Space Research Association, 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, Universities Space Research Association, 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 random error of 35%, multi-point parameter estimation is successful for 90% of polar and 60% of tropical sub-basins.
- Multi-point parameter selection is preferred over single-point parameter selection, offering more robust results with less sensitivity.

Abstract

In situ gauge networks are often used in hydrologic 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 wider than 100 meters. SWOT observation utility for model parameter selection in regions devoid of in situ gauges is assessed using proxy SWOT discharge estimates derived from an observing system simulation experiment and Monte Carlo methods. The sensitivity of the parameter selection to measurement error and observation frequency is also evaluated. Single- and multi-point parameter selection are performed for ten sub-basins within the Susitna 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 measurement error of 35%, parameter estimation is successful for 50% (90%) of sub-basins using single- (multi-) point parameter selection. Decreasing observation frequency to simulate lower latitudes resulted in success for only 10% of midlatitude and tropical sub-basins for single-point selection, whereas multi-point selection was successful in 80% (60%) of midlatitude (tropical) sub-basins. Single-point parameter selection is more sensitive to measurement error than multi-point parameter selection. The results strongly support the use of multi-point over single-point parameter selection, yielding robust results nearly independent of observation frequency. Most importantly, this study suggests SWOT can be used to successfully select hydrologic model parameters in basins without an in situ gauge network.

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 early 2022 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 wider 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 river stage 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 between sites 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.

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. Previous multi-point calibration studies use robust gauge networks for their analysis, but in situ gauges have substantially lower observation error and higher temporal sampling than is expected of SWOT. Thus, this study is essential in understanding whether single- and multi-point parameter selection can be performed solely using SWOT observations. A related study, Nickles et al. (2020), compared hydrologic model multi-site calibration results using daily stream gauge observations, SWOT temporally-sampled discharge (no uncertainty), and SWOT temporally-sampled discharge

111 with uncertainty (systematic and random error), finding that SWOT achieves similar calibration

112 results to in situ daily observations. However, this study differs from Nickles et al. (2020) by

113 expanding results beyond the mid-latitudes and investigating the sensitivity of parameter selection

114 to a potential range of SWOT measurement error.

115

2. Methodology

117 2.1 Experimental Design

118 The design of our fraternal twin parameter selection experiment is shown in Figure 1. This

119 Observing System Simulation Experiment (OSSE) is based on Elmer et al. (2020a,b). The fraternal

120 twins, the "truth run" and "calibration run", simulate model error by employing different

121 hydrologic model representations in the model chain that generates streamflow. The experiment

122 addresses whether the unknown, best parameters for the calibration run can be reliably selected

123 (purple box in Figure 1) from observations of the truth run streamflow imparted with the expected

124 observation error characteristics of the SWOT sensor. Successfully identifying the best parameters

125 from observations is the core of model calibration. In this experiment, because we know the true

126 streamflow values, we can evaluate under what conditions parameter selection is successful.

127 Parameter, model, and observation errors are all ingredients of the experimental design.

128 The parameter error is the quantity we seek to minimize in parameter selection and calibration via

129 the objective function. The truth run was pre-calibrated to a single subdomain of the study and has

130 a single realization (yellow box, Figure 1). In contrast, the calibration runs span the space of

131 thirteen model parameters plus Manning's roughness coefficient using 80 parameter sets. This is

132 represented by the stack of calibration runs (red boxes) in Figure 1. Model error of the calibration

133 runs relative to the truth run is generated by differences summarized by text in the respective boxes

134   in Figure 1. These differences produce errors between the runs in terms of 1) the fixed boundary

135   conditions or geometry for different land surface model (LSM) resolutions and channel routing

136   networks, 2) the LSM and channel parameters, particularly infiltration parameters, which depend

137   on spatial and temporal model resolutions, and 3) streamflow physics. We note that both the

138   atmospheric forcing variables and LSM models are identical between the runs but that errors or

139   differences in the model runs start with and accumulate over time within the soil moisture

140   representation and its two-way coupling to the overland and subsurface runoff models (Gochis and

141   Chen 2003), which feeds back to LSM behavior and parameter differences. The differences

142   between the fraternal twins are described in further detail below.

143   The SWOT observation (discharge) errors are considered both systematic (biased) and

144   random. 10,000 realizations of observation errors are applied to the true states before use in

145   parameter selection to avoid drawing conclusions from a particular set of random errors. This set

146   of 10,000 possible observation realizations is represented by the stack of observations (blue ovals)

147   in Figure 1. Although results using as few as 1000 realizations would have led to similar

148   conclusions, 10,000 realizations are used for this study to ensure robust results. The use of 10,000

149   realizations is also consistent with other studies (e.g., Nickles et al. 2020). Over the 10,000

150   observation sets, the probability of selection (identification as the best parameter set via a version

151   of Nash-Sutcliffe Efficiency (NSE; McCuen et al. 2006) based on the observations) is computed

152   for each of the 80 parameter sets (purple diamond, Figure 1). Finally, in the evaluation step (green

153   diamond, Figure 1), NSE is computed from the true model states and the true rank of each

154   parameter set is assigned, from best (low) to worst (high). The cumulative probability of parameter

155   selection (under observation error) is plotted against rank. Do the true best parameter sets have a

156   reasonably high likelihood of selection in the presence of SWOT observation characteristics (and

157   model error)?

158   This paper focuses on sub-basins within the upper Tanana River and Susitna River basins

159   in Alaska, which will be regularly observed by SWOT (Biancamaria et al. 2016) but have few in

160   situ observations. These watersheds are delineated in Figure 2. The following subsections provide

161   additional details for each step of the experimental design shown in Figure 1.

162

163   2.2 Model configurations and parameters

164   For both truth and calibration model runs in Figure 1, this study uses WRF-Hydro version

165   5.0 (Gochis et al. 2018) configured to mimic the NWM v2.0 configuration (OWP 2020). The Noah

166   land surface model with Multi-Parameterization options (Noah-MP; Niu et al. 2011) with a 1 km

167   spatial resolution is used as the WRF-Hydro land surface model in both models, as well as Global

168   Land Data Assimilation System (GLDAS) Version 2 (Rodell et al. 2004) meteorological forcing.

169   Further details of the model configuration and physics parameterizations used are listed in Table

170   1, which also lists the differences between the truth run and the calibration runs.

171   The truth (calibration) run configuration has a model timestep of 1 (3) hours, performing

172   overland and subsurface routing on a 100m (250m) grid, and uses the Muskingum-Cunge

173   (diffusive wave) routing scheme for simulating streamflow within a channel model. GLDAS

174   forcing data is available at three-hour increments and ingested into both configurations equally.

175   The WRF-Hydro terrain routing grids (100m and 250m) and channel networks were derived from

176   the WRF-Hydro GIS Pre-processing Toolkit v5.1 (Sampson and Gochis 2015) using the Weather

177   Research and Forecasting (WRF; Skamarock et al. 2008) Preprocessing System geogrid file and

178   the National Elevation Dataset (NED; U. S. Geological Survey 2017) Digital Elevation Model

179 (DEM) as inputs. Both the Muskingum-Cunge and diffusive wave schemes represent channels

180 with an infinite depth, preventing overbank flow. However, the diffusive wave scheme allows

181 backwater effects, whereas the Muskingum-Cunge scheme does not. Importantly, the channel

182 networks are derived using different DEM spatial resolutions, leading to different spatial

183 representations of the channel routing.

184   Whereas we calibrate the truth run to in situ streamflow observations (described below),

185 the calibration run configuration is uncalibrated: the point of our experiment being selection of

186 parameters for the calibration run that most accurately simulate the truth run. Eighty calibration

187 parameter sets were created by perturbing Manning's roughness coefficient (as a function of

188 stream order) along with the thirteen most sensitive WRF-Hydro parameters (Cuntz et al. 2016;

189 Elmer 2019). As shown in Table 2, these parameters span the LSM, overland/subsurface routing,

190 groundwater bucket, and channel routing components of the model (model variable names shown

191 in parentheses): the Clapp-Hornberger B exponent (bexp), soil moisture maximum (smcmax),

192 saturated soil conductivity (dksat), soil infiltration parameter (refkdt), soil drainage parameter

193 (slope), retention depth (RETDEPRTFAC), saturated soil lateral conductivity (LKSATFAC),

194 groundwater bucket model max depth (Zmax), groundwater bucket model exponent (Expon),

195 canopy wind parameter (CWPVT), maximum carboxylation at 25°C (VCMX25) which is related

196 to the vegetation height (HVT), the Ball-Berry conductance relationship slope (MP), and the

197 snowmelt parameter (MFSNO). The ranges assigned to these parameters make up the calibration

198 parameter space. From this parameter space, a sample of parameter sets were obtained by randomly

199 assigning values within the valid parameter ranges listed in Table 2 using a uniform distribution.

200 The result is a good representation of parameter space, as shown by depiction of the distribution

201 of the sampled parameter sets using a multidimensional scaling (MDS) transform (Pedregosa et

202 al. 2011) in Figure 3. MDS is a method by which distances in multi-dimensional space (in this

203 case, a thirteen-dimensional parameter space) are transformed to two-dimensional distances while

204 maintaining the true distance in the original multi-dimensional space as closely as possible.

205 The truth model is calibrated using in situ USGS stream gauge observations at the basin

206 outlet using the parameter space described above (Table 2). The mean of the linear NSE and the

207 logarithmic NSE (NSE_{ln}) was used as the calibration metric, denoted as the mean NSE (NSE_{mean}).

208 NSE_{mean} is akin to the metric used in the calibration of the NWM and is given by:

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

210 where w is the weight of 0.5. NSE_{mean} ranges from negative infinity to unity, where a value greater

211 than zero indicates that the model provides a better estimate than the observation mean. Thus, the

212 maximum NSE_{mean} is sought. For single point parameter selection, NSE_{mean} is the metric for

213 evaluation. For multipoint parameter selection, a basin average NSE_{mean} is calculated for

214 evaluation, given by:

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

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

217 As the calibration process is computationally expensive, calibration of the truth simulation

218 was only performed for the Chena River watershed (within the upper Tanana River basin;

219 watershed outlet denoted by point I in Figure 2) and halted after 75 model iterations. The

220 parameters identified using the Chena River calibration were transferred to the full domain.

221 Although the Chena River calibration may not transfer well to the entire upper Tanana River and

222 Susitna River basins, the resulting model output is treated as truth for this experiment and therefore

223 a perfect calibration is not necessary. Certain parameters (e.g., infiltration parameters) are strongly

224 scale dependent, so the calibration of the truth model, in which the model resolution and

225 streamflow physics differ from the 250-m model, is not directly transferrable to the 250-m model.

226 The truth run provides the "true" geolocation and discharge time series (q) for the purposes of this

227 experiment.

228   A spin-up period of eight years (March 2009 - March 2017) using default parameter values

229 (Table 2) was performed, designed to allow for adequate accumulation of groundwater and

230 snowpack and permitting each 250-m simulation to reach equilibrium. The March 2017 restart

231 files from the spin-up were used to restart the 250-m simulations at March 2011 using their

232 respective parameter set and integrated forward in time in an open loop configuration. The periods

233 of March – September 2012 for the Susitna River basin and March – September 2014 for the upper

234 Tanana River basin were used during analysis to determine the utility of SWOT observations in

235 model parameter selection.

236

237 2.3. Generating Proxy SWOT Discharge

238   Since real SWOT data are not yet available, proxy SWOT observations were generated for

239 this analysis. Proxy SWOT data have been used by multiple studies to quantify assimilation

240 impacts on river modeling and reservoir management (Andreadis et al. 2007; Biancamaria et al.

241 2011, Munier et al. 2015; Emery et al. 2020; Revel et al. 2021; Yang et al. 2019; Wongchuig-

242 Correa et al. 2020) and develop procedures for estimating river bathymetry (Durand et al. 2008,

243 2010, 2014; Yoon et al. 2012). Furthermore, Pedinotti et al. (2014) used synthetic SWOT data to

244 optimize Manning roughness coefficients in the Interactions between Soil, Biosphere, and

245 Atmosphere-Total Runoff Integrating Pathways System (ISBA-TRIP) continental hydrologic

246 system using data assimilation, demonstrating that SWOT data can be used for calibration.

247   The truth model q is corrupted with random and systematic error following a Gaussian

248   distribution (N) with mean β and standard deviation ϵ:

249                        $$q' = q\big(1 + N(\beta, \epsilon)\big),   (3)$$

250   where $q = (q_{i,t_0}, q_{i,t_1}, \ldots, q_{i,t_n})$ is a discharge time series for the i-th channel reach, q' is the

251   corrupted discharge (i.e., the proxy SWOT discharge estimate), β is the relative bias, and ϵ is the

252   relative discharge error. For the analysis in Section 3.1, $\epsilon = 0.35$ is used, which is roughly

253   equivalent to the relative root mean squared error (RMSE) of instantaneous discharge estimated

254   by Durand et al. (2016). β is constant with time for each q, but varies in space (across channel

255   reaches). Since an equivalent increase in water surface height will yield a larger increase in

256   discharge for a river with a larger cross-section, uncertainty in q' naturally increases as q increases.

257   However, a single series of q' does not provide an adequate sampling of random error and

258   bias by which to assess calibration potential. Rather, it gives a snapshot of only one possible

259   scenario. Figure 4 illustrates this point. The blue dots indicate a single time series of q', containing

260   some points in which the discharge random errors are small and depart very little from the truth

261   value but also points that extend into the 2-σ error range. A single scenario may be skewed if only

262   small random errors are present ($q' \approx q$), especially at key points along the time series, which

263   would enable superior parameter selection and suggest better results than could actually be

264   expected. Conversely, a scenario containing frequent large errors, parameter selection would

265   underperform. To sample a broad spectrum of the possible outcomes stemming from a SWOT

266   observation set laden with error, Monte Carlo methods were employed to create 10,000 randomly-

267   perturbed sets of q' per sub-basin. Thus, each of the 10,000 error realizations of q' is characterized

268   by random white noise (ϵ) and time-constant systematic bias (β), where the value of β is randomly

269   drawn from N(0.00, 0.20) based on Hagemann et al. (2017) to account for relative bias prior to

270   applying Eq. (3). Following the creation of the 10,000 sets, the probability of successful sub-basin

271   parameter selection was calculated.

272   To obtain q' with appropriate orbit characteristics, it was spatially sampled based on the

273   CNES proxy SWOT orbit (Aviso+ 2015) with a simulated start date of 1 March 2012. First, the

274   cross-track distance of each WRF-Hydro reach from the proxy SWOT orbit at each overpass was

275   calculated. For each pass, only reaches with cross-track distances of 10-60 km (i.e., within the

276   SWOT measurement range) and with a Strahler streamorder greater than or equal to five (used to

277   approximate rivers with widths greater than 50 m) were extracted, following the methodology of

278   Elmer et al. (2020a), which showed that Alaskan rivers with a streamorder greater than or equal to

279   five will generally be observable by SWOT. Figure 3 compares the truth q and q' for sub-basin E,

280   where q' is used to calibrate the 250-m model following the method described in Section 2.3.

281

282   2.3. Parameter selection from proxy SWOT discharge observations

283   We purposely use the term "parameter selection" to differentiate our overall approach from

284   calibration for the following reasons. The parameter sets run through the model and ultimately

285   judged by the objective function are generated a priori through combinations of uniformly sampled

286   distributions on each parameter. As such, the parameters sets being discriminated via the objective

287   function are generally not "close" in parameter space (Figure 3). A true calibration approach would

288   consider points with a similar spacing in parameter space, but would also include parameter sets

289   much closer together in the quest to find minima of the objective function. This paper does not

290   study the ability to accurately find local minima of the objective function using SWOT

291   observations. Doing so would require understanding the relative sensitivities of the objective

292   function to observation error and to distance in parameter space. Rather, we examine the potential

293   for SWOT observations to give a more regional, less detailed picture of the objective function.

294   Given the observation and error characteristics of SWOT, including its spatially distributed nature,

295   and a finite collection of parameter sets, can we accurately select the best parameter set in this

296   collection? Our results provide an encouraging basis for pursuing model calibration using SWOT

297   observations.

298   To review and summarize the experiment design (Figure 1), the 250-m model described in

299   Section 2.1 represents an uncalibrated hydrologic model of an ungauged basin (a single red box,

300   Figure 1). For this basin, an infinite number of parameter sets can be selected from the parameter

301   space for calibration and the correct solution (parameter set) is unknown. The goal is to find the

302   best simulation of streamflow over a finite sample of parameter space. A sample size of 80

303   parameter sets (red stack of boxes, Figure 1) was chosen for this study to minimize computational

304   requirements. While the parameter set can certainly be expanded to more fully represent the whole

305   parameter space, the increased computational requirements may reduce the feasibility of this

306   method for users without access to large computing systems.

307   This study ensures the 250-m model is blind to the calibration of the truth model so that

308   the calibrated parameter set used by the truth model does not inform the selection of the parameter

309   set sampling for the 250-m model. In the absence of in situ gauges, the only source of regular

310   discharge observations for ungauged basins will be from the SWOT mission. The q' values derived

311   in Section 2.3 are representative of the SWOT discharge observations that will be available post-

312   launch, and are used to calibrate the model by finding the best parameter set from the 80 sample

313   parameter sets. A comparison between the 100-m truth model, the 250-m simulations, and the

314   proxy SWOT discharge estimates is provided in Table 3, with example data shown in Figure 4.

315   For this particular point, there are 38 SWOT overpasses from March 1 – September 15, yielding

316   an observation approximately every 5 days, or 4 observations per repeat cycle.

317   Channel reaches were spatially matched between the truth and calibration model channel

318   networks for evaluation, eliminating any matches separated by a 1 km or greater which are

319   considered unrelated. Thus, a total of 10 basins and 991 channel reaches were evaluated. All basins

320   were modeled simultaneously, but evaluated separately. Single-point parameter selection for the

321   uncalibrated 250-m model is performed using NSE_{SWOT}, given as the NSE_{mean} between the

322   discharge for each simulation and q' at each sub-basin outlet in Figure 2. Multi-point parameter

323   selection is evaluated with $\overline{NSE_{SWOT}}$, the basin average NSE_{SWOT} for all observed points P

324   (Equation 2). The 250-m simulation with the maximum NSE_{SWOT} or $\overline{NSE_{SWOT}}$ (the best match

325   parameter set) is chosen, and the parameter set used by that simulation for the thirteen most

326   sensitive WRF-Hydro parameters is selected as the "correct" parameter set.

327

328   2.4. Evaluation of parameter selection

329   The NSE_{mean} was also calculated between each 250-m simulation and the truth model q

330   (NSE_{TRUTH}) at each basin outlet for comparison with NSE_{SWOT} for single point parameter selection

331   evaluation. Multi-point parameter selection is similar to single-point selection except that

332    $\overline{NSE_{SWOT}}$ and $\overline{NSE_{TRUTH}}$ (the basin-average values for all observable channel reaches within each

333   sub-basin) is used. The 250-m simulations are separately ranked according to their NSE_{SWOT}

334   ($\overline{NSE_{SWOT}}$) and NSE_{TRUTH} ($\overline{NSE_{TRUTH}}$) values for single point (multi-point) parameter selection

335   with the member with the best (maximum) value being assigned a ranking of one, and the member

336   with the worst (minimum) value assigned a ranking of 80. Expressing the cumulative rank of

337   NSE_{SWOT} ($\overline{NSE_{SWOT}}$) as a function of NSE_{TRUTH} ($\overline{NSE_{TRUTH}}$) reveals whether single (multi) point

338   parameter selection using SWOT observations is effective. For the purpose of drawing conclusions

339   in this paper, a successful parameter selection is achieved for a watershed if the selected parameter

340   set is contained within the best ten sets according to the $\mathrm{NSE_{TRUTH}}$ ($\overline{NSE_{TRUTH}}$) rank with a >80%

341   probability. This criteria is subjective, thus plots showing the full range of probabilities are

342   included. For example, in Figure 5a the y-axis shows the cumulative probability whereas the x-

343   axis shows the parameter set rank. The cumulative probability is essentially the fraction of

344   observation sets (10,000 sets in total). Thus, we see that for sub-basin G (pink line), approximately

345   0.90 (90%) of the 10,000 observation sets rank the true best parameter set in the top ten (indicated

346   by vertical black dashed line). Since this value is above the 0.8 (80%) threshold, the parameter

347   selection is successful.

348

349   2.5 Sensitivity to measurement error and temporal frequency

350   Additional analysis examines the sensitivity of the parameter selection results to

351   measurement error (Section 3.2) and the temporal frequency of SWOT observations (Section 3.3).

352   Although the measurement error of SWOT instantaneous discharge is estimated to have a relative

353   RMSE of 35%, this error may be higher (Hagemann et al. 2017; Durand et al. 2016). Conversely,

354   the incorporation of ancillary data in the discharge algorithms or using a multialgorithm approach

355   may reduce error further (Durand et al. 2016). Thus, determining the sensitivity of these results to

356   measurement error is useful in evaluating the range of possible impacts for SWOT, in particular

357   with respect to model calibration. Thus, q' is calculated with ϵ values of 0.0, 0.2, 0.35, 0.5, 0.75,

358   and 1.0. The temporal frequency of SWOT observations is inherently tied to latitude due to SWOT

359   orbit characteristics (relatively narrow swath compared to satellite imagers and high inclination

360   angle). Thus, polar regions are observed more frequently than the midlatitudes, and the

361   midlatitudes are observed more frequently than the tropics. Biancamaria et al. (2016) show that

362   SWOT will observe the tropics (0-30°) 1-2 times per repeat cycle, the midlatitudes (30-60°) 2-4

363   times per repeat cycle, and polar regions (60-90°) 3-7 times per repeat cycle.

364   To assess the sensitivity of parameter selection results to temporal frequency and determine

365   whether this process is viable at lower latitudes, the same Alaskan sub-basins are considered but

366   the observation frequency of q' is reduced to mimic SWOT observation of mid- and low-latitudes

367   basins. For the midlatitudes, the observation frequency of q' was halved with respect to that of

368   Alaska. For the tropics, the observation frequency was reduced by a factor of four. The sensitivity

369   of model parameter selection to measurement error and the temporal frequency of observations is

370   calculated by:

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

372   where Y is the probability of selection and X is the measurement error ε or number of observations

373   per repeat cycle.

374

375   **3. Results and discussion**

376   3.1. Parameter selection

377   Figure 5a presents the cumulative probability that the true best parameter set is selected via

378   NSE_{SWOT} at or above (equal or higher ranking) each rank position for a random error ϵ=0.35 using

379   single-point parameter selection. For example, if the rank 10 likelihood for a given basin is 80%,

380   then the parameter set selected by NSE_{SWOT} is one of the ten highest-ranked sets with respect to

381   the truth for 8,000 of the 10,000 Monte Carlo simulations. For five of the ten sub-basins, the

382   selected parameter set is ranked in the top ten with a >80% probability, meeting our criteria for

383   success. These five sub-basins plus sub-basin I also rank in the top five with a >60% probability.

384   Notably, two sub-basins (B and D) display much poorer results than the other sub-basins. Sub-

385   basin B selects a parameter set in the top 50 with less than a 30% probability, indicating that the

386   best parameter set as determined by NSE_{TRUTH} is poorly ranked by this parameter selection

387   approach. Sub-basins B and D experience substantial diurnal oscillations in streamflow

388   presumably due to snowmelt during the warm season resulting in a flashy hydrograph. Table 4

389   provides the Richards-Baker Flashiness Index (R-B Index; Baker et al. 2004) for each sub-basin

390   as a quantitative measure of basin flashiness. Although daily discharge is typically used in the

391   calculation, Baker et al. (2004) suggests the use of hourly discharge when diurnal oscillations are

392   significant, thus hourly discharge is used in the calculation in Table 4. These results suggest that

393   SWOT is not well-suited for parameter selection or calibration in basins with rapidly changing

394   discharge, as the sporadic nature of the SWOT temporal sampling along with measurement

395   uncertainty does not allow a good sampling of the truth hydrograph. Consequently, even the best

396   parameter set has a poor NSE_{SWOT}, making it difficult to extract the correct parameter set using

397   this methodology.

398   Figure 6 compares the NSE_{TRUTH} for each sub-basin shown in Figure 2. The NSE_{TRUTH}

399   curves for both B and D are flatter than for the other sub-basins, indicating that the sample

400   parameter set spread is narrower. Thus, there is less variation among the simulations, resulting in

401   lower ranked parameter sets based on NSE_{TRUTH} to be more frequently ranked highly based on

402   NSE_{SWOT}. Additionally sub-basin B has the lowest NSE_{TRUTH} of any sub-basin for its highest

403   ranked set with a value less than 0.75. Figure 7 maps the true rank of each parameter set selected

404   by NSE_{SWOT} and displays a histogram of these ranks for all SWOT observable channel reaches in

405   the domain. Results are generally very good for the entire upper Tanana basin and most of the

406   Susitna basin, with the selection of a highly-ranked (top 10) parameter set for most of the channel

407 reaches. However, the worst performance occurs in the same outlying sub-basins B and D from

408 Figure 6, with sub-basin B clearly demonstrating the poorest results. Thus, in addition to poor

409 SWOT sampling, it is also apparent that the simulation does not capture the physical processes of

410 these flashy sub-basins as well as for the other sub-basins, indicating that the sample parameter set

411 does not contain the true parameter set and resulting in the inability to achieve a good parameter

412 selection using the approach in this paper. However, this shortcoming highlights several potential

413 issues with parameter selection. Adequate model spin-up and configuration are necessary to ensure

414 significant physical processes are being adequately captured by the model and that physical realism

415 is adequate. Second, parameter sets which appropriately cover the parameter space may also be

416 key to differentiating model simulations (Sharma et al. 2019, Hagedorn et al. 2012, Weigel et al.

417 2008, 2009). A larger parameter set or a parameter sampling strategy that undertakes large searches

418 across parameter space may benefit the parameter selection at several sites in this study.

419   Figure 8a, interpreted in the same manner as Figure 5a, presents the cumulative probability

420 of successful parameter selection for a random error $\epsilon = 0.35$ using multi-point parameter

421 selection. Results improve compared to single-point parameter selection. For nine out of ten (90%)

422 of the sub-basins, the selected parameter set has a true rank in the top ten with a >80% probability,

423 meeting our criteria for successful parameter selection. The only sub-basin that again does not

424 achieve successful calibration is sub-basin B for the same reasons as discussed above. Six basins

425 also identify parameter sets in the top five true parameter sets with >60% probability.

426

427 3.2. Sensitivity to measurement error

428   Reducing the discharge random error ε from 0.35 to 0.20 results in the success probabilities

429 shown in Figures 5b and 8b for single-point and multi-point parameter selection, respectively. As

430   expected, the probability that the selected parameter set is highly ranked increases as random error

431   decreases, with the exception of sub-basin B for reasons discussed in Section 3.1. For single-point

432   parameter selection (Fig. 5b), 50% of sub-basins are assigned a top-ten ranked parameter set with

433   a >95% cumulative probability, while 70% of sub-basins meet the definition of successful

434   parameter selection. 60% of sub-basins also display a >80% probability of selecting a parameter

435   set ranked in the top five. These statistics are improved further using multi-point parameter

436   selection (Fig. 8b) with 90% of sub-basins achieving successful parameter selection and 70% of

437   sub-basins with a >95% cumulative probability. Figure 9 summarizes the effect of decreasing ϵ,

438   showing that as random error decreases the likelihood of successful parameter selection generally

439   increases, regardless of observation frequency. However, decreasing ϵ below 0.35 does not yield

440   improved results for multi-point parameter selection except for the tropics, suggesting that

441   systematic bias is the predominant factor in preventing successful parameter selection when ϵ is

442   less than 0.35. Conversely, as ϵ increases above 0.35 the likelihood of successful parameter

443   selection quickly drops regardless of observation frequency with single (multi-) point parameter

444   selection being successful for 0% (40%) of sub-basins for $\epsilon = 0.50$.

445   The sensitivity of the probability of successful parameter selection is evaluated by

446   considering a 0.10 decrease in observation random error ε within the 0.20 – 0.50 range. The mean

447   sensitivity is calculated over all channel reaches. The mean sensitivity for each rank position is

448   shown in Figure 10 with single-point and multi-point selection sensitivity as blue dashed and solid

449   lines, respectively. For single-point parameter selection, the probability is most sensitive at ranks

450   9-13, with a maximum sensitivity over 6% per 0.10 decrease in ϵ. Sensitivity gradually declines

451   for ranks beyond 13, which is to be expected since the cumulative probability, as shown in Figure

452   5, begins to level off for lower rank positions for most sub-basins. For multi-point parameter

453 selection, random error sensitivity peaks at rank 7 with a maximum sensitivity over 7% per 0.10

454 decrease in ϵ, and rapidly decreasing beyond rank 7. The use of multi-point parameter selection

455 yields a slightly more robust result parameter selection, as the sensitivity is lower than single-point

456 parameter selection at the rank 10 threshold. This is further supported by Figure 9, in which an

457 increase in ϵ from $0.20 - 0.50$ results in a smaller reduction in success for multi-point compared

458 to single point parameter selection. For both single- and multi-point parameter selection, a sizeable

459 improvement in success can be expected if the SWOT observation error can be reduced through

460 the use of ancillary datasets.

461

462 3.3 Sensitivity to temporal frequency of SWOT observations

463   Single-point parameter selection is very sensitive to the temporal frequency of SWOT

464 observations. As shown in Figure 9, for $\epsilon = 0.35$ successful parameter selection is reduced from

465 60% (polar) to 10% (tropics) of sub-basins, whereas for $\epsilon = 0.20$ it is reduced from 70% of polar to

466 30% of tropical sub-basins. For multi-point parameter selection, probability of success is reduced

467 from 90% to 50% of sub-basins for $\epsilon = 0.35$ and from 90% to 70% of sub-basins for $\epsilon = 0.20$ for the

468 same reduction in observation frequency. Figure 11 compares the cumulative probability curves for

469 the midlatitudes and the tropics using multi-point parameter selection, which alongside Figure 8a,

470 shows that 60% of sub-basins were largely unaffected by the four-fold decrease in observation

471 frequency.

472   The sensitivity due to a single additional SWOT observation per repeat cycle for each rank

473 position is shown in Figure 10 with single-point and multi-point parameter selection identified by

474 the orange dashed and solid lines, respectively. In calculating sensitivity to observation frequency,

475 five observations per repeat cycle (21 days) is assigned to the polar region, three observations per

476   repeat cycle is given to the midlatitudes, and 1.5 observations per repeat cycle is used for the

477   tropics. For example, in calculating sensitivity between the polar region and tropics, $\partial X = 5 -$

478     $1.5 = 3.5$. The magnitudes and patterns are similar to that of observation error sensitivity. For

479   single-point parameter selection, maximum sensitivity is observed between ranks 7-10, peaking

480   near 5% per additional observation. For multi-point parameter selection, sensitivity peaks at 2%

481   but drops to approximately 1% beyond rank 10. Clearly, while the number of SWOT observations

482   per repeat cycle noticeably affects the likelihood of successful parameter selection using single-

483   point parameter selection, multi-point selection is much more robust. Similar success using multi-

484   point parameter selection is achieved for polar regions, midlatitudes, and tropics for the full range

485   of evaluated ϵ. Further, a large majority of sub-basins achieved successful parameter selection

486   regardless of observation frequency for $\epsilon \leq 0.35$, demonstrating that reduced observation error

487   can compensate for lower observation frequency.

488

489   **4. Conclusions**

490   Using Monte Carlo methods, we evaluate parameter selection for an uncalibrated 250-m

491   WRF-Hydro model. We examine single- and multi-point objective function parameter selection

492   using simulated SWOT observations in regions unserved or underserved by in situ gauges. The

493   model parameter space is sampled to create an assortment of parameter sets for which the 250-m

494   model is run. Proxy SWOT discharge estimates were derived from an OSSE following the

495   methodology of Elmer et al. (2020a, b). As the true values of streamflow are known, we can

496   evaluate the selection of model parameters based on the comparison of model discharge

497   simulations with proxy SWOT streamflow observations.

498   The results indicate that the use of multi-point parameter selection is advantageous over

499   single-point parameter selection. In effect, the spatially distributed nature of the SWOT

500   observations compensates for systematic and random nature of observation error. In fact,

501   successful parameter selection is largely independent of random error for $\epsilon \leq 0.35$. The high spatial

502   coverage of observations from SWOT also compensates for the lack of their temporal frequency

503   in mid-latitude and tropical basins, perhaps due to spatiotemporal correlations in streamflow (Paiva

504   et al. 2015, Yang et al. 2019, Fisher et al. 2020). Further, this study finds that SWOT can

505   successfully be used for hydrologic model parameter selection despite the non-uniform space-time

506   sampling, agreeing with Nickles et al. (2020). However, whereas Nickles et al. (2020) results are

507   constrained to mid-latitude river basins, this study finds that this conclusion applies to nearly the

508   full range of SWOT temporal sampling. Even with larger errors than in situ gauges, this study

509   shows that SWOT discharge estimates can provide adequate accuracy and temporal sampling to

510   enable parameter selection for SWOT observable river basins globally with the exception of basins

511   experiencing significant variability (flashy basins). In regions devoid of in situ observations or

512   with relatively scarce stream gage networks, this study demonstrates that SWOT will provide

513   valuable observations for calibrating hydrologic models.

514   This study does not account for reprocessing of discharge estimates occasionally

515   throughout the SWOT mission, which is a planned activity to improve accuracy. Actual SWOT

516   observations are expected by mid-2022, so these results are timely in preparing to apply SWOT

517   data immediately following launch. While SWOT has many societal and research applications that

518   rely on near-real-time SWOT measurements (e.g., data assimilation, inundation mapping), the use

519   of SWOT observations for model parameter selection or calibration is not constrained by product

520   latency or mission lifetime, but extend beyond the mission end.

528

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–
    Atmosphere Feedbacks: A Case Study with the WRF-Hydro Coupled Modeling System
    for West Africa. J. Hydrometeor., 17, 1489–1516, doi:10.1175/JHM-D-15-0089.1.

Aviso+ (2015), SWOT orbit, https://www.aviso.altimetry.fr/en/missions/future-
    missions/swot/orbit.html, accessed 8 September 2015.

545   Baker, David B., R. Peter Richards, Timothy T. Loftus, and Jack W. Kramer, 2004. A New

546   Flashiness Index: Characteristics and Applications to Midwestern Rivers and Streams.

547   Journal of the American Water Resources Association (JAWRA) 40(2):503-522.

548   Biancamaria., S., M. Durand, K. M. Andreadis, P. D. Bates., A. Boone, N. M. Mognard, and E.

549   A. Clark (2011), Assimilation of virtual wide swath altimetry to improve Arctic river

550   modeling, Remote Sens. Environ., 115(2), 373-381, doi:10.1016/j.rse.2010.09.008.

551   Biancamaria, S., F. Frappart, A.-S. Leleu, V. Marieu, D. Blumstein, J. D., Desjonqueres, F. Boy,

552   A. Sottolichio, and A. Valle-Levinson (2016), Satellite radar altimetry water elevations

553   performance over a 200 m wide river: Evaluation over the Garonne River, Advances in

554   Space Research, 59(2017), 128-146, doi:10.1016/j.asr.2016.10.008.

555   Biancamaria, S., F. Frappart, A.-S. Leleu, V. Marieu, D. Blumstein, J. D., Desjonqueres, F. Boy,

556   A. Sottolichio, and A. Valle-Levinson (2017), Satellite radar altimetry water elevations

557   performance over a 200 m wide river: Evaluation over the Garonne River, Advances in

558   Space Research, 59(2017), 128-146, doi:10.1016/j.asr.2016.10.008.

559   Boning, C. W., (1992), Policy statement on stage accuracy. Tech. rep., Office of Surface Water

560   Technical Memorandum No. 93.07, available at

561   https://water.usgs.gov/admin/memo/SW/sw93.07.html.

562   Cao, W., W. B. Bowden, T. Davie, and A. Fenemor (2006), Multi-variable and multi-site

563   calibration and validation of SWAT in a large mountainous catchment with high spatial

564   variability. Hydrol. Process., 20, 1057-1073, doi:10.1002/hyp.5933.

565   Cunge, J. A. (1969), On the subject of a flood propagation computation method (Muskingum

566   method), J. Hydraul. Res., 7, 205-230.

567   Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S.

568   Thober (2016), The impact of standard and hard-coded parameters on the hydrologic

569   fluxes in the Noah-MP land surface model, J. Geophys. Res. Atmos., 121, 10,676–

570   10,700, doi:10.1002/2016JD025097.

571   Durand, M. K., K. M. Andreadis, D. E. Alsdorf, D. P. Lettenmaier, D. Moller, and M. Wilson

572   (2008), Estimation of bathymetric depth and slope from data assimilation of swath

573   altimetry into a hydrodynamic model. Geophys. Res. Lett., 35 (L20401), 1-5,

574   doi:10.1029/2008GL034150.

575   Durand, M., E. Rodriguez, D. E. Alsdorf, and M. Trigg (2010), Estimating river depth from

576   remote sensing swath interferometry measurements of river height, slope and width,

577   IEEE. J. Sel. Top. Appl. Earth Observ. Remote Sens., 3(1), 20-31,

578   doi:10.1109/JSTARS.2009.2033453.

579   Durand, M., J. Neal, E. Rodriguez, K. M. Andreadis, L. C. Smith, and Y. Yoon (2014),

580   Estimating reach-averaged discharge for the River Severn from measurements of river

581   water surface elevation and slope, J. Hydrol., 5111, 92-104,

582   doi:10.1016/j.jhydrol.2013.12.050.

583   Durand, M., et al. (2016), An intercomparison of remote sensing river discharge estimation

584   algorithms from measurements of river height, width, and slope, Water Resour. Res.,52,

585   4527–4549, doi:10.1002/2015WR018434.

586   Emery, C. M., Biancamaria, S., Boone, A., Ricci, S., Rochoux, M. C., Pedinotti, V., & David, C.

587   H.(2020). Assimilation of wide-swath altimetry water elevation anomalies to correct

588   large-scale river routing model parameters. Hydrology and Earth System Sciences, 24(5),

589   2207- 2233. https://doi.org/10.5194/hess-24-2207-2020

590   Elmer, N. J. (2019), Using satellite observations of river height and vegetation to improve

591   National Water Model initialization and streamflow prediction, Ph.D. dissertation,

592   Department of Earth and Atmospheric Science, The University of Alabama in Huntsville,

593   Huntsville, Alabama, 113 pp, available online at

594   https://cdm16608.contentdm.oclc.org/digital/collection/p16608coll23/id/64232.

595   Elmer, N. J., C. R. Hain, F. Hossain, D. Desroches, C. Pottier (2020a), Generating proxy SWOT

596   water surface elevations using WRF-Hydro and the CNES SWOT Hydrology Simulator,

597   Water Resour. Res., Early Online Release, doi:10.1002/essoar.10502399.1.

598   Elmer, N. J., C. Hain, J. McCreight, D. Gochis (2020b), SWOT Applications for WRF-Hydro

599   Modeling in Alaska, IGARSS 2020 - 2020 IEEE International Geoscience and Remote

600   Sensing Symposium, pp. 5077-5080, doi: 10.1109/IGARSS39084.2020.9323136.

601   Fisher, C. K., M. Pan, and E. F. Wood (2020), Spatiotemporal assimilation-interpolation of

602   discharge records through inverse streamflow routing, Hydro. Earth. Syst. Sci., 24, 293-

603   305, doi:10.5194/hess-24-293-2020.

604   Fjørtoft R, Gaudin JM, Pourthie N, Lalaurie JC, Mallet A, Nouvel JF, Martinot-Lagarde J, Oriot

605   H, Borderies P, Ruiz C, Daniel S (2014), KaRIn on SWOT: characteristics of near-nadir

606   Ka-band interferometric SAR imagery. IEEE Trans Geosci Remote Sens 52(4):2172–

607   2185, doi:10.1109/TGRS.2013.2258402.

608   Gochis, D.J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J.

609   Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, W. Yu, (2018). The WRF-Hydro

610   modeling system technical description, (Version 5.0). NCAR Technical Note. 107 pages.

611   Available online at https://ral.ucar.edu/sites/default/files/public/WRF-

612   HydroV5TechnicalDescription.pdf. Source Code DOI:10.5065/D6J38RBJ.

613   Gochis, D.J. and F. Chen, 2003: Hydrological enhancements to the community Noah land

614   surface model. NCAR Technical Note, NCAR/TN-454+STR, 68 pp.

615   Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., and Palmer, T. (2012). Comparing

616   TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts.

617   Quarterly Journal of the Royal Meteorological Society, 138(668), 1814–1827.

618   Hagemann, M.W.; Gleason, C.J.; Durand, M.T. BAM: Bayesian AMHG-Manning Inference of

619   Discharge Using Remotely Sensed Stream Width, Slope, and Height. Water Resour. Res.

620   2017, 53, 9692–9707, doi:10.1002/2017WR021626.

621   Hirsch, R. and J. E. Costa (2004), U.S. streamflow measurement and data dissemination

622   improve. Eos Trans. AGU, 85 (20), 197-203, doi:10.1029/2004EO200002.

623   Kouraev, A. V., E. A. Zakharova, O. Samain, N. M. Mognard, and A. Cazenave (2004), Ob river

624   discharge from TOPEX/Poseidon satellite altimetry (1992–2002), Remote Sens.

625   Environ., 93, 238-245.

626   McCuen, R. H., Z. Knight, and A. G. Cutter (2006), Evaluation of the Nash–Sutcliffe efficiency

627   index, J. Hydrol. Eng., 11(6), 597–602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597).

628   Munier, S., A. Polebistki, C. Brown, G. Belaud, and D. P. Lettenmaier (2015), SWOT data

629   assimilation for operational reservoir management on the upper Niger River Basin, Water

630   Resour. Res., 51, 554-575, doi:10.1002/2014WR016157.

631   Nickles, C., Beighley, E., & Feng, D. (2020). The Applicability of SWOT's Non-Uniform Space-

632   Time Sampling in Hydrologic Model Calibration. Remote Sensing, 12(19), 3241.

633   https://doi.org/10.3390/rs12193241.

634 Niraula, R., L. M. Norman, T. Meixner, and J. B. Callegary (2012), Multi-gauge calibration for

635   modeling the semi-arid Santa Cruz watershed in Arizona-Mexico border area using

636   SWAT. Air, Soil and Water Research, 5, 41-57, doi:10.4137/ASWR.S9410.

637 Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning,

638   D. Niyogi, E. Rosero, M. Tewari, and Y. Xia (2011), The community Noah land surface

639   model with multiparameterization options (Noah-MP): 1. Model description and

640   evaluation with local-scale measurements, J. Geophys. Res., 116(D12109),

641   doi:10.1029/2010JD015139.

642 O'Loughlin, F. E., J. Neal, D. Yamazaki, and P. D. Bates (2016), ICESat-derived inland water

643   surface spot heights, Water Resour. Res., 52, 3276-3284, doi:10.1002/2015WR018237.

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

645 Paiva, R. C. D., M. T. Durand, and F. Hossain (2015), Spatiotemporal interpolation of discharge

646   across a river network by using synthetic SWOT satellite data, Water Resour. Res., 51,

647   430–449, doi:10.1002/2014WR015618.

648 Papa, F., F. Durand, W. B. Rossow, A. Rahman, and S. K. Bala (2010), Satellite altimeter-

649   derived monthly discharge of the Ganga-Brahmaputra River and its seasonal to

650   interannual variations from 1993 to 2008, J. Geophys. Res., 115(C12013),

651   doi:10.1029/2009JC006075.

652 Pavelsky, T. M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. D. Paiva, G. H. Allen,

653   and Z. F. Miller (2014), Assessing the potential global extent of SWOT river discharge

654   observations, J. Hydrol., 519, 1516-1525, doi:10.1016/j.jhydrol.2014.08.044.

655 Pedinotti, V., A. Boone, S. Ricci, S. Biancamaria, and N. Mognard (2014), Assimilation of

656   satellite data to optimize large-scale hydrological model parameters: a case study for the

657   SWOT mission, Hydrol. Earth Syst. Sci., 18, 4485-4507, doi:10.5194/hess-18-4485-

658   2014.

659 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011).

660   Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research,

661   12(85), 2825–2830

662 Revel, M., Ikeshima, D., Yamazaki, D., & Kanae, S. (2021). A framework for estimating global-

663   scale river discharge by assimilating satellite altimetry. Water Resources Research, 57,

664   e2020WR027876. https://doi.org/10.1029/2020WR027876

665 Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B.

666   Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D.

667   Toll (2004), The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc.,

668   85(3), 381-394.

669 Rodriguez, E. (2016), Surface Water and Ocean Topography Mission (SWOT) Project, Science

670   Requirements Document, Rev A, 28 pp., https://swot.jpl.nasa.gov/files/swot/D-

671   61923_SRD_Rev%20A_20160318%20with%20signatures1.pdf.

672 Sampson, K. and D. Gochis (2015), WRF-Hydro GIS pre-processing tools: Version 2.2

673   documentation, Tech. rep., Boulder, Colorado, 39 pages.

674 Senatore, A., G. Mendicino, D. J. Gochis, W. Yu, D. N. Yates, and H. Kunstmann (2015), Fully

675   coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of

676   enhanced hydrological parameterization for short and long time scales, J. Adv. Model.

677   Earth Syst., 7, 1693-1715, doi:10.1002/2015MS000510.

678   Sharma, S., Siddique, R., Reed, S., Ahnert, P., and Mejia, A. (2019). Hydrological model

679   diversity enhances streamflow forecast skill at short- to medium-range timescales. Water

680   Resources Research, 55, 1510–1530, doi:10.1029/2018WR023197.

681   Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang,

682   W. Wang, and J. G. Powers (2008), A description of the Advanced Research WRF

683   version 3, NCAR Technical Note, 475, 113 pp.,

684   http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.

685   U. S. Geological Survey (2017), National Elevation Dataset (NED),

686   https://nationalmap.gov/elevation.html.

687   Weigel, A. P., Liniger, M., and Appenzeller, C. (2008). Can multi-model combination really

688   enhance the prediction skill of probabilistic ensemble forecasts? Quarterly Journal of the

689   Royal Meteorological Society, 134(630), 241–260.

690   Weigel, A. P., Liniger, M. A., and Appenzeller, C. (2009). Seasonal ensemble forecasts: Are

691   recalibrated single models better than multimodels? Monthly Weather Review, 137(4),

692   1460–1479.

693   Wongchuig-Correa, S., de Paiva, R. C. D., Biancamaria, S., & Collischonn, W. (2020).

694   Assimilation of future SWOT-based river elevations, surface extent observations and

695   discharge estimations into uncertain global hydrological models. Journal of Hydrology,

696   590, 125473.

697   Yang, Y., Lin, P., Fisher, C. K., Turmon, M., Hobbs, J., Emery, C. M., et al. (2019). Enhancing

698   SWOT discharge assimilation through spatiotemporal correlations. Remote Sensing of

699   Environment, 234, 111450. https://doi.org/10.1016/j.rse.2019.111450

700   Yang, Y., C. Fisher, and J. T. Reager (2019), Enhancing SWOT discharge assimilation through

701   spatiotemporal correlations, Remote Sensing of Environment, 234, 111450,

702   doi:10.1016/j.rse.2019.111450.

703   Yoon, Y., M. Durand, C. J. Merry, E. A. Clark, K. M. Andreadis, and D. E. Alsdorf (2012),

704   Estimating river bathymetry from data assimilation of synthetic SWOT measurements, J.

705   Hydrol., 464-465, 363-375, doi:10.1016/j.jhydro.2012.07.028.

706   Yucel, I., A. Onen, K. K. Yilmaz, and D. J. Gochis (2015), Calibration and evaluation of a flood

707   forecasting system: Utility of numerical weather prediction model, data assimilation, and

708   satellite-based rainfall, J. Hydrol., 523, 49-66, doi:10.1016/j.jhydrol.2015.01.042.

709

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

743   and b) ε=0.20.

744   Figure 9. Probability of successful single and multi-point parameter selection (%) for each

745   latitude zone as a function of relative measurement error (ε).

746   Figure 10. Mean sensitivity (% change per unit) of the probability that the selected parameter set

747   is ranked at or above each rank position with respect to changes in proxy SWOT discharge

748   error ε (blue) and SWOT observation frequency (orange) for single-point (dashed) and multi-

749   point (solid) parameter selection. Since error sensitivity is likely non-linear, note that this

750   evaluation should not be extrapolated beyond 0.20 – 0.50. Units are shown in the legend in

751   parentheses, and rank is determined by NSE_{TRUTH} and $\overline{NSE_{TRUTH}}$. The vertical dashed line

752   marks the 10th-ranked member.

753   Figure 11. As in Figure 8a, but for proxy SWOT mimicking observation frequency for a)

754   midlatitudes and b) tropics as opposed to Alaska. Results are for multipoint parameter

755   selection.

756

Tables

758   Table 1. Noah-MP and WRF-Hydro parameterization options used. More information about these

759   options is 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

760

761

762 Table 2. Parameter table listing the model parameters comprising parameter space. The listed value

763 was applied either as a multiplicative factor (Mult) or as an absolute (substituted) value (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

Component	Parameter (Variable name)	Variable name	Abs Order	Minimum Value	Maximum Value	Default value
Channel	Manning's roughness coefficient	MannN	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

764

765   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

766

767   Table 4. Richards-Baker Flashiness Index (R-B Index) for each sub-basin calculated using

768   hourly discharge from the analysis period.

Sub-basin	**R-B Index**
A	0.66
B	2.02
C	0.76
D	1.30
E	0.66
F	0.26
G	1.06
H	0.90
I	0.54
J	0.20

769
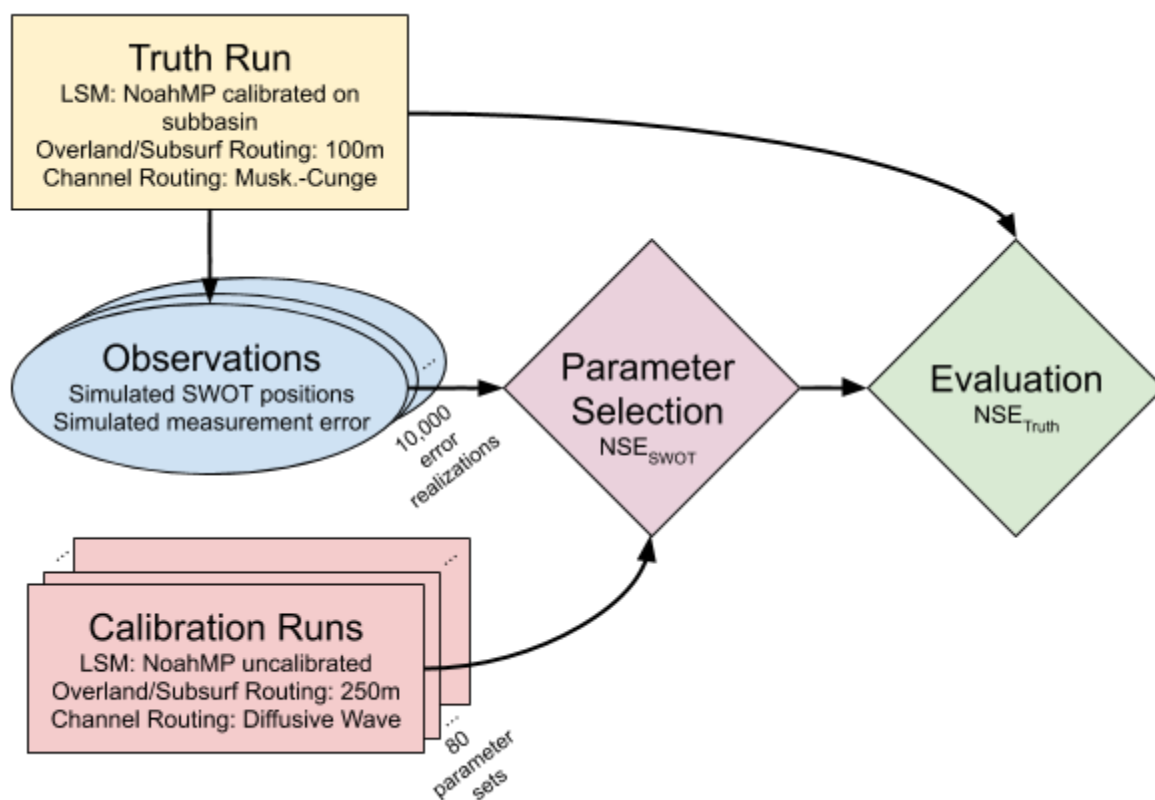
Figures

771

772   Figure 1. Design of "fraternal twin" experiment for evaluating the utility of (simulated) SWOT

773   for hydrologic parameter selection with the WRF-Hydro model: Do the space-time sampling and

774   observation error characteristics of SWOT permit accurate calibration?

775
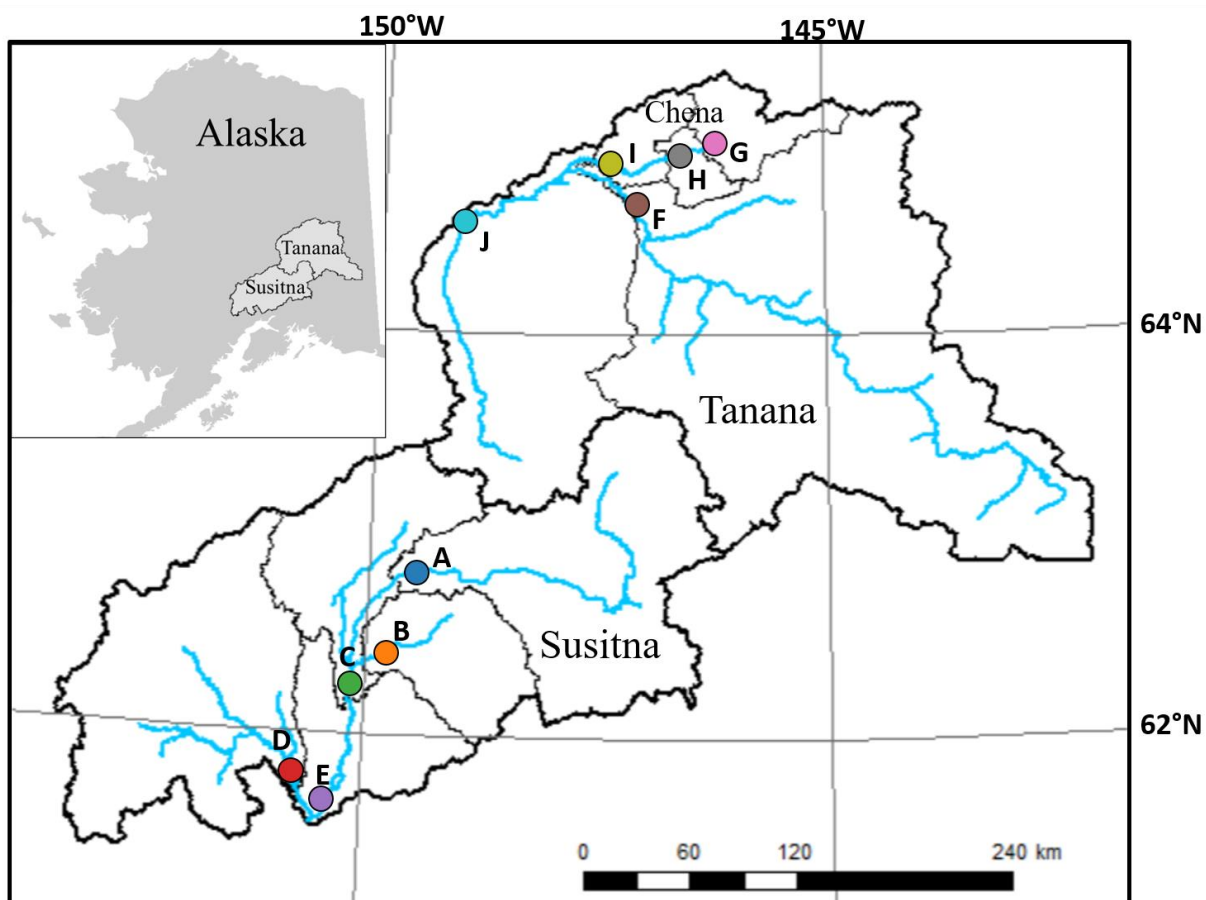
776

777 Figure 2. Study area within the Susitna River and upper Tanana River basins. The letters indicate

778 sub-basins A-J with colors corresponding with Figure 5. Streams of order five and greater are

779 designated in blue.

780

781
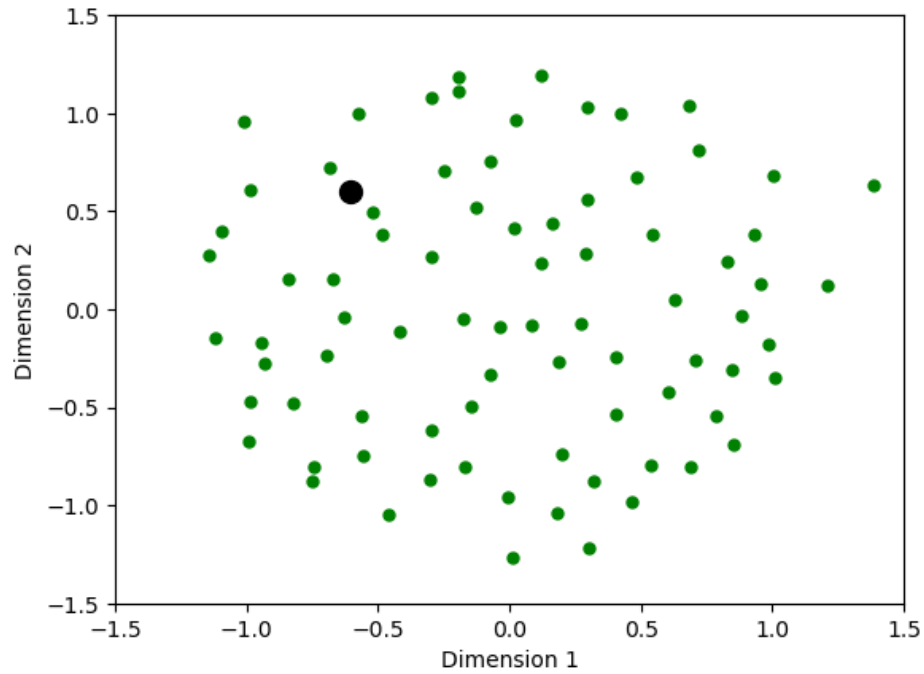
Figure 3. Distribution of the sampled parameter sets (green) used in this study visualized using
two-dimensional multidimensional scaling. These dimensions are generically labeled as they do
not readily correspond to the original parameter space. For reference, the calibrated parameter set
for the truth model is shown in black.

786

787
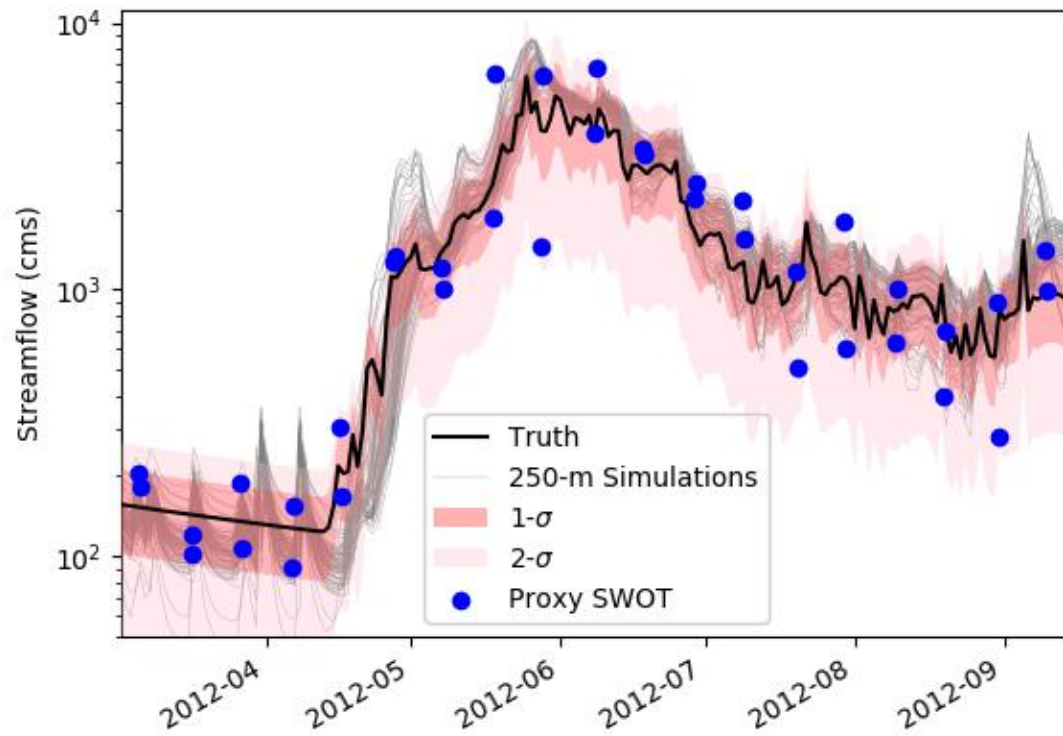
Figure 4. 250-m simulations, truth simulation, 1-σ and 2-σ discharge error ranges, and a sample

proxy SWOT discharge observation set (i.e., hydrograph) for a single point corresponding to the

sub-basin E outlet.

791
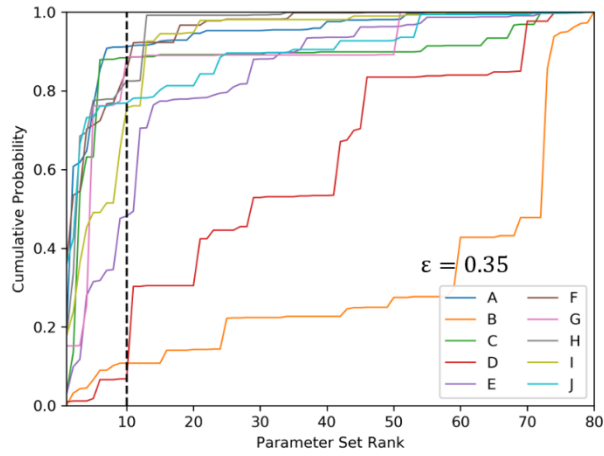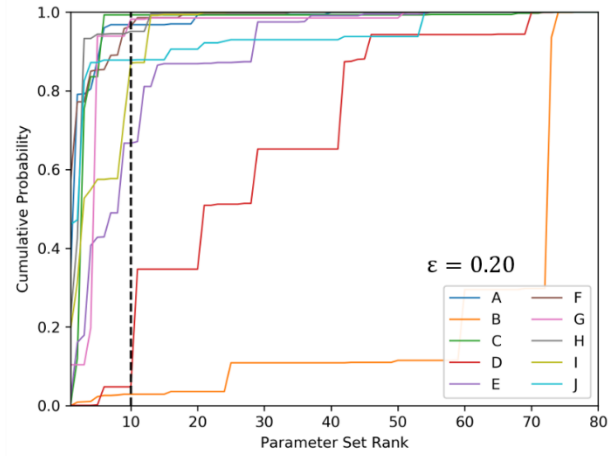
(a)   (b)

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.

798
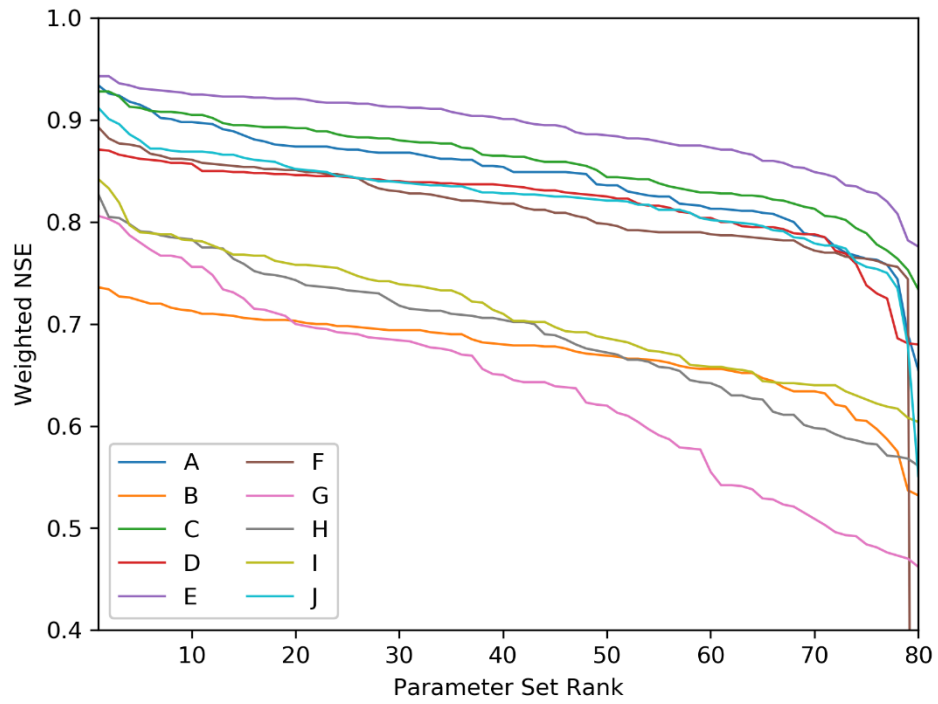
799　Figure 6. NSE$_{TRUTH}$ for each point A-J in Figure 2. Parameter set rank is determined by
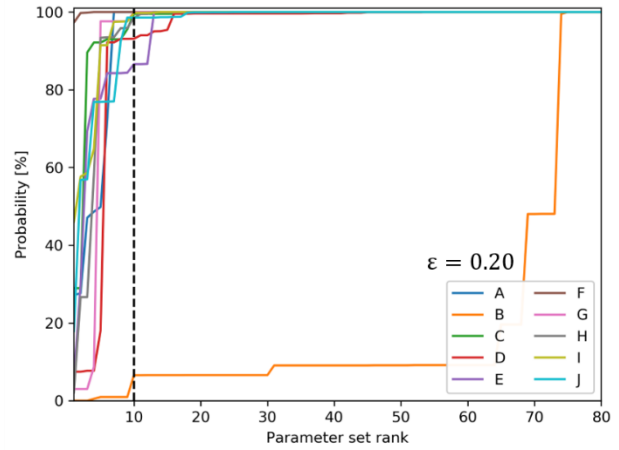
800　NSE$_{TRUTH}$ at the basin outlet (single-point).

801

802

803   (a)   (b)
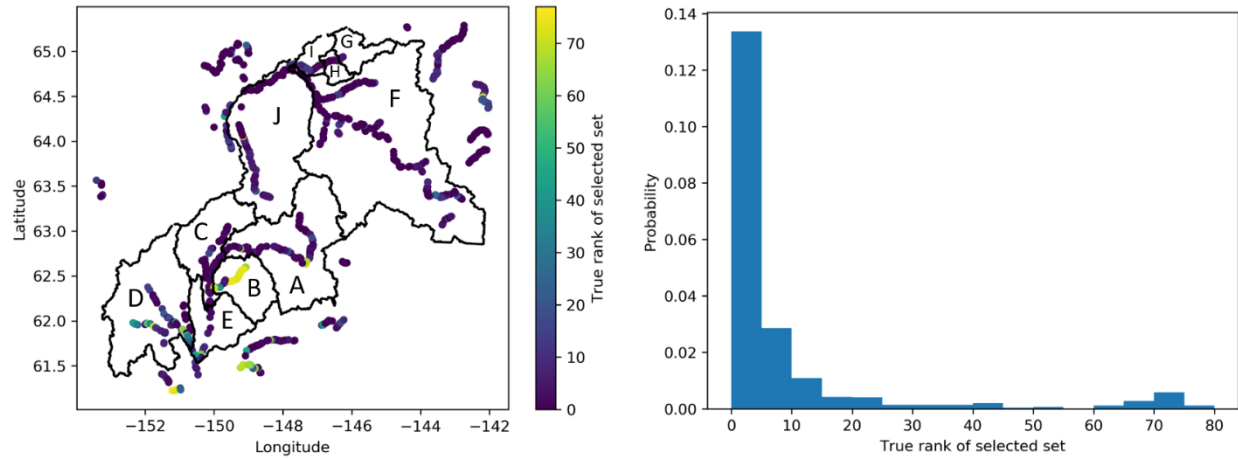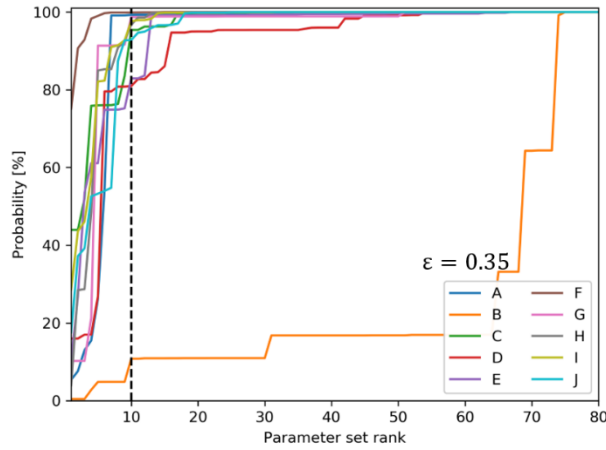
804   Figure 7. a) Map and b) histogram depicting the true rank of selected parameter set for $\epsilon = 0.35$

805   in Alaska. All SWOT observable channel reaches are shown. Basin borders match those in

806   Figure 2. Zero indicates the best true rank.

807

808

809   (a)   (b)

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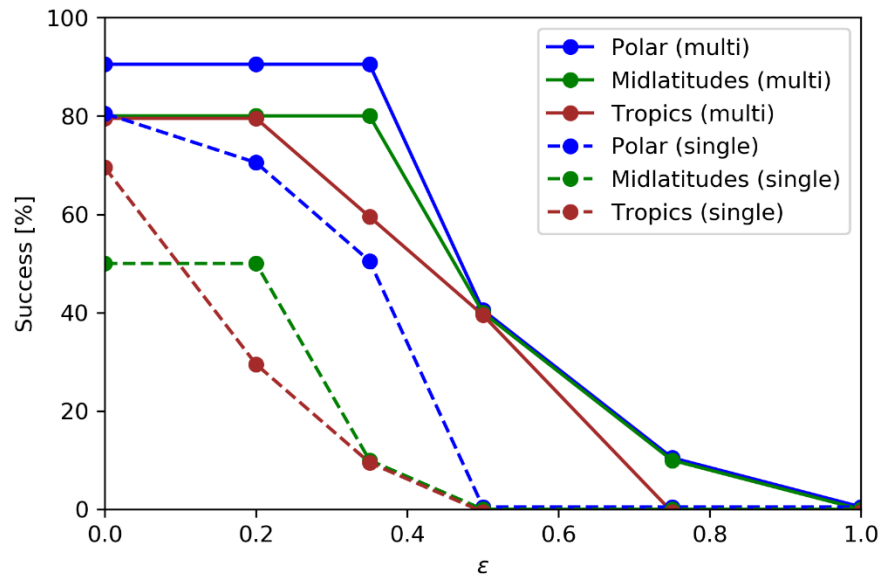
811 and b) ε=0.20.

812

813 Figure 9. Probability of successful single and multi-point parameter selection (%) for each

814 latitude zone as a function of relative measurement random error (ε).
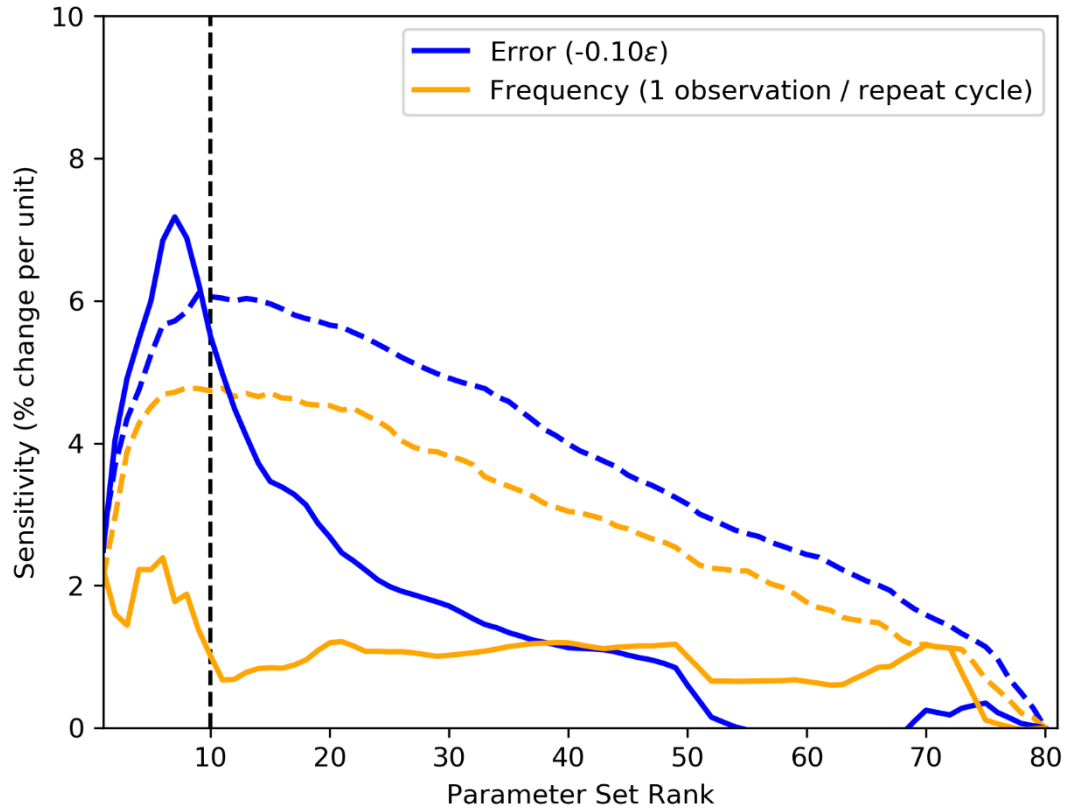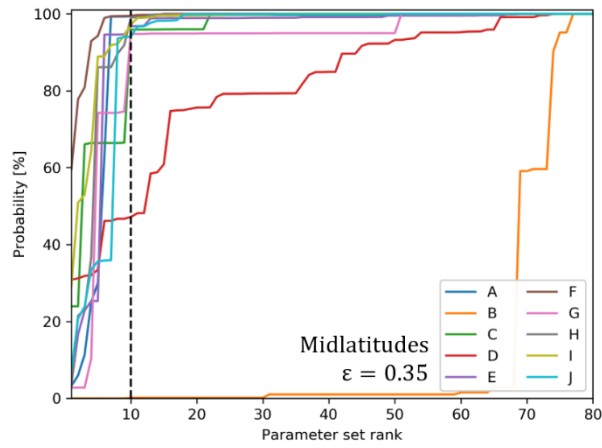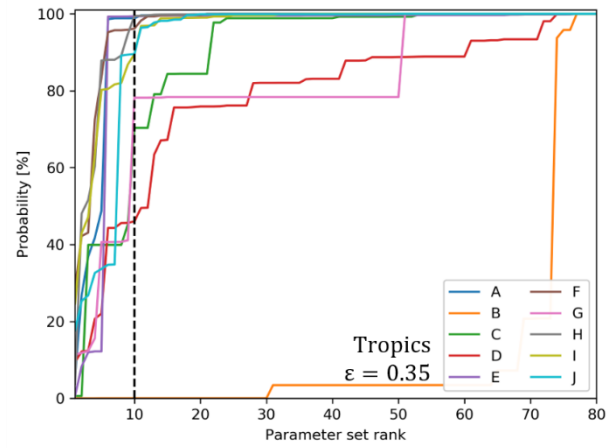
815

46

816

Figure 10. 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 (dashed) and multi-point

(solid) parameter selection. Since error sensitivity is likely non-linear, note that this evaluation

should not be extrapolated beyond 0.20 – 0.50. 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.

824

825

826   (a)   (b)

827   Figure 11. As in Figure 8a, but for proxy SWOT mimicking observation frequency for a)

828   midlatitudes and b) tropics as opposed to Alaska. Results are for multipoint parameter selection.