

1 Moving land models towards actionable science: A novel application of the Community
2 Terrestrial Systems Model across Alaska and the Yukon River Basin
3
4 Yifan Cheng^{1,*}, Andrew Newman¹, Keith N. Musselman², Sean Swenson¹, David Lawrence¹,
5 Joseph Hamman^{1,3}, Katherine Dagon¹, Daniel Kennedy¹
6

7 1. National Center for Atmospheric Research, Boulder, CO 80307, USA
8 2. Institute of Arctic and Alpine Research, University of Colorado, Boulder CO, 80309, USA
9 3. CarbonPlan, San Francisco CA, 94115, USA

10 * Corresponding author: Yifan Cheng (yifanc@ucar.edu)

11 Key points (140 characters):

- 12 • This study provides a high-resolution, high-fidelity Arctic hydrologic simulation and
13 evaluation of 15 major Alaskan river basins
- 14 • We develop an optimization workflow and co-design application relevant evaluations to
15 move CTSM toward an actionable Earth Science paradigm
- 16 • The final model leads to better snow, overall and high flow, yet worse low flow
17 simulation and mixed improvement in flow climate sensitivity

18

19 **Abstract (250 words)**

20 The Arctic hydrological system is an interconnected system that is experiencing rapid
21 change. It is comprised of permafrost, snow, glacier, frozen soils, and inland river systems.
22 Permafrost degradation, trends towards earlier snow melt, a lengthening snow-free season, soil
23 ice melt, and warming frozen soils all challenge hydrologic simulation under climate change in
24 the Arctic. In this study, to lower the barrier of using complex land models in regional
25 applications, we develop a generalizable optimization methodology and workflow for the
26 community. We applied the Community Terrestrial Systems Model (CTSM) across Alaska and
27 the Yukon River Basin at 4-km spatial resolution. We highlighted several potentially useful high-
28 resolution CTSM configuration changes. Additionally, we performed a multi-objective
29 optimization using snow and river flow metrics within an adaptive surrogate-based model
30 optimization scheme. Four representative river basins across our study domain were selected for
31 optimization based on observed streamflow and snow water equivalent observations at ten
32 SNOTEL sites. Fourteen sensitive parameters were identified for optimization with half of them
33 not directly related to hydrology or snow processes. Across fifteen out-of-sample river basins,
34 thirteen had improved flow simulations after optimization and the mean Kling-Gupta Efficiency
35 of daily flow increased from 0.43 to 0.63 in a 30-year evaluation. In addition, we adapted the
36 Shapley Decomposition to disentangle each parameter's contribution to streamflow performance
37 changes, with the seven non-hydrological parameters providing a non-negligible contribution to
38 performance gains. The snow simulation had limited improvement, likely because snow
39 simulation is influenced more by meteorological forcing than model parameter choices.

40

41 **1 Introduction**

42 The Arctic is experiencing rapid change across all Earth system components including
43 Arctic hydrology (Fox-Kemper et al., 2021; Yang & Kane, 2020). Specifically, Arctic Alaska is
44 experiencing a multitude of changes. Abrupt increases in permafrost degradation and increasing
45 active layer depth greatly influence the subsurface runoff process (Jorgenson et al., 2006;
46 Lawrence et al., 2012; Lawrence & Slater, 2005; Osterkamp & Romanovsky, 1999). Larger
47 surface energy fluxes due to increased atmospheric temperatures and moisture lead to earlier
48 snow melt, lengthening of the snow-free season, reduced river ice, frozen soil warming,
49 permafrost degradation, and related shifts in the fluvial freshwater seasonality(Cox et al., 2017;
50 Hamman et al., 2017; Pavelsky & Zarnetske, 2017; Stone et al., 2002). These anthropogenic
51 climate-driven transformations in hydrology and river ice in the Alaskan and Yukon rivers will
52 likely have substantial impacts on Indigenous community members who rely heavily on inland
53 river systems for subsistence fishing and river-ice road transportation (Knoll et al., 2019;
54 Pavelsky & Zarnetske, 2017; Sharma et al., 2019).

55 We urgently need “actionable science” to support policy and decision-making towards
56 adapting or mitigating the potential climate impacts on arctic hydrology. In this study, we
57 adopted the definition of “actionable science” from the Advisory Committee on Climate Change
58 and Natural Resource Science, appointed to advise the Secretary of the Interior. *Actionable*
59 *science provides data, analyses, projections, or tools that can support decisions regarding the*
60 *management of the risks and impacts of climate change.* We specifically focused on enhancing
61 the actionability of process-based modeling in this study.

62 Hydrologic modeling of Arctic rivers is challenging due to the aforementioned complex
63 and interacting terrestrial processes. However, recent developments in advanced land models

64 (LMs) are now enabling us to simulate complex land surface processes and their subsequent
65 impacts on hydrology (Clark et al., 2015; Hamman et al., 2016). Additionally, advances in
66 computationally frugal optimization methods and improvements in LM agility (i.e., the
67 capability to adjust model equations and parameters to faithfully represent observed processes;
68 Mendoza et al., 2015), allow for parameter sensitivity and application-oriented optimization
69 studies of these advanced LMs.

70 In this study, we focused on a state-of-the-science land model, the Community Terrestrial
71 Systems Model (CTSM). CTSM includes complex vegetation and canopy representation, a
72 multi-layer snow model, as well as hydrology and frozen soil physics necessary for the
73 representation of streamflow and permafrost in the Arctic (Oleson et al., 2010). More recent
74 updates to parameterizations and model structures for hydrology and snow (Lawrence et al.,
75 2019) further improve the physical representation related to freshwater cycles in cold regions,
76 including spatially explicit soil depth (Pelletier et al., 2016), representation of soil organic matter
77 (Lawrence et al., 2008), revised canopy interception and canopy snow processes, and updated
78 fresh snow density (van Kampenhout et al., 2017). Finally, a representative hillslope hydrology
79 capability has recently been implemented into CTSM, which enables parameterization of the
80 impacts of slope and aspect on lateral water transfer and incident radiation and subsequent
81 impacts on hydrology (Fan et al., 2019; Swenson et al., 2019).

82 Earth System models are being applied at an increasingly higher resolution to improve
83 the physical representation like convection or orographic impacts (Bierkens et al., 2015; Singh et
84 al., 2015). Higher-resolution models can more faithfully represent varied and complex
85 topography, and thus often more realistically simulate seasonal snow, orographic precipitation
86 patterns, and potentially heterogeneous permafrost (Jafarov et al., 2012; Newman et al., 2021;

87 Rasmussen et al., 2011). A more realistic physical representation of the landscape and land-
88 atmosphere interactions increases the credibility of a model in regional applications, which can
89 help to build stakeholder trust in model results and can help to facilitate a move toward a more
90 actionable Earth Science paradigm (Giorgi, 2019).

91 This study is supported by the Arctic Rivers Project, which is guided by an 11-member
92 Indigenous Advisory Council. The Council helps project investigators make decisions about
93 research design, analyses, and deliverables to ensure that Indigenous knowledge and perspectives
94 are included, valued, and protected, and that the project benefits the Indigenous peoples the
95 project is intended to serve. A climate information survey was co-developed by the research
96 team and Council (Herman-Mercer, 2021) and distributed to decision-making bodies in the
97 communities of our study domain. Survey responses were received from 23 (10% response rate)
98 Tribal Councils, Traditional Councils, First Nation Governments, City Councils, and Regional
99 Indigenous Organizations. Among the survey respondents, there was agreement that the most
100 useful information for Indigenous decision-makers would be sub-watershed scale (or high-
101 resolution) streamflow and other land-surface and sensible (i.e., relatable) weather variables such
102 as 2-m air temperature and precipitation. While the survey respondents cannot be considered a
103 representative sample of decision-makers, the consensus among responses plus Council
104 concurrence gives us confidence in applying this information to guide modeling efforts.
105 Configuring a high-resolution model is multi-faceted, which not only means a finer grid but also
106 requires corresponding meteorological forcing data and land surface datasets that are often more
107 difficult to work with if they even exist. In addition, high-resolution LMs require substantially
108 more computational resources, which decreases the potential to run ensembles of simulations.

109 Therefore, high-resolution modeling limits the ability to account for uncertainties in the modeled
110 system.

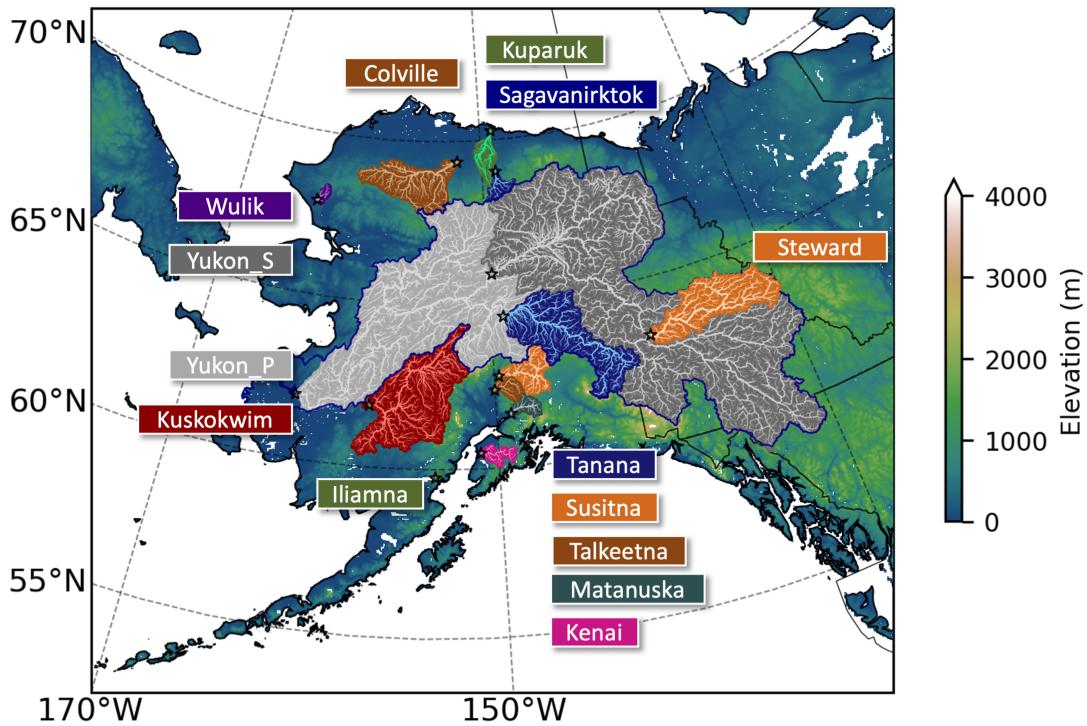
111 Even with improved process representation and hydrologically focused model
112 configurations (Choi & Liang, 2010; Jiao et al., 2017; Singh et al., 2015), optimization of
113 parameters within complex LMs is often necessary because of uncertainty in model parameters,
114 model structural errors, and missing process representations (Lehner et al., 2019; Mendoza,
115 Clark, Barlage, et al., 2015; Sankarasubramanian et al., 2001). In addition, as common practice
116 for modelers and stakeholders, a model needs to be calibrated before it can be trusted to provide
117 useful information and support decision-making. Optimization of complex LMs like CTSM is a
118 substantial challenge given the high computational costs, and this challenge limits the usage of
119 CTSM and similar models in large-scale hydrological or other stakeholder-specific applications.
120 Although several sensitivity analyses have been conducted to examine the hydrological
121 responses to CTSM model parameters (Jefferson et al., 2015; Ren et al., 2016; Srivastava et al.,
122 2014), their limited spatial coverage or number of parameters cast few insights on sensitive
123 parameters to Arctic terrestrial hydrology.

124 The sophisticated land process representations in CTSM with high spatial resolution
125 make it a potentially robust tool in projecting climate impacts on hydrology, yet its complexity
126 undermines its useability for real-world applications. Therefore, in this study, we provide: 1) a
127 methodology for efficient optimization of CTSM to lower the barrier of using CTSM in real-
128 world applications and enhance the actionability of CTSM; 2) a high-resolution Arctic CTSM
129 configuration focused on improved hydrologic simulation fidelity; 3) an evaluation of the
130 performance of the regional CTSM configuration and its actionability; 4) tools available to the
131 scientific community to apply our methodology to other regions and applications. We evaluated

132 the model performance, or actionability, by using hydrology-related metrics that are related to
133 events of concern from our IAC and the climate information survey available in Section 4.5.
134 Additionally, this study lays the foundation for knowledge co-production research with
135 Indigenous communities on a range of topics, including improving our understanding of climate-
136 induced impacts on the rivers and fishes, and communities necessary to inform adaptation efforts
137 in our study domain. We aim to lower the barriers of using CTSM and other complex, process-
138 rich land models in regional applications by developing an optimization workflow, providing an
139 example evaluation for a specific application, and therefore moving them toward a more
140 actionable Earth Science paradigm (Findlater et al., 2021) with this exemplar hydrologic
141 application.

142 **2 Study domain**

143 Our study domain includes the Yukon River Basin (dark blue boundaries in Figure 1) and
144 Alaska. Over 200 Indigenous tribes and First Nations reside in this area and their culture and
145 livelihood are deeply rooted in inland freshwater systems. Figure 1 highlights key river basins
146 and gaging stations along the Yukon River that have minimal diversions and enough
147 observations to be used for model calibration or validation. The Tanana River and Steward River
148 are two major tributaries to the Yukon River. Along the North Slope, four river basins with
149 quality flow observations are highlighted in Figure 1; the Colville River, Kuparuk River,
150 Sagavanirktoq River, and Wulik River. Six river basins south of the Yukon River Basin also
151 have enough quality flow observations for our purposes; the Kuskokwim River, Iliamna River,
152 Susitna River, Talkeetna River, Matanuska River, and Kenai River. We also used observations
153 from two gauges along the main stem of the Yukon River, i.e., one at the Pilot station and one
154 near Stevens Village denoted as Yukon_P and Yukon_S in Figure 1.



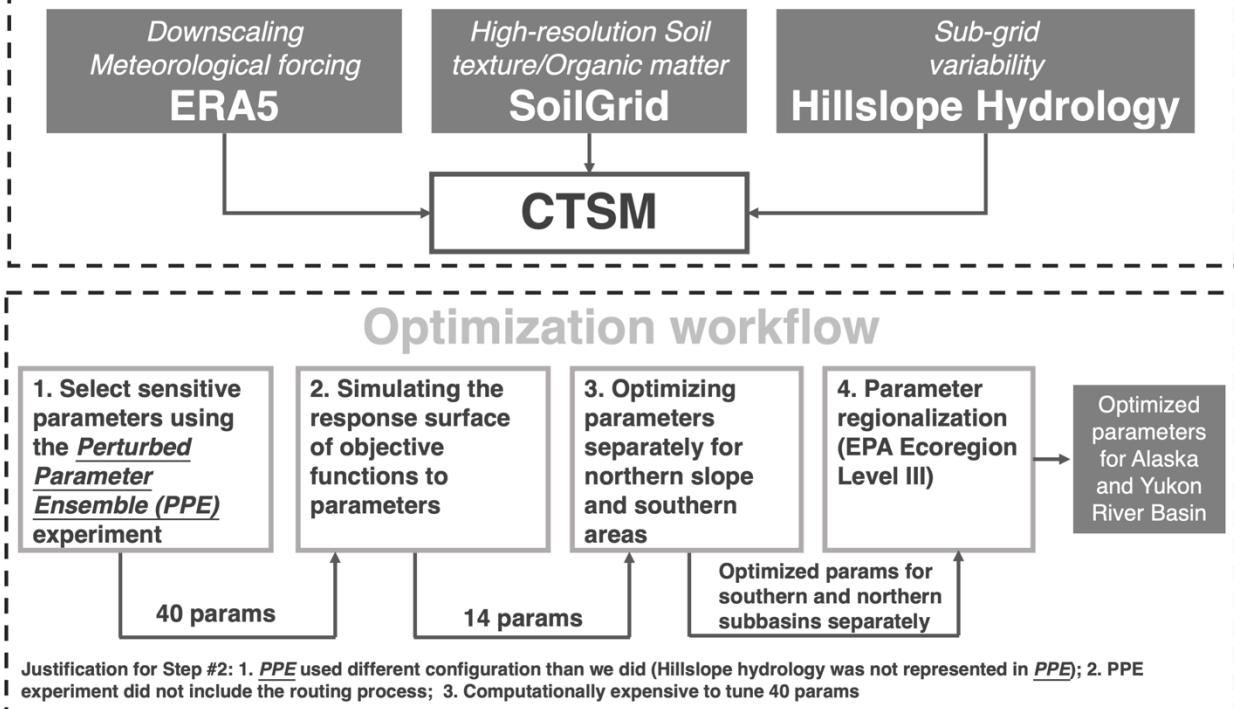
155

156 *Figure 1: Study domain. The dark blue line denotes the boundary of the Yukon River Basin and*
 157 *black stars denote the outlets of the highlighted river basins.*

158 **3 Baseline CTSM configuration**

159 To configure a high-resolution CTSM application, we downscaled the available coarse
 160 meteorological forcing data (Section 3.1) and used finer-than-default soil texture data (Section
 161 3.2). In addition, we used the hillslope hydrology scheme in CTSM to account for the remaining
 162 sub-grid topographic variability (Section 3.3) and used the satellite phenology CTSM
 163 configuration with default model parameter values. We used the vector-based mizuRoute to
 164 route runoff (Mizukami et al., 2016, 2021) and we extracted the river network from a high-
 165 resolution global hydrography map, i.e., MERIT Hydro (Yamazaki et al., 2019). This constitutes
 166 our baseline CTSM model (Figure 2).

Baseline Scenario – High-resolution application



167

168 Figure 2: CTSM baseline scenario and workflow for optimization

169 3.1 Downscaling meteorological forcing data - ERA5

170 We used the fifth generation of ECMWF atmospheric reanalysis of the global climate
171 (ERA5) as the meteorological forcing data (European Centre for Medium-Range Weather
172 Forecasts, 2019). The forcing is at an hourly timestep and on a 0.25-degree (~14 km) latitude-
173 longitude grid. While a quarter degree resolution is a substantial improvement over previous
174 global reanalysis, it is still too coarse to fully resolve complex topography and small-scale
175 variations in near-surface meteorology, e.g., orographic precipitation, altitudinal temperature
176 gradients (Monaghan et al., 2018; Rasmussen et al., 2011). Therefore, we performed a simple
177 downscaling to add high-resolution information to our hourly forcing data. We used the monthly
178 climatology from a 4 km simulation of coupled WRF and Noah-MP (Monaghan et al., 2018) to
179 downscale the ERA5 data. This simulation was shown to represent historical observations well

180 (Monaghan et al. 2018) and is available from September 2002 to August 2016 (14 years), which
 181 we use to calculate ERA5 correction factors. For precipitation, we used a monthly multiplicative
 182 correction. Precipitation varies by orders of magnitude across regions and is bounded by zero so
 183 a multiplicative correction method is more appropriate than a delta method (Maraun &
 184 Widmann, 2018).

$$\Pi_{M,H,g}^P = \frac{\overline{P}_{M,H,g}^{WRF}}{\overline{P}_{M,H,g}^{ERA5}} \quad (1)$$

$$P_{m,h,g}^{ds} = P_{m,h,g}^{ERA5} \times \Pi_{M,H,g}^P \quad (2)$$

185 where P denotes precipitation. Π denotes the multiplicative correction factor, which has three
 186 dimensions, i.e., month (M), hour of the day (H), and grid (g). For each combination of month
 187 and hour, we averaged the values across 14 years to calculate the correction factor. Lower-case
 188 m and h denote the month and day for the to-be-corrected precipitation time series. We used a
 189 simple delta method to downscale the remaining meteorological forcing variables.

$$\Sigma_{M,H,g}^v = \overline{v}_{M,H,g}^{WRF} - \overline{v}_{M,H,g}^{ERA5} \quad (3)$$

$$v_{m,h,g}^{ds} = v_{m,h,g}^{ERA5} + \Sigma_{M,H,g}^v \quad (4)$$

190 v denotes the meteorological forcing variables, i.e., air temperature, specific humidity, surface
 191 pressure, wind speed, longwave and shortwave radiation. Σ denotes the additive correction
 192 factor. In addition, corrected specific humidity was capped by its physically plausible upper
 193 limit, i.e., the specific humidity when air temperature equals the dew point.

194 3.2 Soil texture and organic matter – SoilGrids

195 Soil texture and organic matter directly affect the soil thermal and hydrologic properties
 196 and thus the hydrologic cycle. The spatial resolution of the default soil texture data in CTSM is
 197 very coarse, so we replaced it with the high-resolution soil property products from the SoilGrids

198 system (Hengl et al., 2017). The SoilGrids prediction model utilized over 230,000 soil profile
199 observations from the WoSIS database (Batjes et al., 2020) and environmental covariates to
200 generate global soil property maps at 250-m resolution for six standard depth intervals.

$$PCT_{SAND} = \frac{\gamma_{sand}}{\gamma_{sand} + \gamma_{silt} + \gamma_{clay}} \times 100\% \quad (5)$$

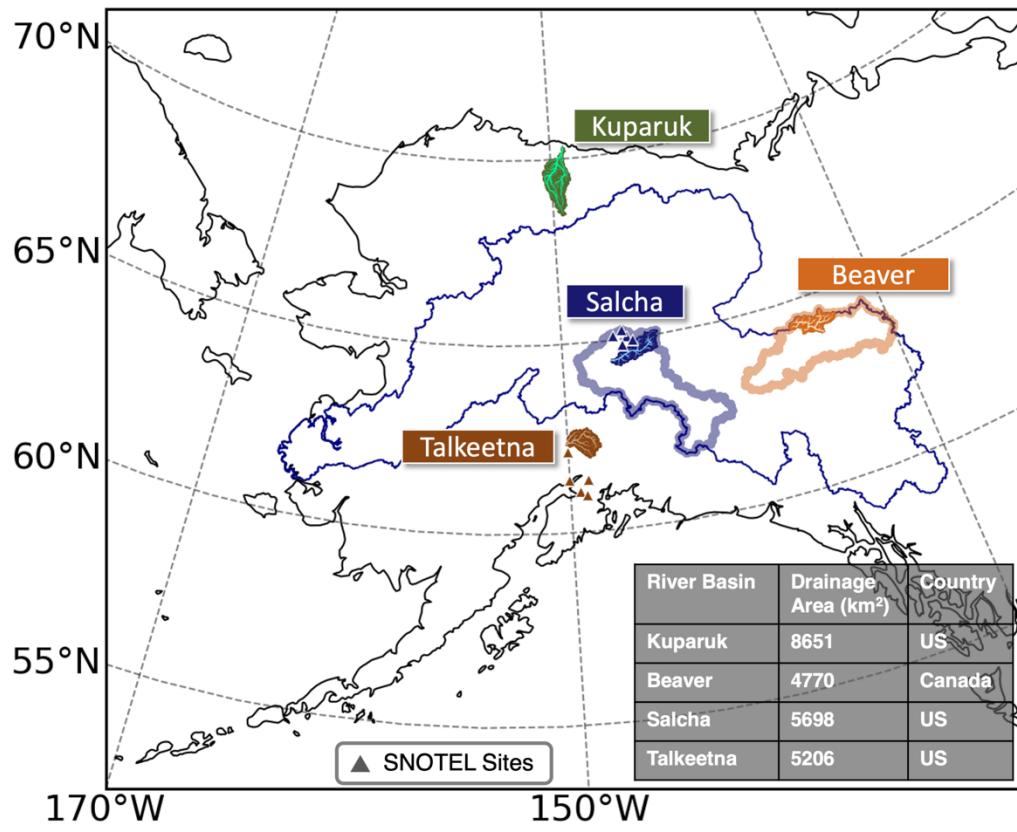
$$PCT_{CLAY} = \frac{\gamma_{clay}}{\gamma_{sand} + \gamma_{silt} + \gamma_{clay}} \times 100\% \quad (6)$$

$$\rho_{OM} = \frac{\gamma_C \cdot \rho_{bulk} \cdot 10^{-3}}{0.58} \quad (7)$$

201 Percentages of sand and clay, PCT_{SAND} and PCT_{CLAY} , were calculated based on the sand, silt, and
202 clay contents (γ_{sand} , γ_{silt} , γ_{clay} , unit: g/kg). Organic matter density (ρ_{OM} , unit: kg/m³) was
203 calculated using the soil organic carbon (γ_C , unit: dg/kg) and bulk density (ρ_{bulk} , cg/cm³) with
204 the assumption of carbon content 0.58gC per gOM.

205 3.3 Sub-grid variability – Hillslope Hydrology

206 Explicitly resolving hillslope-scale features can better capture the sub-grid distribution of
207 water and energy within an LM grid cell (Fan et al., 2019), and has been implemented into
208 CTSM (Swenson et al., 2019). The hillslope configuration used in this study consisted of four
209 hillslopes per grid cell, each representing a different aspect (i.e., north, east, south, west), with
210 each hillslope comprised of an upland column and a lowland column to explicitly simulate the
211 flow of soil water along topographic gradients. In low-relief grid cells, only one column was
212 specified.



213

214 *Figure 3: Selected representative medium-sized basins for parameter estimation. Triangles*
 215 *denote SNOTEL sites with snow observations. Salcha River Basin is a subbasin of the Tanana River*
 216 *Basin (thick blue line) and Beaver River Basin is a subbasin of the Steward River Basin (thick orange*
 217 *line).*

218 **4 Optimization framework**

219 We utilized a surrogate-based modeling optimization machine learning method to
 220 optimize CTSM parameters to provide improved hydrologic simulations across our study region.
 221 We specifically focused on river flow and snow and their objective functions were defined in
 222 Section 4.1. As a state-of-the-science land model, CTSM is computationally expensive to run,
 223 and it has over 200 tunable parameters. To constrain the computational cost, we first selected
 224 four representative medium-sized river basins for optimization: the Talkeetna, Salcha, Beaver,
 225 and Kuparuk river basins (Figure 3). Second, we determined the most sensitive parameters that

226 impact the simulation of Arctic hydrology (Section 4.2). In addition, we used a computationally
227 frugal optimization method to reduce the total number of CTSM runs (Section 4.3). Based upon
228 a preliminary optimization experiment for each basin, we found the optimized parameters
229 showed substantial differences for the basin in the northern slope, i.e., Kuparuk, as opposed to
230 the three southern basins. A simple parameter regionalization method was adopted with
231 corresponding modifications to CTSM to accommodate the spatial heterogeneity of model
232 parameters (Section 4.4). The optimization workflow is shown in Figure 2.

233 **4.1 Multi-objective functions for flow and snow conditions**

234 We aimed to provide optimized simulations of multiple components of the water budget.
235 Given the limited observations in the region, we chose to optimize streamflow and snowpack as
236 these two components of the water budget have direct measurements across our study domain
237 over multiple years and locations. The flow objective function (O_Q) is the Kling-Gupta
238 Efficiency (KGE, Gupta et al., 2009) using daily mean streamflow.

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (8)$$

$$O_Q = 1 - KGE \quad (9)$$

239 KGE is a comprehensive metric that integrates the linear correlation (r), a measure of flow
240 variability error (α), and a bias term (β). For all USGS flow observations, we only used the data
241 with a qualifier equal to A, which corresponds to the ice-free period. For snow, we designed an
242 objective function (O_S) that aggregates three bias terms in snow simulations; relative errors in
243 annual peak SWE (rE_{pS}), snow persistence time (rE_{tSp}), and snow melting rate (rE_{vSm}). Snow
244 persistence time is defined as the annual number of days with SWE larger than 0.1 mm. If it is
245 perennial snow, the melting rate is calculated based upon the annual peak SWE and the SWE on
246 August 31st, which is close to the date with the lowest annual SWE. If it is not perennial snow,

247 the melting rate is calculated based upon the annual peak SWE and the first day when SWE falls
248 below 0.1 mm. The snow objective function is the quadratic mean of the three relative error
249 terms,

$$O_S = (rE_{pS}^2 + rE_{tSp}^2 + rE_{vSm}^2)^{\frac{1}{2}} \quad (10)$$

250 The aggregated snow metric O_S is unitless.

251 4.2 Parameter sensitivity

252 We were able to leverage ongoing CTSM parameter sensitivity experiments to inform our
253 parameter optimization experiments. Dagon et al. (2020) established the most sensitive CTSM
254 parameters for global surface energy balance and hydrology among a subset of 34 parameters.
255 An ongoing experiment, the CTSM Perturbed Parameter Ensemble (henceforth PPE), extends
256 this work to a larger set of CTSM parameters. This work is ongoing, but we were able to access
257 their one-at-a-time experiment, which varied over 200 parameters across expert-derived ranges.
258 Data and description are available via <https://github.com/djk2120/clm5ppe>.

259 We adopted a two-step method to select sensitive parameters for optimization. First, we
260 selected the top 40 parameters that exert a strong influence on Arctic hydrology from over 200
261 parameters that were varied within the PPE. Because the CTSM configuration for the PPE did
262 not utilize the hillslope hydrology nor did it include river routing, routed flow is not available in
263 the PPE experiment and we thus performed an additional filtering step. We used the unrouted
264 runoff as a substitute for routed flow when selecting sensitive parameters in the first step. While
265 moving from over 200 to 40 parameters is a substantial simplification of the potential
266 optimization space, it is still computationally expensive to tune 40 parameters within CTSM.
267 Therefore, we further identified the most sensitive parameters by training a surrogate model to
268 simulate the response surface of objective functions to each parameter. The top 14 out of the 40

269 pre-screened parameters were selected for optimization. Both steps are explained in detail as
270 follows.

- 271 • ***Step 1:*** We used the PPE one-at-a-time experiment to select which parameters exert the
272 most control on total runoff (QRUNOFF) and snow water equivalent (SWE). To
273 constrain computational costs, the PPE was run at 400 grid cells globally to represent the
274 parameter sensitivities at different land cover types and climatologies. Seven of those
275 grid cells fall in our study domain and we used the mean response across the seven grid
276 cells to evaluate parameter sensitivity for Arctic hydrology. For QRUNOFF, we
277 evaluated the mean, seasonality, and amplitude; for SWE, we evaluated the snow
278 persistence duration, maximum monthly SWE, and snowmelt rate, which leads to a total
279 of six variable-metric combinations. For each combination, we selected the top 15 most
280 sensitive parameters and assigned a higher score to more sensitive parameters (e.g., 15
281 points to the most sensitive parameter, 1 point to the least sensitive parameter). As a pre-
282 screen step, we would like to include as many sensitive parameters as possible within our
283 capacity to handle complexity and we selected 15 after experimenting with different
284 numbers. The scores for each parameter were summarized across all six variable-metric
285 combinations and the total score represents the general uncertainty of the parameters to
286 runoff and snow conditions in our study domain. A total of 40 parameters across all
287 variable-metric combinations were pre-screened as candidate parameters and would be
288 further selected in Step 2.

289

290 Table 1: Summary of 14 parameters selected for optimization, their categories, relevant physical processes, ranking based on
 291 scores in Step 2, parameter default values, ranges, as well as optimized values for northern and southern basins

Category	Parameters	Relevant Physical process	Rank	Default value	Range	Optimized value in south	Optimized value in north
Acclimation parameters	vcmaxha	Photosynthesis, activation energy for $V_{c,max}$	11	72000	[20000, 250000]	235175	155394
	om_frac_sf	Scalar adjustment for organic matter fraction	6	100%* DV	[25%,200%]* DV	52.551%* DV	85.813%* DV
	slopebeta	Surface water storage	9	-3	[-10,-0.5]	-6.936	-8.131
	fff	Delay factor for fractional saturated area	2	0.5	[0.01,10]	0.010	5.494
	e_ice	Ice impedance factor	4	6	[1,8]	7.994	2.335
Hydrology	liq_canopy_storage_scalar	Maximum storage of liquid water on leaf surface	11	0.1	[0.025, 4]	3.695	2.372
	krmax [‡]	Root segment max conductance	11	1.223×10^{-9}	$[5.827 \times 10^{-11}, 6.896 \times 10^{-9}]$	3.626×10^{-9}	1.280×10^{-10}
Sensible, latent heat and momentum fluxes	d_max	Heat and momentum flux for non-vegetated surface, dry surface layer (DSL) thickness	9	15	[5,100]	49.808	27.065
	frac_sat_soil_dsl_init	Heat and momentum flux for non-vegetated surface, Fraction of saturated soil for moisture value at which DSL initiates	4	0.8	[0.25,2]	0.250	1.782
	cv	Turbulent transfer coefficient between canopy surface and canopy air	11	0.01	[0.0025,0.04]	1.708×10^{-2}	1.801×10^{-2}

	a_coef	Drag coefficient under less dense canopy	8	0.13	[0.05,0.15]	5.009×10^{-2}	1.086×10^{-1}
Snow processes	upplim_destruct_metamorph	Upper limit for snow densification through destructive metamorphism	1	175	[10,500]	86.023	321.095
	n_melt_coef	Parameter controlling shape of snow covered area	2	200	[25,600]	25.000	232.078
	Medlyn_intercept [‡]	Medlyn intercept of conductance-photosynthesis relationship	6	100	[1,20000]	7.326×10^3	1.200×10^4

292

293 [‡] denotes that the parameter is plant functional type (PFT) dependent and the value shown in the table is the mean value across all

294 PFTs.

295 DV is short for default values.

296 Hydrologic parameters are highlighted using blue (Hydrology) and navy (Snow processes) and non-hydrologic parameters are

297 highlighted using red (Sensible, latent heat, and momentum fluxes) and green (plant parameters).

298

- 299 • **Step 2:** To select the most sensitive parameters, we simulated the response of flow and
300 snow objectives to the CTSM model parameters using surrogate models. For each river
301 basin, we trained one surrogate model from 200 samples generated using the LHS
302 method across the 40-dimension parameter space. Because the response of the objective
303 function to one parameter in a multi-variate surrogate model is affected by other
304 parameters, we can get a mean response by fixing the target parameter while perturbing
305 the remaining 39 parameters. For example, to get the response to *fff* (Table 1) when *fff*
306 equals 1, we utilized the 200 samples that were generated using LHS and fixed *fff* to 1,
307 using the surrogate model to predict the response of the 200 modified samples, and
308 average the responses to get a mean response. For one parameter, we calculated the mean
309 responses at multiple points to get a two-dimensional response curve (Figure S1). The
310 amplitude of the response curve was used to evaluate each parameter's sensitivity. For
311 the basins where we conducted multi-objective optimization, we aggregated the
312 amplitudes of the response surface across both objectives to make sure that the selected
313 parameters should be generally sensitive for both objectives.

314
315 We used a simple weighting algorithm to select the final parameter list for optimization.
316 For each river basin, the most sensitive 10 parameters were assigned non-zero scores, i.e.,
317 5, 3, 3, 2, 2, 2, 1, 1, 1, 1. In any single basin, parameters with ranks lower than 10 barely
318 show sensitivity to the objective functions. In addition, this weighting algorithm
319 emphasizes the most sensitive parameters in any single basin, which may not be sensitive
320 elsewhere. In total, nineteen parameters were in the top 10 most sensitive across all
321 basins. We selected all parameters with a total score higher than 1, meaning they were at

least one of the 6 most sensitive parameters in any one basin, or somewhat sensitive in multiple basins. This resulted in fourteen parameters being selected for full optimization (Table 1). It is possible that multiple parameters shared the same scores and therefore the same ranks, e.g., *om_frac_sf*, *medlynintercept* (rank 6) and *d_max*, *slopebeta* (rank 9) in Table 1.

4.3 Adaptive Surrogate Based Modeling Optimization (ASMO)

Adaptive Surrogate Based Modeling Optimization (ASMO) is an emerging optimization method that can be used for tuning hydrologic model parameters (Wang et al., 2014). Compared to the widely used Shuffled Complex Evolution global optimization method (Duan et al., 1994), ASMO is much more efficient, which is especially important in this application because CTSM is more computationally expensive than most hydrologic models due to its comprehensive suite of processes. We adopted the workflow developed in Gong et al (2016) for a multi-objective optimization, which is summarized below:

- **Initial Sampling:** 200 samples were generated using the Latin Hypercube Sampling (LHS, McKay et al., 2000) method for the selected parameters. In this study, one sample denotes one set of parameter values. We ran CTSM using the 200 sets of parameter values and calculated their corresponding objective functions.
- **Main Loop (Iteration):** We used the Gaussian Process Regression (GPR) model to train a surrogate model, which mimics the response of the objective functions to parameters. In the first iteration, we used all 200 initial samples and corresponding objectives to train the surrogate model. In each subsequent iteration, all samples from the initial sampling and previous iterations were used to train a new surrogate model. Then we used a multi-objective optimization, i.e., Non-dominated Sorting Genetic Algorithm II (NSGA-II, Deb

345 et al., 2002), on the surrogate model, and obtained N (N=20) Pareto optimal sets of
346 parameters values. We then ran CTSM using the N sets of parameter values and
347 calculated their objective functions.

348 The trained surrogate model better mimicked the response curves as the number of
349 samples increased via iterating the Main Loop. In Wang et al. (2014), a 13-parameter
350 optimization case converges in roughly 400 runs. In this study, we optimized 14 parameters and
351 stopped after the fifteenth iteration given the limited improvement in the last iteration runs (500
352 runs in total). We used k-fold cross validation to evaluate the accuracy of the surrogate model
353 (k=5). We calculated the root-mean-square error (RMSE) of the simulated objectives from
354 surrogate models versus the objectives calculated from CTSM runs.

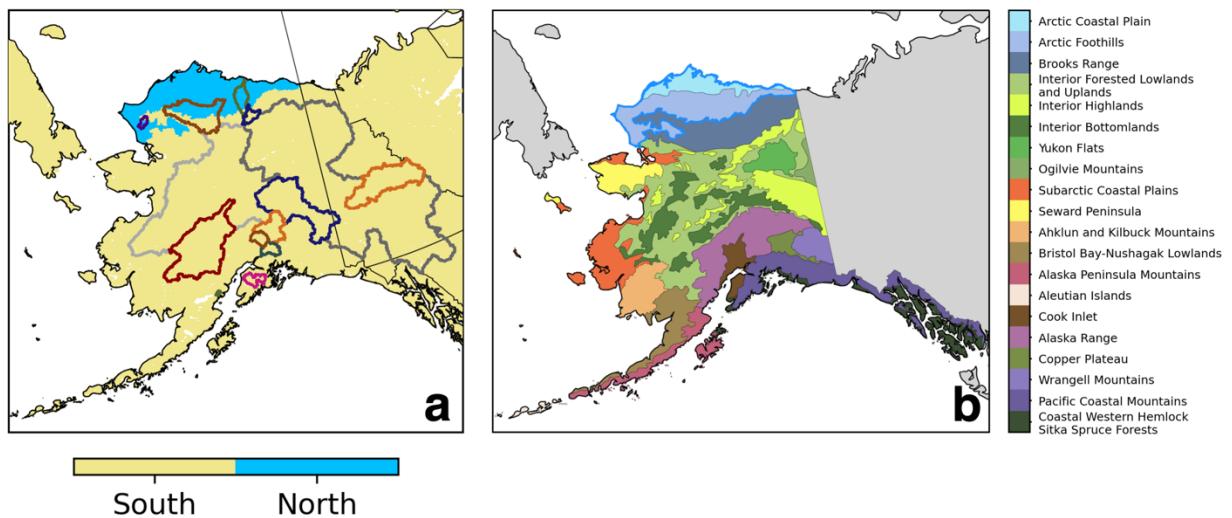
355 The optimization run ranges from 1 September 2002 to 1 September 2009. The first two
356 years are used for spin-up, with data from 2004 to 2009 used for optimization. Prior to the
357 optimization simulations, as it takes a relatively long time for deep-layer soil moisture to reach
358 equilibrium from the default initial condition, we used a 52-year spin-up forced by ERA5 data to
359 generate the initial state for 1 September 2002 using the default CTSM parameters. For the
360 optimization runs, a 2-year spin-up is sufficient as we already have an equilibrium soil moisture
361 state. All simulations were performed on the NCAR Cheyenne supercomputer (Computational
362 and Information Systems Laboratory, 2019).

363 **4.4 Parameter regionalization**

364 Many parameters within CTSM are spatially uniform by default, which can be a limiting
365 assumption when optimizing a model as many parameters within hydrologic and land models
366 should vary spatially to account for the heterogeneity across the landscape (Mizukami et al.,
367 2017; Rakovec et al., 2019; Samaniego et al., 2010). For plant parameters, parameter spatial

368 heterogeneity might result from different plant traits in different dominant plant species. We
369 conducted preliminary single basin optimizations which showed large optimal parameter
370 discrepancies between the northern river basin, i.e., Kuparuk, and southern river basins, i.e.,
371 Beaver, Salcha, and Talkeetna (not shown). The Kuparuk River Basin is located north of the
372 Arctic Circle, much farther north than the other three basins. Therefore, we conducted two
373 optimization runs in this study, one for the northern river basin, and one for the three southern
374 river basins. Note that no SNOTEL sites near the Kuparuk have records overlapping with our
375 optimization period, thus we conducted a single-objective optimization on river flow for the
376 Kuparuk River. For the southern basins, we conducted a dual-objective optimization by
377 averaging the flow objectives across the three basins and snow objectives across ten SNOTEL
378 sites, which serve as the representative objectives across the entire southern region so the
379 selected SNOTEL sites and basins do not necessarily need to overlap (triangles in Figure 3).

380 We leveraged the ecohydrology region classification level III by Environmental
381 Protection Agency for our simple parameter regionalization (Gallant et al., 1995). The EPA
382 ecohydrology region classification helps determine the boundary between the southern and
383 northern regions. Because the Kuparuk mostly resides north of the arctic circle, optimized
384 parameters for the Kuparuk are applied to the two Arctic ecohydrology regions, Arctic Coastal
385 Plain and Arctic Foothills (highlighted in blue lines, Figure 4b). The remaining area uses the
386 optimized parameters for the southern basins. Three out of the 15 basins intersect both southern
387 and northern parameter regions, the Colville, Wulik, and Kuparuk rivers. The Colville is
388 comprised of 54% northern and 46% southern areas, the Wulik contains 72% northern and 28%
389 southern areas. Also, even though we optimized the Kuparuk to represent northern basins, 10%
390 of the area in the Kuparuk watershed is located in our southern region (Figure 4a).



391

392 *Figure 4: Parameter regionalization based on ecohydrology region classification. In Figure 4a,*
 393 *background colors denote the selection of optimized parameters, and river basins are highlighted using*
 394 *solid lines with colors corresponding to Figure 1. In Figure 4b, regions using optimized parameters for*
 395 *northern regions are highlighted in blue boundaries.*

396 **4.5 Actionability of CTSM**

397 The definition of actionable science above suggests the necessity to set a performance
 398 benchmark that a model has to hit before it could be used to inform decision-making. We
 399 selected the following two hydrologic metrics, percent bias in flow duration curve (FDC) high-
 400 segment volume (%BiasFHV, Yilmaz et al., 2008) given the interest of the IAC and the results
 401 of the climate information survey focused on flooding, and percent bias in FDC low-segment
 402 volume during summer season (%BiasFLV_{summer}, May to October) because summer low flows
 403 have significant impacts on the juvenile production for salmon (Ohlberger et al., 2018). No
 404 universally accepted benchmarks exist for these metrics so we used the uncalibrated model
 405 performance as our benchmark. In addition, we selected an additional hydrologic metric, the
 406 Nash Sutcliffe Coefficient (NSE) for daily flow, because it is a widely used metric in water

407 management authorities in the United States. We adopted a benchmark of daily NSE at 0.5
408 (Moriasi et al., 2015).

409 Complementary to solely evaluating flow simulations against observations, we conducted
410 a climate sensitivity analysis to assess whether CTSM captures the response of flow simulations
411 to changes in precipitation and air temperature. We followed the technique developed in Wood et
412 al. (2004). For each river basin, regionally averaged flow (\bar{Q}), air temperature (\bar{T}), and
413 precipitation (\bar{Pr}) were calculated for each hydrologic year and for the observations, baseline,
414 and optimized simulations. To quantify the uncertainties in flow responses to climate variables,
415 we conducted bootstrapping 300 times with each bootstrapping sample generating a new series
416 of precipitation, air temperature, and streamflow by resampling the available hydrological years
417 with replacement (n = 5000 times, Brunner et al., 2020). For each new series, we fitted a simple
418 linear regression between the flow and climate variables and the slope denotes the corresponding
419 responses. By assessing the baseline and optimized responses to the observed response, we can
420 evaluate whether the optimization improves estimated climate sensitivities.

421 **4.6 Parameter performance contributions**

422 For each optimization region, we applied the Shapley decomposition to quantify the
423 contribution of each parameter to the total change in the objective functions (Roth, 1988). The
424 Shapley decomposition originated from cooperative game theory, where it was applied to
425 determine each player's unique contribution to a total surplus generated by a coalition of all
426 players. Recently, this method has also been applied in energy and environmental analyses (Ang
427 et al., 2003; Yu et al., 2014). We performed the analysis on the 14 optimized parameters for
428 southern and northern regions separately. The change in the objective function is calculated as

$$O_s = f(P) \quad (11)$$

$$\Delta O = O_{optz} - O_{base} \quad (12)$$

$$\Delta O = \sum_{\gamma} \varphi_{\gamma}(f) \quad (13)$$

429 where O denotes objective functions, f denotes the trained surrogate model for one region,
 430 subscript s denotes scenarios ($s=optz, base$, denoting the optimized and baseline scenarios
 431 respectively), P denotes the list of all parameters for optimization, and $\varphi_{\gamma}(f)$ denotes the unique
 432 contribution of parameter γ for the selected region. For one selected parameter γ , the unique
 433 contribution $\varphi_{\gamma}(f)$ is calculated as

$$\varphi_{\gamma}(f) = \frac{1}{n} \sum_{S \subseteq P \setminus \{\gamma\}} \binom{n-1}{|S|}^{-1} (O(S \cup \{\gamma\}) - O(S)) \quad (14)$$

$$\binom{n-1}{|S|} = \frac{(n-1)!}{|S|! (n-1-|S|)!} \quad (15)$$

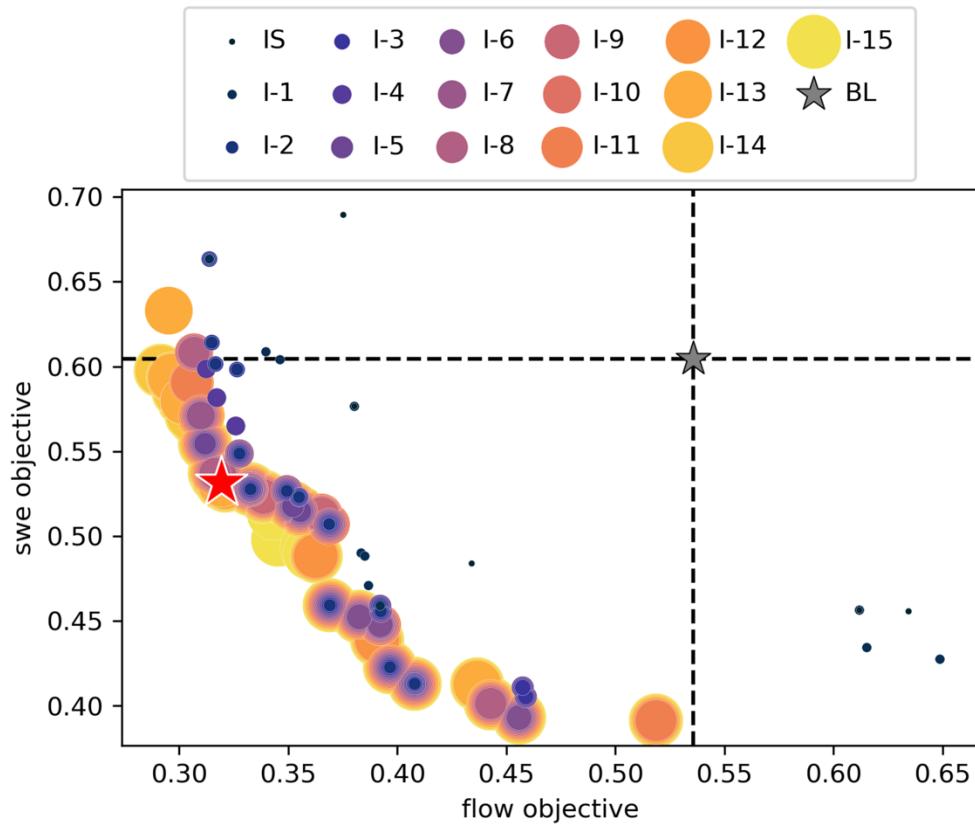
434 where n is the total number of parameters for optimization, i.e., 14, $P \setminus \{\gamma\}$ denotes all parameters
 435 except the selected one γ , S denotes the subset of $P \setminus \{\gamma\}$, $|S|$ denotes the length of the subset,
 436 $O(S)$ denotes the objective function when we replace the baseline value using the optimized
 437 value for all parameters in subset S .

438 5 Results

439 5.1 Optimization

440 For the southern basins, flow simulation is improved substantially while snow simulation
 441 only sees minor improvements (Figure 5). Dots with the same color in Figure 5 constitute the
 442 simulated Pareto front for a given optimization iteration. A Pareto front consists of simulated
 443 Pareto optimal, i.e., if none of the objective functions can be improved in value without
 444 degrading some of the other objective values. In general, the simulated Pareto front shifts

445 towards the origin, signifying improved model performance. Overlapping dots indicate the new
 446 iteration failed to improve the Pareto front at that point. The two-dimensional Pareto front serves
 447 as the basis for choosing our optimal parameter set. Future work could explore using an
 448 ensemble of optimal parameter sets along the Pareto front, but that is outside the scope of this
 449 initial investigation. We choose the set of optimized parameters that correspond to the minimum
 450 averaged flow and snow objectives, highlighted using a red star in Figure 5. For this parameter
 451 set, the corresponding flow and snow objective functions for the southern basins are 0.319 (0.681
 452 KGE) and 0.531 respectively, while the baseline flow and snow objectives are 0.536 (0.464
 453 KGE) and 0.604 respectively. Flow simulation in the Kuparuk is also significantly improved
 454 through optimization. Since we conducted a single-objective optimization for Kuparuk, we
 455 simply selected the set of parameters resulting in the best flow simulation. The optimized flow
 456 objective is 0.311 (0.689 KGE) while the default flow objective is 0.739 (0.261 KGE).



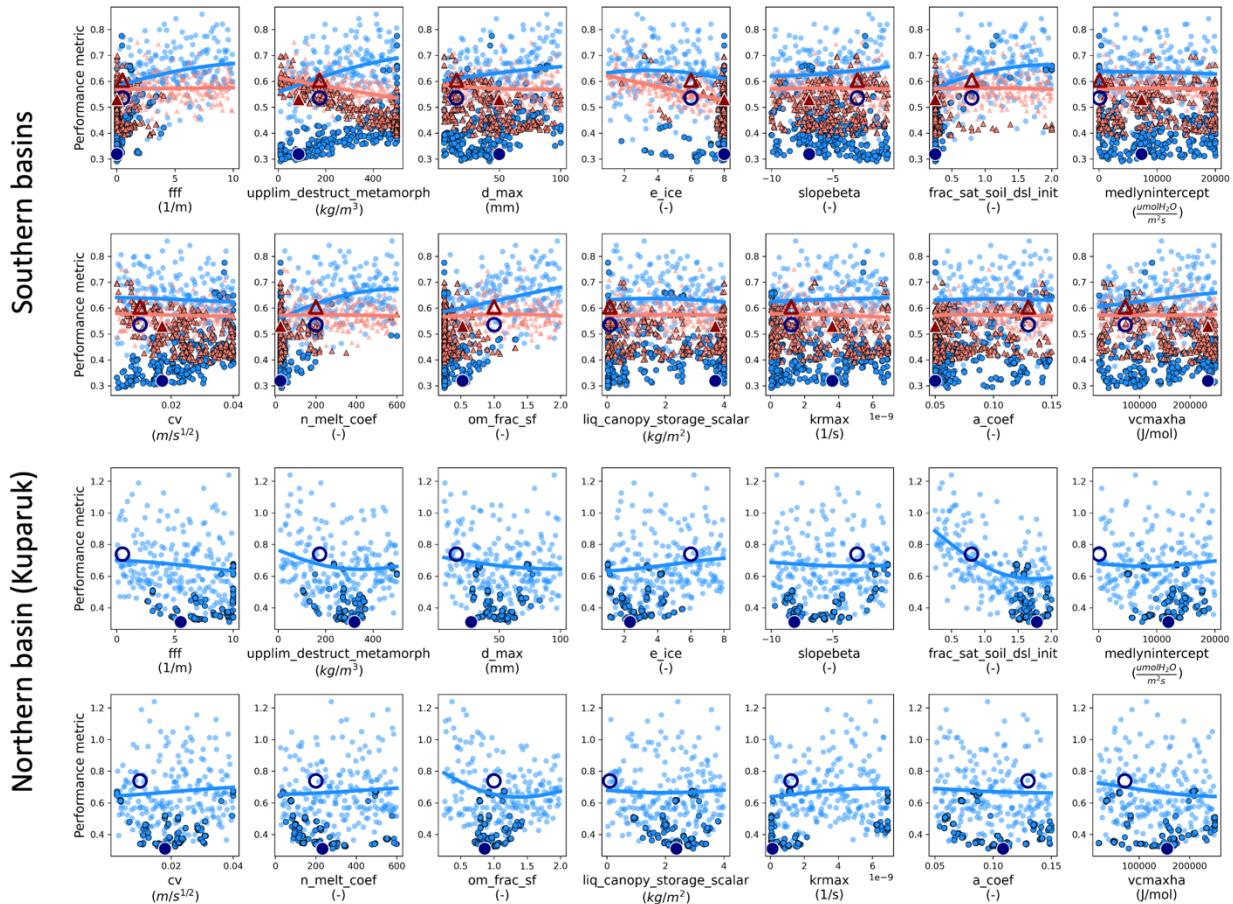
457

458 *Figure 5: Simulated Pareto front of optimization for southern basins. Each colored dot*
459 *corresponds to a Pareto optimal set of parameters. IS denotes initial sampling, I-I denotes the first*
460 *iteration, and so on, and BL denotes the baseline configuration. The red star denotes the selected*
461 *optimized parameters.*

462 Interestingly, the northern and southern basins show very different hydrological
463 responses to parameter perturbations as noted above. The mean response curve of flow (blue
464 dots) and snow (red triangles) objectives to model parameters are shown in Figure 6. We used
465 the method in Section 4.2 (Step 2) to calculate the mean response curves. Transparent dots
466 denote the initial samples, while solid dots denote samples during optimizations and large dots
467 correspond to the selected optimized parameters (optimized parameter values are shown in Table
468 1). The parameter sensitivity differs across basins. For example, *upplim_destruct_metamorph*,
469 which affects snow densification through destructive morphism, shows greater sensitivity on
470 flow simulations in the southern basins and is only marginally sensitive in the Kuparuk. In
471 addition, the flow performance in the south degrades as *upplim_destruct_metamorph* increases
472 while the opposite trend was observed in the north. Some other parameters also show the
473 opposite sensitivity across regions, including, *d_max*, *e_ice*, *frac_sat_soil_dsl_init*, and
474 *om_frac_sf*. This intrinsic sensitivity difference leads to the divergence in optimized parameters
475 across regions. In some extreme cases, the optimized parameters approach the upper and lower
476 limits for the northern and southern basins, respectively, e.g., *frac_sat_soil_dsl_init* and *e_ice*,
477 which might result from differences in physical processes across the domain. For example, *e_ice*
478 together with soil ice content affects the hydraulic conductivity in frozen soils and therefore has
479 impacts on the vertical distribution of soil moisture and runoff (Swenson et al., 2012). There is
480 ice-rich permafrost in the north while not in the south (Saito et al., 2020), so the differences in
481 soil ice content might affect the optimized value of the ice impedance factor differently. In

482 addition, parameter values approaching the specified limits could indicate that the ranges are not
483 wide enough due to model structural or forcing data errors that are compensated for during
484 parameter optimization.

485 The responses of flow and snow objectives may diverge for the same parameter
486 perturbation. For the southern basins, as *upplim_destruct_metamorph*, *n_melt_coef*, and
487 *om_frac_sf* increase, flow simulation becomes worse while snow simulation improves (Figure
488 6). These parameter divergences could be the result of compensating errors from model structure
489 (either a lack of or incorrectly parameterized processes), meteorological forcing, or indicative of
490 the true CTSM parameter sensitivities for our study domain (Clark & Vrugt, 2006; Vrugt et al.,
491 2005). In addition, the spread of the flow objective (blue dots) is much larger than that of the
492 snow objective (red triangles) in Figure 6. The SWE simulation is likely more controlled by
493 meteorological forcing than parameter perturbations. Therefore, runoff and flow simulations
494 might show a stronger sensitivity to the parameter perturbations than SWE.



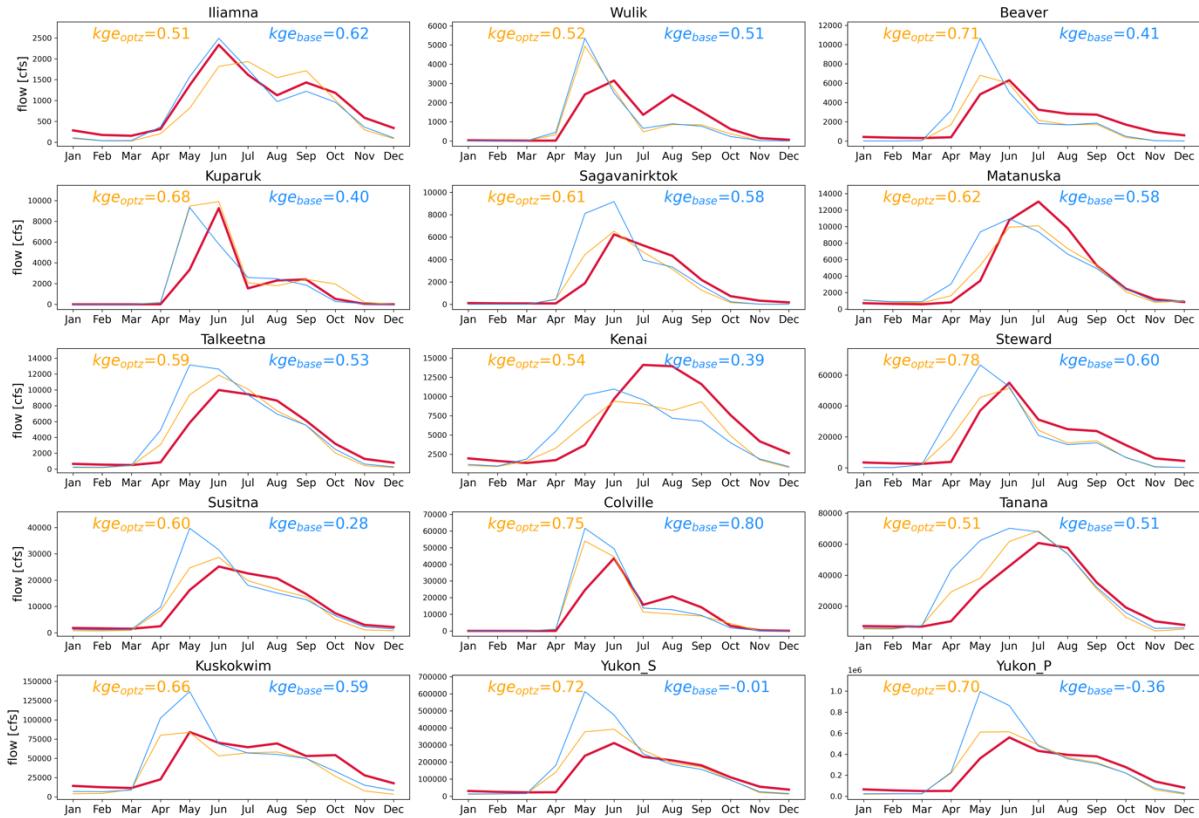
495

496 *Figure 6: Mean response curve of flow (blue dots) and snow (red triangles) objectives to*
 497 *parameters. Transparent dots and triangles denote initial samples, solid dots, and triangles with black*
 498 *edges denote samples during optimization iterations, the large hallow dot and triangle with darker colors*
 499 *denote the default parameters, and the large dot and triangle with darker colors denote the selected*
 500 *optimized parameters.*

501 5.2 A 30-year out-of-sample evaluation of optimized parameters

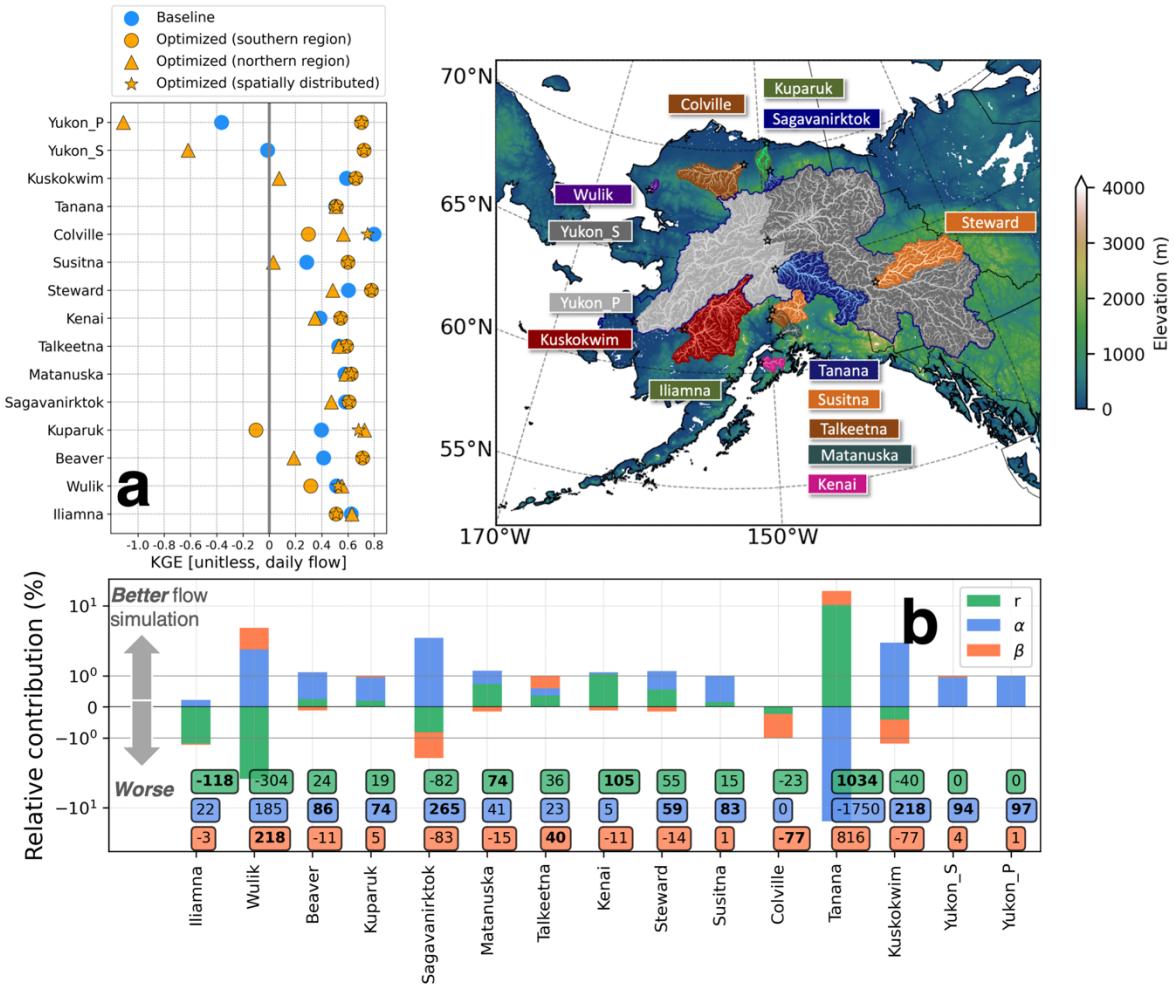
502 We further evaluate the optimized CTSM at 15 major river basins and 12 of them are out-
 503 of-sample from 1991 to 2020 water year (WY). The daily KGE improves at 13 out of 15 basins
 504 and the mean KGE across the 15 basins increases from 0.43 to 0.63 after optimization.
 505 Furthermore, even though we only conducted optimizations for four medium-size river basins
 506 with a total confluence area of around 16,500 km², 2.1% of the total out-of-sample simulated

area, the optimized flow simulations for the largest basins still substantially improved (Figure 7).
 For example, the daily KGE for the Yukon River at Pilot Station ($824,393 \text{ km}^2$) increases from -0.36 to 0.70 and the daily KGE for the Yukon River at Stevens Village ($502,458 \text{ km}^2$) increases from -0.01 to 0.72. Only the Iliamna and Colville river basins show slightly worse performance, with daily KGE decreasing from 0.62 and 0.80 to 0.51 and 0.75 respectively.



512

513 *Figure 7: Model evaluation for mean monthly flow time series (from 1991 to 2020 WY). KGE is*
 514 *based on qualified observed flow at a daily time step.*



515

516 Figure 8: A 30-year evaluation from 1991 to 2020 WY a) default model performance (blue dot),
 517 and model performance using southern optimized parameters (orange dot), northern optimized
 518 parameters (orange triangles), and spatially distributed parameters by parameter regionalization
 519 (orange stars) and b) contribution of r , α , and β to KGE changes using the spatially distributed
 520 parameters by parameter regionalization. Percent contribution values are noted below and the main
 521 contributing factors are highlighted in bold fonts.

522 Improved model performance in cross-regional basins highlights the necessity of spatially
 523 variable parameters and parameter regionalization schemes. For Colville and Wulik, two
 524 uncalibrated basins, their model performance using spatially distributed parameters is better or
 525 barely worse than that of any single optimized parameter set (Figure 8a). In addition, their

526 default flow simulations are similar to the ones using spatially distributed parameters. Compared
527 to the baseline, optimized flows in Wulik and Colville both show a smaller wet bias in spring, a
528 larger dry bias in summer, and a smaller dry bias in fall (Figure 7). The optimized parameters in
529 Kuparuk represent the northern region while our regionalization algorithm categorized 10% of
530 the area in Kuparuk to the southern region, which explains the slightly worse performance in
531 Kuparuk using the spatially distributed parameters than that using only the northern optimized
532 parameters.

533 Improved flow variability contributes the most to better flow simulation. KGE combines
534 three components in model errors, i.e., the linear correlation (r), a measure of flow variability
535 error (α), and a bias term (β), so we decompose the KGE increment to the three components and
536 calculate their relative contribution (RC) as follows

$$RC = \frac{(m_{base} - 1)^2 - (m_{optz} - 1)^2}{|(KGE_{base} - 1)^2 - (KGE_{optz} - 1)^2|}, m = r, \alpha, \beta \quad (16)$$

537 Since we used the absolute value of KGE difference as the denominator, regardless of
538 KGE_{optz} being higher or lower than KGE_{base} , a positive RC value always denotes better flow
539 simulation and a negative RC value always denotes worse flow simulation. Additionally, when
540 the sum of RC is positive, the optimized flow simulation is improved, and vice versa. Improved
541 flow variability, linear correlation, and volume bias contribute the most to the improved flow
542 simulations in eight, three, and two river basins respectively (Figure 8b). Poorly simulated flow
543 volume and correlation mostly contribute to the poorer flow simulation in Colville and Iliamna,
544 respectively.

545

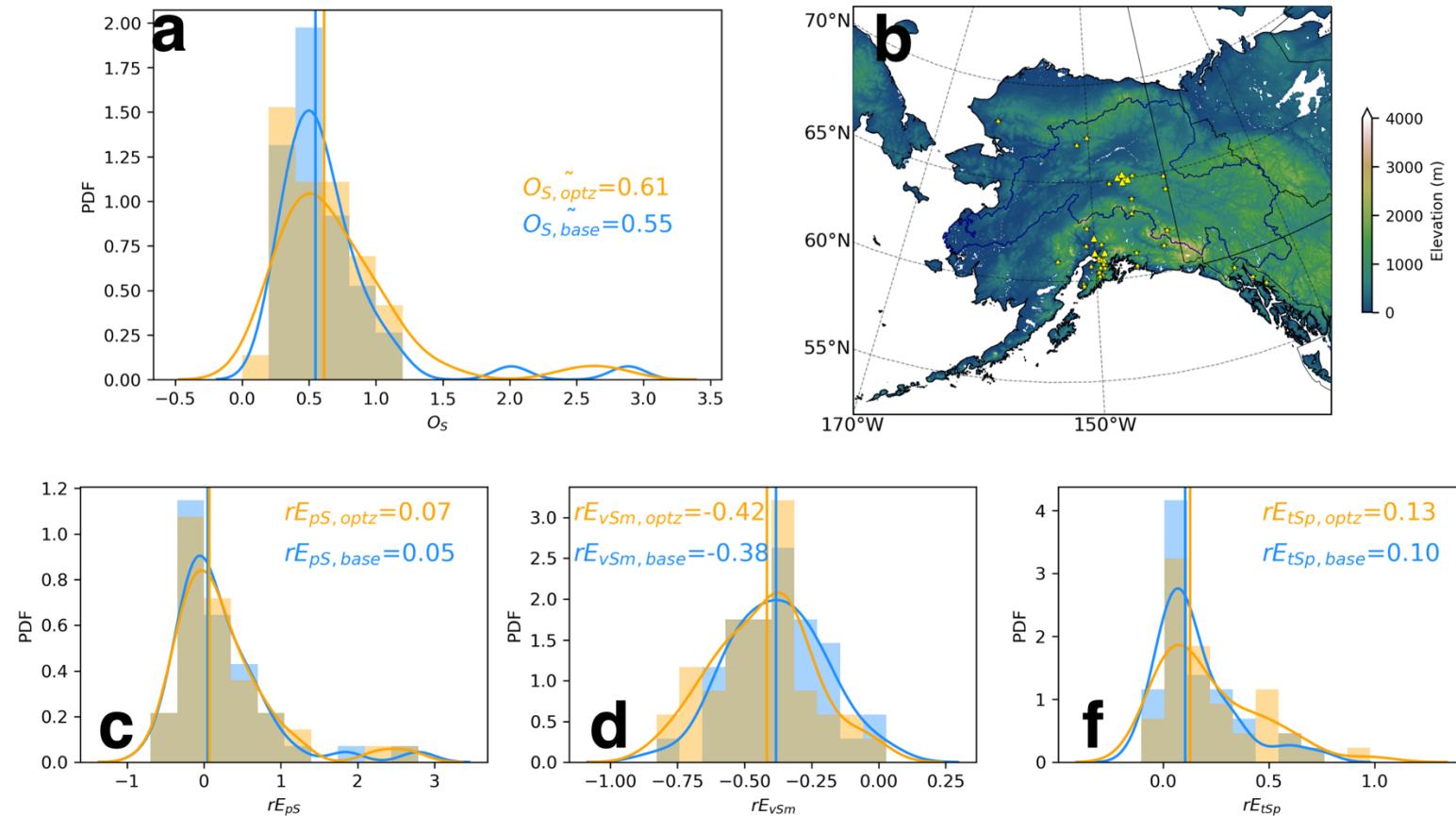
546

Basin Name	Nation	NSE optz [-]	NSE base [-]	%BiasFHV optz [%]	%BiasFHV base [%]	%BiasFLV_{summer} optz [%]	%BiasFLV_{summer} base [%]
		Benchmark: NSE=0.5		Benchmark: Baseline		Benchmark: Baseline	
Iliamna	US	0.32	0.55	-14.0	-26.8	-415.4	-504.5
Wulik	US	0.25	0.20	-5.6	-5.2	32.8	-56.4
Beaver	Canada	0.51	-0.21	-5.1	32.3	-496.9	-272.0
Kuparuk	US	0.35	0.22	-26.4	-36.6	55.9	67.7
Sagavanirktok	US	0.53	0.01	-15.9	29.2	-15.9	-42.1
Matanuska	US	0.59	0.55	-12.2	-22.0	-188.7	-134.0
Talkeetna	US	0.55	0.37	15.4	16.4	-158.5	-145.0
Kenai	US	0.43	0.26	3.3	-12.5	-466.3	-180.7
Steward	Canada	0.64	0.17	-20.8	4.2	-113.0	-124.0
Susitna	US	0.61	0.18	17.1	47.3	-45.4	-4.0
Colville	US	0.47	0.35	-1.6	2.4	7.2	-160.8
Tanana	US	0.56	0.17	31.9	35.5	-114.6	-78.9
Kuskokwim	US	-0.03	-0.33	24.2	85.7	-34.1	32.5
Yukon_S	US	0.50	-0.57	1.4	61.8	-78.3	-36.2
Yukon_P	US	0.50	-0.55	8.4	81.4	-11.2	18.2
Mean Value		0.45	0.09	0.0	19.5	-136.2	-108.0

547

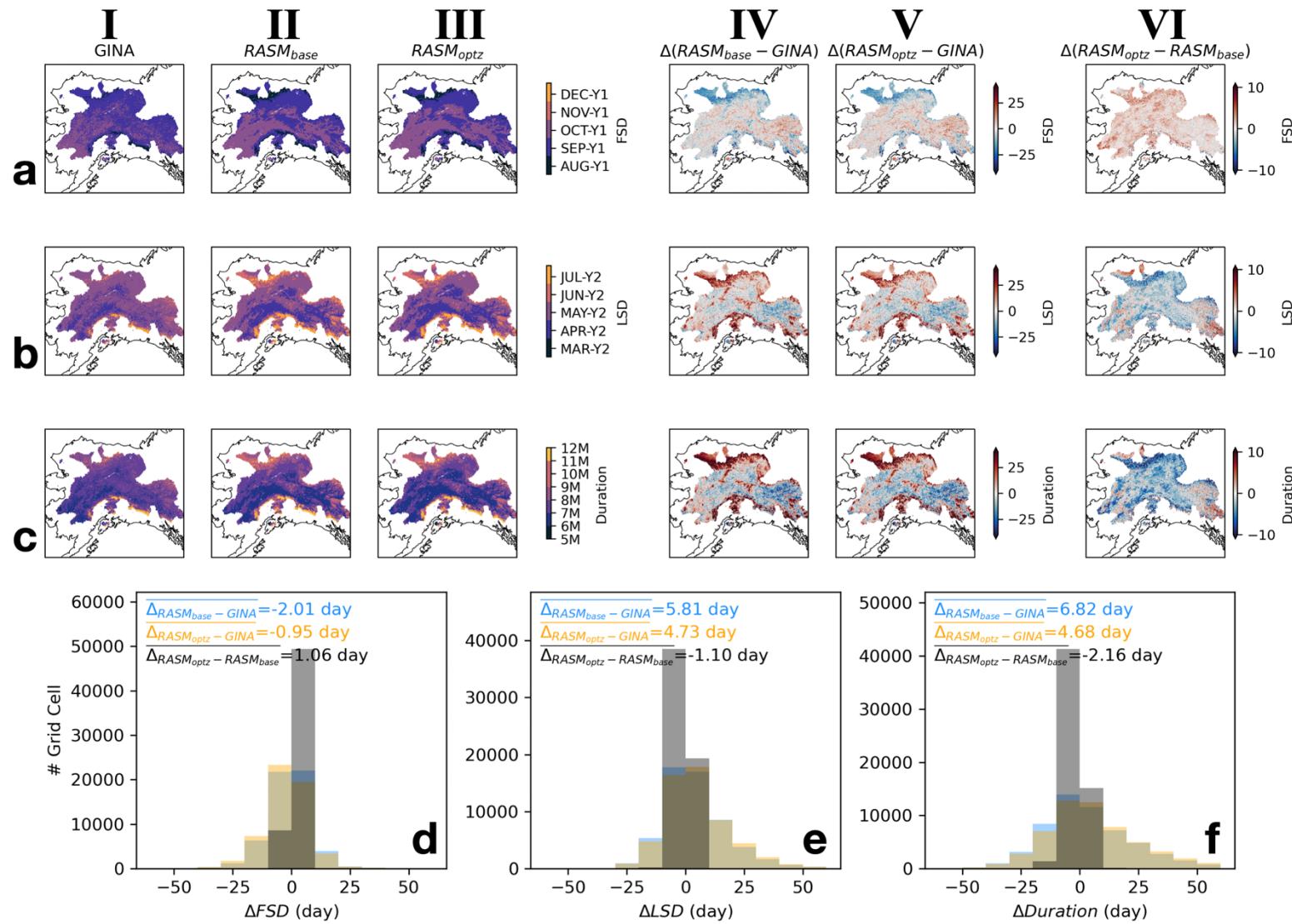
548 Table 2: Nash-Sutcliffe Coefficient, percent bias in FDC high-segment volume (%BiasFHV), and percent bias in FDC low-segment
 549 volume during summer season (%BiasFLV_{summer}) for all river basins (a 30-year evaluation from 1991 to 2020 WY). If models outperform
 550 benchmarks, the metric is highlighted in bold font.

551



552

553 Figure 9: A 30-year evaluation of model performance on snow simulations. Panel a shows the distribution of aggregated snow metrics
 554 (O_S) across all SNOTEL sites in Alaska, 10 of them used in optimization are highlighted in yellow stars and the remaining 30 are highlighted in
 555 yellow triangles in Panel b. Panels c, d, f summarize the distribution of individual snow metrics, including relative errors in annual peak SWE
 556 (rE_{pS}), snow melting rate (rE_{vSm}), and snow persistence time (rE_{tSp}). In Panels a, c, d, f, yellow corresponds with the model runs using optimized
 557 parameter values and blue corresponds with model runs using default parameter values.



560 *Figure 10: Evaluation of the first snow date (FSD, row **a**), last snow date (LSD, row **b**), and snow duration (row **c**) using the GINA
561 dataset (2001-2020). Columns **I**, **II**, **III** denote the GINA dataset, baseline simulation, and optimized simulations, columns **IV**, **V** denote the biases
562 in baseline and optimized simulations compared with the GINA dataset, and column **VI** denotes the difference between optimized and baseline
563 simulations. Figures *d*, *e*, *f* show the histogram of biases across all grid cells for FSD, LSD, and duration, respectively.*

564

565 Seasonal snowpack simulation performance was not greatly improved by optimization
566 across the final Pareto front, or with our final optimized parameter set by choice. The median
567 value of the aggregated snow metric (O_s , *Equation 10*) across all SNOTEL sites is 0.61 using
568 our optimized parameters and 0.55 using default parameters (Figure 9a). We also examined each
569 component of the snow metric. Out of 40 SNOTEL sites, optimization reduced relative biases in
570 peak SWE, snowmelt rate, and snow persistence duration at 18, 13, and 12 sites respectively
571 (Figures 9b, 9c, 9d). The median values for the relative bias in peak SWE, snowmelt rate, and
572 snow persistence duration are -0.07, -0.42, 0.13 using optimized parameters, and -0.05, -0.38,
573 and 0.10 using default parameters. In general, the snow simulation shows slightly worse
574 performance against the 40 SNOTEL observations used for validation. However, forty SNOTEL
575 sites may not be spatially representative for the entire study domain and about 60% of them are
576 located in the southern coastal and mountainous regions.

577 Given the limited number of SNOTEL sites and their uneven spatial distribution, we
578 further evaluate the snow simulation against the satellite-derived Geographic Information
579 Network of Alaska (GINA) dataset, available from 2001 to 2020. We specifically evaluated three
580 snow metrics, including the first snow date (FSD), last snow date (LSD), and snow duration. In
581 general, the default and optimized snow simulations show similar performance and spatial
582 patterns. For example, they show an earlier first snow date (Figures 10a.IV, 10a.V), a later last
583 snow date (Figures 10b.IV, 10b.V), and thus longer snow duration (Figures 10c.IV, 10c.V) in the
584 northern and southern mountainous regions. Additionally, the optimized snow simulation shows
585 slightly smaller biases than the baseline simulation. The biases in FSD, LSD, and duration are -
586 0.95 days, 4.73 days, 4.68 days for optimized snow simulations, and -2.01 days, 5.81 days, and
587 6.82 days for baseline snow simulations.

588 **5.3 Evaluation of CTSM actionability**

589 As discussed in section 5.2, the optimized model has improved out of sample overall flow
590 simulation. Specifically using NSE, our actionability flow metric, the mean NSE of daily flow
591 increases from 0.09 to 0.45 after optimization. Nine out of fifteen basins meet the 0.5 daily NSE
592 benchmark of flow performance after optimization while only two met the benchmark before
593 optimization. In addition, 13 out of 15 basins have improved high flow (a proxy for flooding)
594 magnitudes after optimization. The mean %BiasFHV across all basins is 0% and 19.5% for
595 optimized and baseline simulations respectively. However, the dry bias increases for low-flow
596 events. After optimization, the mean %BiasFLV_{summer} across all basins decreases from -108.0%
597 to -136.2%, and only 7 out of 15 basins show better performance in capturing low flow events.

598 Streamflow climate sensitivities for large river basins are improved while climate
599 sensitivities for small river basins are slightly degraded from the baseline simulation. Subplots in
600 Figure 11 are organized based on the confluence area, with the smallest basin on the top left and
601 the largest basin on the bottom right. When optimization improves the sensitivity to precipitation
602 change, we marked “P” in the upper right corner. Similarly, we marked “T” when optimization
603 improves the sensitivity to air temperature change. For the eight largest basins, at least one
604 climate sensitivity is improved and both climate sensitivities are improved in four of them.
605 However, the climate sensitivity for smaller river basins is generally slightly worse in the
606 optimized simulation (Fig. 11).

607 Interestingly, this implies that a better flow simulation does not necessarily lead to a
608 better streamflow climate sensitivity. For example, the KGE for Kuparuk increases from 0.40 to
609 0.68 after optimization (Fig. 8a) while both climate sensitivities are worse in the optimized
610 simulation (Fig. 11). Additionally, a worse flow simulation does not necessarily lead to worse

611 climate sensitivity. For example, Colville has slightly lower daily streamflow performance after
612 optimization with KGE decreasing from 0.80 to 0.75 (Fig. 8b) but its climate sensitivities to
613 precipitation and air temperature are both improved (Fig. 11). This feature may be related to the
614 fact that our optimization metrics are focused on daily flow and daily snowpack, leaving other
615 components of the water budget that could impact climate sensitivities less constrained. It has
616 been shown that metrics not closely related to the optimization metrics result in larger variations
617 in optimized model performance (Mendoza, Clark, Mizukami, et al., 2015). This suggests that
618 metrics more directly related to climate sensitivity should be included in optimization objective
619 functions.

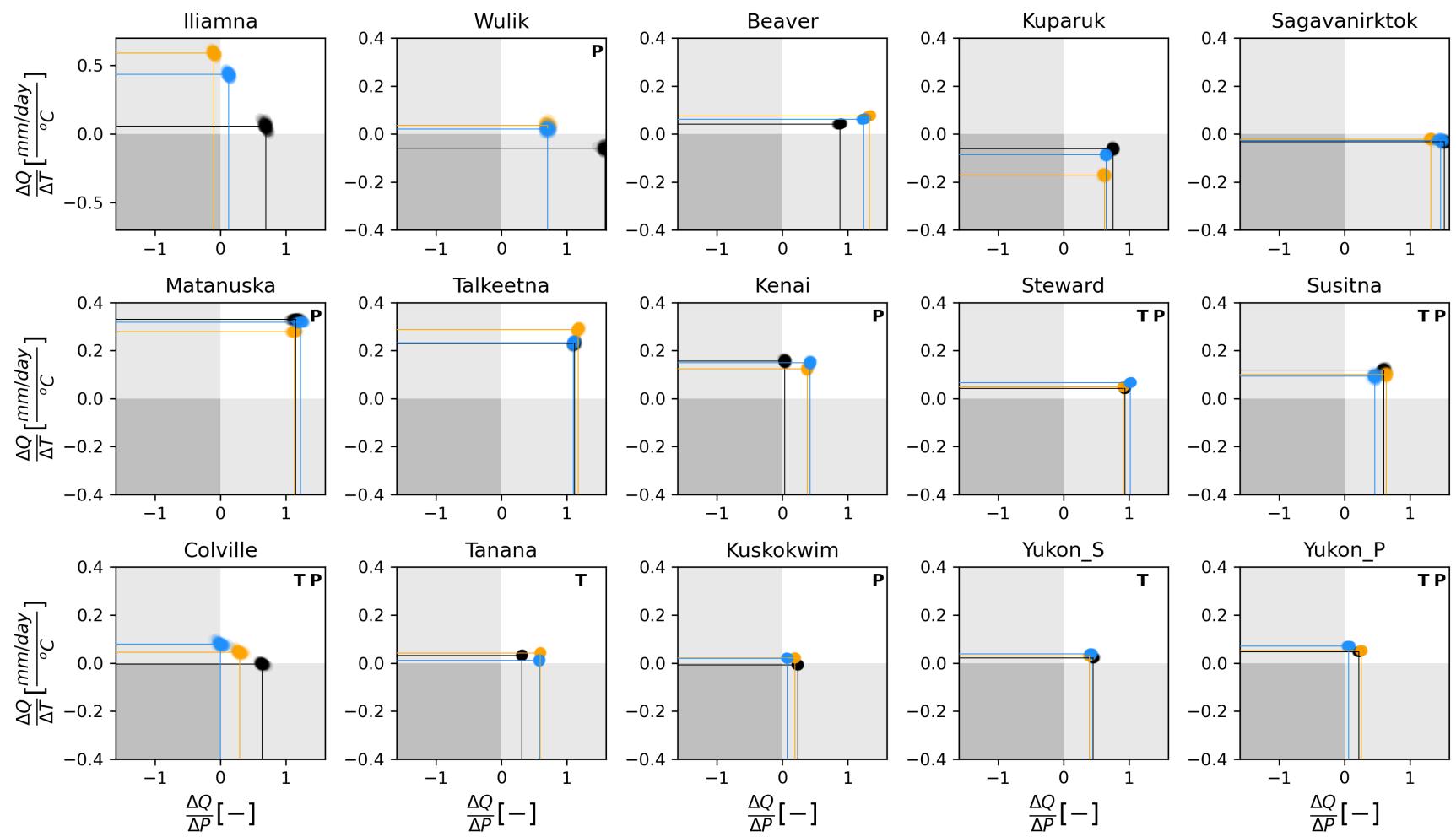
620 **5.4 Shapley decomposition**

621 In the northern region, large variations exist in the contribution of individual parameter
622 perturbations to simulation performance changes. For example, reducing the
623 *frac_sat_soil_dsl_init* value greatly improves flow performance, which contributes over 40% of
624 the KGE increment (Figure 12). Five other parameters made noticeable contributions to the
625 improving flow simulation, i.e., *e_ice*, *slopebeta*, *liq_canopy_storage_scalar*, *fff*, and
626 *medlynintercept*. These parameters belong to multiple categories, including acclimation
627 parameters relevant to photosynthesis, hydrology, and parameters affecting sensible, latent heat,
628 and momentum fluxes.

629 In the southern region, the variation of parameter contributions is much smaller than that
630 in the northern region. The perturbation of *n_melt_coef* contributes the most to the flow
631 improvement in the southern region, which only accounts for 30% of the KGE increment (Figure
632 12). Other than *n_melt_coef*, the top 6 parameters that contribute to the improved flow
633 simulation include *frac_sat_soil_dsl_init*, *liq_canopy_storage_scalar*,

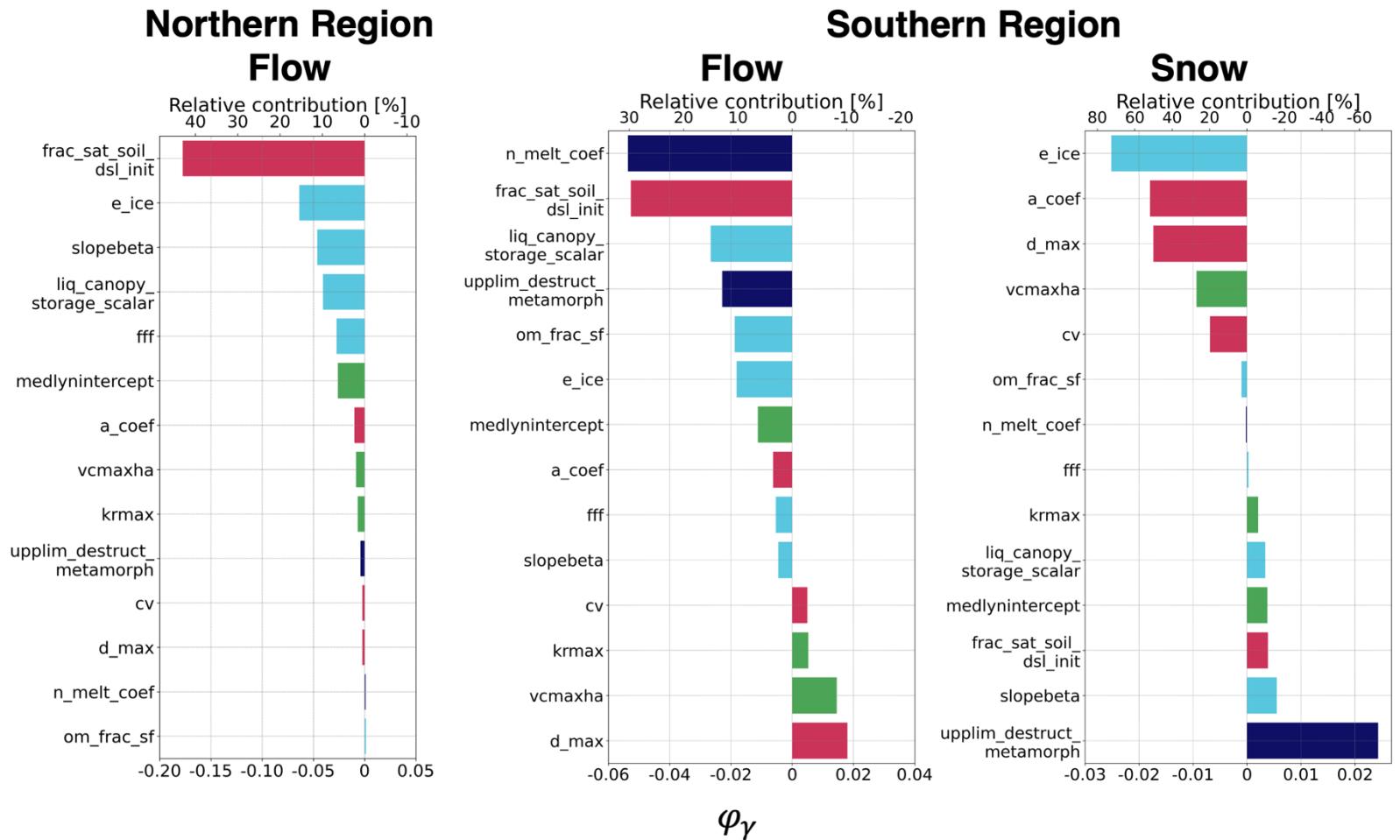
634 *upplim_destruct_metamorph, om_frac_sf, e_ice*. In addition, the same parameter perturbation
635 leads to opposite contributions in our flow and snow simulations. The perturbation of
636 *upplim_destrct_metamorph, frac_sat_soil_dsl_init, and liq_canopy_storage_scalar* improves
637 flow simulation while degrading snow simulation, while the perturbation of *d_max and vcmmaxha*
638 worsens flow simulation while improving snow simulation.

639



640

641 Figure 11: A climate sensitivity analysis. The x-axis denotes the rate of basin-averaged discharge change with precipitation change, and
 642 the y-axis denotes the rate of basin-averaged discharge change with air temperature change. Black, orange, and blue denote observation,
 643 optimized simulation, and baseline simulation respectively.



Hydrologic Parameters: Hydrology, Snow processes

Non-Hydrologic Parameters: Sensible, latent heat and momentum fluxes, Plant parameters

647 **6 Discussion and conclusions**

648 We have developed the first high-resolution application and optimization of CTSM for
649 Arctic hydrology. River flow simulations are significantly improved after optimization, while the
650 optimized snow simulation as compared to SNOTEL sites remains similar. The limited
651 improvement in snow simulations depends more on the meteorological forcing such as
652 precipitation than on model parameter choices (Günther et al., 2019; Raleigh et al., 2015). The
653 mean NSE of daily flow increases from 0.09 to 0.45 across 15 river basins for a 30-year
654 evaluation. For the Yukon River at Pilot Station, the USGS site with the largest confluence area
655 in Alaska, and the fourth-largest river in North America, the NSE of daily flow increased
656 from -0.55 to 0.50. In addition, the optimization is highly efficient given that the total area of the
657 four optimized river basins only occupies 2% of the confluence area at the Pilot Station. To our
658 knowledge, this study provides the most comprehensive evaluation and optimization of
659 hydrological simulations across Alaska and the Yukon River Basin, which can be used as a
660 benchmark for future Arctic hydrological modeling studies.

661 The optimization in this study generally improves the actionability of CTSM as defined
662 using our application specific metrics. The overall daily flow simulation using NSE (or KGE),
663 high flow, and streamflow climate sensitivities for large basins are all improved, but there is a
664 degradation for low flow simulation and climate sensitivities for small basins. Optimizing for
665 KGE has been shown to possibly negatively impact low flow simulation (Althoff & Rodrigues,
666 2021; Gupta et al., 2009). In addition, low flow events during the warm season are largely
667 affected by subsurface flow regimes and thermal characteristics of frozen soils and permafrost.
668 Sensitive parameters to these physical processes might not be identified in our existing
669 workflow, and CTSM does not include many inter-grid cell lateral flow processes, which is an

670 area of needed future research. In addition, we acknowledge that the model results have not yet
671 been used to inform decision-making but this will be further investigated as part of the Arctic
672 Rivers Project.

673 For expensive land models, we should use computationally frugal optimization methods.
674 As a state-of-the-science land model, CTSM is expensive to run, let alone optimize. To reduce
675 the computational cost, we selected representative basins and used surrogate modeling
676 optimization. Since the representative basins only occupy a small portion (2%) of the study
677 domain. Therefore, using the representative basin approach reduces the optimization cost to
678 roughly 2% of the full domain cost. Additional savings a realized using an efficient optimization
679 algorithm. For example, if we optimized CTSM using the widely used Shuffled Complex
680 Evaluation algorithm (Duan et al., 1994), it would take roughly four (4) times the computational
681 resources to reach a similar model performance (Wang et al., 2014), which is impractical for
682 complex LMs. The detailed computational cost is discussed in Supplemental Information Text
683 S1.

684 The optimization model framework is transferrable to other CTSM applications and can
685 be informative when developing optimization workflows for complex land models. The
686 transferability largely results from the global availability of the datasets used in this study, i.e.,
687 the ERA5 meteorological forcing data (European Centre for Medium-Range Weather Forecasts,
688 2019), soil texture from SoilGrid (Hengl et al., 2017), MERIT Hydro vector-based river network
689 (Yamazaki et al., 2019), and especially the CTSM PPE global parameter sensitivity analysis.
690 There also exist high-resolution WRF runs available for CONUS (Liu et al., 2017; Rasmussen &
691 Liu, 2017) and ongoing efforts for global coverage, which can be used for downscaling
692 meteorological forcing data outside Alaska. Correctly selecting sensitive parameters provides the

693 foundation for the success of parameter optimization. The PPE experiment can be extensively
694 used in selecting sensitive CTSM parameters not only to runoff and SWE but to any other
695 variable simulated by CTSM. Finally, hydrologic modeling of Arctic Alaska and the Yukon is
696 one of the most challenging regions in hydrological modeling due to the complicated land
697 surface processes that are important in this region. The improvements in hydrological simulation
698 achieved by our optimization framework in the challenging Arctic environment may motivate
699 exploration of the effectiveness of this optimization framework for regions outside the Arctic.

700 We also show that Arctic hydrology is not only influenced by hydrological parameters
701 but also parameters related to vegetation and thermal conductance. Previous hydrological studies
702 using CTSM mostly focused on hydrological parameters (Ren et al., 2016; Zhang et al., 2021). In
703 this study, out of the 14 optimized parameters, half are not directly related to hydrology and
704 snow processes (Table 1), which reveals the strong influence of non-hydrological
705 parameterization on Arctic hydrology. In the northern region, according to the Shapley
706 decomposition, perturbations of the 7 non-hydrological parameters contribute a total of 56.5% to
707 the KGE increment, and 2 out of the top 6 sensitive parameters are non-hydrological, including
708 *frac_sat_soil_dsl_init* and *medlynintercept* (Figure 10). In the southern region, the non-
709 hydrological parameter perturbations contribute 15.5% of the flow KGE increment and a
710 decrease of *frac_sat_soil_dsl_init* alone contributes 29.7% increment.

711 The Shapley decomposition analysis showed the different parameter contributions across
712 regions, reflecting the spatial heterogeneity of parameter sensitivities. The heterogeneity
713 manifests primarily in two ways. First, similar parameter perturbations lead to the opposite
714 direction of effects across regions. For example, an increase in activation energy for $V_{c,max}$ in
715 photosynthesis (*vcmaxha*), i.e., from the default value to the upper limit, contributes 2.0% of

716 flow KGE increment in the northern region but -8.2% in the south. Second, the opposite
717 parameter perturbation leads to the same direction of effects. A decrease of
718 *frac_sat_soil_dsl_init* in the southern region, i.e., 0.8 to 0.25, and an increase of
719 *frac_sat_soil_dsl_init* in the northern region, i.e., 0.8 to 1.78, both contribute positively to their
720 flow KGE increments, with the RC value of 29.7% and 43.0% respectively. Because CTSM by
721 default uses many spatially constant model parameters, we have modified CTSM to read in
722 distributed parameters when they are available. This effort provides an important example of
723 utilizing spatially distributed parameters, which should be informative to future CTSM
724 development for allowing this feature.

725 The parameter regionalization in this study is simple and effective, yet can still be
726 improved. Spatially distributed parameters in Colville and Wulik, i.e., basins overlapping both
727 southern and northern regions, generated better flow simulations than the parameters optimized
728 for either region. However, for Iliamna, a southern basin, its flow simulation using northern
729 optimized parameters is better than the one using southern optimized parameters, with daily flow
730 KGE of 0.63 and 0.51, respectively (Figure 8a). The similarities between Iliamna and northern
731 regions are neglected, likely because of either the oversimplified regionalization method or
732 compensating errors. In addition, the large discrepancies in optimized parameters across regions
733 only slightly affect the flow simulations in Tanana and Matanuska (Figures 7, 8a), which
734 indicates that the selected parameters may not be very sensitive for those out-of-sample basins.
735 Therefore, for future improvement of regional applications, it may be helpful to include more
736 representative basins for optimization and to implement a more sophisticated parameter
737 regionalization algorithm.

738 The surrogate model can only mimic the true response surface. For the southern region,
739 the root-mean-square error (RMSE) of the simulated flow and snow objectives are 0.04 and 0.03
740 respectively, and the RMSE of the simulated flow objective is 0.09 for the northern region. In
741 addition, the Shapley decomposition analysis is based upon the surrogate model, so the
742 contribution of each parameter perturbation reflects the simulated response surface. However, it
743 is infeasible to disentangle each parameter's contribution without a surrogate model. We would
744 need to run CTSM 16,384 (2^{14}) times for the Shapley decomposition while in this study we only
745 ran CTSM 500 times. Additional benefits from using surrogate models might arise by
746 incorporating other observational constraints, e.g., Active Layer Thickness, snow depth, or
747 evapotranspiration.

748 Finally, the development of the optimization framework and example application specific
749 evaluation lowers the barrier of applying complex land models in regional applications and
750 therefore enhances the actionability of the robust scientific tools. The authors hope this work lays
751 the foundation for a process-focused, stakeholder useful, high-resolution coupled land and
752 atmospheric modeling for cold regions both historically and under future projections to quantify
753 climate change impacts on inland freshwater systems.

754 **7 Acknowledgments**

755 This project was funded by NSF Navigating the New Arctic grant 1928078 and supported
756 by the National Center for Atmospheric Research, which is a major facility sponsored by the
757 National Science Foundation under Cooperative Agreement No. 1852977. We would like to
758 acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX)
759 provided by NCAR's Computational and Information Systems Laboratory, sponsored by the
760 National Science Foundation. We thank our Indigenous Advisory Council and numerous Tribal

761 and First Nation decision-makers for providing important insights to help inform the CTSM
762 configuration design. We also thank the Arctic Rivers Summit investigator team including
763 Nicole Herman-Mercer, Karen Cozzetto, and the Institute for Tribal Environmental Professionals
764 for facilitating the Indigenous Advisory Council and survey efforts. We thank Ryan Toohey for
765 informing our representative river basin selection. Furthermore, we greatly benefited from the
766 following open-source libraries to perform analyses presented in this study: NumPy (Van Der
767 Walt et al., 2011), pandas (McKinney, 2010), geopandas (Jordahl et al., 2020), xarray (Hoyer &
768 Hamman, 2017), matplotlib (Hunter, 2007), and cartopy (Met Office, 2015).

769 The optimization framework is available on Github
770 (https://github.com/NCAR/ctsm_optz). The CTSM version used in this study is available on
771 Github (https://github.com/YifanCheng/CTSM/tree/hh.ppe.n08_ctsm5.1.dev023).
772

773 **8 Reference**

- 774 Althoff, D., & Rodrigues, L. N. (2021). Goodness-of-fit criteria for hydrological models: Model
775 calibration and performance assessment. *Journal of Hydrology*, 600, 126674.
776 <https://doi.org/10.1016/J.JHYDROL.2021.126674>
- 777 Ang, B. W., Liu, F. L., & Chew, E. P. (2003). Perfect decomposition techniques in energy and
778 environmental analysis. *Energy Policy*, 31(14), 1561–1566. [https://doi.org/10.1016/S0301-4215\(02\)00206-9](https://doi.org/10.1016/S0301-4215(02)00206-9)
- 780 Batjes, N. H., Ribeiro, E., & Van Oostrum, A. (2020). Standardised soil profile data to support
781 global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data*, 12(1),
782 299–320. <https://doi.org/10.5194/ESSD-12-299-2020>
- 783 Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A.,
784 Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune,
785 J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., ... Wood, E. F. (2015). Hyper-
786 resolution global hydrological modelling: what is next? *Hydrological Processes*, 29(2),
787 310–320. <https://doi.org/10.1002/HYP.10391>
- 788 Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future
789 streamflow regime changes in the United States: Assessment using functional classification.
790 *Hydrology and Earth System Sciences*, 24(8), 3951–3966. <https://doi.org/10.5194/HESS-24-3951-2020>
- 792 Choi, H. I., & Liang, X. Z. (2010). Improved Terrestrial Hydrologic Representation in Mesoscale
793 Land Surface Models. *Journal of Hydrometeorology*, 11(3), 797–809.
794 <https://doi.org/10.1175/2010JHM1221.1>
- 795 Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P.,
796 Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., &
797 Zeng, X. (2015). Improving the representation of hydrologic processes in Earth System
798 Models. *Water Resources Research*, 51(8), 5929–5956.
799 <https://doi.org/10.1002/2015WR017096>
- 800 Clark, M. P., & Vrugt, J. A. (2006). Unraveling uncertainties in hydrologic model calibration:
801 Addressing the problem of compensatory parameters. *Geophysical Research Letters*, 33(6),
802 6406. <https://doi.org/10.1029/2005GL025604>
- 803 Computational and Information Systems Laboratory. (2019). *Cheyenne: HPE/SGI ICE XA
804 System (University Community Computing)*. Boulder CO: National Center for Atmospheric
805 Research. <https://doi.org/10.5065/D6RX99HX>
- 806 Cox, C. J., Stone, R. S., Douglas, D. C., Stanitski, D. M., Divoky, G. J., Dutton, G. S., Sweeney,
807 C., George, J. C., & Longenecker, D. U. (2017). Drivers and Environmental Responses to
808 the Changing Annual Snow Cycle of Northern Alaska. *Bulletin of the American
809 Meteorological Society*, 98(12), 2559–2577. <https://doi.org/10.1175/BAMS-D-16-0201.1>
- 810 Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning
811 approach to emulation and biophysical parameter estimation with the Community Land
812 Model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography*,
813 6(2), 223–244. <https://doi.org/10.5194/ascmo-6-223-2020>
- 814 Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective
815 genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–
816 197. <https://doi.org/10.1109/4235.996017>
- 817 Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global

- 818 optimization method for calibrating watershed models. *Journal of Hydrology*, 158(3–4),
819 265–284. [https://doi.org/10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4)
- 820 European Centre for Medium-Range Weather Forecasts. (2019). *ERA5 Reanalysis (0.25 Degree*
821 *Latitude-Longitude Grid)*. Research Data Archive at the National Center for Atmospheric
822 Research, Computational and Information Systems Laboratory.
823 <https://doi.org/10.5065/BH6N-5N20>
- 824 Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D.,
825 Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J.,
826 Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., ...
827 Yamazaki, D. (2019). Hillslope Hydrology in Global Change Research and Earth System
828 Modeling. *Water Resources Research*, 55(2), 1737–1772.
829 <https://doi.org/10.1029/2018WR023903>
- 830 Findlater, K., Webber, S., Kandlikar, M., & Donner, S. (2021). Climate services promise better
831 decisions but mainly focus on better data. *Nature Climate Change* 2021 11:9, 11(9), 731–
832 737. <https://doi.org/10.1038/s41558-021-01125-3>
- 833 Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aðalgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L.,
834 Golledge, N. R., Hemer, M., Kopp, R. E., Krinner, G., Mix, A., Notz, D., Nowicki, S.,
835 Nurhati, I. S., Ruiz, L., Sallée, J.-B., Slangen, A. B. A., & Yu, Y. (2021). Ocean,
836 Cryosphere and Sea Level Change. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L.
837 Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K.
838 Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, &
839 B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of*
840 *Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate*
841 *Change*. Cambridge University Press.
- 842 Gallant, A. L., Binnian, E. F., Omernik, J. M., & Shasby, M. B. (1995). *Ecoregions of Alaska*.
843 US Government Printing Office.
- 844 Giorgi, F. (2019). Thirty Years of Regional Climate Modeling: Where Are We and Where Are
845 We Going next? *Journal of Geophysical Research: Atmospheres*, 124(11), 5696–5723.
846 <https://doi.org/10.1029/2018JD030094>
- 847 Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., & Dai, Y. (2016). Multiobjective
848 adaptive surrogate modeling-based optimization for parameter estimation of large, complex
849 geophysical models. *Water Resources Research*, 52(3), 1984–2008.
850 <https://doi.org/10.1002/2015WR018230>
- 851 Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in Snowpack
852 Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing
853 Data Error on Point-Scale Energy Balance Snow Model Performance. *Water Resources*
854 *Research*, 55(4), 2779–2800. <https://doi.org/10.1029/2018WR023403>
- 855 Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean
856 squared error and NSE performance criteria: Implications for improving hydrological
857 modelling. *Journal of Hydrology*, 377(1–2), 80–91.
858 <https://doi.org/10.1016/J.JHYDROL.2009.08.003>
- 859 Hamman, J., Nijssen, B., Brunke, M., Cassano, J., Craig, A., DuVivier, A., Hughes, M.,
860 Lettenmaier, D. P., Maslowski, W., Osinski, R., Roberts, A., & Zeng, X. (2016). Land
861 Surface Climate in the Regional Arctic System Model. *Journal of Climate*, 29(18), 6543–
862 6562. <https://doi.org/10.1175/JCLI-D-15-0415.1>
- 863 Hamman, J., Nijssen, B., Roberts, A., Craig, A., Maslowski, W., & Osinski, R. (2017). The

- 864 coastal streamflow flux in the Regional Arctic System Model. *Journal of Geophysical*
865 *Research: Oceans*, 122(3), 1683–1701. <https://doi.org/10.1002/2016JC012323>
- 866 Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
867 Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,
868 Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I.,
869 Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on
870 machine learning. *PLOS ONE*, 12(2), e0169748.
871 <https://doi.org/10.1371/JOURNAL.PONE.0169748>
- 872 Herman-Mercer, N. M. (2021). *Guiding the Arctic Rivers Project Climate Model Development: Results from the Climate Information Survey*. https://www.colorado.edu/research/arctic-rivers/sites/default/files/attached-files/arp_modelsurveyresults_report_final.pdf
- 873 Hoyer, S., & Hamman, J. J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal*
874 *of Open Research Software*, 5(1). <https://doi.org/10.5334/JORS.148>
- 875 Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(03), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- 876 Jafarov, E. E., Marchenko, S. S., & Romanovsky, V. E. (2012). Numerical modeling of
877 permafrost dynamics in Alaska using a high spatial resolution dataset. *Cryosphere*, 6(3),
878 613–624. <https://doi.org/10.5194/TC-6-613-2012>
- 879 Jefferson, J. L., Gilbert, J. M., Constantine, P. G., & Maxwell, R. M. (2015). Active subspaces
880 for sensitivity analysis and dimension reduction of an integrated hydrologic model.
881 *Computers & Geosciences*, 83, 127–138. <https://doi.org/10.1016/J.CAGEO.2015.07.001>
- 882 Jiao, Y., Lei, H., Yang, D., Huang, M., Liu, D., & Yuan, X. (2017). Impact of vegetation
883 dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology
884 coupled model. *Journal of Hydrology*, 551, 116–131.
885 <https://doi.org/10.1016/J.JHYDROL.2017.05.060>
- 886 Jordahl, K., Bossche, J. Van den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J.,
887 Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran,
888 M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., ... Leblanc,
889 F. (2020). *geopandas/geopandas: v0.8.1*. <https://doi.org/10.5281/ZENODO.3946761>
- 890 Jorgenson, M. T., Shur, Y. L., & Pullman, E. R. (2006). Abrupt increase in permafrost
891 degradation in Arctic Alaska. *Geophysical Research Letters*, 33(2).
892 <https://doi.org/10.1029/2005GL024960>
- 893 Knoll, L. B., Sharma, S., Denfeld, B. A., Flaim, G., Hori, Y., Magnuson, J. J., Straile, D., &
894 Weyhenmeyer, G. A. (2019). Consequences of lake and river ice loss on cultural ecosystem
895 services. *Limnology and Oceanography Letters*, 4(5), 119–131.
896 <https://doi.org/10.1002/LOL2.10116>
- 897 Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G.,
898 Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzeck, E., Lawrence, P. J., Li,
899 F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., ... Zeng,
900 X. (2019). The Community Land Model Version 5: Description of New Features,
901 Benchmarking, and Impact of Forcing Uncertainty. *Journal of Advances in Modeling Earth*
902 *Systems*, 11(12), 4245–4287. <https://doi.org/10.1029/2018MS001583>
- 903 Lawrence, D. M., & Slater, A. G. (2005). A projection of severe near-surface permafrost
904 degradation during the 21st century. *Geophysical Research Letters*, 32(24), 1–5.
905 <https://doi.org/10.1029/2005GL025080>
- 906 Lawrence, D. M., Slater, A. G., Romanovsky, V. E., & Nicolsky, D. J. (2008). Sensitivity of a
907

- model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter. *Journal of Geophysical Research: Earth Surface*, 113(F2). <https://doi.org/10.1029/2007JF000883>
- Lawrence, D. M., Slater, A. G., & Swenson, S. C. (2012). Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4. *Journal of Climate*, 25(7), 2207–2225. <https://doi.org/10.1175/JCLI-D-11-00334.1>
- Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., & Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change* 2019 9:12, 9(12), 926–933. <https://doi.org/10.1038/s41558-019-0639-x>
- Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G., & Yates, D. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, 49(1–2), 71–95. <https://doi.org/10.1007/S00382-016-3327-9/FIGURES/16>
- Maraun, D., & Widmann, M. (2018). *Statistical downscaling and bias correction for climate research*. Cambridge University Press.
- McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1), 55–61. <https://doi.org/10.1080/00401706.2000.10485979>
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *THE 9th PYTHON IN SCIENCE CONFERENCE*, 56–61. <https://doi.org/10.25080/MAJORA-92BF1922-00A>
- Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, 51(1), 716–728. <https://doi.org/10.1002/2014WR015820>
- Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., & Arnold, J. R. (2015). Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts. *Journal of Hydrometeorology*, 16(2), 762–780. <https://doi.org/10.1175/JHM-D-14-0104.1>
- Met Office. (2015). *Cartopy: a cartographic python library with a matplotlib interface*.
- Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., Beck, H. E., & Yamazaki, D. (2021). A Vector-Based River Routing Model for Earth System Models: Parallelization and Global Applications. *Journal of Advances in Modeling Earth Systems*, 13(6), e2020MS002434. <https://doi.org/10.1029/2020MS002434>
- Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., & Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, 53(9), 8020–8040. <https://doi.org/10.1002/2017WR020401>
- Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R. J., Markstrom, S. L., Hay, L. E., Woods, R., Arnold, J. R., & Brekke, L. D. (2016). mizuRoute version 1: a river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, 9(6), 2223–2238. <https://doi.org/10.5194/gmd-9-2223-2016>
- Monaghan, A. J., Clark, M. P., Barlage, M. P., Newman, A. J., Xue, L., Arnold, J. R., & Rasmussen, R. M. (2018). High-Resolution Historical Climate Simulations over Alaska.

- 956 *Journal of Applied Meteorology and Climatology*, 57(3), 709–731.
957 <https://doi.org/10.1175/JAMC-D-17-0161.1>
- 958 Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and Water Quality
959 Models: Performance Measures and Evaluation Criteria. *Transactions of the ASABE*, 58(6),
960 1763–1785. <https://doi.org/10.13031/TRANS.58.10715>
- 961 Newman, A. J., Monaghan, A. J., Clark, M. P., Ikeda, K., Xue, L., Gutmann, E. D., & Arnold, J.
962 R. (2021). Hydroclimatic changes in Alaska portrayed by a high-resolution regional climate
963 simulation. *Climatic Change*, 164(1–2), 1–21. <https://doi.org/10.1007/S10584-021-02956-X/FIGURES/12>
- 964 Ohlberger, J., Buehrens, T. W., Brenkman, S. J., Crain, P., Quinn, T. P., & Hilborn, R. (2018).
965 Effects of past and projected river discharge variability on freshwater production in an
966 anadromous fish. *Freshwater Biology*, 63(4), 331–340. <https://doi.org/10.1111/FWB.13070>
- 967 Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., Levis, S.,
968 Swenson, S. C., Thornton, E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C.
969 L., Lamarque, J., Niu, G., Qian, T., Running, S., Sakaguchi, K., ... Zeng, X. (2010).
970 *Technical Description of version 4 . 0 of the Community Land Model (CLM)*. April.
- 971 Osterkamp, T. E., & Romanovsky, V. E. (1999). Evidence for warming and thawing of
972 discontinuous permafrost in Alaska. *PERMAFROST AND PERIGLACIAL PROCESSES*,
973 10, 17–37. <https://onlinelibrary-wiley-com.cuucar.idm.oclc.org/doi/epdf/10.1002/%28SICI%291099-1530%28199901/03%2910%3A1%3C17%3A%3AAID-PPP303%3E3.0.CO%3B2-4>
- 974 Pavelsky, T. M., & Zarnetske, J. P. (2017). Rapid decline in river icings detected in Arctic
975 Alaska: Implications for a changing hydrologic cycle and river ecosystems. *Geophysical
976 Research Letters*, 44(7), 3228–3235. <https://doi.org/10.1002/2016GL072397>
- 977 Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., Williams, Z.,
978 Brunke, M. A., & Gochis, D. (2016). A gridded global data set of soil, intact regolith, and
979 sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of
980 Advances in Modeling Earth Systems*, 8(1), 41–65. <https://doi.org/10.1002/2015MS000526>
- 981 Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., Clark, M. P.,
982 & Samaniego, L. (2019). Diagnostic Evaluation of Large-Domain Hydrologic Models
983 Calibrated Across the Contiguous United States. *Journal of Geophysical Research:
984 Atmospheres*, 124(24), 13991–14007. <https://doi.org/10.1029/2019JD030767>
- 985 Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error
986 characteristics on physically based snow simulations within a global sensitivity analysis
987 framework. *Hydrology and Earth System Sciences*, 19(7), 3153–3179.
988 <https://doi.org/10.5194/HESS-19-3153-2015>
- 989 Rasmussen, R., & Liu, C. (2017). *High Resolution WRF Simulations of the Current and Future
990 Climate of North America*. Research Data Archive at the National Center for Atmospheric
991 Research, Computational and Information Systems Laboratory.
992 <https://doi.org/10.5065/D6V40SXP>
- 993 Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M.,
994 Dudhia, J., Yu, W., Miller, K., Arsenault, K., Grubišić, V., Thompson, G., & Gutmann, E.
995 (2011). High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over
996 Colorado: A Process Study of Current and Warmer Climate. *Journal of Climate*, 24(12),
997 3015–3048. <https://doi.org/10.1175/2010JCLI3985.1>
- 998 Ren, H., Hou, Z., Huang, M., Bao, J., Sun, Y., Tesfa, T., & Ruby Leung, L. (2016).
- 999
1000

- 1002 Classification of hydrological parameter sensitivity and evaluation of parameter
1003 transferability across 431 US MOPEX basins. *Journal of Hydrology*, 536, 92–108.
1004 <https://doi.org/10.1016/J.JHYDROL.2016.02.042>
- 1005 Roth, A. E. (1988). *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge
1006 University Press. <https://doi.org/10.1017/cbo9780511528446>
- 1007 Saito, K., Machiya, H., Iwahana, G., Ohno, H., & Yokohata, T. (2020). Mapping simulated
1008 circum-Arctic organic carbon, ground ice, and vulnerability of ice-rich permafrost to
1009 degradation. *Progress in Earth and Planetary Science*, 7(1), 1–15.
1010 <https://doi.org/10.1186/S40645-020-00345-Z/FIGURES/7>
- 1011 Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-
1012 based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), 5523.
1013 <https://doi.org/10.1029/2008WR007327>
- 1014 Sankarasubramanian, A., Vogel, R. M., & Limbrunner, J. F. (2001). Climate elasticity of
1015 streamflow in the United States. *Water Resources Research*, 37(6), 1771–1781.
1016 <https://doi.org/10.1029/2000WR900330>
- 1017 Sharma, S., Blagrave, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., Magee, M. R.,
1018 Straile, D., Weyhenmeyer, G. A., Winslow, L., & Woolway, R. I. (2019). Widespread loss
1019 of lake ice around the Northern Hemisphere in a warming world. *Nature Climate Change*
1020 2019 9:3, 9(3), 227–231. <https://doi.org/10.1038/s41558-018-0393-5>
- 1021 Singh, R. S., Reager, J. T., Miller, N. L., & Famiglietti, J. S. (2015). Toward hyper-resolution
1022 land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0
1023 simulations over the Southwestern U.S. *Water Resources Research*, 51(4), 2648–2667.
1024 <https://doi.org/10.1002/2014WR015686>
- 1025 Srivastava, V., Graham, W., Muñoz-Carpena, R., & Maxwell, R. M. (2014). Insights on geologic
1026 and vegetative controls over hydrologic behavior of a large complex basin – Global
1027 Sensitivity Analysis of an integrated parallel hydrologic model. *Journal of Hydrology*,
1028 519(PB), 2238–2257. <https://doi.org/10.1016/J.JHYDROL.2014.10.020>
- 1029 Stone, R. S., Dutton, E. G., Harris, J. M., & Longenecker, D. (2002). Earlier spring snowmelt in
1030 northern Alaska as an indicator of climate change. *Journal of Geophysical Research: Atmospheres*, 107(D10), ACL 10-1. <https://doi.org/10.1029/2000JD000286>
- 1031 Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., & Perket, J. (2019). Representing
1032 Intrahillslope Lateral Subsurface Flow in the Community Land Model. *Journal of Advances*
1033 in Modeling Earth Systems, 11(12), 4044–4065. <https://doi.org/10.1029/2019MS001833>
- 1034 Swenson, S. C., Lawrence, D. M., & Lee, H. (2012). Improved simulation of the terrestrial
1035 hydrological cycle in permafrost regions by the Community Land Model. *Journal of*
1036 *Advances in Modeling Earth Systems*, 4(8), 8002. <https://doi.org/10.1029/2012MS000165>
- 1037 Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for
1038 efficient numerical computation. *Computing in Science and Engineering*, 13(2), 22–30.
1039 <https://doi.org/10.1109/MCSE.2011.37>
- 1040 van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M.,
1041 Slater, A. G., & van den Broeke, M. R. (2017). Improving the Representation of Polar Snow
1042 and Firn in the Community Earth System Model. *Journal of Advances in Modeling Earth*
1043 *Systems*, 9(7), 2583–2600. <https://doi.org/10.1002/2017MS000988>
- 1044 Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved
1045 treatment of uncertainty in hydrologic modeling: Combining the strengths of global
1046 optimization and data assimilation. *Water Resources Research*, 41(1), 1–17.
1047

- 1048 https://doi.org/10.1029/2004WR003059
1049 Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive
1050 surrogate modeling based optimization with two benchmark problems. *Environmental*
1051 *Modelling & Software*, 60, 167–179. https://doi.org/10.1016/J.ENVSOFT.2014.05.026
1052 Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic Implications
1053 of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Climatic*
1054 *Change 2004* 62:1, 62(1), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9E
1055 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019).
1056 MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography
1057 Dataset. *Water Resources Research*, 55(6), 5053–5073.
1058 https://doi.org/10.1029/2019WR024873
1059 Yang, D., & Kane, D. L. (Eds.). (2020). *Arctic Hydrology, Permafrost and Ecosystems*. Springer
1060 Nature.
1061 Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to
1062 model evaluation: Application to the NWS distributed hydrologic model. *Water Resources*
1063 *Research*, 44(9), 9417. https://doi.org/10.1029/2007WR006716
1064 Yu, S., Wei, Y. M., & Wang, K. (2014). Provincial allocation of carbon emission reduction
1065 targets in China: An approach based on improved fuzzy cluster and Shapley value
1066 decomposition. *Energy Policy*, 66, 630–644. https://doi.org/10.1016/j.enpol.2013.11.025
1067 Zhang, C., Abbaszadeh, P., Xu, L., Moradkhani, H., Duan, Q., & Gong, W. (2021). A Combined
1068 Optimization-Assimilation Framework to Enhance the Predictive Skill of Community Land
1069 Model. *Water Resources Research*, e2021WR029879.
1070 https://doi.org/10.1029/2021WR029879
1071