Moving land models towards actionable science: A novel application of the Community Terrestrial Systems Model across Alaska and the Yukon River Basin

Yifan Cheng[1, *], Andrew Newman[1], Keith N. Musselman[2], Sean Swenson[1], David Lawrence[1], Joseph Hamman[1,3], Katherine Dagon[1], Daniel Kennedy[1]

1. National Center for Atmospheric Research, Boulder, CO 80301, USA
2. Institute of Arctic and Alpine Research, University of Colorado, Boulder CO, 80309, USA
3. CarbonPlan, San Francisco CA, 94115, USA
* Corresponding author: Yifan Cheng (yifanc@ucar.edu)

Key points (140 characters):

- This study provides a high-resolution, high-fidelity Arctic hydrologic simulation and evaluation for 15 major Alaskan river basins
- This work moves CTSM towards a more actionable Earth Science paradigm through optimization for hydrology-related applications
- The optimization framework developed in this study is transferable to other CTSM applications and is informative for land models generally

Abstract (250 words)

The Arctic hydrological system is an interconnected system that is experiencing rapid change. It is comprised of permafrost, snow, glacier, frozen soils, and inland river systems. Permafrost degradation, trends towards earlier snow melt, a lengthening snow-free season, soil ice melt, and warming frozen soils all challenge hydrologic simulation under climate change in the Arctic. In this study, we provide an improved representation of the hydrologic cycle across a regional Arctic domain using a generalizable optimization methodology and workflow for the community. We applied the Community Terrestrial Systems Model (CTSM) across the US state of Alaska and the Yukon River Basin at 4-km spatial resolution. We highlight several potentially useful high-resolution CTSM configuration changes. Additionally, we performed a multi-objective optimization using snow and river flow metrics within an adaptive surrogate-based model optimization scheme. Four representative river basins across our study domain were selected for optimization based on observed streamflow and snow water equivalent observations at ten SNOTEL sites. Fourteen sensitive parameters were identified for optimization with half of them not directly related to hydrology or snow processes. Across fifteen out-of-sample river basins, thirteen had improved flow simulations after optimization and the median Kling-Gupta Efficiency of daily flow increased from 0.40 to 0.63. In addition, we adapted the Shapley Decomposition to disentangle each parameter's contribution to streamflow performance changes, with the seven non-hydrological parameters providing a non-negligible contribution to performance gains. The snow simulation had limited improvement, likely because snow simulation is influenced more by meteorological forcing than model parameter choices.

1 Introduction

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

Hydrologic modeling of Arctic rivers is challenging due to the aforementioned complex and interacting terrestrial processes. However, recent developments in advanced land models (LMs) are now enabling us to simulate complex land surface processes and their subsequent impacts on hydrology (Clark et al., 2015; Hamman et al., 2016). Additionally, advances in computationally frugal optimization methods and improvements in LM agility (i.e., the capability to adjust model equations and parameters to faithfully represent observed processes; Mendoza et al., 2015), allow for parameter sensitivity and application-oriented optimization studies of these advanced LMs.

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

Earth System models are being applied at an increasingly higher resolution to improve accuracy and increase actionability (Bierkens et al., 2015; Singh et al., 2015). Higher-resolution models can more faithfully represent varied and complex topography, and thus often more realistically simulate seasonal snow, orographic precipitation patterns, and potentially heterogeneous permafrost (Newman et al., 2021; Rasmussen et al., 2011). A more realistic physical representation of the landscape and land-atmosphere interactions increases the credibility of a model in regional applications, which can help to build stakeholder trust in model results and can help to facilitate a move toward a more actionable Earth Science paradigm (Giorgi, 2019).

As part of the Arctic Rivers Project, we are guided by an 11-member Indigenous Advisory Council. The Council helps project investigators make decisions about research design,

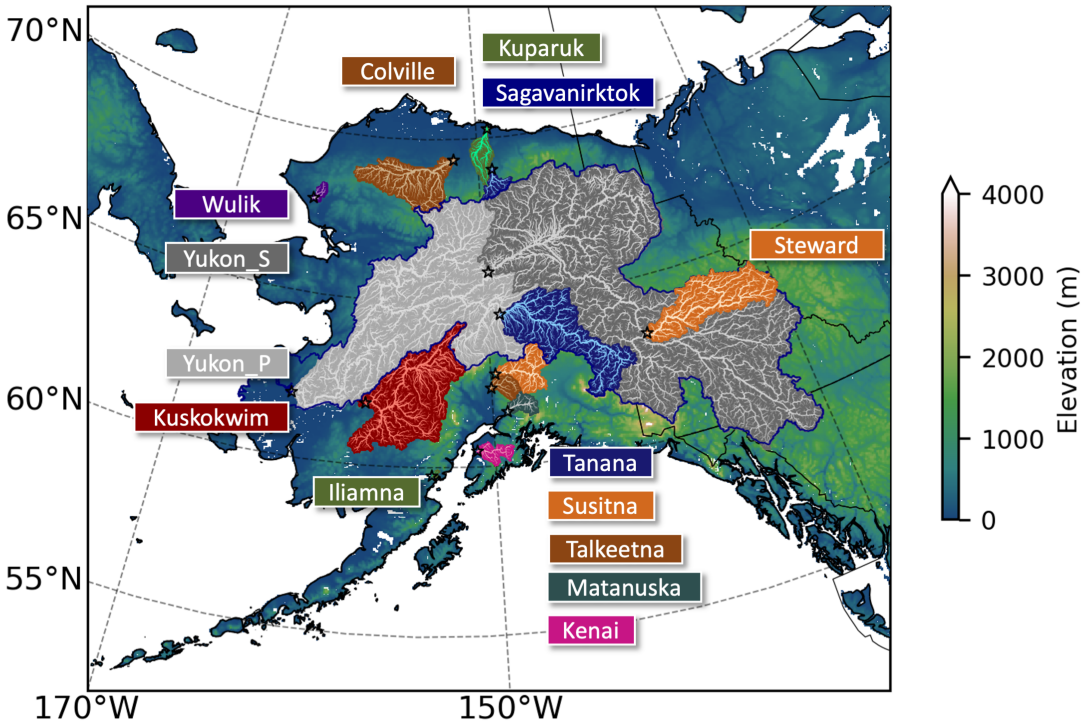
86 analysis, and deliverables to ensure that Indigenous knowledge and perspectives are included,
87 valued, and protected, and that the project benefits the Indigenous peoples the project is intended
88 to serve. We co-developed a climate information survey (Herman-Mercer, 2021) completed by
89 23 Tribal Councils, Traditional Councils, First Nation Governments, City Councils, and
90 Regional Indigenous Organizations. From the survey results, there was consensus that the most
91 useful information for Indigenous decision-makers would be sub-watershed scale (or high-
92 resolution) streamflow and other land-surface and sensible (i.e., relatable) weather variables such
93 as 2-m air temperature and precipitation. Configuring a high-resolution model is multi-faceted,
94 which not only means a finer grid but also requires corresponding meteorological forcing data
95 and land surface datasets that are often more difficult to work with if they even exist. In addition,
96 high-resolution LMs require substantially more computational resources, which decreases their
97 potential to be optimized.

98 Even with improved process representation and hydrologically focused model
99 configurations (Choi & Liang, 2010; Jiao et al., 2017; Singh et al., 2015), optimization of
100 parameters within complex LMs is often necessary because of uncertainty in model parameters,
101 model structural errors, and missing process representations (Lehner et al., 2019; Mendoza et al.,
102 2015; Sankarasubramanian et al., 2001). Optimization of complex LMs like CTSM is a
103 substantial challenge given the high computational costs, and this challenge limits the usage of
104 CTSM and similar models in large-scale hydrological or other stakeholder specific applications.
105 Although several sensitivity analyses have been conducted to examine the hydrological
106 responses to CTSM model parameters (Jefferson et al., 2015; Ren et al., 2016; Srivastava et al.,
107 2014), their limited spatial coverage or number of parameters cast few insights on sensitive
108 parameters to Arctic terrestrial hydrology.

In this study, we provide: 1) a methodology for efficient optimization of CTSM for regional to large-scale hydrologically focused applications; 2) a high-resolution Arctic CTSM configuration focused on improved hydrologic simulation fidelity; and 3) tools available to the scientific community to apply our methodology to other regions and applications. Additionally, this study lays the foundation for knowledge co-production research with Indigenous communities for a range of topics, including improving our understanding of climate-induced impacts on the rivers and fishes, and communities necessary to inform adaptation efforts. We aim to move CTSM and other complex, process-rich land models toward a more actionable Earth Science paradigm (Findlater et al., 2021) through this regional hydrologic application.

2 Study domain

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



131

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

134 # 3 Baseline CTSM configuration

135 To configure a high-resolution CTSM application, we downscaled the available coarse
136 meteorological forcing data (Section 3.1) and used finer-than-default soil texture data (Section
137 3.2). In addition, we used the hillslope hydrology scheme in CTSM to account for the remaining
138 sub-grid topographic variability (Section 3.3) and used the satellite phenology CTSM
139 configuration with default model parameter values. We used the vector-based mizuRoute to
140 route runoff (Mizukami et al., 2016, 2021) and we extracted the river network from a high-
141 resolution global hydrography map, i.e., MERIT Hydro (Yamazaki et al., 2019). This constitutes
142 our baseline CTSM model (Figure 2).

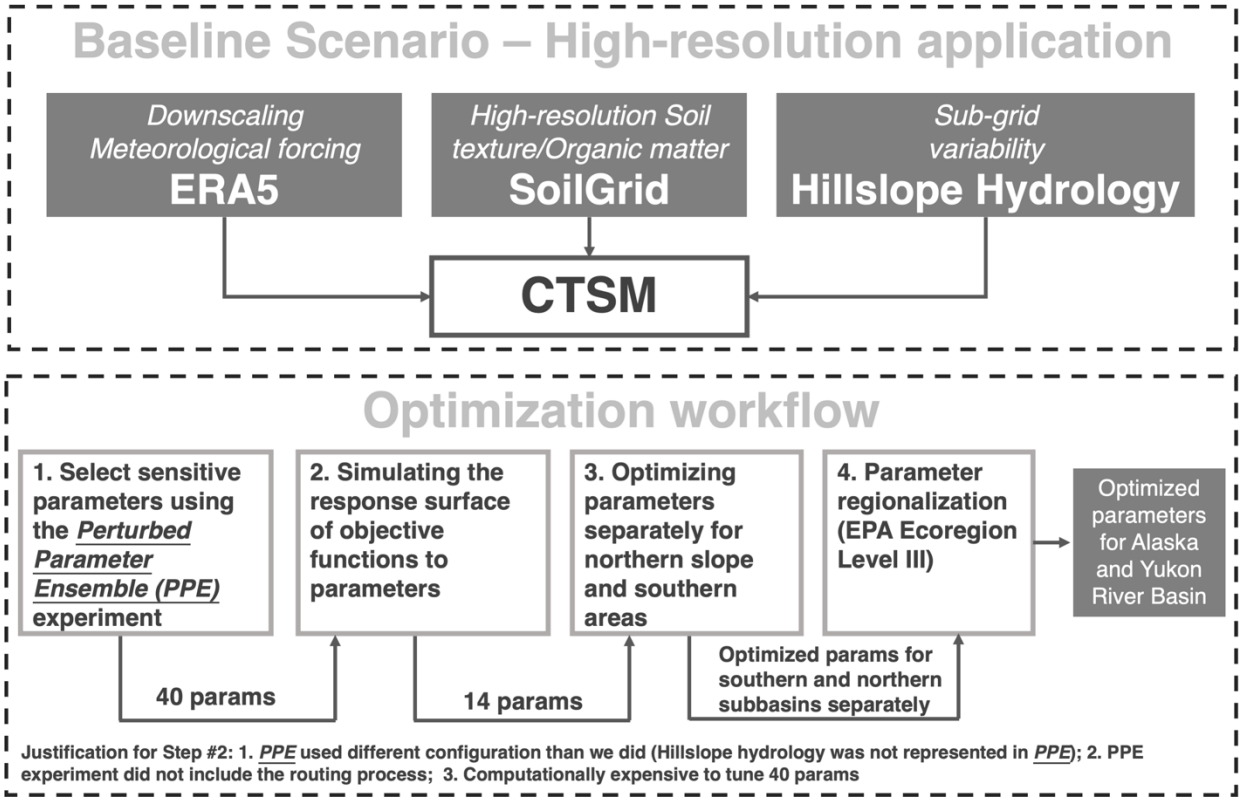


Figure 2: CTSM baseline scenario and workflow for optimization

3.1 Downscaling meteorological forcing data - ERA5

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

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

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

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

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

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

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

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

3.2 Soil texture and organic matter – SoilGrids

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

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

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

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

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

Percentages of sand and clay, PCT_{SAND} and PCT_{CLAY}, were calculated based on the sand, silt, and clay contents (γ_{sand}, γ_{silt}, γ_{clay}, unit: g/kg). Organic matter density (ρ_{OM}, unit: kg/m^3) was calculated using the soil organic carbon (γ_C, unit: dg/kg) and bulk density (ρ_{bulk}, cg/cm^3) with the assumption of carbon content 0.58gC per gOM.

3.3 Sub-grid variability – Hillslope Hydrology

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

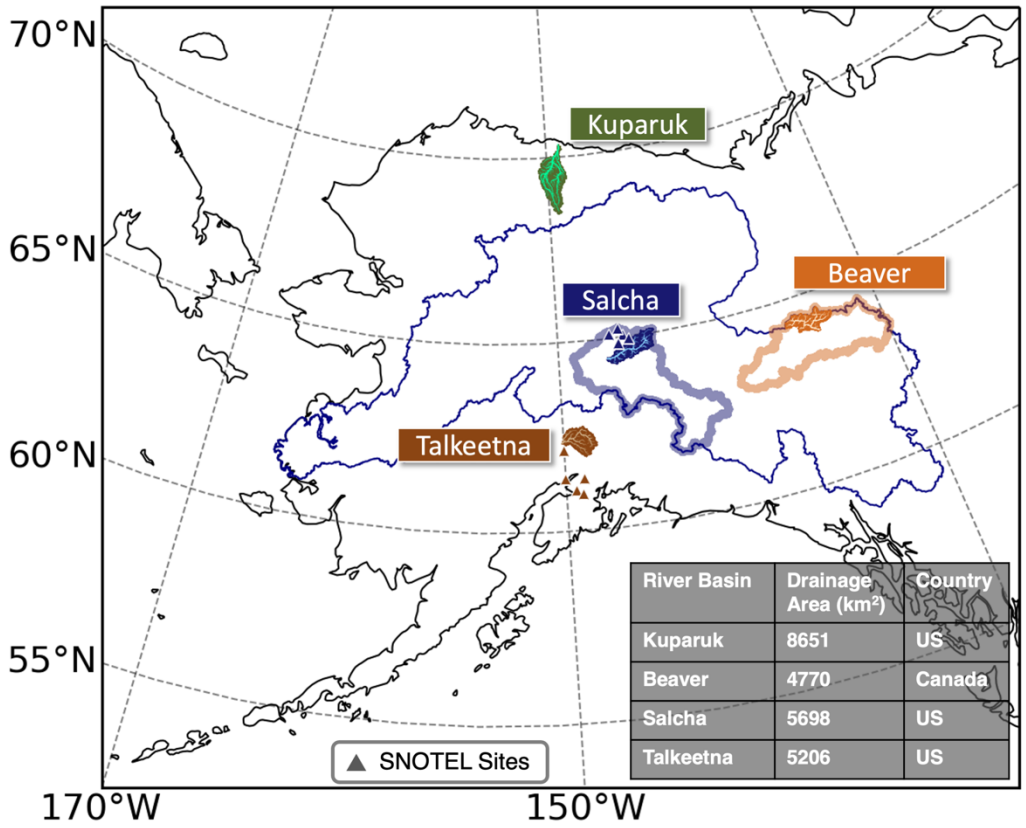


Figure 3: Selected representative medium-sized basins for parameter estimation. Triangles denote SNOTEL sites with snow observations. Salcha River Basin is a subbasin of the Tanana River Basin (thick blue line) and Beaver River Basin is a subbasin of the Steward River Basin (thick orange line).

4 Optimization framework

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

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

4.1 Multi-objective functions for flow and snow conditions

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

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

$$O_Q = 1 - KGE \tag{9}$$

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

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

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

226 The aggregated snow metric O_S is unitless.

227 **4.2 Parameter sensitivity**

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

235 We adopted a two-step method to select sensitive parameters for optimization. First, we
236 selected the top 40 parameters that exert a strong influence on Arctic hydrology from over 200
237 parameters that were varied within the PPE. Because the CTSM configuration for the PPE did
238 not utilize the hillslope hydrology nor did it include river routing, an additional filtering step was
239 performed. While moving from over 200 to 40 parameters is a substantial simplification of the
240 potential optimization space, it is still computationally expensive to tune 40 parameters within
241 CTSM. Therefore, we further identified the most sensitive parameters by training a surrogate
242 model to simulate the response surface of objective functions to each parameter. The top 14 out
243 of the 40 pre-screened parameters were selected for optimization. Both steps are explained in
244 detail as follows.

- ***Step 1***: We used the PPE one-at-a-time experiment to select which parameters exert the most control on total runoff (QRUNOFF) and snow water equivalent (H2OSNO). To constrain computational costs, the PPE was run at 400 grid cells globally to represent the parameter sensitivities at different land cover types and climatologies. Seven of those grid cells fall in our study domain and we used the mean response across the seven grid cells to evaluate parameter sensitivity for Arctic hydrology. For QRUNOFF, we evaluated the mean, seasonality, and amplitude; for SWE, we evaluated the snow persistence duration, maximum monthly SWE, and snowmelt rate, which leads to a total of six variable-metric combinations. For each combination, we selected the top 15 most sensitive parameters and assigned a higher score to more sensitive parameters (e.g., 15 points to the most sensitive parameter, 1 point to the least sensitive parameter). As a pre-screen step, we would like to include as many sensitive parameters as possible within our capacity to handle complexity and we selected 15 after experimenting with different numbers. The scores for each parameter were summarized across all six variable-metric combinations and the total score represents the general uncertainty of the parameters to runoff and snow conditions in our study domain. A total of 40 parameters across all variable-metric combinations were pre-screened as candidate parameters and would be further selected in Step 2.

264 Table 1: Summary of 14 parameters selected for optimization, their categories, relevant physical processes, ranking based on

265 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 $V_{c,max}$ (maximum rate of Rubisco-mediated carboxylation)	14	72000	[20000, 250000]	20364	160155
Hydrology	om_frac_sf	Scalar adjustment for organic matter fraction	2	100%* DV	[25%,200%]* *DV*	26.876%* DV	95.786%* DV
	slopebeta	Surface water storage	11	-3	[-10,-0.5]	-4.162	-6.713
	fff	Decay factor for fractional saturated area	6	0.5	[0.01,10]	0.298	8.553
	e_ice	Ice impedance factor	6	6	[1,8]	7.016	1
Plant hydraulics	krmax‡	Root segment maximum conductance	5	1.223×10^{-9}	$[5.827\times10^{-11}, 6.896\times10^{-9}]$	2.046×10^{-9}	2.735×10^{-9}
Sensible, latent heat and momentum fluxes	d_max	Heat and momentum flux for non-vegetated surface, dry surface layer (DSL) thickness	8	15	[5,100]	27.744	8.957
	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]	0.25	0.628
	cv	Turbulent transfer coefficient between canopy surface and canopy air	11	0.01	[0.0025,0.04]	0.0165	0.04

	a_coef	Drag coefficient under less dense canopy	8	0.13	[0.05,0.15]	5.003×10^{-2}	0.121
Snow processes	upplim_destruct_metamorph	Upper limit for snow densification through destructive metamorphism	1	175	[10,500]	10	500
	n_melt_coef	Parameter controlling shape of snow-covered area	3	200	[25,600]	93.702	526.216
	snw_rds_refrz	Effective radius of re-frozen snow	11	1000	[500,1500]	526.434	500
Stomatal resistance and photosynthesis	Medlynslope‡	Medlyn slope of stomatal conductance-photosynthesis relationship	8	4.954	[3.173, 6.934]	4.287	3.196

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

DV is short for default values.

Hydrologic parameters are highlighted using blue (Hydrology) and navy (Snow processes) and non-hydrologic parameters are highlighted using red (Sensible, latent heat, and momentum fluxes) and green (plant parameters).

- ***Step 2***: To select the most sensitive parameters, we simulated the response of flow and snow objectives to the CTSM model parameters using surrogate models. For each river basin, we trained one surrogate model from 200 samples generated using the LHS method across the 40-dimension parameter space. Because the response of the objective function to one parameter in a multi-variate surrogate model is affected by other parameters, we can get a mean response by fixing the target parameter while perturbing the remaining 39 parameters. For example, to get the response to *fff* (Table 1) when *fff* equals 1, we utilized the 200 samples that were generated using LHS and fixed *fff* to 1, using the surrogate model to predict the response of the 200 modified samples, and average the responses to get a mean response. For one parameter, we calculated the mean responses at multiple points to get a two-dimensional response curve (Figure S1). The amplitude of the response curve was used to evaluate each parameter's sensitivity.

We used a simple weighting algorithm to select the final parameter list for optimization. For each river basin, the most sensitive 10 parameters were assigned non-zero scores, i.e., 5, 3, 3, 2, 2, 2, 1, 1, 1, 1. In any single basin, parameters with ranks lower than 10 barely show sensitivity to the objective functions. In addition, this weighting algorithm emphasizes the most sensitive parameters in any single basin, which may not be sensitive elsewhere. In total, nineteen parameters were in the top 10 most sensitive across all 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., *fff*, *e_ice* (rank 6) and *d_max, a_coef, medlynslope* (rank 8) 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 et al., 2002), on the surrogate model, and obtained N (N=20) Pareto optimal sets of parameters values. We then ran CTSM using the N sets of parameter values and calculated their objective functions.

318 The trained surrogate model better mimicked the response curves as the number of
319 samples increases via iterating the Main Loop. In this study, we stopped after the tenth iteration
320 given the limited improvement in the last iteration runs. We used k-fold cross validation to
321 evaluate the accuracy of the surrogate model (k=5). We calculated the root-mean-square error
322 (RMSE) of the simulated objectives from surrogate models versus the objectives calculated from
323 CTSM runs.

324 The optimization run ranges from 1 September 2002 to 1 September 2009. The first two
325 years are used for spin-up, with data from 2004 to 2009 used for optimization. Prior to the
326 optimization simulations, we used a 58-year spin-up forced by ERA5 data to generate the initial
327 state for 1 September 2002 using the default CTSM parameters. All simulations were performed
328 on the NCAR Cheyenne supercomputer (Computational and Information Systems Laboratory,
329 2019).

330 ### 4.4 Parameter regionalization

331 Many parameters within CTSM are spatially uniform by default, which can be a limiting
332 assumption when optimizing a model as many parameters within hydrologic and land models
333 should vary spatially to account for the heterogeneity across the landscape (Mizukami et al.,
334 2017; Rakovec et al., 2019; Samaniego et al., 2010). For plant parameters, parameter spatial
335 heterogeneity might result from different plant traits in different dominant plant species. We
336 conducted preliminary single basin optimizations which showed large optimal parameter
337 discrepancies between the northern river basin, i.e., Kuparuk, and southern river basins, i.e.,
338 Beaver, Salcha, and Talkeetna (not shown). The Kuparuk River Basin is located north of the
339 Arctic Circle, much farther north than the other three basins. Therefore, we conducted two
340 optimization runs in this study, one for the northern river basin, and one for the three southern

river basins. Note that no SNOTEL sites near the Kuparuk have records overlapping with our optimization period, thus we conducted a single-objective optimization on river flow for the Kuparuk River. For the southern basins, we averaged the flow objectives across the three basins and snow objectives across ten SNOTEL sites (triangles in Figure 3).

We leveraged the ecohydrology region classification level III by Environmental Protection Agency for our simple parameter regionalization (Gallant et al., 1995). Optimized parameters for the northern basin are applied to the two Arctic ecohydrology regions, Arctic Coastal Plain and Arctic Foothills (highlighted in blue lines, Figure 4b). The remaining area uses the optimized parameters for the southern basins. Three out of the 15 basins intersect both southern and northern parameter regions, the Colville, Wulik, and Kuparuk rivers. The Colville is comprised of 54% northern and 46% southern areas, the Wulik contains 72% northern and 28% southern areas. Also, even though we optimized the Kuparuk to represent northern basins, 10% of the area in the Kuparuk watershed is located in our southern region (Figure 4a).

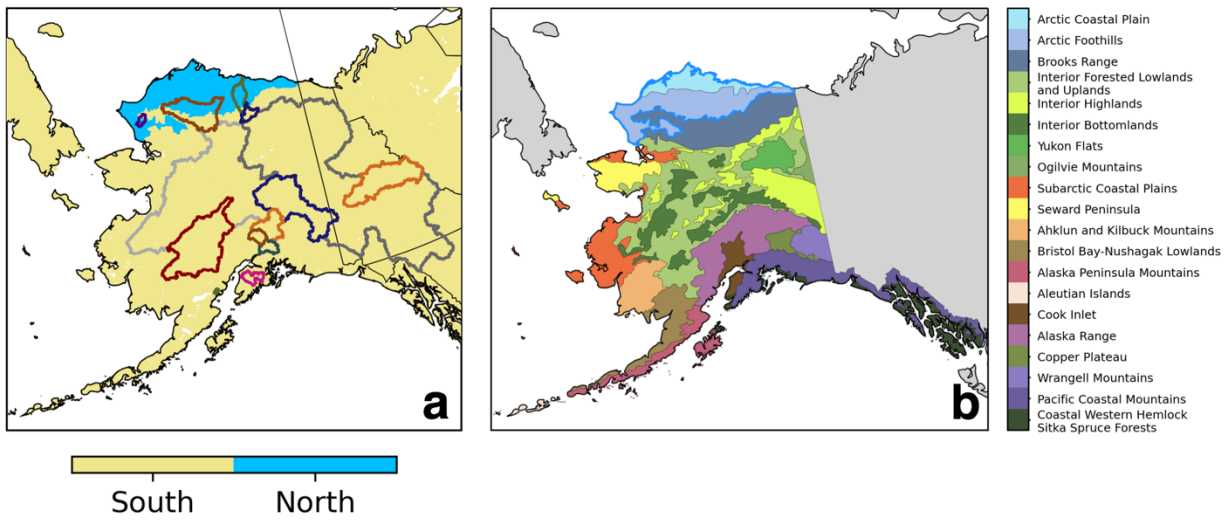


Figure 4: Parameter regionalization based on ecohydrology region classification. In Figure 4a, background colors denote the selection of optimized parameters, and river basins are highlighted using solid lines with colors corresponding to Figure 1. In Figure 4b, regions using optimized parameters for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

4.5 Parameter performance contributions

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

$$O_s = f(P) \tag{11}$$

$$\Delta O = O_{optz} - O_{base} \tag{12}$$

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

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

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

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

where n is the total number of parameters for optimization, i.e., 14, $P\backslash\{\gamma\}$ denotes all parameters except the selected one γ, S denotes the subset of $P\backslash\{\gamma\}$, $|S|$ denotes the length of the subset,

374 $O(S)$ denotes the objective function when we replace the baseline value using the optimized
375 value for all parameters in subset *S*.

376 5 Results

377 5.1 Optimization

378 For the southern basins, flow simulation is improved substantially while snow simulation
379 only sees minor improvements (Figure 5). Dots with the same color in Figure 5 constitute the
380 simulated Pareto front for a given optimization iteration. A Pareto front consists of simulated
381 Pareto optimal, i.e., if none of the objective functions can be improved in value without
382 degrading some of the other objective values. In general, the simulated Pareto front shifts
383 towards the origin, signifying improved model performance. Overlapping dots indicate the new
384 iteration failed to improve the Pareto front at that point. The two-dimensional Pareto front serves
385 as the basis for choosing our optimal parameter set. Future work could explore using an
386 ensemble of optimal parameter sets along the Pareto front, but that is outside the scope of this
387 initial investigation. We choose the set of optimized parameters that correspond to the minimum
388 averaged flow and snow objectives, highlighted using a red star in Figure 5. For this parameter
389 set, the corresponding flow and snow objective functions for the southern basins are 0.324 (0.676
390 KGE) and 0.490 respectively, while the baseline flow and snow objectives are 0.696 (0.304
391 KGE) and 0.489 respectively. Flow simulation in the Kuparuk is also significantly improved
392 through optimization. Since we conducted a single-objective optimization for Kuparuk, we
393 simply selected the set of parameters resulting in the best flow simulation. The optimized flow
394 objective is 0.189 (0.811 KGE) while the default flow objective is 0.574 (0.426 KGE).

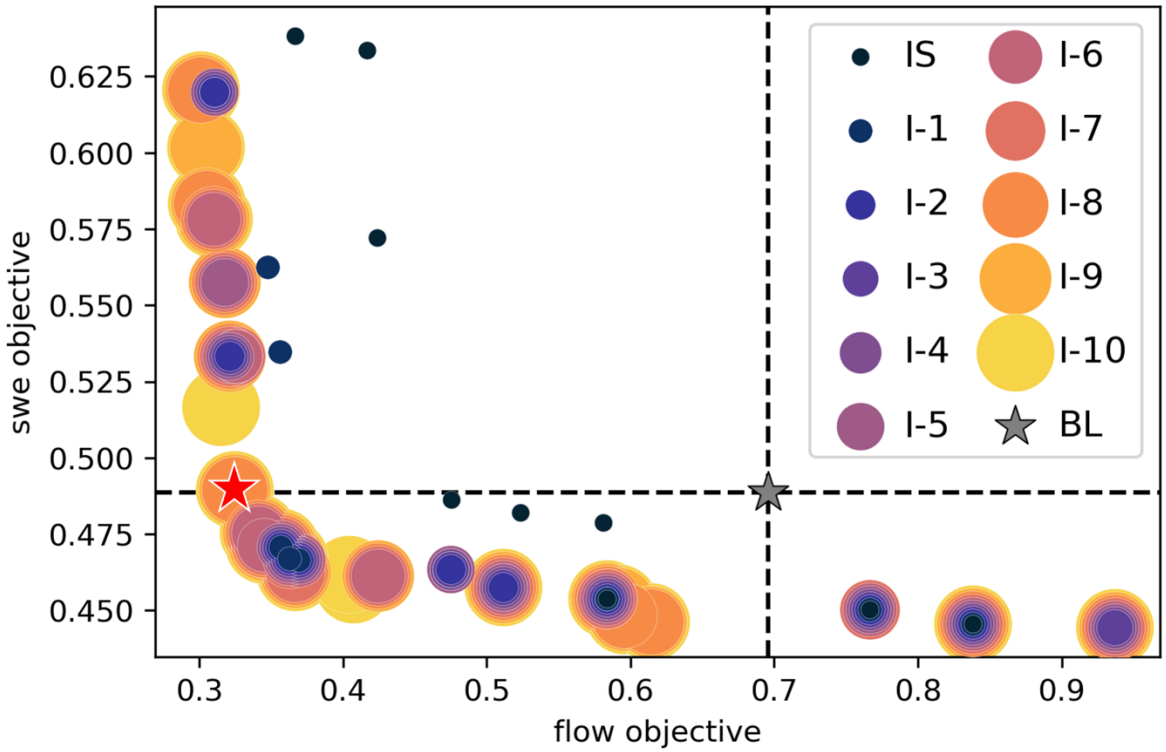


Figure 5: Simulated Pareto front of optimization for southern basins. Each colored dot corresponds to a Pareto optimal set of parameters. ***IS*** *denotes initial sampling,* ***I-1*** *denotes the first iteration, and so on, and* ***BL*** *denotes the baseline configuration. The red star denotes the selected optimized parameters.*

Interestingly, the northern and southern basins show very different hydrological responses to parameter perturbations as noted above. The mean response curve of flow (blue dots) and snow (red triangles) objectives to model parameters are shown in Figure 6. We used the method in Section 4.2 (Step 2) to calculate the mean response curves. Transparent dots denote the initial samples, while solid dots denote samples during optimizations and large dots correspond to the selected optimized parameters (optimized parameter values are shown in Table 1). The parameter sensitivity differs across basins. For example, *upplim_destruct_metamorph,* which affects snow densification through destructive morphism, shows greater sensitivity on flow simulations in the southern basins and is only marginally sensitive in the Kuparuk. In

addition, the flow performance in the south degrades as *upplim_destruct_metamorph* increases while the opposite trend was observed in the north. Some other parameters also show the opposite sensitivity across regions, including, *d_max, e_ice, frac_sat_soil_dsl_init,* and *om_frac_sf.* This intrinsic sensitivity difference leads to the divergence in optimized parameters across regions. In some extreme cases, the optimized parameters approach the upper and lower limits for the northern and southern basins, respectively, e.g., *fff, upplim_destruct_metamorph, e_ice, n_melt_coef*, which might result from differences in physical processes across the domain. For example, *e_ice* together with soil ice content affects the hydraulic conductivity in frozen soils and therefore has impacts on the vertical distribution of soil moisture and runoff (Swenson et al., 2012). There is ice-rich permafrost in the north while not in the south (Saito et al., 2020), so the differences in soil ice content might affect the optimized value of the ice impedance factor differently. In addition, parameter values approaching limits could indicate that the ranges are not wide enough due to model structural or forcing data errors that are compensated for during parameter optimization.

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

meteorological forcing than parameter perturbations. Therefore, runoff and flow simulations might show a stronger sensitivity to the parameter perturbations than SWE.

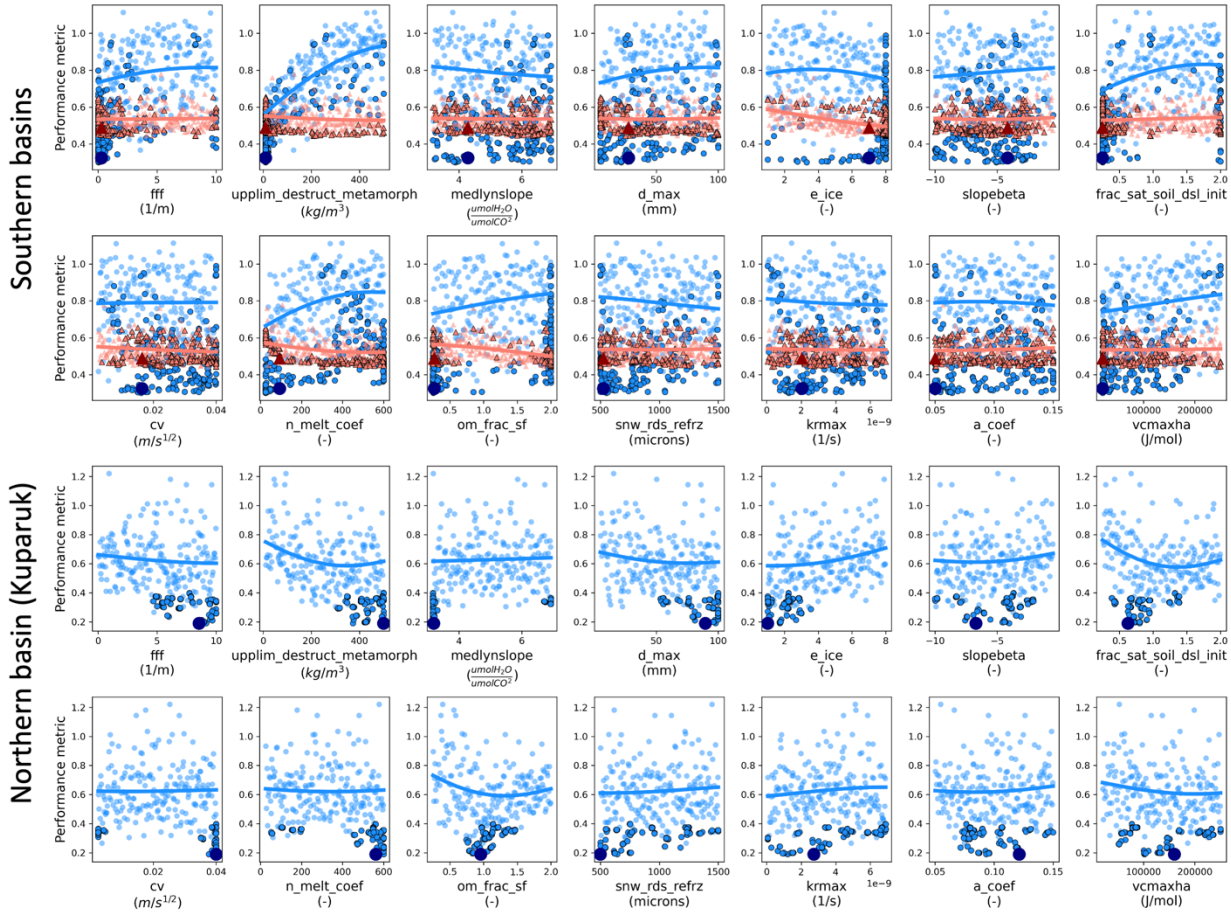


Figure 6: Mean response curve of flow (blue dots) and snow (red triangles) objectives to parameters. Transparent dots and triangles denote initial samples, solid dots and triangles with black edges denote samples during optimization iterations, and the large dot and triangle with darker colors denote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

5.2 Out-of-sample evaluation of optimized parameters

We further evaluate the optimized CTSM at 15 major river basins and 12 of them are out-of-sample. The daily KGE improves at 13 out of 15 basins and the median KGE across the 15 basins increases from 0.40 to 0.63 after optimization. Furthermore, even though we only conducted optimizations for four medium-size river basins with a total confluence area of around

16,500 km^2, 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 km^2) increases from -0.76 to 0.51 and the daily KGE for the Yukon River at Stevens Village (502,458 km^2) increases from -0.80 to 0.27. Only the Matanuska and Wulik river basins show slightly worse performance, with daily KGE decreasing from 0.63 and 0.57 to 0.60 and 0.51 respectively.

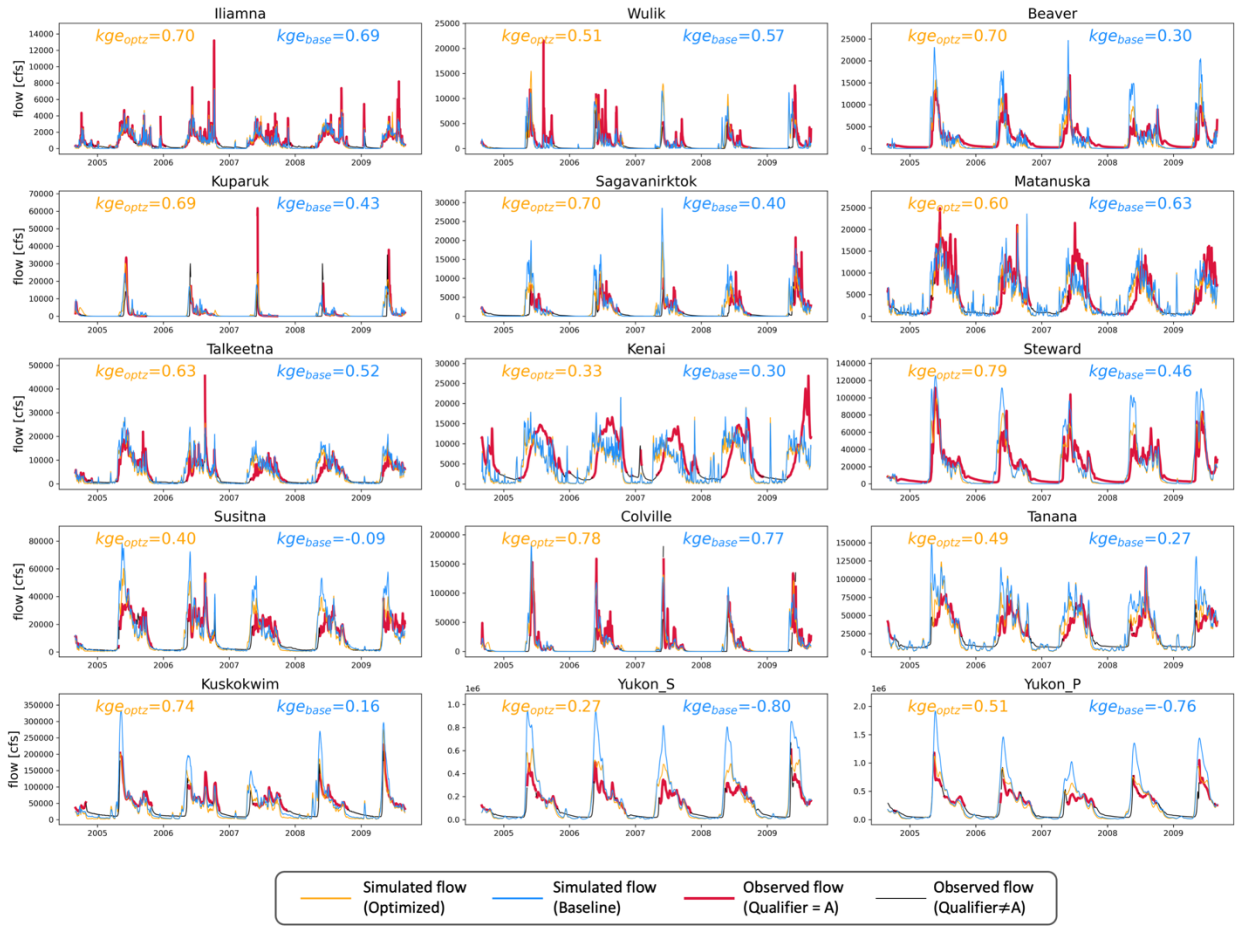


Figure 7: Model evaluation for flow time series. KGE is based on qualified observed flow at a daily time step.

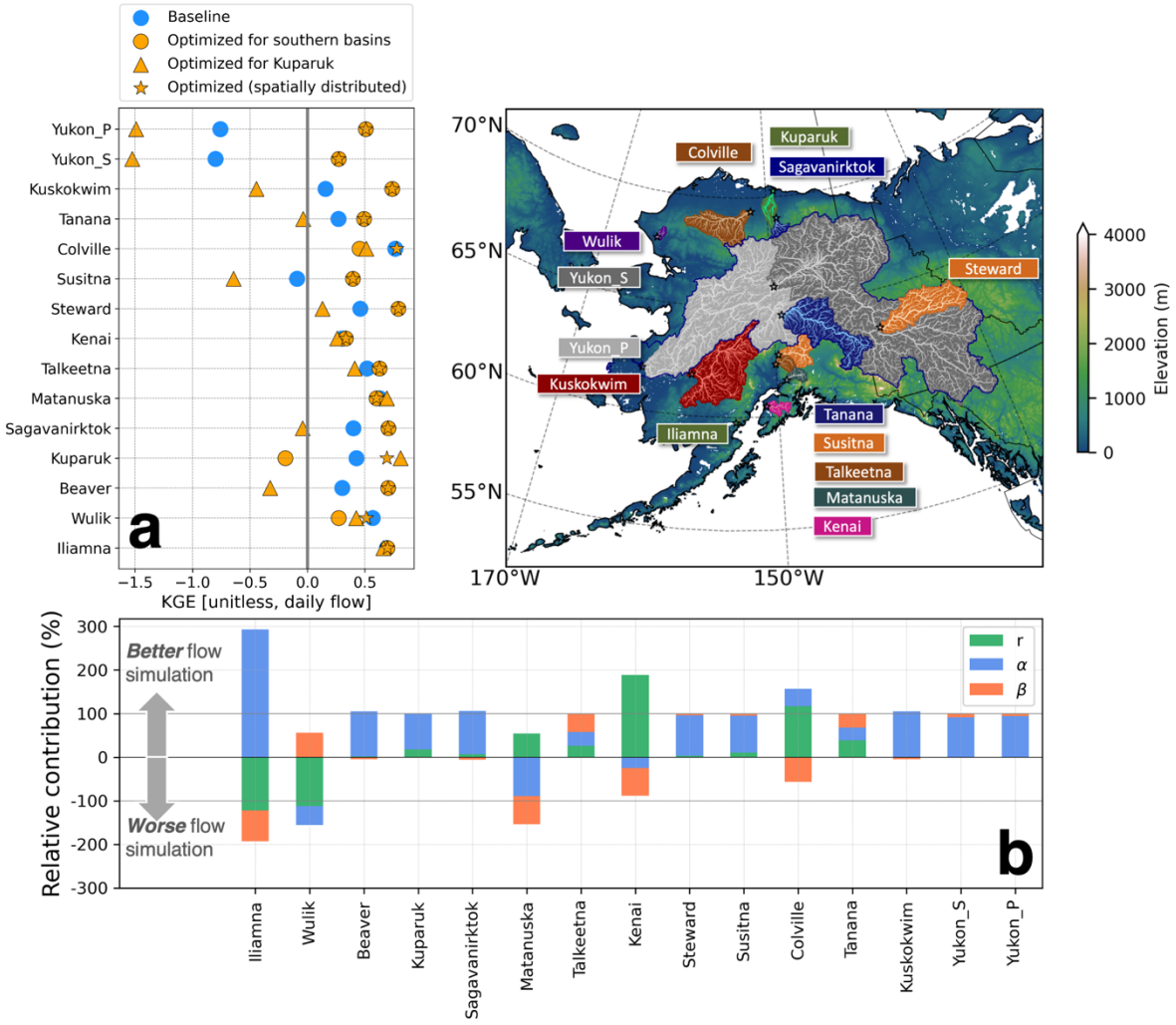


Figure 8: a) Default model performance (blue dot), and model performance using southern optimized parameters (orange dot), northern optimized parameters (orange triangles), and spatially distributed parameters by parameter regionalization (orange stars) and b) contribution of r, α, and β to KGE changes using the spatially distributed parameters by parameter regionalization

Improved model performance in cross-regional basins highlights the necessity of spatially variable parameters and parameter regionalization schemes. For Colville and Wulik, two uncalibrated basins, their model performance using spatially distributed parameters is better than that of any single optimized parameter set (Figure 8a). The optimized parameters in Kuparuk represent the northern region while our regionalization algorithm categorized 10% of the area in

Kuparuk to the southern region, which explains the slightly worse performance in Kuparuk using the spatially distributed parameters than that using only the northern optimized parameters. Improved flow variability contributes the most to better flow simulation. KGE combines three components in model errors, i.e., the linear correlation (r), a measure of flow variability error (α), and a bias term (β), so we decompose the KGE increment to the three components and calculate their relative contribution (RC) as follows

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

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

We also compute the Nash-Sutcliffe Efficiency (NSE) as many hydrologists and stakeholders are more familiar with this metric (Table 2). Even though river flow with ice cover (qualifier = A e) was not included in our optimization, it is worthwhile reporting the all-year NSE given the frozen river flow from the USGS are the best estimates available. For the unfrozen flow observation only (qualifier = A), the median NSE across all 15 basins increases from -0.07 to 0.43 after optimization. For all-year flow observation, the median NSE increased from 0.17 to 0.57.

Basin Name	Nationality	NSE-optz (all-year)	NSE-base (all-year)	NSE-optz (qualifier=A)	NSE-base (qualifier=A)	# days (qualifier=A) per year	Yellow (qualifier = A) Gray (qualifier != A)											
							Winter			Spring			Summer			Autumn		
							D	J	F	M	A	M	J	J	A	S	O	N
Iliamna	US	0.601	0.625	0.473	0.501	221												
Wulik	US	0.03	0.009	0.034	0.377	132												
Beaver	Canada	0.541	-0.069	0.541	-0.069	365												
Kuparuk	US	0.527	0.259	0.718	0.493	114												
Sagavanirktok	US	0.272	-0.749	0.434	-0.12	115												
Matanuska	US	0.599	0.578	0.437	0.415	212												
Talkeetna	US	0.587	0.466	0.218	-0.061	183												
Kenai	US	0.212	0.166	-0.242	-0.317	226												
Steward	Canada	0.652	0.228	0.652	0.228	365												
Susitna	US	0.578	0.172	-0.169	-1.88	150												
Colville	US	0.571	0.462	0.654	0.6	118												
Tanana	US	0.585	0.07	-0.062	-1.275	157												
Kuskokwim	US	0.502	0.174	0.578	-0.115	171												
Yukon_S	US	0.422	-0.946	-0.182	-4.323	158												
Yukon_P	US	0.697	-0.166	0.329	-4.091	128												
Median Value		0.571	0.172	0.434	-0.069													

Table 2: Nash-Sutcliffe Coefficient for all river basins.

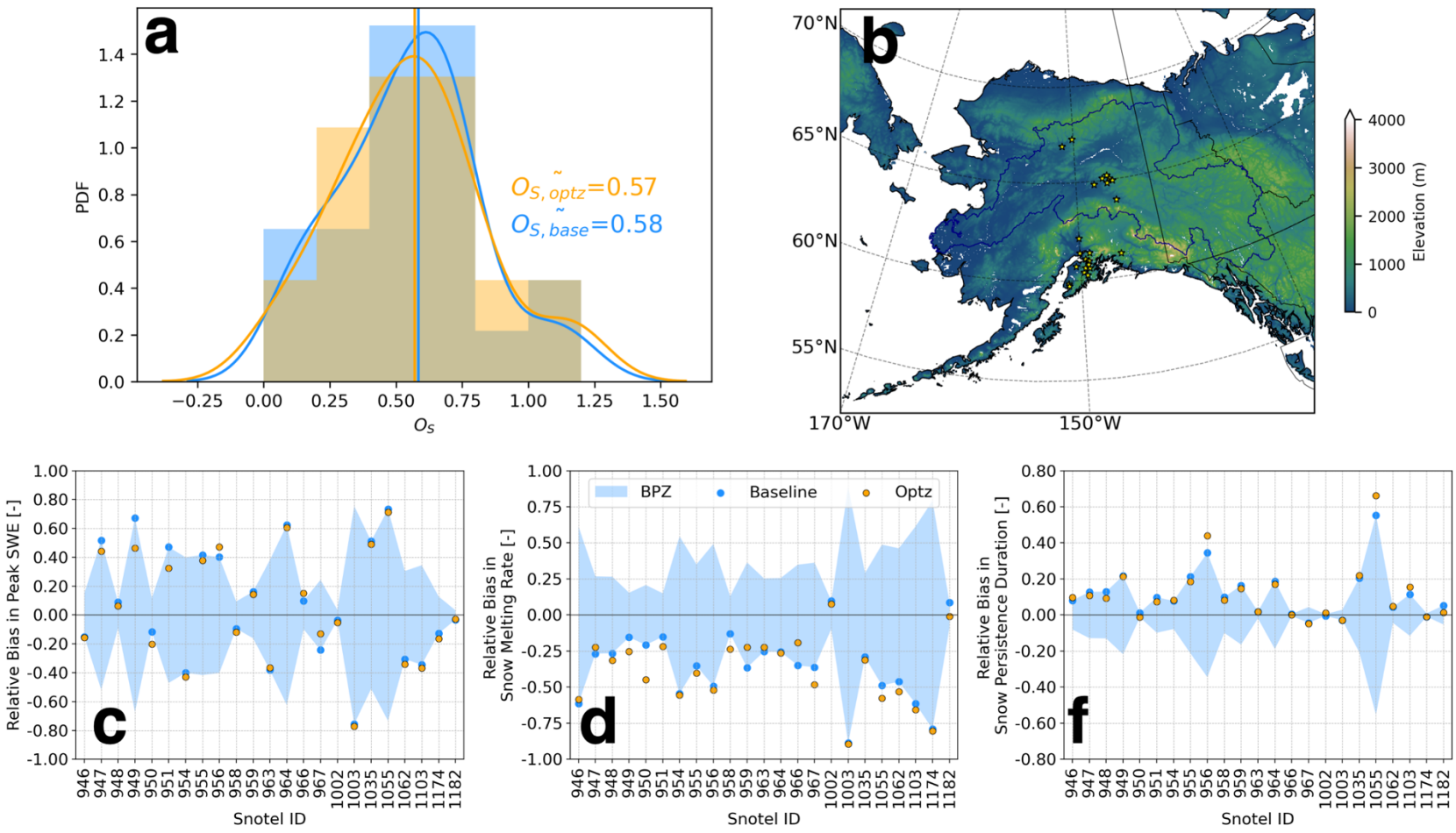


Figure 9: Evaluation of model performance on snow simulations. Panel a shows the distribution of aggregated snow metrics (O_S) across all SNOTEL sites in Alaska, whose locations are highlighted in yellow stars in Panel b. Panels c, d, f summarize the performance of snow simulation based upon individual metrics, and BPZ in the legend of Panel d is short for a better performance zone. If orange dots are located in BPZ, it means optimization improves snow performance. In Panels a, c, d, f, yellow corresponds with the model runs using optimized parameter values and blue corresponds with model runs using default parameter values.

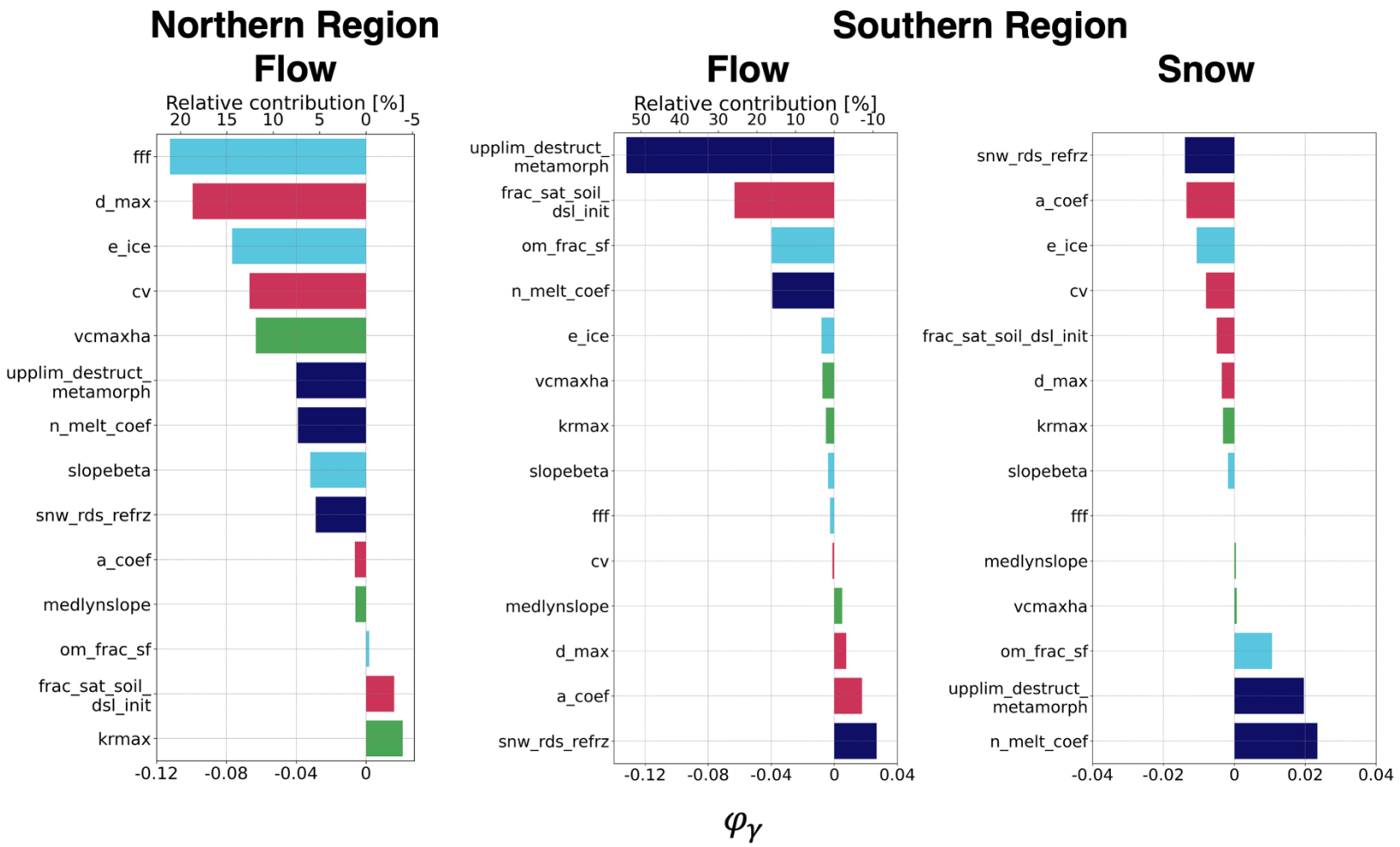
493 Seasonal snowpack simulation performance was not greatly improved by optimization
494 across the final Pareto front, or with our final optimized parameter set by choice. The median
495 value of the aggregated snow metric ($O_s, Equation\ 10$) across all SNOTEL sites is 0.57 using
496 our optimized parameters and 0.58 using default parameters (Figure 9a). We also examined each
497 component of the snow metric. If the orange dots fall in the blue area in Figures 9b, 9c, 9d, it
498 means that optimization improved the snow simulation. Out of 23 SNOTEL sites, optimization
499 reduced relative biases in peak SWE, snowmelt rate, and snow persistence duration at 12, 7, and
500 12 sites respectively (Figures 9b,9c, 9d). The median values for the relative bias in peak SWE,
501 snowmelt rate, and snow persistence duration are -0.03, -0.31, 0.08 using optimized parameters,
502 and -0.03, -0.35, and 0.10 using default parameters.

503 **5.3 Shapley decomposition**

504 In the southern region, large variations exist in the contribution of individual parameter
505 perturbations to simulation performance changes. For example, reducing the
506 *upplim_destruct_metamorph* value greatly improves flow performance, which contributes over
507 50% of the KGE increment (Figure 10). Three other parameters made noticeable contributions to
508 the improving flow simulation, i.e., *frac_sat_soil_dsl_init*, *om_frac_sf*, *n_melt_coef*. In addition,
509 the same parameter perturbation leads to opposite contributions in our flow and snow
510 simulations. The perturbation of *upplim_destrct_metamorph, om_frac_sf*, and *n_melt_coef*
511 significantly improves flow simulation while degrading snow simulation, while the perturbation
512 of *a_coef* and *snw_rds_refrz* worsens flow simulation while improving snow simulation.

513 In the northern region, the variation of parameter contributions is much smaller than that
514 in the southern region. The perturbation of *fff* contributes the most to the flow improvement in
515 the southern region, which only accounts for 21% of the KGE increment (Figure 10). Other than

516 *fff*, the top 5 parameters that contribute to the improved flow simulation include *d_max*, *e_ice*,
517 *cv*, and *vcmaxha*. These parameters belong to multiple categories, including acclimation
518 parameters relevant to photosynthesis, hydrology, and parameters affecting sensible, latent heat,
519 and momentum fluxes.
520



521

522 *Figure 10: Contribution of each parameter to the changes in objective function using Shapley decomposition*

6 Discussion and conclusions

We have developed the first high-resolution application and optimization of CTSM for Arctic hydrology. We used a high-resolution configuration because higher resolution models can more faithfully represent the complex ridge-valley patterns and high peaks across Alaska, and thus often more realistically simulate seasonal snow, orographic precipitation distributions, and potentially heterogeneous permafrost conditions (Newman et al., 2021; Rasmussen et al., 2011). River flow simulations are significantly improved after optimization, while the optimized snow simulation as compared to SNOTEL sites remains similar. The limited improvement in snow simulations depends more on the meteorological forcing such as precipitation than model parameter choices (Günther et al., 2019; Raleigh et al., 2015). The median NSE of daily flow increases from 0.17 to 0.57 across 15 river basins. For the Yukon River at Pilot Station, the USGS site with the largest confluence area in Alaska, and the fourth-largest river in North America, the NSE of daily flow increased from -0.17 to 0.70. In addition, the optimization is highly efficient given that the total area of the four optimized river basins only occupies 2% of the confluence area at the Pilot Station. To our knowledge, this study provides the most comprehensive evaluation and optimization of hydrological simulations across Alaska and the Yukon River Basin, which can be used as a benchmark for future Arctic hydrological modeling studies.

The optimization model framework is transferrable to other CTSM applications and can be informative when developing optimization workflows for complex land models. The transferability largely results from the global availability of the datasets used in this study, i.e., the ERA5 meteorological forcing data (European Centre for Medium-Range Weather Forecasts, 2019), soil texture from SoilGrid (Hengl et al., 2017), MERIT Hydro vector-based river network

546 (Yamazaki et al., 2019), and especially the CTSM PPE global parameter sensitivity analysis.
547 Correctly selecting sensitive parameters provides the foundation for the success of parameter
548 optimization. The PPE experiment can be extensively used in selecting sensitive CTSM
549 parameters not only to runoff and SWE but to any other variable simulated by CTSM. Finally,
550 hydrologic modeling of Arctic Alaska and the Yukon is one of the most challenging regions in
551 hydrological modeling due to the complicated land surface processes that are important in this
552 region. The improvements in hydrological simulation achieved by our optimization framework in
553 this environment imply potential efficacy in regions beyond the Arctic.

554 We also show that Arctic hydrology is not only influenced by hydrological parameters
555 but also parameters related to vegetation and thermal conductance. Previous hydrological studies
556 using CTSM mostly focused on hydrological parameters (Ren et al., 2016; Zhang et al., 2021). In
557 this study, out of the 14 optimized parameters, half are not directly related to hydrology and
558 snow processes (Table 1), which reveals the strong influence of non-hydrological
559 parameterization on Arctic hydrology. In the northern region, according to the Shapley
560 decomposition, perturbations of the 7 non-hydrological parameters contribute a total of 38.5% to
561 the KGE increment and 3 out of the top 5 sensitive parameters are non-hydrological, including
562 *d_max*, *cv*, and *vcmaxha* (Figure 10). In the southern region, the non-hydrological parameter
563 perturbations contribute 19.0% of the flow KGE increment and a decrease of
564 *frac_sat_soil_dsl_init* alone contributes 25.8% increment.

565 The Shapley decomposition analysis showed the different parameter contributions across
566 regions, reflecting the spatial heterogeneity of parameter sensitivities. The heterogeneity
567 manifests primarily in two ways. First, similar parameter perturbations lead to the opposite
568 direction of effects across regions. For example, a decrease in effective radius of re-frozen snow

(snw_rds_refrz), i.e., from the default value to the lower limit, contributes 5.4% of flow KGE increment in the northern region but -11.0% in the south. Second, the opposite parameter perturbation leads to the same direction of effects. A decrease of *n_melt_coef* in the northern region, i.e., 200 to 94, and an increase of *n_melt_coef* in the southern region, i.e., 200 to 526, both contribute positively to their flow KGE increments, with the RC value of 7.3% and 16.0% respectively. Because CTSM by default uses many spatially constant model parameters, we have modified CTSM to read in distributed parameters when they are available. This effort should be informative to future CTSM development for allowing spatially distributed parameters.

The parameter regionalization in this study is simple and effective, yet can still be improved. Spatially distributed parameters in Colville and Wulik, i.e., basins overlapping both southern and northern regions, generated better flow simulations than the parameters optimized for either region. However, for Matanuska, a southern basin, its flow simulation using northern optimized parameters is better than the one using southern optimized parameters, with daily flow KGE of 0.69 and 0.60, respectively (Figure 8a). The similarities between Matanuska and northern regions are neglected, likely because of either the oversimplified regionalization method or compensating errors. In addition, the large discrepancies in optimized parameters across regions only slightly affect the flow simulations in Kenai, Iliamna, and Matanuska (Figures 7, 8a), which indicates that the selected parameters may not be very sensitive for those out-of-sample basins. Therefore, for future improvement of regional applications, it may be helpful to include more representative basins for optimization and to implement a more sophisticated parameter regionalization algorithm.

The surrogate model can only mimic the true response surface. For the southern region, the root-mean-square error (RMSE) of the simulated flow and snow objectives are 0.05 and 0.03

592 respectively, and the RMSE of the simulated flow objective is 0.09 for the northern region. In
593 addition, the Shapley decomposition analysis is based upon the surrogate model, so the
594 contribution of each parameter perturbation reflects the simulated response surface. However, it
595 is infeasible to disentangle each parameter's contribution without a surrogate model. We would
596 need to run CTSM 16,384 (2^{14}) times for the Shapley decomposition while in this study we only
597 ran CTSM 400 times. Additional benefits from using surrogate models might arise by
598 incorporating other observational constraints, e.g., Active Layer Thickness, snow depth, or
599 evapotranspiration.

600 Finally, this work lays the foundation for a process-focused, stakeholder useful, high-
601 resolution coupled land and atmospheric modeling for cold regions both historically and under
602 future projections to quantify climate change impacts on inland freshwater systems.

603 # 7 Acknowledgments

604 This project was funded by NSF Navigating the New Arctic grant 1928078 and supported
605 by the National Center for Atmospheric Research, which is a major facility sponsored by the
606 National Science Foundation under Cooperative Agreement No. 1852977. We would like to
607 acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX)
608 provided by NCAR's Computational and Information Systems Laboratory, sponsored by the
609 National Science Foundation. We thank our Indigenous Advisory Council and numerous Tribal
610 and First Nation decision-makers for providing important insights to help inform the CTSM
611 configuration design. We also thank the Arctic Rivers Summit investigator team including
612 Nicole Herman-Mercer, Karen Cozzetto, and the Institute for Tribal Environmental Professionals
613 for facilitating the Indigenous Advisory Council and survey efforts. We thank Ryan Toohey for
614 informing our representative river basin selection. Furthermore, we greatly benefited from the

following open-source libraries to perform analyses presented in this study: NumPy (Van Der Walt et al., 2011), pandas (McKinney, 2010), geopandas (Jordahl et al., 2020), xarray (Hoyer & Hamman, 2017), matplotlib (Hunter, 2007), and cartopy (Met Office, 2015).

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

8 Reference

Ang, B. W., Liu, F. L., & Chew, E. P. (2003). Perfect decomposition techniques in energy and environmental analysis. *Energy Policy*, *31*(14), 1561–1566. https://doi.org/10.1016/S0301-4215(02)00206-9

Batjes, N. H., Ribeiro, E., & Van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data*, *12*(1), 299–320. https://doi.org/10.5194/ESSD-12-299-2020

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., … Wood, E. F. (2015). Hyper-resolution global hydrological modelling: what is next? *Hydrological Processes*, *29*(2), 310–320. https://doi.org/10.1002/HYP.10391

Choi, H. I., & Liang, X. Z. (2010). Improved Terrestrial Hydrologic Representation in Mesoscale Land Surface Models. *Journal of Hydrometeorology*, *11*(3), 797–809. https://doi.org/10.1175/2010JHM1221.1

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., & Zeng, X. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015WR017096

Clark, M. P., & Vrugt, J. A. (2006). Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. *Geophysical Research Letters*, *33*(6), 6406. https://doi.org/10.1029/2005GL025604

Computational and Information Systems Laboratory. (2019). *Cheyenne: HPE/SGI ICE XA System (University Community Computing)*. Boulder CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6RX99HX

Cox, C. J., Stone, R. S., Douglas, D. C., Stanitski, D. M., Divoky, G. J., Dutton, G. S., Sweeney, C., George, J. C., & Longenecker, D. U. (2017). Drivers and Environmental Responses to the Changing Annual Snow Cycle of Northern Alaska. *Bulletin of the American Meteorological Society*, *98*(12), 2559–2577. https://doi.org/10.1175/BAMS-D-16-0201.1

Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 223–244. https://doi.org/10.5194/ascmo-6-223-2020

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, *158*(3–4), 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

European Centre for Medium-Range Weather Forecasts. (2019). *ERA5 Reanalysis (0.25 Degree Latitude-Longitude Grid)*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/BH6N-5N20

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D.,

Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., … Yamazaki, D. (2019). Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resources Research*, *55*(2), 1737–1772. https://doi.org/10.1029/2018WR023903

Findlater, K., Webber, S., Kandlikar, M., & Donner, S. (2021). Climate services promise better decisions but mainly focus on better data. *Nature Climate Change 2021 11:9*, *11*(9), 731–737. https://doi.org/10.1038/s41558-021-01125-3

Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aðalgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L., Golledge, N. R., Hemer, M., Kopp, R. E., Krinner, G., Mix, A., Notz, D., Nowicki, S., Nurhati, I. S., Ruiz, L., Sallée, J.-B., Slangen, A. B. A., & Yu, Y. (2021). Ocean, Cryosphere and Sea Level Change. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Gallant, A. L., Binnian, E. F., Omernik, J. M., & Shasby, M. B. (1995). *Ecoregions of Alaska*. US Government Printing Office.

Giorgi, F. (2019). Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next? *Journal of Geophysical Research: Atmospheres*, *124*(11), 5696–5723. https://doi.org/10.1029/2018JD030094

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., & Dai, Y. (2016). Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models. *Water Resources Research*, *52*(3), 1984–2008. https://doi.org/10.1002/2015WR018230

Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance. *Water Resources Research*, *55*(4), 2779–2800. https://doi.org/10.1029/2018WR023403

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003

Hamman, J., Nijssen, B., Brunke, M., Cassano, J., Craig, A., DuVivier, A., Hughes, M., Lettenmaier, D. P., Maslowski, W., Osinski, R., Roberts, A., & Zeng, X. (2016). Land Surface Climate in the Regional Arctic System Model. *Journal of Climate*, *29*(18), 6543–6562. https://doi.org/10.1175/JCLI-D-15-0415.1

Hamman, J., Nijssen, B., Roberts, A., Craig, A., Maslowski, W., & Osinski, R. (2017). The coastal streamflow flux in the Regional Arctic System Model. *Journal of Geophysical Research: Oceans*, *122*(3), 1683–1701. https://doi.org/10.1002/2016JC012323

Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748.

https://doi.org/10.1371/JOURNAL.PONE.0169748
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
Hoyer, S., & Hamman, J. J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/JORS.148
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
Jefferson, J. L., Gilbert, J. M., Constantine, P. G., & Maxwell, R. M. (2015). Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. *Computers & Geosciences*, *83*, 127–138. https://doi.org/10.1016/J.CAGEO.2015.07.001
Jiao, Y., Lei, H., Yang, D., Huang, M., Liu, D., & Yuan, X. (2017). Impact of vegetation dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology coupled model. *Journal of Hydrology*, *551*, 116–131. https://doi.org/10.1016/J.JHYDROL.2017.05.060
Jordahl, K., Bossche, J. Van den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *geopandas/geopandas: v0.8.1*. https://doi.org/10.5281/ZENODO.3946761
Jorgenson, M. T., Shur, Y. L., & Pullman, E. R. (2006). Abrupt increase in permafrost degradation in Arctic Alaska. *Geophysical Research Letters*, *33*(2). https://doi.org/10.1029/2005GL024960
Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., … Zeng, X. (2019). The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583
Lawrence, D. M., & Slater, A. G. (2005). A projection of severe near-surface permafrost degradation during the 21st century. *Geophysical Research Letters*, *32*(24), 1–5. https://doi.org/10.1029/2005GL025080
Lawrence, D. M., Slater, A. G., Romanovsky, V. E., & Nicolsky, D. J. (2008). Sensitivity of a 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
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
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. *Journal of Applied Meteorology and Climatology*, *57*(3), 709–731. https://doi.org/10.1175/JAMC-D-17-0161.1
Newman, A. J., Monaghan, A. J., Clark, M. P., Ikeda, K., Xue, L., Gutmann, E. D., & Arnold, J. R. (2021). Hydroclimatic changes in Alaska portrayed by a high-resolution regional climate simulation. *Climatic Change*, *164*(1–2), 1–21. https://doi.org/10.1007/S10584-021-02956-X/FIGURES/12
Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., Levis, S., Swenson, S. C., Thornton, E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Lamarque, J., Niu, G., Qian, T., Running, S., Sakaguchi, K., … Zeng, X. (2010). *Technical Description of version 4 . 0 of the Community Land Model (CLM)*. *April*.
Osterkamp, T. E., & Romanovsky, V. E. (1999). Evidence for warming and thawing of discontinuous permafrost in Alaska. *PERMAFROST AND PERIGLACIAL PROCESSES*, *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
Pavelsky, T. M., & Zarnetske, J. P. (2017). Rapid decline in river icings detected in Arctic Alaska: Implications for a changing hydrologic cycle and river ecosystems. *Geophysical Research Letters*, *44*(7), 3228–3235. https://doi.org/10.1002/2016GL072397
Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., Williams, Z., Brunke, M. A., & Gochis, D. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, *8*(1), 41–65. https://doi.org/10.1002/2015MS000526
Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., Clark, M. P.,

& Samaniego, L. (2019). Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13991–14007. https://doi.org/10.1029/2019JD030767

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences*, *19*(7), 3153–3179. https://doi.org/10.5194/HESS-19-3153-2015

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J., Yu, W., Miller, K., Arsenault, K., Grubišić, V., Thompson, G., & Gutmann, E. (2011). High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *Journal of Climate*, *24*(12), 3015–3048. https://doi.org/10.1175/2010JCLI3985.1

Ren, H., Hou, Z., Huang, M., Bao, J., Sun, Y., Tesfa, T., & Ruby Leung, L. (2016). Classification of hydrological parameter sensitivity and evaluation of parameter transferability across 431 US MOPEX basins. *Journal of Hydrology*, *536*, 92–108. https://doi.org/10.1016/J.JHYDROL.2016.02.042

Roth, A. E. (1988). *The Shapley Value: Essays in Honor of Lloyd S. Shapley.* Cambridge University Press. https://doi.org/10.1017/cbo9780511528446

Saito, K., Machiya, H., Iwahana, G., Ohno, H., & Yokohata, T. (2020). Mapping simulated circum-Arctic organic carbon, ground ice, and vulnerability of ice-rich permafrost to degradation. *Progress in Earth and Planetary Science*, *7*(1), 1–15. https://doi.org/10.1186/S40645-020-00345-Z/FIGURES/7

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 5523. https://doi.org/10.1029/2008WR007327

Sankarasubramanian, A., Vogel, R. M., & Limbrunner, J. F. (2001). Climate elasticity of streamflow in the United States. *Water Resources Research*, *37*(6), 1771–1781. https://doi.org/10.1029/2000WR900330

Singh, R. S., Reager, J. T., Miller, N. L., & Famiglietti, J. S. (2015). Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations over the Southwestern U.S. *Water Resources Research*, *51*(4), 2648–2667. https://doi.org/10.1002/2014WR015686

Srivastava, V., Graham, W., Muñoz-Carpena, R., & Maxwell, R. M. (2014). Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin – Global Sensitivity Analysis of an integrated parallel hydrologic model. *Journal of Hydrology*, *519*(PB), 2238–2257. https://doi.org/10.1016/J.JHYDROL.2014.10.020

Stone, R. S., Dutton, E. G., Harris, J. M., & Longenecker, D. (2002). Earlier spring snowmelt in northern Alaska as an indicator of climate change. *Journal of Geophysical Research: Atmospheres*, *107*(D10), ACL 10-1. https://doi.org/10.1029/2000JD000286

Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., & Perket, J. (2019). Representing Intrahillslope Lateral Subsurface Flow in the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4044–4065. https://doi.org/10.1029/2019MS001833

Swenson, S. C., Lawrence, D. M., & Lee, H. (2012). Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *4*(8), 8002. https://doi.org/10.1029/2012MS000165

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for

efficient numerical computation. *Computing in Science and Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., & van den Broeke, M. R. (2017). Improving the Representation of Polar Snow and Firn in the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *9*(7), 2583–2600. https://doi.org/10.1002/2017MS000988

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, *41*(1), 1–17. https://doi.org/10.1029/2004WR003059

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environmental Modelling & Software*, *60*, 167–179. https://doi.org/10.1016/J.ENVSOFT.2014.05.026

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Yang, D., & Kane, D. L. (Eds.). (2020). *Arctic Hydrology, Permafrost and Ecosystems*. Springer Nature.

Yu, S., Wei, Y. M., & Wang, K. (2014). Provincial allocation of carbon emission reduction targets in China: An approach based on improved fuzzy cluster and Shapley value decomposition. *Energy Policy*, *66*, 630–644. https://doi.org/10.1016/j.enpol.2013.11.025

Zhang, C., Abbaszadeh, P., Xu, L., Moradkhani, H., Duan, Q., & Gong, W. (2021). A Combined Optimization-Assimilation Framework to Enhance the Predictive Skill of Community Land Model. *Water Resources Research*, e2021WR029879. https://doi.org/10.1029/2021WR029879